

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/242692090>

Strengthening the Role of ICT in Development

Article

CITATION
1

READS
375

5 authors, including:



Joseph Kizza

University of Tennessee at Chattanooga

128 PUBLICATIONS 559 CITATIONS

[SEE PROFILE](#)



Janet Aisbett

The University of Newcastle, Australia

91 PUBLICATIONS 889 CITATIONS

[SEE PROFILE](#)



Vir V. Phoha

Louisiana Tech University

214 PUBLICATIONS 2,972 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Guide to Computer Network Security [View project](#)

SPECIAL TOPICS IN COMPUTING AND ICT RESEARCH

Strengthening the Role of ICT in Development

Editors

Joseph Migga Kizza,
Kathy Lynch,
Ravi Nath,
Janet Aisbett
Phoha Vir

Volume V

FOUNTAIN PUBLISHERS
Kampala

Fountain Publishers
P.O. Box 488
Kampala
E-mail: sales@fountainpublishers.co.ug;
publishing@fountainpublishers.co.ug
Website: www.fountainpublishers.co.ug

Distributed in Europe and Commonwealth countries outside Africa by:
African Books Collective, P.O. Box 721, Oxford OX1 9EN, UK.
Tel: 44(0) 1869 349110, Fax:+44(0)1869 349110
E-mail: orders@africanbookscollective.com
Website: www.africanbookscollective.com

Distributed in North America by:
Michigan State University Press
1405 South Harrison Road
25 Manly Miles Building
East Lansing, MI 48823-5245
Tel: +1 517 3559543, 800 6782120
Fax: +1 517 4322611
E-mail: msupress@msu.edu
Website: www.msupress.msu.edu

© Makerere University 2009

All rights reserved. No part of this publication may be reprinted or reproduced or utilised in any form or by any means, electronic, mechanical or other means now known or hereafter invented, including copying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

ISBN 978-9970-02-738-5

Contents

<i>Preface</i>	<i>vi</i>
<i>Keynote speaker biographies</i>	<i>viii</i>
<i>Author Biographies</i>	<i>x</i>
Part One: Information Technology	1
1. Integration of Biometrics with Cryptographic Techniques for Secure Authentication of Networked Data Access <i>Abanti Cyrus Makori</i>	1
2. A Requirements Analysis Framework for Human Activity Systems (HAS):The Case of Online Learning <i>Philip O Ayoo and Jude T Lubega</i>	14
3. Knowledge Management Frameworks: A Review of Conceptual Foundations and a KMF for IT-based Organizations. <i>Kyoratungye Karemente, Jennifer Rose Aduwo, Emmanuel Mugejjera and Jude Lubega</i>	35
Part Two: Information Systems	59
1. Managing University Research: Key Policy Issues <i>Anthony. J. Rodrigues</i>	60
2. Application of Grid Computing for on line Learning Resources <i>Sadeque Imam Shaikh</i>	67
3. Relationship between Information Systems Development Paradigms and Methods <i>Peter Nabende, Benjamin Ahimbisibwe, Jude T. Lubega</i>	75
4. From the Ground Up: User Involvement in the Development of an Information System for use in HIV Voluntary Counselling and Testing <i>Kathy Lynch and Stacey Lynch</i>	85
5. Translating Transliterations <i>Jörg Tiedemann and Peter Nabende</i>	97
6. Decision Support in the Operating Theatre – Usability Aspects <i>John Kizito</i>	109
7. Implementing Successful E-health Implementations within Developing Countries <i>Stella Ouma, Marlien E. Herselman and Van Greunen</i>	118

8.	The Evaluation of Information Visualisation Techniques Using Eye Tracking <i>André P. Calitz, Marco C. Pretorius, Darelle Van Greunen.....</i>	135
9.	A Model for the Adaptation of Contact Centre Computer User Interfaces <i>Bronwin Jason and André Calitz.....</i>	152
10.	A Flexible Biomedical Ontology Selection Tool <i>Gilbert Maiga.....</i>	171
11.	Web Content Filtration According to Context of Use: Case Study of Accessibility Guidelines <i>Rebema Baguma, Jude T. Lubega, Roger G. Stone and Th.P. van der Weide.....</i>	190
Part Three: ICT Sustainable Development.....		211
1.	SIM or Application Layer? An Implementation-Level Analysis on the use of Mobile Phones for ICT Development <i>Hannah Thinyane.....</i>	203
2.	A Gender Critique of Uganda’s Rural ICT Access Policy: Opportunities and Challenges <i>Aramanzan Madanda, Dorothy Okello, Grace Bantebya – Kyomubendo.....</i>	213
3.	Pedagogical and Conceptual Design of An E-learning Environment for HIV/AIDS Education <i>Bada Joseph Kizito, Jarkko Suhonen.....</i>	228
4.	Assessing Appropriate ICT with ARIS case in Mozambique <i>Markus Pscheidt, Victor Van Reijswoud and Theo Van Der Weide.....</i>	238
5.	The Effect of Cultural Differences on Software Development <i>D Patel, C Lawson-Johnson & S Patel.....</i>	250
6.	Service-oriented Architectures as a Vehicle for ICT in developing Countries: An Awareness Campaign <i>Agnes F. N. Lumala, Benjamin Kanagwa, José Ghislain Quenum and Jude T. Lubega.....</i>	264
Part Four: Data Communication and Computer Networks.....		277
1.	(Inter)net Neutrality: Your Voice Matters <i>Joseph Kiziza Migga.....</i>	279
2.	Security Issues in Ambulatory Wireless Sensor Networks (AWSN): Security Vs Mobility <i>Drake Patrick Mirembe and Maybin Muyebe.....</i>	289

3.	Measurement and Analysis of Copper Access Network for ISDN Basic Rate Interface <i>Grace Bakaki and Idris A. Rai</i>	302
Part Five: Computer Science		309
1.	Apprenticeship Environment and Co-operation Based on the Grid and Web 2.0 Designed for Training Communities with Common Interest Centres <i>Bernabe Batchakui , Claude Tangha , Roger Nkambou And Thomas Ndie Djotio</i> ...	311
2.	Developing a Set of Requirements for Algorithm Animation Systems <i>Jean Greyling</i>	329
3.	Multi-Scale Angiography Filters: Techniques Today <i>Fred N. Kivanuka</i>	337
4.	Security Analysis of an Agent-Mediated Book Trading Application <i>Richard Ssekibuule and Jose Ghislain Quenum</i>	347

Preface

Makerere University has so far successfully organized and hosted four Annual International Conferences since 2005. The Conference provides an international forum for researchers and practitioners to present and discuss the state-of the-art research and practices in Computer Science, Software Engineering, Information Systems, Information Technology, Computer Engineering, Data Communications and Computer Networks and ICT for Sustainable Development.

This is the fifth conference in the ICCIR series. While the original theme of ICT for sustainable development is still strong, for the last years the conference has been held as the International Conference on Computing and ICT Research (ICCIR). This name reflects the broad nature of ICT research that, in the end, contributes to sustainable development.

A brief look through these proceedings will convince you of the explicit mix of applied and basic research across the disciplines. Keynote speakers and contributing authors alike address both generic developments and applications of state-of the art research to specific African problems.

Thus, in the Computer Science stream, basic research contributions range across semi-supervised learning, dynamic resource allocation, automatic auction mechanisms, and more. In the Information Systems stream, researchers look at appropriate selection of applied grid computing, information systems paradigms, language translations, decision and bio-health applications. In Data Communications and Computer Networks, basic research ranges from security issues in Wireless Sensor Networks, network path optimization, and access network for ISDN.

On the applied research side, we see investigations into ICT adoption amongst grassroots NGOs; mobile phone use for development, gender and ICT, service oriented architectures, culture effects on software development and e-learning for rural communities. There are studies into applications of ICT to support interaction amongst SME clusters, between lecturers and students, and amongst stakeholders in small scale farming.

The Faculty of Computing and Information Technology at Makerere University again hosts ICCIR, with the generous support from both the Netherlands Organization for International Cooperation in Higher Education and IBM. The conference starts off with a pre-conference workshop on Innovative Mobile Technology and services for Developing countries, PhD colloquium sessions which give trainee researchers the valuable opportunity to expose their research to kindly but searching critique. The main conference has one guest speaker, five key note speakers and the 28 contributed papers that each survived review by two independent reviewers.

A conference like this one can only succeed with an excellent team effort. We would like to acknowledge the great contribution from the technical committee members and paper reviewers who helped in the review of the submitted papers, select high

quality papers, and provide valuable comments to the authors. Special thanks go to the organizing committee and the ICCIR secretariat who worked hard to bring together the many strands needed to create a successful conference. As in previous years, the publishing teams at Fountain Publishers, Kampala, were professional and prompt, and the Conference Secretariat worked hard and long.

ICCIR Secretariat

Keynote Speaker Biographies

Henk SOL is a Professor at the Faculty of Economics and Business (Business and IT) at Groningen Nederland Netherlands and the Technical University Delft. He is a Commissioner for Groningen Airport Eelde NV te Eelde; Director / owner Sol Information Management BV, Haren, President Foundation Board PAO Informatica, Informatica seminars Foundation Board President, Foundation Board Member Allersmaborg; Member of the Board of Trustees Nicolaes Mulerius; Advisory Professor The Expertise Centre. Chaired Professor of Business Engineering and ICT at Delft and Groningen, Supervisor responsible for over 60 Ph.D. dissertations and some 20 dissertations are presently being supervised. He has highly published in both Journals and conferences in the discipline of Business and IT.

Jean GREYLING is an Associate-Professor at the Department of Computer Science and Information Systems at the Nelson Mandela Metropolitan University in Port Elizabeth, South Africa. He obtained his PhD in 2000, focusing on the selection of first year IT students. Since then he has been actively involved in various projects related to Computer Education. As lecturer he specializes in teaching programming at various levels from First Year to Honours. His current interest relates to the use of technology in the development of communities. Some of his current postgraduate students are researching the use of mobile phones for serving various communities. This includes skills provision as well as supporting new students on campus.

André Paul CALITZ is currently a Professor in the Department of Computer Science and Information Systems at the Nelson Mandela Metropolitan University in Port Elizabeth, South Africa. He has supervised 46 Honours treatises, 14 master dissertations and 6 PhD's. He has 19 refereed journal publications, 46 refereed conference publications and two book contributions. He has been an ICT consultant on projects for businesses in S.A. such as VW, Firestone and NMBM.

He served on the UPE and NMMU Council (Elected by Academics), and his research interests include ICT education research, ICT personnel research, HCI and data visualisation, ICT Project Management, ICT strategic advantage and E/M –Commerce. He was awarded the Eastern Cape Computer Society of South Africa: ICT Computer Person of the Year, Eastern Cape and was the first Professional Member of the CSSA in the Eastern Cape, South Africa.

Joseph Kizza MIGGA is currently a professor of Computer Science and Director of the Center for Information Security and Assurance at the University of Tennessee - Chattanooga, Tennessee, USA. His research interests are in Social Computing and Computer Network Security in which he has so far published numerous conference and journal articles and more than seven books. He was appointed a UNESCO expert in Information Technology in 1994. He is the chief editor for the Journal of Computing and Information Research.

Anthony J. RODRIGUES is Professor of Computer Science at the School of Computing and Informatics, University of Nairobi. Research interests include Scientific Computing, Approximation Theory, Error Analysis and Systems Modeling. He has also been involved in the design and development of sustainable integrated management information systems at various universities and is currently studying the developmental impact of various ICT policies and strategies (or the lack thereof) in the region.

Author Biographies

Benjamin AHIMBISIBWE is a PhD student at the Faculty of Computing and Information Technology, Makerere University. He holds a MSc. in Information Science and Bachelors in Library and Information Science from Makerere University.

Philip Ouma Ayoo is the head of ICT at the Inter-University Council for East Africa, and is currently a PhD student in Information Systems at Makerere University, Uganda. His research interests are in the areas of e-learning and ICT policy development, where he has published and presented a number of papers. Before joining the IUCEA he was teaching at the Department of Information and Media Technology, Moi University, Kenya, where he had previously served as a Senior Technician.

Rehema BAGUMA is a PHD student in Information Systems researching on Web accessibility. She holds a BLIS Hons, a PGD CS and an MSc in Computer Applications. She is employed in the Faculty of Computing and IT, Makerere University as an Assistant Lecturer and has skills in Development Informatics, Networking and System Administration

Grace BAKAKI a holder of a Bachelor of Information and Computing degree from Kyambogo University, Uganda and a Student of MSc in Computer Science from Makerere University, Department of Computer Science, Faculty of Computing and IT, Makerere University and he has finished his MSc project of Analysis of Copper Access Networks for Data Connectivity Optimization in ISDN, Basic Data Rate Interface. His main areas of research interest are in the wireless and wired Networks.

Bernabé BATCHAKUI is a post graduate Computer Science Engineer and PhD student at LIMMS at the Ecole Nationale Supérieure Polytechnique of the University of Yaoundé 1. He is a junior Lecturer at the same Engineering School teaching object-oriented programming, unified modeling language, software engineering and expert system. His research areas are e-learning Grid and Multi-agents Systems.

Thomas Ndie DJOTIO is a post graduate Computer Science Engineer and PhD student at LIMMS at the Ecole Nationale Supérieure Polytechnique (ENSP) of the University of Yaounde 1 – Cameroon. He has been working as software developer, system and network administrator, entrepreneur, author and trainer. Thomas is a MCP, CCNA and Security +Certified and now a junior Lecturer at the ENSP engineering school teaching information and network security, operating system, network technology and administration and system programming. His research areas are intelligent systems for network administration and security, threats investigation and mitigation, network and telecommunication protocol definition.

Bronwin Anastasia JASON received her BSc Computer Science degree in 2005, her BSc (Hons) degree in 2006 and recently completed her MSc Computer Science and Information Systems degree (um Lauda) from NMMU. Her present position is “project / integration manager” at a section called Technology Strategy and Integration (TSI) at Telkom South Africa. The section manages the end-to-end introduction of new technology into Telkom’s network.

Aramanzan MADANDA is an Assistant Lecturer in the Department of Women and Gender Studies, Makerere University. He is currently a PhD candidate by research on Gender Relations and ICT Adoption in Uganda focusing on Computing and Mobile Telephony. Madanda has published and Co-authored a number of publications examining Women's Empowerment through ICT as well as examining the role of ICT policy in accessibility by especially women.

Dorothy OKELLO is a Lecturer in the Department of Electrical Engineering, Makerere University and also coordinates an NGO Women of Uganda Network (WOUGNET) whose mission is to promote and support the use of ICTs by women and women organisations in Uganda. Dr. Okello has over 10 years of experience in teaching, researching, as well as conducting projects and contributing to policies in the telecommunications sector. She is an activist in the area of getting more women, small-scale enterprises and rural communities engaged in the information society for development – via policy advocacy, programme implementation, monitoring and evaluation.

Grace Bantebya-KYOMUHENDO is a Professor in the Department of Women and Gender Studies Makerere University. Grace has done extensive research and published widely. Her most recent publications being a book entitled “Women, Work and Domestic Virtue in Uganda” which got an award from African Studies Association 2007, and a chapter entitled “The Mobile “Village” Phone Business: A Vehicle for Rural Women's Empowerment and Gender Equality in Uganda” (2009) in Ineke Buskens and Anne Webb (Eds), African Women and ICTs: Investigating Technology, Gender and Empowerment Published by Zed publishers UK.

MD.Sadeque Imam SHAIKH is an asst. professor & Head, ICT section, Dept. of Computer Science and Engineering (CSE), University of Science & Technology Chittagong, Bangladesh, he has been serving this Department for about nine years. He has completed his BSc (Hon's) in Computer Science and MSc in Computer Science from National University of Bangladesh after then he has completed his Msc in Computer Networking from University of Bedfordshire of United Kingdom. He is also the Chief Executive Editor of the International Journal of University of Science & Technology Chittagong (IJUSTC). He was also the part time asst. professor Head of IT Department of London Reading College of UK.

John KIZITO is a PhD student in the framework of the project “Strengthening Teaching and Research in the four Public Universities in Uganda”; collaboration with University of Groningen (The Netherlands) and Makerere University Kampala (Uganda). He is specifically working with the *Fundamental Computing* and *Computational Linguistics* research groups. He is also an Assistant Lecturer in the Department of Computer Science and working with the *Health Informatics* Research group, Faculty of Computing and IT at Makerere University.

Kyoratungye KAREMENTE is currently a part time lecturer at the Faculty of Computing and Information Technology, Makerere University and also a doctoral student in Information Technology at the same faculty. He earned his MBA at Stirling University, United Kingdom and an M. Sc. in Computer Science at Makerere University, Uganda. His research interests include the electronic business, knowledge management,

strategic information systems and information system security. He has published a paper on Dynamic Multi-Media Information Provision for Promotion of Tourism in the Thematic proceedings of ATLAS Africa Conferences Volume 5 – New avenues for tourism and wealth creation in Africa, October 2008.

Hannah THINYANE is a lecturer in the Computer Science Department at Rhodes University. She has been a member of the Siyakhula Living Lab since it began. Her research interests are in Human Computer Interaction, in ICT for Development projects.

Drake Patrick MIREMBE is a Second year PhD at the Faculty of Computing and IT, Makerere University in collaboration with Radboud University Nijmegen, under the supervision of Prof. Fisseha Mekura, Dr. Maybin Muyeba and Dr. Richard Brinkman. Also he works with the Department of Networks, Faculty of Computing and IT, Makerere University as an Assistant Lecturer. His research interests include; Digital Security and privacy enhancing technologies, Low cost Mobile communication technologies, wireless technologies, Artificial intelligence and health informatics.

Maybin MUYEBA is a member of BCS, IEEE and a fellow of the Higher Education Academy (FHEA). He holds a PhD from University of Manchester Institute of Science and Technology (UMIST) in the area of Mining Generalisation-based induction rules from databases and their algorithmic performance enhancements. Before joining MMU, Dr. Maybin worked as a lecturer at Bradford and Liverpool Hope Universities and has industrial experience, at Granada Learning (Manchester) as a Software Engineer. He is involved in many conferences and journals as a reviewer, presenter and session chair. Maybin Muyeba's work can be summarised as Data Engineering, in particular Data Mining. He also has collaborative research links with the University of Manchester, University of Liverpool, Montana University (USA), Makerere University (Uganda) and Macau University (China). His research interests include; applying fuzzy logic (type-1, type-2) to Data Mining algorithms such as association rules, classification (decision trees, support vector machines), clustering, rule induction and other areas in machine learning and soft computing. Recently areas such as classification association rules, data Stream mining and privacy preservation and security assurance interest him as well.

Peter NABENDE is a member of academic staff at the Faculty of Computing and Information Technology, Makerere University, Kampala. Currently, he is a PhD student in the Computational Linguistics Research group, University of Groningen. His interests include: Information Systems Theory, Development, Processing and Management; Transliteration and Translation generation, and Cross Lingual Information Extraction; Machine Learning and Artificial Intelligence application to Information Systems.

Jörg TIEDEMANN is a member of academic staff at the Department of Alfa Informatica, Faculty of Arts, University of Groningen. He holds a PhD in Computational Linguistics and MSc. in Computer Science. His research interests include: Question Answering; Parallel corpora and extraction of multilingual data; machine translation; Machine Learning and Empirical Natural Language Processing.

Abanti Cyrus MAKORI is a PhD Student in Information System at Nairobi University; has MSc IT (Nkumba University), MSc Computer Based Information System (Sunderland University UK), B. Ed Science (Egerton University) with eight

years of teaching experience, four of which at college and university level. Currently a lecturer at Kabarak University and visiting lecturer KCA University. He has presented two papers at international conference.

Gilbert MAIGA Holds an MSc. In Computer Science [MAK], MSc. Animal Physiology [UON], BVM [MAK], and Diploma in Computer Systems Design [WLC]. The MSc. Computer Science project was on “Information systems integration in HIV-AIDS Management: A web-based database approach.” He a PhD student in the Department of Information Systems, Faculty of Computing and IT, Makerere University, where he is also an assistant lecturer. Bio medical ontology evaluation is currently his main area of interest for research.

Bada Joseph KIZITO is a final year PhD student at the department of Computer Science and Statistics, Faculty of Science, University of Joensuu, Finland. His major area of studies in doctoral programme is Computer Science; he has membership with ACM and IEEE. He has taught at undergraduate and postgraduate levels in a number of Ugandan Universities since 2002.

Fred N. KIWANUKA is a holder of a Bachelor of Science (Mathematics), Makerere University, Uganda, PGD. Data Communications (London), Msc Digital Communication Networks (London Metropolitan University, UK). He is an Assistant Lecturer in the department of Networks, Makerere University and also a PhD student at the University of Groningen, The Netherlands at the Institute of Mathematics and Computing Science, Intelligent Systems Group under the sponsorship of NUFFIC and Makerere university. His research is about automated Image processing with a focus on automated thresholding, Clustering systems, attribute and connected filters and algorithms for medical Images.

Dr. Jarkko SUHONEN holds an Assistant Professor position at the Department of Computer Science, University of Joensuu, Finland. Dr. Suhonen is the coordinator of two e-learning programmes offered by the University of Joensuu: ViSCoS (<http://cs.joensuu.fi>) and IMPDET (<http://www.impdet.org>). In his PhD thesis, Dr. Suhonen constructed and evaluated the multi-threaded and agile FODEM method which was used to design and develop the ViSCoS studies. Dr. Suhonen has co-authored and published over 40 peer reviewed papers in scientific journals and conferences.

Stella OUMA is currently a lecturer of programming units at Tshwane University of Technology in the Faculty of Information Communication Technologies in Pretoria, South Africa. Previously, she obtained her Bsc. Information Technology from Jomo Kenyatta University of Agriculture and Technology. She is through with her Masters degree and currently she is pursuing her doctoral studies.

Marlien HERSELMAN is a Principal Researcher at Meraka Institute at the CSIR in Pretoria, South Africa. In this capacity, she assists in various projects and also supervises various students from different universities on their masters and doctoral studies. In 1999 she obtained a PhD at the University of Pretoria. Her PhD studies focused on the use of computer games. She is currently busy with a funded research project on Technology assessment models in rural communities in South Africa, funded by the National Research Foundation. She is also an adjunct professor at Nelson Mandela Metropolitan University in Port Elizabeth.

In 2000, Prof Herselman was awarded a scholarship of the Deutsche Akademischer Austauschdienst (DAAD) University of the International Women's University (IFU) to attend a summer semester at the University of Hamburg, in the project field of information. In 2005, she was named Women Researcher of the Year at TUT. In 2006 she obtained NRF rating at level C3 in South Africa. She has already written over 40 accredited research articles and has 27 accredited conference contributions at international conferences. Her research focuses on ICT for business enhancement, rural community, and informatics and e-health projects.

Kathy LYNCH is an Associate Professor in ICT Research and Development at the University of the Sunshine Coast, Queensland Australia. Her research interest in ICTs transcend their development, implementation and evaluation in promoting efficiencies and effectiveness in business and organisations. She is a research associate in the Faculty of Computing and IT at Makerere University. Dr Lynch has worked as a volunteer for VOLSET in Uganda.

Stacey LYNCH is a virologist undertaking a PhD at the University of Melbourne, Victoria Australia. She has volunteered for VOLSET in Uganda, and continues to support their goals.

Marco C. PRETORIUS is the Usability Manager and Researcher at the UNISA School of Computing, in Pretoria, South Africa where he conducts usability and eye tracking tests and research. He is a Certified Usability Analyst (CUA) through Human Factors International. Marco completed a Masters degree (MSc) Cum Laude in Computer Science and Information Systems at the Nelson Mandela Metropolitan University, Port Elizabeth, South Africa.

Darelle van GREUNEN is a Senior Lecturer in the School of ICT at the Nelson Mandela Metropolitan University in Port Elizabeth. She is also contracted as Senior Researcher for SAP CEC Research. Darelle has researched a variety of IT systems but particularly focused on the users of southern Africa. She serves on the Board and Executive Councils of several national and international professional bodies. Darelle has published extensively in the area of Human Computer Interaction. Her research focuses mainly on User Experience, User Interface Design and cultural issues that impact on interaction design and then ensuring that applications are usable.

Martin MUBANGIZI holds a BSc With Education (Physics and Mathematics) from Mbarara University of Science and Technology, Uganda and a Masters of Computer Applications from University of Pune, India. He is currently a PhD Student in Computer Science at the Faculty of Computing & I.T., Makerere University.

Martin has been a Lecturer at Nkumba University and a Part-Time Lecturer in the Department of Information Technology, Faculty of Computing & I.T. Makerere since 2005.

Ernest MWEBAZE holds a BSc (Electrical Engineering) and an MSc (Computer Science) from Makerere University. Ernest has been with the Faculty of Computing & I.T., Makerere University as an Assistant Lecturer since 2006 and is currently pursuing a PhD in the field of Causal Machine Learning on a sandwich program between Makerere University, Uganda and the University of Groningen in The Netherlands.

Idris A. RAI is an Associate Professor in Communication Networks, and Deputy Dean Academic Affairs Faculty of Computing and Information Technology, Makerere University. He spent around 15 years pursuing Higher education in different countries in Europe. During his PhD work in France, he proposed new mechanisms that significantly change the way traditional Internet performs under Web traffic. He also designed new architecture that allows autonomic multimedia content delivery over the Internet during his Post-doctorate in UK. Some of his research results have been published in some of the world's best journals and conferences, and his publications have been widely cited by academic world over.

He decided to come back to Africa to contribute in research capacity building. At Makerere University, he has supervised a number of MSc students to successful completion, and he is currently supervising tens more and some PhD students. His current research activities span networking areas such as Internet protocols, wireless mesh networks, overlay networks, and mobile computing. His research generally investigates how advanced networking can be exploited to impact on communities in developing countries.

Agnes F. N. LUMALA holds an MSc in Computer Science of Makerere University, Uganda. He is currently an Assistant Lecturer in the department of Networks, FCIT, Makerere University. Her research interests include formal aspects of software engineering and service-oriented systems.

Benjamin KANAGWA holds a PhD in Science of Radboud University, Nijmegen, Netherlands. He is currently an Assistant Lecturer in the department of Computer Science, FCIT, Makerere University. His research interests include systematic design of software systems. SOA and cloud computing.

Jude T. LUBEGA holds a PhD degree in Computer Science (E-Learning) from the Department of Computer Science, School of Systems Engineering of the University of Reading, UK. He is a member of the British Computer Society and Hong Kong Web Society. He is a Deputy Dean in charge of Graduate Studies and Research and lecturer in the Faculty of Computing and Information Technology, Makerere University, Uganda. His research interests include Tracking and Assessment in e-learning, Content Authoring, Multimedia, Multi-Agent Systems, Data Warehousing, Knowledge Representation, ICT for Development, Web-based Systems and Mobile Learning. He has authored numerous papers in international journals and conferences in the mentioned areas.

Richard SSEKIBUULE is an Assistant Lecturer in the Department of Computer Science, Makerere University. He holds a BSc (Computer Science, Mathematics) from Makerere University Kampala, and a MSc (Computer Science) from Radboud University Nijmegen. Richard is a PhD student in the department of Computer Science. His research interests are in the area of agents systems and digital security for open distributed computing.

Claude TANGHA is currently the Chief of the Department of the Computer Engineering at the Ecole Nationale Supérieure Polytechnique (ENSP) of the University of Yaounde 1 – Cameroon and the Director of the Laboratoire d'Informatique, de Modélisation et Simulation des Systèmes (LIMMS, ex-LABORIMA). He has been teaching since 1976 at the ENSP. His research areas are software engineering, e-learning

and artificial intelligence including expert systems, multiagent systems, and formal systems. He has directed many research projects, contributed to many conferences and seminars nationwide and at the international level.

Roger NKAMBOU is currently a Professor in Computer Science at the University of Quebec at Montreal, and Director of the GDAC (Knowledge Management Research) Laboratory (<http://gdac.dinfo.uqam.ca>). He received a Ph.D. (1996) in Computer Science from the University of Montreal. His research interests include knowledge representation, intelligent tutoring systems, intelligent software agents, ontology engineering, student modeling and affective computing. He also serves as member of the program committee of the most important international conferences in Artificial Intelligence and Education.

Dilip PATEL holds the chair of Information Systems at London South Bank University. He is the head of the Centre for Information Management and EBusiness, which consist of four research groups: E-Business and The Digital Society, Health Informatics, Information Management and Modelling and Knowledge Based Systems. Professor Dilip Patel is currently on the editorial board for two international journals: The International Journal of Cognitive Informatics and Natural Intelligence (IJCiNi) and the International Journal of Information Technology and Web Engineering. He is also Editor-in- Chief of The International Journal of Computing and ICT Research.

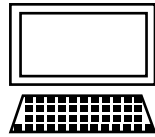
Markus PSCHIEDT has finished his Master of Science in the area of Computer Science in Austria and then worked several years in the field of software development. Currently, he is a development worker with HORIZONT3000, an Austrian NGO for development cooperation, at the Catholic University of Mozambique. In this assignment he is responsible for the design and construction of the university's management information systems and has been introducing computer based systems at the university in the departments of bookkeeping, human resources and academic registry. Currently he also is an external PhD student at the Institute for Computing and Information Sciences (ICIS) at the Radboud University in the Netherlands. He is interested in personal, organizational and technical factors which influence the sustainability of information systems in general and in the context of developing countries in particular. He investigates these issues in his PhD thesis which is supervised by Prof. Theo van der Weide at the Institute for Computing and Information Sciences at Radboud University Nijmegen.

John QUINN is a lecturer in Computer Science at the Faculty of Computing & I.T., Makerere University. He received a PhD in machine learning from the University of Edinburgh (2007), in the Institute for Adaptive and Neural Computation, School of Informatics. He is currently researching applications of machine learning in Uganda.

Jos'e G. QUENUM earned a Ph.D. in Computer Science from University Pierre et Marie Curie (UPMC), France, in 2005. He held a post doctoral position from early 2006 to mid 2007. He is currently Senior Lecturer at Makerere University, Faculty of Computing and It in the department of Computer science. His research interests include complexity and adaptiveness, interaction and interaction protocols, the Web and service-oriented computing systems. He has published many research papers in areas related to intelligent systems.

Ddembe WILLIAMS Is a Senior Lecturer in Information Systems at Faculty of Computing and It Makerere University. For seven years, Ddembe headed the Centre for Advanced Systems Modeling and Simulation , at London South Bank University, U.K. Ddembe has an MSc and a PhD in Computer Science from LSBU. Ddembe's practical and professional work in teaching and consultancy focuses on applied computing to systems requirements engineering, process analysis, and problem structuring and model-based decision support systems. His research work centers on the contribution that system dynamics can make to the advancement of systems modeling and simulation. Ddembe has published over 20 refereed conference and journal papers.

PART ONE



Information Technology

1

Integration of Biometrics with Cryptographic Techniques for Secure Authentication of Networked Data Access

Abanti Cyrus Makori

The modern Information Technology evolution demands the use of computer networks with strict security performance. The password-based authentication system and the token-based systems that are current deployed, are not able to meet this performance. Achieving higher levels of security requires authentication. Today's standard security mechanism is a password that you create, remember it, and change it frequently. These passwords are not foolproof because they can be stolen, forgotten, cracked, sniffed and even tampered with. In fact it is quicker to gain access to electronically held data, copy or print it out than it is to search through or photocopy manually held files.

The problems of traditional personal authentication systems may be solved by biometric systems. Biometric identification is extremely effective authentication. It is an important weapon to protect against credit card fraud and phantom withdrawals. Biometrics can identify a person's unique physical characteristics, including fingerprints, facial features, voice pattern, retinal, irises DNA and keystroke.

The problem with biometrics approach is that the biometric properties cannot reasonably be kept secret. Although biometric systems have advantages over traditional system, one of the unsolved issues is how we can combine cryptography with biometrics to increase overall system security The objectives of this paper are:

- 1. To critically examine the attacks on traditional personal identification authentication systems in order to determine the risks of using password based authentication..*
 - 2. To examine the biometric technologies and how they are used in authentication with an aim to improvise a secure model.*
 - 3. To apply cryptographic techniques in biometric system with a view to increase its security on networked data access.*
-

1. Introduction

The Information Technology (IT) revolutions such as the Internet, wireless communication, and e-commerce have led to organisational data to be accessed online and offline. The major headache is how to securely authenticate identity of the people they are dealing with. Identification and authentication requirements are steadily increasing in both online and offline. There is the urgent need of both the public and private sector entities to know who they are dealing with. Andrew [2003] has defined authentication as a technique by which a process verifies that its communication

partner is who it is supposed to be and not an impostor. Technology demands that either you fully automate or semi-automate the operations. This means that the data is electronically stored

Authentication is the binding of an identity to the principal. Network-based authentication mechanisms require a principal to authenticate to single system either local or remote. According to Bishop [2003], subjects act on behalf of some other external entity. The identity of that subjects control the action that subjects may take. The subject must bind the identity of that external entity. The trick lies on the proving that the identity supplied for authentication indeed proves the subject to be the real subject.

The current security model for verification of identity, protection of information and authentication to access data or services is based on using a token or password, tied to and thereby representing an individual to either authenticate identity or allow access to information [Ann et al, 2007]. This token may be password or shared secret (something you know), an identity card (something you have) or biometric (something you are). In all this cases, the details of the token are held by a third party whose functions is to authorise and at times allow the transaction to proceed if the details of an individual's token match those stored in a database.

According to Andrew [2003], William [2000], Simson [1997], Kaufman et al [2002], O'Brien [2004], Turban and Wetherbe [2000] and Haag et al [2002], have identified noticeable weaknesses of the traditional-based authentication. The password which is the most commonly used technology for authentication may be guessed by hackers, eavesdropped, forgotten, stolen, wired among others. Bishop [2003] in particular identified a number of password categories that are easy to guess. Kessler [1996] in his paper that appeared in the internet and internetworking security published in 1997 identified the password weaknesses as password guessing, blacklisting, password theft, login spoofing, monitoring the traffic between the user and the computer, and replay attack.

To secure the password-based authentication, a secret that has 64 bit randomness is desirable. This means that the users are required to have much long password that is difficulty to remember. Research has also shown that many organisations are becoming sceptical on the other person they are dealing with. Turban and Wetherbe [2000] noted that vulnerability of information system is increasing day by day as we move towards a highly networked and distributed computing.

Kaufman et al [2002] identified authentication systems such as password-based, address-based and cryptographic authentication all of which have some weaknesses. Many researchers have proposed the use of biometric-based authentication as the most secure and privacy way to access data on the network. [Haag et al 2004, William 2003, Bishop 2003, Ann et at 2007, Umit 2006].

2. Authentication

Authentication is a technique by which a process verifies that it's communicating partner is who it is supposed to be and not an impostor [Tanenbaum 2003]. According

to Bishop [2003] authentication is the binding of an identity to a principal. Verifying the identity of a remote process in the face of a malicious active intruder is surprising difficult and requires complex protocol based on cryptography. Authentication deals with the question of whether you are actually communicating with a specific process. Network based authentication mechanism requires a principal to authenticate to a single system either local or remote [Bishop 2003]. On such a case subjects act on behalf of some other external entity. The identity of that entity controls the actions that its associated subject must bind to the identity of that external entity.

Authentication is used to verify the identity of users in order to control access to resources, to prevent unauthorized users from gaining access to the system and to record the activities of the users in order to hold them accountable for their activities [Mathew 2002]. This is the reason why external entity must provide information to enable the system to confirm its identity. Authentication process consists of obtaining the authentication information from an entity, analyzing the data and determines if it is associated with that entity. The information comes from (i).What the entity knows (Password, secret information), (ii). What the entity has (Badge or card), (iii).Where the entity is (Terminal), and (iv).What the entity is (Fingerprint, odour, retina, hand geometry).

2.1. Password-Based Authentication

Password is information associated with an entity that confirms the entity's identity [Bishop 2003]. This is an example of authentication mechanism based on what the people know. The user supplies a password, and the computer validates it. If the password is one associated with the user's identity, the identity is authenticated. If not the password is rejected and authentication fails. Kaufman et la [2002] argues that password based authentication is not who you know, its what you know. In the middle ages castles and fortress were building to protect the people and the valuable properties inside [Pfleeger and Pfleeger 2007]. This traditional way of providing security was characterised by strong gate or door to repel invaders; heavy walls to withstand objects thrown or projected against them; surrounding moats, to control access; arrow slit to let arches shoot approaching enemies; crenulations to allow inhabitants to lean out from the roof and pour hot or vile liquids on attackers; draw bridge to limit access to authorised people; and gatekeepers to verify that only authorised people and goods could enter.

Turban et al [2003] has defined access control as the restriction of unauthorised user to access to a portion of a computerised user or to the entire system. Access to a computer consist of physical, access to system and access to specific commands, control transaction privileges , programs and data. Password cracking is a technical vulnerability attacks to a system [Scott et al 2003]. Each user account represents a potential vulnerability. In spite of all efforts, even difficult-to-guess password become essentially worthless with the advent of fast cheap computer, utilities like crack and network sniffers. Breaking into the system is no longer a big deal anyone can do that, even a script kiddie [Fadia 2003]. According to Pfleeger and Pfleeger [2007] network

environments needs authentication, but this is difficult to achieve securely because of the possibility of eavesdropping and wiretappings.

Password authentication has a number of weaknesses making authentication to fail. This is due to the way to control password distribution, password is simple and easy, naïve implementation, brute force attack, eavesdropping and guessing [Mathew 2003, Comer 2004, William 2006, Tere and Null 1999, Pflieger and Pflieger 2007]. The passwords that are easily guessed as put forward by Bishop [2003] are those based on Account names; Usernames; Computer names Dictionary words; Reversed dictionary words; Pattern from keyboard; Shorter than six characters and Containing only digits. All these passwords can be learned by: (i). Try default passwords used with standard account that are shipped, (ii). Exhaustively try all short passwords, (iii). Try words in the systems online dictionary or list likely passwords, (iv). Collect information about users. (v). Use Trojan horse, and (v). Tap line between remote use and the host [William 2006].

In the context of communication across networks the attackers that have been by William [2006] are disclosure, traffic analysis, masquerade, content modification, timing, sequence source and destination repudiation. The main disadvantage of password is that they can be stolen and forgotten. Hence to help cap this problem, biometric technologies are used as discussed in the following section.

3. Biometric Technologies

The problems of password authentication can be solved by biometric [Laudon and Laudon 2006]. Biometric has been describe by [Bishop 2003] as the identification by physical characteristics. Using such a feature for computer to authenticate would eliminate errors in authentication. These feature are physiological (fingerprint, hand geometry, eye (iris and retina), face and ear), behavioural biometric (such as voice, signature, keystroke, and gait) and esoteric biometric (facial thermographs, DNA, odour and palm vein). These biometric advantage and disadvantages are discussed in the section that follows.

3.1. Physiological biometric

a. Fingerprint

When you touch something with your fingers, you leave a specific impression on the touched item. This is called a finger print. A fingerprint has been defined by Oxford Dictionary as an impression on a surface of the curves formed by ridges on fingertip, especially such an impression made in ink and used as a means of identification. One of the most common forms of biometrics available is the fingerprint. The strengths associated to this are that it is more widely accepted, convenient and reliable.

A foetus' fingerprints are normally fully developed already after seven months. Except for big injury, disease or decomposition after death the specific characteristics on one's finger does not change throughout a lifetime [Henning, 2005].

During the years of working with fingerprint matching, examiners have come up with three levels of detail in fingerprint [Bolle et al 2001 and Bishop 2003]: **Level 1, Global or Galton Level:** If you have a look at your fingerprint you will see it is a

“landscape” full of papillary lines. The higher and lower parts of the papillary lines are called ridges and the valleys respectively. According to [Henning, 2005] the formation of these ridges and valleys are combination of several environmental and genetic factors. The direction in the skin formation is given in the DNA structure but final structure of the fingerprint is formed by different random events such as the position of the foetus in the womb, and the composition and density of the surrounding amniotic fluid. This is why fingerprints, unlike DNA, are different on identical twins. The flow of the ridges and valleys, together with singular points, core and delta, ridge count and orientation all belong to the set of features that can classify and index a fingerprint at the first level [Henning, 2005]. The pattern is classified using the Henry classification system. **Level 2, Local Level:** At the local level the examination process looks closer at different local ridge characteristic, so called minutiae. A minutiae characteristic is a ridges termination, where a ridge ends or a ridge bifurcation, where a ridge diverges into two new branch ridges. The NIST standard for Forensic Identification definition of minutiae is “friction ridge characteristics that are used to individualise that print. Minutiae occur at points where a single friction ridges deviate from an uninterrupted flow. Deviation may take the form of ending division or immediate origination and termination”. **Level 3, Very Fine Level:** At this level, intra-ridge details can be detected. These are essentially the shape and position of the sweat pores which are considered highly distinctive and can help identify a person. It requires high resolution.

The fingerprint scanners are fairly cheap. The FBI certified fingerprint scanners are Indentix, DBI, Crossmatch, Veridom, Infineon and Authentec [Bolle et al, 2001]. However the error rate associated with this is approximately one in one hundred. The fingerprint biometric technology merits are: Long tradition of use as immutable identification in law enforcement; large existing database; Good for forensic investigation; can be collected using low-technology- means and converted into digital forms; and Low cost readers. However fingerprint have problems including Presentation of fingerprint, Elasticity of the skin, Pressure, Bad quality fingerprint, Impostor attacks, Public un acceptance in some countries as it is associated with criminal activities, Contact based sensing and Can be hard to get a good read with old, cold, greasy, cut, or bruised fingers.

b. Eyes (Iris and Retina)

Current research suggests that it might be possible to use iris scan to determine not only that a woman is pregnant but also the sex of the unborn child. According to Haag et al [2007] future transaction processing systems will be integrated with biometric processing system (BPS). The BPS will capture and process physiological characteristics of the person performing the transaction. The physiological characteristics may include the presence of alcohol, illegal drug, hair loss, weight gain, low blood sugar, vitamin deficiencies, cataracts and even pregnancy.

c. Hand Geometry

Hand geometry is the measurement and comparison of the different physical characteristics of the hand. According to Bolle et al [2001], Hand geometry is one way of identifying a person. It involves computing the widths and lengths of the fingers

at various locations, using the capture image. These metrics define the features of the user's hand. Hand geometry has been used for physical access and time attendance at a wide variety of location. For example it has been used at Citibank data centre, 1996 Atlanta Olympics, New York university dorms, University of Georgia to verify students when they use their meal cards, Walt Disney World and US Department of immigration and naturalisation has installed it at Fastgate Bermuda International Airport.

The hand scan is manufactured by Recognition System Inc. (RSI)-patent a division of Giant Ingersoll-Rand. Founded in 1986. It is a popular means of biometric authentication due to public user acceptance, good for verification, and easy for self administration.

. Although hand geometry doesn't have the high degree of performance or individuality, they are limited by poor for identification, no international database, and contact based sensing

d. *Face Recognition*

Face recognition is the process of authentication of a person based on different characteristics on his or her face. Humans often recognise each other by their faces, but no one knows which are the most significant characteristics used when a human recognises another human face. This is the reason why there is no unified theory on how to best represent and recognise a face in an automated biometric authentication system. However the fundamental structure of the face is mostly used and most systems are invariant to variables like position, pose, expression, facial hair or glasses.

Face recognition software can operate in different environments, from well controlled environments to uncontrolled environments. The controlled environment is when a person sits in front of the camera and is looking straight into the camera without any special expressions. This method is usually used for verification (confirming claimed identity, such as with a computer logon, or an ATM). The surveillance cameras, at a football match; scanning the faces of the crowd, looking for known hooligans is an example of face recognition system. This is usually used for identification (picking one person out in a database of many). Face application include surveillance, database lookup, video indexing, secure computer logon, airport and banking security. The main advantages of face recognition are: Public acceptance; No intrusive or contact less; Works with photographs, videotapes or other image source; and Good for verification. The disadvantages of face recognition are: need good lighting, poor for identification, and individuals have option of disguising the face.

It is to understand that the face recognition has some challenges. A face is detected according to shape and features in the image, such as eyes, ears, and mouth. To cope with some of these problem, neural networks is often used in face recognition software. This allows the software to "learn" how to perform classification tasks base directly on patterns in data [Orlan, 2003]. The top suppliers of facial recognition system are Viisager Technology, and Visionic

3.1. Behavioural biometrics

a. Voice recognition

Voice recognition is a very common biometric technology. The goal of voice recognition is to understand spoken words- that the contents of what is being said [Orlan et al 2003]. The voice recognition technology will be valuable in systems that require hands free system, such as hand free set for mobile phones and voice command interpretation in automated telephone call centre. Other potential uses include computers, cars, consumer electronics and even appliances [Orlan et al 2003].

Voice recognition has unique advantage over other biometric because it relies on human speech, which is primary modality in human-to-human communication, and provides a non-intrusive method for authentication. By extracting appropriate features from a person's voice and modelling the voiceprint, the uniqueness of the physiology of the vocal tract and the articulatory properties can be captured to high degree and used very effectively for recognising the identity of the person. Recognising the user based on voiceprint is commonly known as speaker recognition in academic community, encompassing speaker verification, speaker identification, speaker classification, speaker segmentation and speaker clustering.

Speaker recognition has improved over the years hence it is very favourable with respect to fingerprint recognition and other biometric [Mansfield et al 2001]. IBM has developed conversational biometrics which combines acoustic voiceprint recognition with knowledge-based recognition and more than 60 invention disclosures have been filed, covering various aspects of robust acoustic voiceprint recognition knowledge-based recognition [Ramaswamy2003].

The major vendors are T-Netix, Naunce, and Veritel of America. There are two modes which voice recognition can operate in. The most common is the constrained mode or text-dependent mode, where the user is restricted to predetermined single words or short phrases. In unconstrained verification mode where the speech input is free, or text-independent, the user is not required to say the same sentence during each access, but this mode has a higher error rate than constrained mode [Orlan et al 2003]. The advantages of voice recognition are: Very high accuracy and flexibility when combined with knowledge verification; Non-intrusive authentication; Incremental authentication-waits for more voice/knowledge data when higher degree of recognition confidence is needed; Continuous authentication, maybe embedded in natural dialogue; Background authentication; Public acceptance; and Inexpensive hardware, suitable for pervasive security management. The disadvantages of voice recognition are: Performance degrades under severe environmental noise; Lack of public awareness; and Not robust enough to determine an identity by itself as is vulnerable to tape recorders mimicry by humans [Malt et al, 2003].

When using voice authentication error mainly occur due to the following factors like age, sickness, acoustic, misread/misspoken utterance, emotional states and placement or distance to microphone or use of different microphone.

b. Signature

Signature dynamics is how a personal signature is generated what features it holds. Geometry, curvature and shape information of words and characters are all features provided by signature itself, while pressure metrics, stroke direction, speed and pen up and pen down events say something about how signature was generated [Orlan et al 2003]. Signature verification may be divided into two groups. (i) Off-line signature verification: These are signatures that only have a static visual record, such as signature on traditional paper, painting etc, often written with ink and (ii). On-line/digitised signature verification: This signature is where pen trajectory and /or dynamics are captured by electronic devices and digitised.

Transformation and atomisation of off-line signatures to digitised media is a complex process, and hence a reliable verification of the signature is not possible. The verification of the on-line signature is on the other hand feasible, and is more and more used authentication in the business world. The signature has weaknesses. It is mostly used for one-to-one verification, different signatures collected from the same person may vary, in shape and feature, shape and weight of the pen, surface on which the signature is written, personal and emotional factors at time of signing, and the signing is routine or not.

c. Keystroke

The keystroke dynamics is to identify users based on his/her typing techniques using traditional pattern recognition and neural network techniques [Bolle et al 2001]. One of the advantages of keystroke dynamics compared to signature is that no traditional equipment is required. Keystroke dynamics recognition system can either be used for single authentication or for continuous monitoring. For single authentication the user typically is required to type a phrase as he/she normally would do, and the software compares this provided template with the one previously stored for this user. In a continuously monitoring system, the software monitors the keystroke dynamics detected on the keyboard. If the user for example left his working station unattended and another person started using the computer (typing on the keyboard), the system could immediately recognise this as a different user, lock the system and ask for re-authentication.

One of the purposes of using the keystroke dynamics for authentication is to make passwords more secure. Because keystroke dynamics require the user to type the password in a certain way, with regard to speed, hold time, press and release pattern among others. It would be more difficult for an impostor to falsely authenticate to the system, even if he/she knows the password.

3.1. Esoteric Biometrics

a. Facial Thermographs

Facial thermography biometric recognition uses cameras sensitive in the infrared spectrum to recognise patterns of facial heat. Facial heat is caused by the blood flow under the skin, and makes a distinct pattern. Facial thermograms yields the same blood vessel pathways that are the underlying vein and tissue structure, but the dynamics nature

of the blood flow causes fluctuation due to environment conditions such as variation in temperature, ingestion of alcohol, drugs and cigarette smoke.

Facial thermography has special feature that other biometric characteristics cannot provide, the image can tell if the person is present or absent, alive or death, attentive or not attentive, physically rested or fatigued, relaxed or anxious [Orlan et al 2003]. It functions from either short distance. It is in place in several airport terminals and at any border crossing to help determine the identity of individuals at a distance who may be involved in criminal activities without alerting the individual that they being monitored.

b. Deoxyribonucleic Acid (DNA)

This is a molecule that carries the genetic information necessary for organisation and functioning of most living cells and control the inheritance of characteristics. DNA is a way of biometric characteristics, but differs from standard biometric characteristics in the following ways: DNA requires a tangible physical sample as opposed to an impression, image or recording. DNA matching is not done in real-time and currently not all stages of comparisons are automated. DNA matching does not employ template or feature extraction, but rather represent the comparison of actual samples.

c. Palm Vein Pattern

This system uses infrared beam to penetrate the user's hand as it is waved over the system. It relies on using a special camera together with an infrared light. The camera captures the image of the vascular pattern made by the blood vessels everyone has on the back of their hand. These patterns are developed at the foetus stage; differ even between identical twins except from overall size, consistent throughout life. It has high level of authentication accuracy due to the complexity of vein pattern.

4. Integration of Biometric with Cryptography

4.1. Attacks on biometrics system

Authentication with biometric has advantages over password because biometric cannot be stolen, lost, forgotten, lent or forged and is always available, always at hand to speak (Fadia 2003). However biometric have problems such as: Are relatively new and some people may find their use intrusive (resisting), Devices are costly, Readers use sampling and establish a threshold for match is close enough to accept (variation reduces accuracy), can become sing point for failure, False reading, speed limits accuracy, and forgeries. These attack areas are (i). Type one attacks: involves presenting fake biometric e.g. fingerprint made from silicon, face mask, fake iris texture to the sensor, (ii). Type Two attacks: Replay attack-because an interpreted biometric is submitted to the feature extractor, bypassing the sensor, (iii). Type three attacks: The feature extractor module is replaced with a Trojan horse program that functions according to its designer's specification, (iv). Type four attacks: Genuine feature values are replaced with values selected by the attacker, (v). Type five attacks: the matcher is replaced with Trojan horse, (vi). Type six attack: template database attack e.g. addition, modification, or removal of

template, (vii). Type seven attacks: templates are tampered with i.e. stolen, replaced or altered in transition medium between the template and the database, and (viii). Type eight attack: The matcher result (accept or reject) can be overridden by the attacker.

Biometric security is a fast growing area of computer security [O'Brien 2004]. Vulnerability of information systems is increasing as we move towards a networked computing [Turban et al 2000]. The need for security is now increasing because organisations are now recognising as corporate data is increasing being stored on computer and acceptance that any loss or unavailability of these data could prove to be disastrous [Cannony and Begg 2004]. To enhance biometric security, encryption techniques are examined in the next section.

4.2. Biometric Encryption

Cryptography is the science and art to transform message to make them secure and immune to attacks [Behrouz 2006]. Encryption is the encoding of the data by special algorithm that renders the data unreadable by any program without decryption key. Biometric encryption authentication is a strong, certainly far superior to passwords. The technology is mature, products exist standards define product's interface, reliability rate are acceptable and costs are reasonable [Pfleeger and Pfleeger 2007]. The cryptographic techniques those are available like public key infrastructure, hash algorithm, message encryption and Kerberos. These though have problems according to Simson [1997] such as: Private keys are not people; distinguished names are not people; there are too many Robert Smith; Today's digital certificate doesn't tell much; and X.509V3 does not allow selective disclosure.

4.3. Survey Results Analysis

Various studies and survey have been done as examined in the previous section in this paper. These studies however are from the developed countries with that have the latest technologies. In less developed countries like Kenya and others in Africa, the issue of biometric security is different. In a survey we did at Kabarak University Kenya reveals new challenges facing biometric application in less developed countries. Most active users of the electronic are between the age of 21 years and 30 years as shown on figure. This age bracket is aggressive in having an authorised access.

Figure 1: Demographic Data

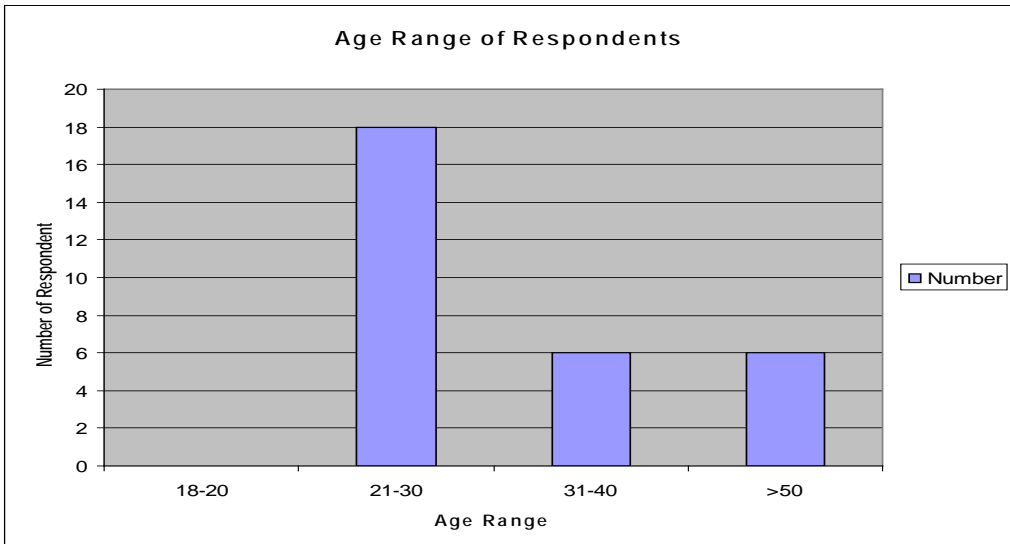
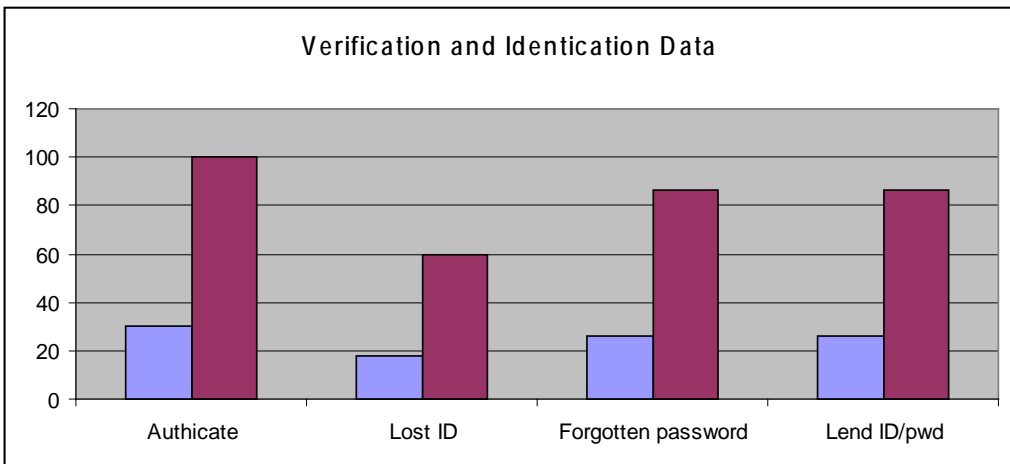
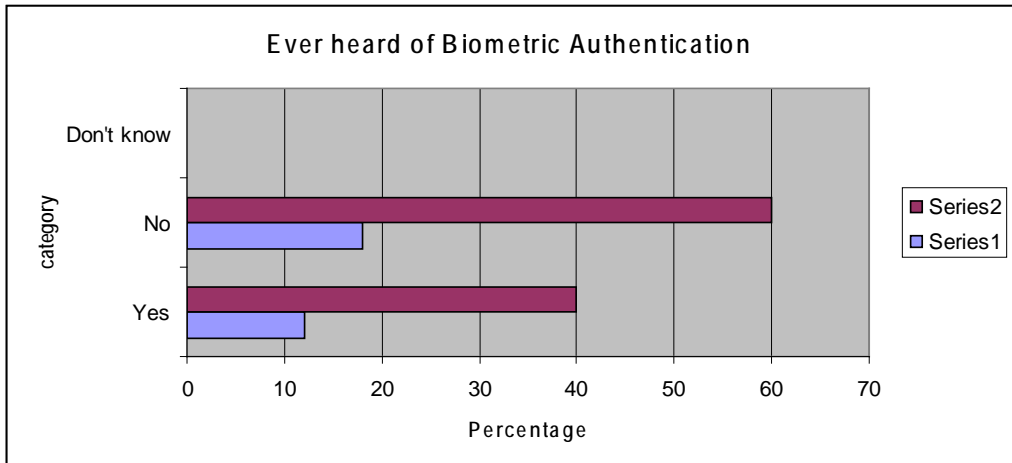


Figure 2: Authentication and Passwords



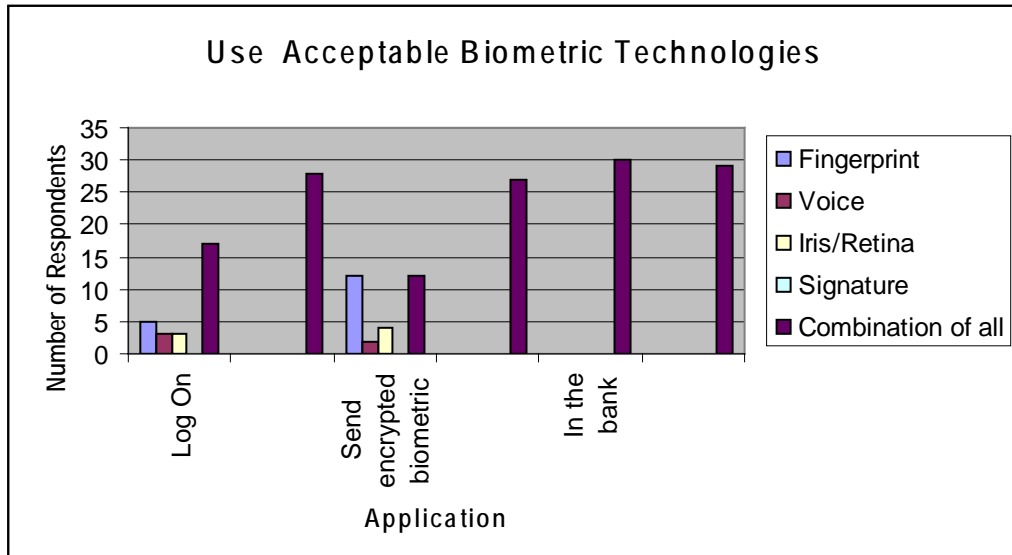
The users of the electronic resources at Kabarak University have to be authenticated before they can access the resources. This is probably due to the password requirement the system administrators. Figure 2 show that 100% of these users must supply their username and passwords before they can access the computer resources, of this 60% had lost their passwords, 84% have ever forgotten the password and similar percentage had given out their password. This presentation is shows clearly that the use of password is a threat to electronic resources.

Figure 3: Participants with Biometric Authentication Information



It is interesting to note that while no biometric system had been installed in the institution before this survey, 40 percent of the respondent had a theoretical understanding of biometric technologies as show in figure 3. The 60%, who had no idea about, are like to pause danger to authentication process. However, when asked which biometric they will prefer to use combined technologies as shown in figure 4.

Figure 4: User Acceptance Level



The survey done shows that security is of great concern for most of the electronic user. What the less developed countries need is brainstorming, availing the information. From 4 there are various reservations on biometric technologies. The signature and fingerprints have low acceptance level compared combined technologies. What the users need is system the assures the of the security of their data.

5. Conclusion

Modern Information Technology revolution that is heavily networked requires authentication that is the binding of an identity to the principal. The password which is the most commonly used technology for authentication has weaknesses. These weaknesses can be solved by biometric technologies such as fingerprints, hand geometry, facial thermographs, face recognition and DNA. The survey done also shows that there is need for combined biometric technologies. Though biometrics has problems; they can be enhanced by cryptographic techniques. An integration of biometrics with encryption provides an enhanced secure authentication system for our electronic resources. In less developed countries like Kenya and other African are willing to implement new authentication system. What are needed in the less developed countries are the willingness and knowledge availabilities.

References

- ANKIT FADIA (2003). Network Security A Hacker's Perspective. Macmillan. New Delhi.
- BOLLE R., JAIN A., PANKANTI S. O. (2001). Biometric-Personal Identification in Networked Society. Kluwer academic publication.
- THOMAS CANNONY AND CAROLYN BEGG (4th ed) (2004). Database System a Practical approach to design, implementation and Management. Addison Wesley. New York.
- CHARLES P. PFLEEGER AND SHARI LAWRENCE PFLEEGER (4th ed) (2007). Security in Computing. Pearson Education. India.
- DOUGLAS E. COMER [2004]. Computer Networks with Application. Pearson Education. India
- STEPHEN DOYLE (3rd Ed) (2001). Information System for you. Nelson Thornes Ltd. Uk.
- STEPHEN HAAG, MAERE CUMMINGS, AND DONALD J. MCCUBBREY (3rd Ed)(2002). Management for Information Age. McGraw-Hill Irwin. Boston
- CHARLIE KAUFMAN, RADIA PERMAN and MIKE SPECINER (2nd Ed) (2002). Network
- KENNETH C. LAUDON AND JANE P. LAUDON (9th Ed) (2006). Management Information System Managing the Digital Firm. Printice-Hall New Delhi
- STREBE MATHEW (2002). Network Security Jumpstart. San Francisco. London.
- MATT BISHOP (2003). Computer Security Art and Science. Pearson Education. India.
- SIMSON GARFINKEL (1997). Web Security and Commerce. O'Relly and Associate. USA
- TANENBAAUN S.ANDREW (4th Ed) (2003). Computer Networks. Prentice Hall. India.
- PARNEL TERE AND CHRISTOPHER NULL (1997). Network Adminstrator's reference. McGraw-Hill. Osborne.
- MCLEAN,TURBAN AND WETHERBE (2nd ed) (2003). Information Technology fo management. Jonh Wiley and sons. New york.
- STALLINGS WILLIAM (2000). Network Security Essentials Application and Standards. Pearson Education. Singapore.
- WOODWARD JR. ORLAN AND N M. HIGGINGS P. T. (2003). Biometrics. McGraw-Hill Osborne. California.

2

A Requirements Analysis Framework for Human Activity Systems (HAS): The Case of Online Learning

Philip O Ayoo and Jude T Lubega

The task of designing information systems is clearly interdisciplinary, since it requires domain knowledge in business process development within the social environment, and the processes management of technological applications. This paper explains e-learning as a human activity system, which requires soft methodologies that deal with the analysis of evolving and ill-defined needs, as well as traditional hard approaches to the design of physical solutions to meet those needs. Consequently, a requirements analysis framework is proposed for constructivist online learning systems in which soft systems thinking is integrated as the essential strategy of requirements elicitation and analysis. This framework is capable of capturing both the formal and informal, as well as the hard and soft aspects of the requirements within a social environment.

1. Introduction

Requirements drive almost every activity, task, and deliverable in a software development project [McEwen, 2004], since accurate requirements form the most essential part of the formula for successful software product development. On surveys of failed projects, the one cause of failure that shows up most often is having missed or poorly defined requirements. Consequently, poor requirements are the biggest problem that end up increasing the budgeted engineering cost and not matching with the expected quality. Without good quality of specifications the people charged with developments will have no firm idea of the needs of the would-be users of the systems [Liu, 1992].

According to Rose [2002], the software engineering approach to information systems development (ISD), in association with structured methods, and the more recent generation of object-oriented development methods, remain the most significant influence on contemporary systems development. Many other influences (prototyping, RAD, extreme programming) share the same primarily technical standpoint. However, given the nature of the IS discipline, a technical standpoint is reasonable and inevitable, but many commentators have pointed out that it tends to exclude or minimise many other important factors: social, cultural, political, semantic, managerial and so on [Rose, 2002]. For this reason, a lot of research is currently on-going to better understand the social aspects of IS development.

2. Purpose of Requirements Analysis

Systematic requirements analysis is also known as *requirements engineering*, and is sometimes referred to loosely by names such as *requirements gathering*, *requirements capture*, or *requirements specification* [McConnell, 1996]. The term can also be applied specifically to the analysis proper (as opposed to elicitation or documentation of the requirements).

A requirement describes what a system should do. According to *Technology Blueprint*, requirements are enumerated specifications that list characteristics that identify the accomplishment levels needed to achieve specific objectives for a given set of conditions. Requirements include business goals, objectives, processes, and all other business and system requirements whose purpose is to alter the “*as is*” view of the world in some way [Bleistein, *et al.*, 2006]. A requirement is a vital tool to help clarify client needs, the business objectives, and methods required to fulfilling that need. However, requirements must be actionable, measurable, testable, related to identified business needs or opportunities, and defined to a level of detail sufficient for system design.

Some people separate “requirements analysis” from “analysis” [The Object Agency, 1995]. Within this categorization, “requirements analysis” involves the establishment of system and/or project requirements, and “analysis” (e.g. structured analysis and object-oriented analysis) involves the proposal of a solution (going no deeper than the user interface) that satisfies the established requirements. According to *Merriam-Webster Dictionary*, analysis is the separation of a thing into the parts or elements of which it is composed; it is the examination of a thing to determine its parts or elements, including a statement showing the results of such an examination. Analysis often begins with the recognition of a need/want, and goes no deeper (in detail) than the “user interface” of the delivered product [The Object Agency, 1995]. Within this perspective, therefore, requirements analysis is the investigation of a problem that focuses on what functionality is required but not on how to provide that functionality. Beyond this (i.e. providing functionality) is design, which starts at the “user interface,” and defines the internal software architecture for the delivered product.

Sudhakar [2005] identifies the types of requirements needed for capturing all aspects of software. They are:

1. *Business Vision and Business Requirements*: These are described as the “why” requirements. They capture the high level objectives of the organization or customer requesting the system or the product. They can be considered as “why we are doing this” requirements and should capture the fundamental reason for product existence.
2. *Functional Requirements*: Functional requirements can be considered as the “what” requirements. These capture the functionality that must be built into the system to satisfy business requirements. Use-cases are a great way to document functional requirements.
3. *Non-Functional Requirements*: They are the “how” or “how well” requirements. They capture the technology specific and “-ility” requirements of the system like compatibility, usability, availability, performance, reliability, etc.

4. *UI Prototypes*: UI prototyping helps in reducing risk, decreasing system size and complexity, and reducing requirements change. The UI prototypes (or screenshots) are the most effective way of communicating requirements to the development team, as they provide a great visualization tool for both development and test engineers to understand the exact software flow.

The requirements analysis for the construction of requirements specifications is the key activity for successful developments of systems [Liu, 1992; Wieggers, 2003; Vat, 2005]. Without good quality of specifications the people charged with developments will have no firm idea of the needs of the would-be users of the systems. Naik [2004] observes that getting the requirements right in the first place costs 50 to 200 times less than correcting code. According to this observation, each hour spent on review avoids an average of 33 hours of maintenance. Ideally, therefore, any project over two weeks duration should necessarily include the requirements analysis process [Naik, 2004].

The ultimate aim of understanding end-users and their requirements is to develop a product that is both useful and usable, the combination of which determine the usability of a system [Löwgren 1995, cited in Neale *et al.*, 2007]. According to the International Organisation for Standardisation (ISO), usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. Indeed, users have been shown to be more satisfied with system design arising from thorough requirements analysis [Neale *et al.*, 2007], and lack of user input has been cited as one of the major reasons for the failure of system development (Standish Group, 1994). To prevent this disaster, the clients have to be involved in the whole process of the developments of systems for consultation by the developers [Liu, 1992].

3. Classifying and Documenting Requirements

Requirements analysis can be a long and arduous process during which many delicate psychological skills are involved [Wikipedia]. Conceptually, requirements analysis includes three types of activity:

1. *Eliciting requirements*: the task of communicating with customers and users to determine what their requirements are. This is sometimes also called requirements gathering;
2. *Analyzing requirements*: determining whether the stated requirements are unclear, incomplete, ambiguous, or contradictory, and then resolving these issues;
3. *Recording requirements*: Requirements may be documented in various forms, such as natural-language documents, use cases, user stories, or process specifications.

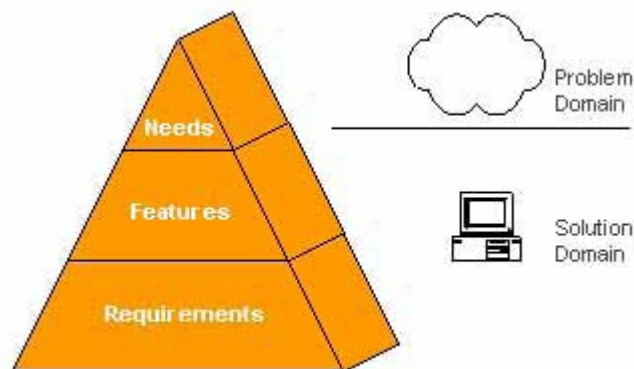
As observed by McEwen [2004], requirements are not requirements unless they are written down. Sudhakar [2005] states that the most critical factor of successful software product development is the right requirements documentation. Taking the time at the beginning of a project to define and document needs, features, and requirements enables one to establish traceability to ensure that software requirements specification aligns with business objectives and continues to do so throughout the project [McEwen,

2004]. Poor requirements are the biggest problem that end up increasing the budgeted engineering cost and not matching with the expected quality.

According to Pohl [1993] (cited in Williams and Kennedy, 2006), the term *requirements engineering* is used to describe a systematic process of developing requirements through an iterative co-operative process of analysing the problem, documenting the resulting observations in a variety of representation formats, and checking the accuracy of the understanding gained. McEwen [2004] explains three separate documents within which requirements are usually captured (as shown in **figure 1**):

1. Stakeholder needs
2. Software features
3. Software requirements specification

Figure 1: Requirements Categories [McEwen, 2004]



Stakeholder Needs

Stakeholder needs, which are part of the problem domain, describe what stakeholders require for a successful project. In other words, *needs describe what the application should do to help improve or lower the cost of a business process, increase revenue, or meet regulatory or other obligations*. Documenting stakeholder needs involves identifying, understanding, and representing different viewpoints. Often, users and stakeholders do not know how to solve the entire problem but are experts at explaining what they need to do their job better. Each stakeholder sees the problem from a different perspective. Therefore, one must understand the needs of *all* stakeholders in order to understand the entire problem domain.

Software Features

After defining stakeholder needs, they must be translated into a set of distinct system features. Needs do not indicate a particular solution; they simply describe the business need. It is perfectly fine for stakeholders to express themselves in any way they wish; often, an analyst will want to ask additional questions to clearly understand both needs and features. Needs are part of the *problem domain*, and features are part of the *solution domain*. It is critically important to fully understand the problem domain before deciding

on a solution; often, an analyst will find opportunities to generalize the solution once they fully understand the problem.

Software Requirements Specification

A requirement is a software capability that must be met or possessed by a system or a system component to satisfy a contract, standard, or desired feature. Requirements may be classified into functional and non-functional requirements. Functional requirements present a *complete* description of how the system will function from the user's perspective. They should allow both business stakeholders and technical people to walk through the system and see every aspect of how it should work before it is built. Non-functional requirements, in contrast, dictate properties and impose constraints on the project or system. They specify *attributes* of the system, rather than what the system will do.

4. Traditional Requirements Analysis Techniques

During requirements analysis energy should be directed towards ensuring that the final system or product conforms to client needs rather than attempting to mould user expectations to fit the requirements [Liu, 1992]. A trained software practitioner called the "requirements analyst" communicates with knowledgeable user(s) to understand what the requirements are. Analysts can employ several techniques to elicit the requirements from the customer. Historically, this has included such things as holding interviews, or holding focus groups (sometimes referred to as requirements workshops) and creating requirements lists [Wikipedia]. More modern techniques include prototyping, and use-cases. Where necessary, the analyst will employ a combination of these methods to establish the exact requirements of the stakeholders, so that a system that meets the business needs is produced.

Noting that the quality of requirements specifications directly affects the quality of the developed systems, Liu [1992] and Urquhart [2001] outline some of the more commonly used techniques in requirements analysis:

4.1 Structured System Analysis Methodology

Two different sorts of methods for system analysis have been developed under the name of structured system analysis methodology: *informal methods* and *formal methods*. The specifications produced using *informal methods* (such as Structured Analysis Method) are comprehensible due to the notations such as diagrams, graphics and natural language, etc, but lack the formality. *Formal methods* try to overcome this weakness by adopting mathematical notations. The specifications produced using formal methods are concise and precise in general, but lack comprehensibility. Therefore, there is a great need for developing a language and a method for producing comprehensible and precise requirements specifications.

4.2 DeMarco's Approach

Tom DeMarco first proposed the model of data flow diagrams to be a method and language for requirements analysis. A data flow diagram shows the connections between processes, data flows and files. DeMarco's data flow diagram provides the hierarchy mechanism to fit into complex requirements analysis. The hierarchical structure in data

flow diagrams is constructed by decomposing each upper level bubble into a sub data flow diagram which reflects the details of how to transform their input data flows into their output data flows. The function of each bottom level process is usually specified using structured English. In addition to the data flow diagrams, a data dictionary is usually designed for defining data flows, components of data flows, files and processes occurring in the data flow diagrams.

4.3 JSD

JSD, short for Jackson Software Development, is a systematic method for specifying and implementing computer systems, especially for information systems. JSD starts with the description of the model of the real world and then specifies the functions of the system to be developed based on the model of the real world. The implementation step is therefore centrally concerned with transforming the specification to make it convenient to execute. To be clear and understandable, JSD provides a series of diagrams for the development steps.

4.4 SSADM

Structured Systems Analysis and Design Methodology (SSADM) breaks down the work into phases, which are then divided into stages. The three phases of SSADM are: feasibility, analysis, and design. The feasibility phase consists of two stages: problem definition and project definition. These stages assess the scale of the problem and the costs and likelihood of improving the situation. The analysis phase includes three stages: analysis of the current system, specification of the required system, and select service level for new system. The design phase consists of three stages: detailed data design, detailed process design, and physical design control. Initially design is done at a logical level, and then converted to a physical design. In a SSADM project, data flow diagrams (DFDs) are used to describe the existing physical system, the logical equivalent of the existing system, and the required system, while logical data structuring technique (LDST) and entity life histories (ELHs) diagrams are employed to describe the data requirements.

4.5 Prototyping

In the mid-1980s, *prototyping* was seen as the solution to the requirements analysis problem. Prototypes are mock-ups of an application that allow users to visualize an application that has not yet been constructed. The technology of rapid prototyping presents an effective approach of portraying functionality. But its power is limited as rapid prototypes are produced based on the informal requirements specifications of the proposed systems. The possible misunderstanding of the specifications during the rapid prototyping often affects the quality of the rapid prototypes and therefore a lot of modifications for the rapid prototypes are required in order to fit clients' true requirements. Eventually the rapid prototyping is not rapid and the cost is not so low.

4.6 Use Cases

The “use case” is introduced as a more formal method for requirements capture and initial planning. Use cases typically avoid technical jargon, preferring instead the language of the end user or *domain expert*. According to McEwen [2004], organizations are rapidly adopting use cases as a means to communicate requirements because they:

1. are easier to create, read, and understand than traditional functional specifications;
2. show how the system will work from the users’ perspective rather than the system’s perspective;
3. force us to think about the end-game (what is the user trying to accomplish by using the system?);
4. require us to define how the system should work, step-by-step;
5. Provide an excellent basis for building test cases and helping to ensure that these are built before the code is written;
6. provide a common requirements “language” that is easy for stakeholders, users, analysts, architects, programmers, and testers to understand.

Fundamentally, use cases provide a way of depicting and structuring the interaction of a real world Actor (person, organisation or external system) with the application being studied. Use cases do not describe any internal workings of the system, nor do they explain how that system will be implemented, but simply show the steps that a user follows to perform a task. Use cases are often co-authored by requirements engineers and stakeholders.

4.7 Object-Oriented Requirements Analysis

Object-oriented requirements are requirements where the information (content) is localized around, and expressed in terms of, objects and the interactions and interrelationships among these objects [The Object Agency, 1995]. Object-oriented requirements analysis is a process whereby two important activities occur: one, we study user (customer) wants/needs to arrive at an object-oriented definition of the implicit or explicit criteria that must, should, or might be met by any delivered solution to these wants/needs; and then we define project needs, priorities, and constraints (preferably in an object-oriented manner), and define an object-oriented solution to the user (customer) wants/needs that, preferably, goes no deeper in detail than the user interface(s).

Object-oriented analysis aims to overcome some of the problems of the structured methodologies (such as SSADM), which can become a rigid straitjacket on systems development [Urquhart, 2001]. The high degree of consistency within object-oriented approaches, coupled with the recursive/parallel life-cycle, greatly blurs the distinction between what has been traditionally thought of as “analysis,” and what has been traditionally thought of as “design.”

4.8 CRC Cards

One similar structured approach to user involvement in the process of object-oriented development is the use of Class, Responsibility and Collaboration (CRC) cards. The

technique uses small teams, including users who are “domain experts”, with the aim of discovering the objects, the classes that the system comprises, their responsibilities (what they do) and their collaboration, how they interact with other objects in the system. Brainstorming is used to derive provisional classes and behaviour and role-playing is then used to work through some scenarios, “living the system”. The technique is aimed at a better, deeper understanding of the problem situation so that the information system developed provides a proper solution.

5. Problems With Traditional RA Techniques

The software crisis of the 1960s embodied a growing realisation that systems took too long to build, cost too much and did not work too well [Lycett, 1999]. According to a survey by the *Standish Group* [McEwen, 2004], of more than 352 companies reporting on more than 8,000 software projects: over 40% to 60% of defects in software are caused by poor requirements definition; 31% of all software projects are cancelled before they are completed (a waste of \$81 billion); customers do not use 20% of their product features; while 53% of projects cost 189% of their original estimate. The survey also revealed that in large companies, only 9% of projects are on time and within budget; in small companies, 16% of projects are on time and within budget; while the average time overrun is 222% of the original estimate [Williams and Kennedy, 2006].

According to Williams and Kennedy [2006], systems failure has been blamed on poor requirements engineering process, resulting to poor understanding of domain knowledge and poor use of methods, techniques and tools. Most requirements analysis writings either advise general guidelines which are short of any concrete operational aspects, or they are not simply at the scale of complex open systems construction [Bastani, 2007]. The Standish Group survey also asked respondents to identify the causes of these failures. Table 1 below shows the top three reasons why projects are “impaired.”

Table 1: Top Three Project Impairment Factors (Standish Group)

Project impairment factors	% of Responses
Lack of user input	12.8%
Incomplete requirements and specifications	12.3%
Changing requirements and specifications	11.8%

As this table shows, poor requirements are the biggest problem. McEwen [2004] observes:

If it isn’t clear what you are supposed to build, how can you estimate the cost of building it? How can you create a project plan, assign resources, design system components, or create work orders? You need accurate requirements to perform these

activities. Of course, requirements evolve as a project proceeds, but carefully worded basic requirements provide a starting point. Then, as the project progresses, you can fill in details and update planning documents as the requirements evolve.

In their study, Williams and Kennedy [2006] established that errors in information systems have the following distribution: incomplete requirements 56%, design 27%, coding 7% and other 10%. The high error percentage due to incomplete requirements is attributed to the poor methods used to elicit and analyse requirements. Further evidence that errors due to incomplete requirements analysis take a disproportionate larger effort share has been provided: incomplete requirements 82%, design 13%, coding 1% and others 4% (Williams and Kennedy, 2006).

In *Software Project Survival Guide*, Steve McConnell [McConnell, 1996] comments that “fixing upstream defects downstream can cost 50 to 200 times as much as fixing them upstream”. This is compelling motivation for taking the time to plan, and is a pretty logical statement when one considers that, during requirements analysis (an upstream activity), a feature is described using a handful of words. Once that handful of words makes it to code (a downstream activity), it can be hundreds or thousands of lines of program code.

Although substantial progress has been made in terms of methods, techniques and tools used within the requirements engineering (RE) phase of systems development, little attention has been paid to the understanding of the RE process effectiveness itself [Williams and Kennedy, 2006]. The designers of information systems (IS) and programmers often begin designing and programming the incumbent system too early, before they actually understand the users’ or stakeholders’ requirements.

In recognition of this defective methodology, the Garmisch Conference [1968] recommended the application of a systematic, disciplined and quantifiable approach to the development, operation and maintenance of software (the discipline of software engineering). This provided a somewhat technical orientation to information systems development, which also later demonstrated the inadequacy of computer science to address socially based problems associated with the use of computers in an organization or business context [Lycett, 1999]). For the future system to be effective it has to have a balance between the technical worldview of designers and programmers, and the social worldview of users and customers [Urquhart, 2001; Castro, *et al.*, 2002; Williams and Kennedy, 2006].

Checkland and Holwell [1998] argue that information systems are more than “data manipulation” systems and that the study of the human activity systems is vital to an understanding of what an effective information system might look like. A human activity system (HAS) is an assembly of people and other resources organized into a whole in order to accomplish a purpose [Nardi, 1996; Banathy, undated]. The people in the system are affected by being in the system and, by their participation in the system, they affect the system. According to Vat [2005], IS developers must therefore understand the intertwined regularities and idiosyncrasies of human activities and create systems that support the right moves and degrees of agility at the right times and places and for specific purposes.

6. Systems Thinking

The design of IS must therefore be examined more closely from a systems perspective. Checkland [1981] draws attention to two alternative paradigms to explain the nature and significance of systems thinking. In the first paradigm, the world is considered to be systemic and is studied systematically; while in the second paradigm, the world is problematic, that is, it admits to many different interpretations and we study it systemically. Following this continuum, the first paradigm reflects the notion of hard systems thinking, while the second paradigm reflects the notion of soft systems thinking.

Hard systems thinking refers to systems engineering thinking where a systematic process of problem solving is followed [Goede, 2005]. Within this thinking, information systems development is traditionally seen as a hard approach, where stages of a lifecycle can be identified to simplify the development process. As an example, the system development lifecycle (SDLC) for traditional information systems, can be classified as a hard systems approach to systems development. Typical phases of the SDLC according to the ‘waterfall’ model [Royce, 1970] include requirements analysis, system and software design, implementation and unit testing, systems testing, and operation and maintenance. In this mode, user participation is normally restricted to the first phase and testing is done according to user specifications.

During the analysis phase of a software development effort, however, the analyst(s) must accomplish two goals: first, achieve an adequate, complete, and verifiable understanding of the problem at hand, i.e., the problem that the new/modified system is supposed to solve (about 60% of the analysis effort according to researchers); and second, define a solution to this problem, i.e., describe how the new/modified system will appear to the user(s). Hard systems rely more on the technical view in the design of computer-based information systems, and this approach usually leads to failure in such projects [Goede, 2005]. Jackson [1997] [in Cropley and Cook, 2005] identifies, among the reasons for failure of information technology systems, the fact that most of them give inadequate attention to the human and organizational factors which are vital in determining the ultimate effectiveness of new systems. Cropley and Cook [2005] emphasize that the lack of success to date, results from the limited scope of systems engineering methodology as currently applied and the inability of hard systems engineering, by itself, to adequately deal with this complex systems domain.

Action research carried out at Lancaster revealed that traditional approaches from systems engineering were not suitable to all environments, particularly those experiencing some kind of “problem” which could not easily be defined [Bond and Kirkham, 1999]. Since the purpose of information systems is to facilitate and mediate social interaction, the social aspects are a crucial element in their development and use [Lycett, 1999]. Traditional hard techniques are therefore of little use in the initial analysis of a human activity system, which is likely to have problem areas that are not clearly defined, and which are unstructured [Lehane and Paul, 1996; Williams and Kennedy, 2006].

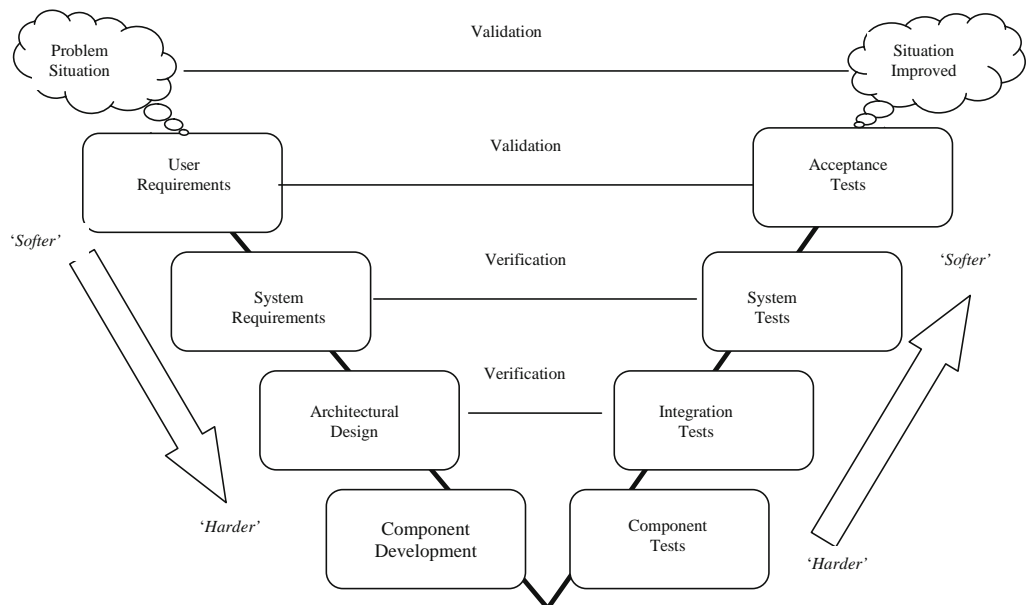
Soft systems methodology (SSM) was therefore developed to provide a structured methodology, which could be used to explore such problems, initially as a precursor to

information systems design. SSM is a general problem structuring approach that seeks to incorporate multiple stakeholder views in the analysis of a given problem [Moore and Gregory, 2000]. When applied to IS development, the method requires negotiation and debate between the stakeholders when exploring the feasibility of developing an information system.

Effective requirements engineering, therefore, is most effectively seen as a form of heterogeneous engineering in which technical, social, economic and institutional factors are brought together in a current solution space that provides the baseline for construction of proposed new solution spaces [Bergman, et al., 2002]. The rationale for the suggested methodological pluralism is to maximize our understanding in elicitation of domain knowledge that can be specified and validated to improve the effectiveness of the RE process.

Thus, according to Hutton [2006], it is sensible that a soft systems analysis precedes the definition of the user requirements and the transition into the hard world of solutions. In this respect, Hutton [2006] recommends the introduction of an additional layer to the familiar ‘V’ model of systems development, with a transition from soft to hard thinking as the problem situation is refined into a statement of requirements, and then from hard to soft again as we ensure the system delivers the requirements within the operational context and does improve the situation. This is illustrated on figure 6 below.

Figure 2: Adding another Layer to the Systems Engineering V-Model [Adapted from Arnold *et al*, 1998]



This model shows that SSM can be applied as an additional layer to a systems engineering framework. The root definitions and conceptual modeling encourage stakeholder

involvement and buy-in during the early stages of a project, and provide an abstract representation of what is needed to develop and justify the user requirements. Maintaining the rich picture and the conceptual models provides a vehicle to demonstrate that the system has improved the problem situation, hence another layer to the ‘V’. Eva [2004] believes that the end result in SSM is an understanding of the problem domain so that a hard study can then be applied to specify a solution.

7. E-Learning Requirements Analysis

Learning involves real life situations, and real life involves ill-formulated and ill-structured problems and conditions: real life problems have context, depth, complexity and duration; involve cooperative situations and shared consequences; and are worth solving and can provide benefits when solved [Fitzsimmons, 2001]. In the analysis of any learning environment, whether small or large, one observes within it a set of human activities related to each other so they can be viewed as a whole [Checkland, 1981].

The analysis of learners’ requirements is vital for identifying what they want and what they expect from a learning environment [Sun *et al.*, 2003]. A learning environment can be seen as a complex socio-technical system, since it can be idealised as an open system that depends on the technology, the sentiments of the members, and the organisational environment [Checkland, 1981]. It is also essential for the instructional designer to know learners’ characteristics in order to build a system that maps learners’ preferences.

Although the system is organised to focus on a primary task (the learning) this cannot be separated from the environment and the social factors, including cultural ones [Slay, 2002]. For active, authentic learning, then, activities should be ill-defined, comprise a single complex task to be investigated over a sustained period of time, provide the opportunity to examine the task from different perspectives, promote reflection to enable abstractions to be formed, promote articulation to enable tacit knowledge to be made explicit, and provide coaching and scaffolding by faculty at critical times [Fitzsimmons, 2001].

The design of traditional e-learning environments has been criticized for being technology-driven rather than problem-driven or learner-driven [Tynjala and Hakkinen, 2005]. In designing e-learning systems, however, von Brevern (2004) argues that one needs first to define the psychological (i.e. cognitive) and social rationales, and then incorporate them into any instructional design model. Indeed, von Brevern believes that applying the pedagogical context of e-learning (cognitive, behavioural or social) is key to understanding and planning educational technology systems that are interrelated to “subject heavy domains” or “context specific” environments.

The situated aspects of learning, according to Bell [2006], could be explored better if learning was regarded as taking place within a human activity system (HAS), and therefore, rather than focusing on ‘hard system’ thinking, Peter Checkland’s soft systems methodology (SSM) has been found to be quite appropriate during the initial phases in developing e-learning artefacts. Hutton [2006] argues that SSM would encourage the analyst to explore the problem situation rigorously, rather than to develop a solution to a perceived problem, or even to define a problem that fits the preferred solution.

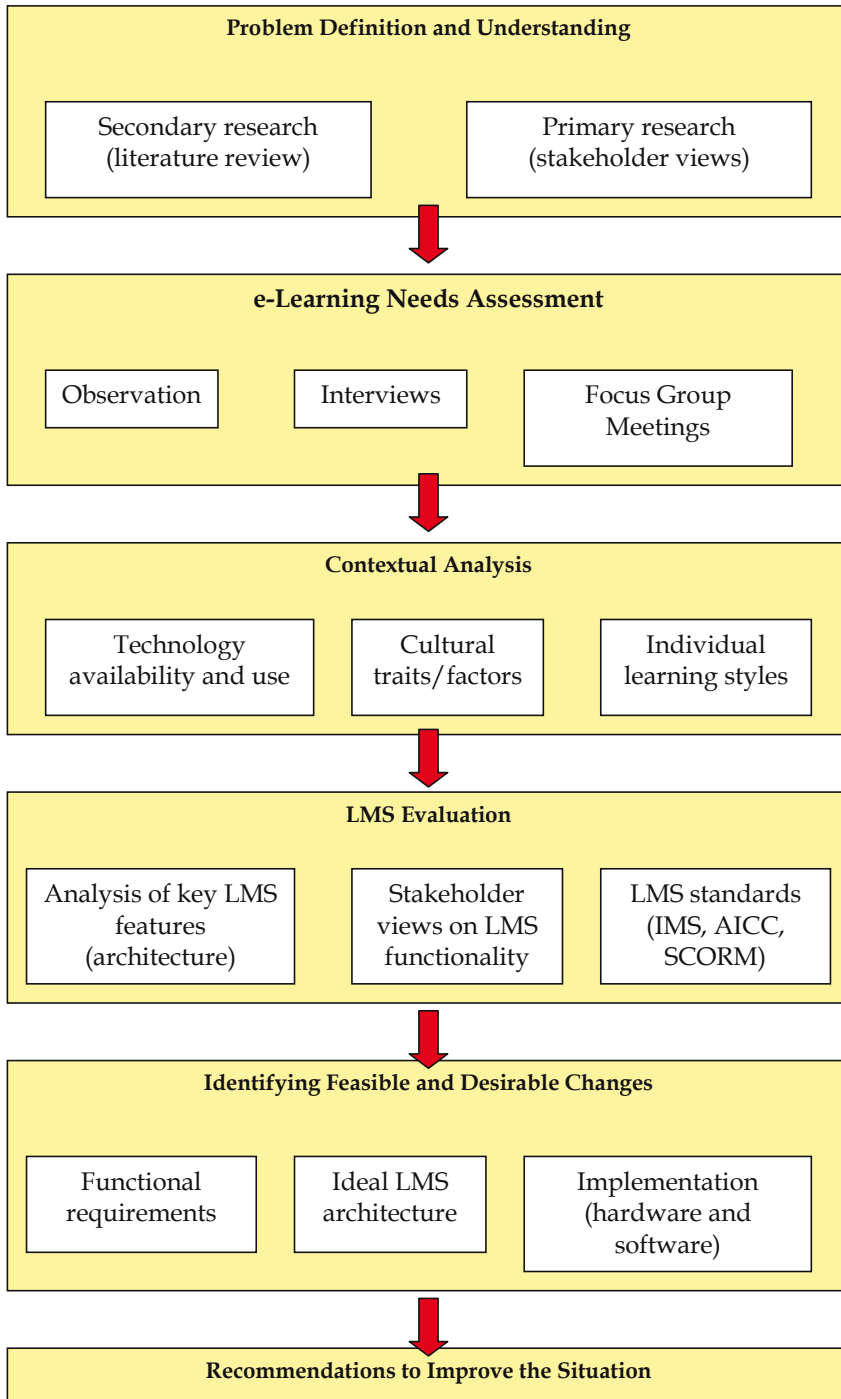
For this reason, e-learning solutions must include features related both to software design issues, and pedagogical aspects [Tynjala and Hakkinen, 2005; Epstein, 2006]. Because software design and pedagogical solutions are inter-dependent, the design process itself has to be carried out as a collaborative process between experts in learning and experts in software design.

8. Proposed Requirements Analysis Framework

From the foregoing discussion, a framework may now be proposed for analysing requirements for a constructivist online learning system. Carey and Carlson [2002] advise that in order to develop a framework that addresses a set of applications in a domain (or domains), one needs to have requirements from each of these applications (for each domain). Requirements are best gained from experts in the domain, for example, from business analysts for business applications (or education/e-learning experts for e-learning requirements). These domain experts do not have to be technically savvy—that is the technical team’s job—but they need to know the applications (or more specifically, the business processes) in the domain (or domains) very well.

A six-step framework has been proposed for this analysis (figure 3), and the process is explained in the sections that follow.

Figure 3: Proposed Requirements Analysis Framework for Constructivist Online Learning System



8.1 Problem Definition and Understanding

Preceding requirements analysis, a comprehensive structured approach is applied in understating the challenges faced and the gaps that currently exist in the provision of personalised e-learning. As observed by Talavera *et al.* [2001], successful technology-based environments design must start with a compelling vision statement where the stakeholders of the project reach an agreement on what problem need to be solved and which are the boundaries and most important features of the system. Ideas from Checkland's 7-stage soft systems methodology (SSM) will be applied in developing this understanding.

First, the researcher must seek to develop a proper understanding of the e-learning environment with respect to the learning environment that they seek to address. Apart from literature review, as many perceptions of the problem situation as possible are captured from a wide range of people to ensure a balanced view of the situation. This is done through consultations (in the form of structured interviews) with education/e-learning experts, LMS system administrators, and learners, especially those with some experience in using the web as a teaching and learning resource. More preferably, this sample will be selected from instructors who have at least shared an online version of a syllabus, posted an instructor profile, or reviewed and critiqued online resources on the web.

8.2 (e-)Learning Needs Assessment

The exact needs of learners in this environment are established, and then the main e-learning platforms that are currently being used to address these needs are identified. Respondents here include learners, education/e-learning experts, instructors, content developers, and LMS system administrators.

A case study approach is employed in developing an understanding of the e-learning needs, and the three East African Partner States of Kenya, Tanzania and Uganda are included. Data is collected from three universities in this region, which are currently experimenting with online learning in a meaningful way – Makerere University (Uganda), University of Dar es Salaam (Tanzania), and University of Nairobi (Kenya). Sampling from the three countries in the region is expected to give a relatively diverse set of views based on the experience of users.

Various strategies are employed when collecting facts, ranging from very informal, unstructured approaches to very formal, structured tools employed in traditional systems analysis. Participant observation, for instance, is used to identify tasks performed (what learning activities are undertaken), identify tools employed (the e-learning systems/platforms used), and establish interactions between people/systems (the extent to which users appreciate the systems). A consensus workshop (in the lines of focus group meetings) is conducted within each of the three universities to examine the various models and attempt to reach an accommodation between the viewpoints reflected. All relevant stakeholders – learners, instructors, education/e-learning experts, content developers, LMS system administrators, etc – are participants in these meetings.

8.3 Contextual Analysis

There must be a deliberate attempt to analyse the environmental and situational contexts within which learning occurs, and which may have some impact on learning – including technology use and challenges (e.g. access to Internet (bandwidth), cultural traits that influence learning (e.g. reading culture), and individual predisposition towards learning (e.g. learning styles). The target group for this category are mainly the learners themselves, supplemented by instructors and LMS system administrators. Over a period of time, a database of the user characteristics and the corresponding preferences would allow for the automatic generation and recommendations of the design features of a constructivist framework. Personalisation rules are established based on the database of the relationship between user characteristics and their preferred design features.

8.4 LMS Evaluation

Before these systems are considered effective, the user experience is studied and analyzed to provide the optimum solution to meet pedagogical needs of both faculty and students. This not only by studying the systems themselves, but also seeking views of users – including learners, instructors, education/e-learning experts, etc. These platforms are studied in detail in order to understand their core features, and how they implement the various learning theories/models as informed by the identified needs. In another study [Ayoo, 2008], it was established that East African universities use mainly an open source system called Moodle for their e-learning activities, while popular commercial products are Blackboard and WebCT. Still others (used occasionally/minimally) include WEDUSOFT, KEWL and A-Tutor.

Constructivist evaluation offers a very explicit methodological framework, grounded in the epistemology of social constructivism, with clear points of convergence and useful pointers to appropriate validation criteria for research [Levy, 2003]. Constructivist evaluation is employed to consider *claims* (assertions favourable to what is being evaluated); *concerns* (unfavourable assertions); and *issues* (items on which reasonable people might disagree). This type of evaluation aims to reach consensus between different stakeholders and the evaluators in organising the negotiation.

Based on the evaluation criterion, both structured and semi-structured questionnaires are used to capture the views of users on the present online learning environments. Various features are taken into account when evaluating the online learning platforms, starting from the function and usability of the overall learning system in the context of the human, social and cultural contexts where they are used, and developing an understanding of how the various features are integrated to facilitate personalised learning. Both pedagogical and technological aspects are carefully examined.

When evaluating learning management systems, one hears a lot of terms that end in “-ity”: high availability, usability, scalability, interoperability, stability and security. Hall [2003] provides a framework within which each of these issues may be examined to establish why they are critical to the function of any enterprise management system.

1. *High availability*: The LMS must be robust enough to serve the diverse needs of thousands of learners, administrators, content builders and instructors simultaneously.
2. *Scalability*: The infrastructure should be able to expand—or “scale”—to meet future growth, both in terms of the volume of instruction and the size of the student body.
3. *Usability*: To support a host of automated and personalized services, such as self-paced and role-specific learning, the access, delivery and presentation of material must be easy-to-use and highly intuitive—like surfing on the Web or shopping on Amazon.com.
4. *Interoperability*: To support content from different sources and multiple vendors’ hardware/software solutions, the LMS should be based on open industry standards for Web deployments (XML, SOAP or AQ) and support the major learning standards (AICC, SCORM, IMS and IEEE).
5. *Stability*: The LMS infrastructure can reliably and effectively manage a large enterprise implementation running 24x7.
6. *Security*: As with any outward-facing collaborative solution, the LMS can selectively limit and control access to online content, resources and back-end functions, both internally and externally, for its diverse user community.

From an operational point of view, the extent to which the LMS and its key components are web-deployable – including content management, user administration and system administration – is established. Further, the evaluation targets the extent to which the LMS is built on an open architecture that supports the emerging learning standards – IMS, AICC and SCORM. These industry-standard groups are creating technical specifications to enable and support a unified, standardized content model for web-based learning.

8.5 Identifying Feasible and Desirable Changes

The focus at this stage is to seek ways of improving the situation, by identifying possible changes in the system structure, system procedures and attitudes of the users. Here, care is taken to ensure that the changes are systemically desirable, that they are as a result of the insight gained from selection of root definitions and conceptual model building, and that they are culturally feasible given the characteristics of the situation, the people in it, their shared experiences and their prejudices.

Stakeholder views are sought on what their expectations are from an ideal/improved environment. The aim is to collect worldviews on the question “What is your view of a pedagogically effective online learning environment?” In applying a systems engineering perspective to e-learning, a number of issues are considered:

1. What are all the functional requirements an e-learning system must fulfill?
2. What should an ideal e-learning architecture look like?
3. How does one pull together all the hardware and software pieces to get the best solution for the identified needs?

8.6 Recommendations to Improve the Situation

The job at this stage is to implement the changes identified and put them into action. This is the stage at which a hard methodology now takes over in ensuring the development of a system to address the problem that will now have been identified and structured.

7. Conclusion

The emphasis in information systems development must be a collaborative effort between technical experts and those who truly understand the purposeful action served. Models of purposeful human activities can be used as scenarios to initiate and structure sensible discussion about information support for the people undertaking the real-world problem situations.

Hard systems engineering methods return to the foreground when the ill-structured problem of the human activity system has been refined to give satisfactory requirements for the realization of the components, in an evolving, complex environment. The central feature of a system methodology for real-time information systems, therefore, calls for the judicious application of hard systems methods to well-defined needs, in the global framework of a soft systems approach. Hard systems engineering must take place in, and be guided by, the nature of the overarching human activity system. As a result, the requirement for a complete systems approach, using both hard and soft concepts, is recommended for e-learning systems.

References

- AYOO, P. O. 2008. Are East African universities ready for e-learning? *Proceedings of the 1st Regional Conference on e-Learning*, Kenyatta University, Nairobi, November 18 – 20 2008.
- BANATHY, B. website. A taste of systemics: characteristics of a human activity system. In Mandel, T. (ed.) *ISSS Integrated Systemic Inquiry Primer Project (ISIPP)*. <http://www.iss.org/primer/bela6.html> (retrieved 05/10/2008)
- BASTANI, B. 2007. **A requirements analysis framework for open systems requirements engineering.** In *ACM SIGSOFT Software Engineering Notes*. Vol. 32, Issue 2, pp. 1 – 19. <http://portal.acm.org/citation.cfm?id=1234753> (retrieved 30/07/2008)
- BENJAMIN, L. K. 1998. *Practical software requirements: a manual of content and style*. Greenwich, CT: Manning Publications Co.
- BERGMAN, M., KING, J. L. AND LYYTINEN, K. 2002. Large-scale requirements analysis revisited: the need for understanding the political ecology of requirements engineering. *Requirements Engineering Journal*, Springer London. Vol. 7, No. 3. <http://www.springerlink.com/content/79xlamrhc5691ex8/> (retrieved 14/05/2008).
- BLEISTEIN, S., *et al.* 2006. B-SCP: A requirements analysis framework for validating strategic alignment of organizational IT based on strategy, context, and process. *Information and Software Technology*, 48 (2006), pp. 846–868.
- CAREY, J. AND CARLSON, B. 2002. Introduction to developing object-oriented frameworks. <http://www.awprofessional.com/articles/article.asp?p=27123&rtl=1> (retrieved 27/07/07)

- CASTRO, J., KOLP, M. AND MYLOPOULOS, J. 2002. Towards requirements-driven information systems engineering: The Tropos Project. <http://citeseer.ist.psu.edu/castro02towards.html> (retrieved 30/07/2008)
- CHECKLAND, P. 1981. *Systems thinking, systems practice*. Chichester, UK: Wiley.
- CHECKLAND, P. AND HOLWELL, S. 1998. *Information, systems and information systems: making sense of the field*. Chichester: Wiley.
- DE CROOCK M. *et al.*, 2002. Active learning for adaptive Internet (Alfanet IST-2001-33288). <ftp://ftp.cordis.europa.eu/pub/ist/docs/ka3/eat/ALFANET.pdf> (retrieved 11/09/2007)
- GREGORY, F. H. 1995. Soft systems models for knowledge elicitation and representation. *J. Opl Res. Soc.*, 46, pp. 562 - 578.
- GREGORY, F. H. 1993. SSM to information systems: a Wittgensteinian approach. *Journal of Information Systems* 3, 149 - 168.
- GREGORY, F. H. 1993. Cause, effect, efficiency and soft systems models. *J. Opl Res. Soc.*, 44, pp. 333 - 344.
- HALL, J. 2003. Assessing learning management systems. http://www.clomedia.com/content/templates/clo_feature.asp?articleid=91 (retrieved 14/05/2008)
- IEEE 1998. IEEE recommended practice for software requirements specifications. *IEEE STD-830*.
- LIU, S. 1992. Formal requirements analysis method based on data flow analysis and rapid prototyping. A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Science.
- LYCETT, M. 1999. The development of component-based evolutionary information systems. A dissertation submitted for the degree of Doctor of Philosophy. Department of Information Systems and Computing at St John's Brunel University.
- MACHADO, M. AND TAO, E. 2007. **Blackboard vs. moodle: comparing user experience of learning management systems**. *Frontiers in Education Conference - Global Engineering: Knowledge without borders, opportunities without passports, 2007 FIE '07 37th annual*. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=4417910&isnumber=4417795 (retrieved 14/05/2008).
- Merriam-Webster Dictionary, 1995.
- McConnell, S. 1996. *Rapid development: taming wild software schedules*, 1st ed., Redmond, WA: Microsoft Press.
- McEwen, S. 2004. Requirements: an introduction. http://www.ibm.com/developerworks/rational/library/4166.html?S_TACT=105AGX78&S_CMP=HP (retrieved 30/07/2008).
- MERCKER, J. 2007 A requirements analysis primer**. <http://www.techlinks.net/CommunityPublishing/tabid/92/articleType/ArticleView/articleId/3849/A-Requirements-Analysis-Primer.aspx> (retrieved 11/09/2007).

- MÖDRITSCHER, F. *et al.*, 2006. The first AdeLE prototype at a glance. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006*. pp. 791-798. <http://coronet.iicm.tugraz.at/denis/pubs/edmedia06b.pdf>.
- MOORES, T. T. AND GREGORY, F. H. 2000. Cultural problems in applying SSM for IS development. *Journal of Global Information Management*, vol. 8, no. 1. http://www.geocities.com/Athens/Oracle/6925/Cultprobs_ssm.htm (retrieved 31/07/2008).
- NG, PAN-WEI 2003. Adopting use cases, part 1: understanding types of use cases and artifacts. http://www.ibm.com/developerworks/rational/library/1809.html?S_TACT=105AGX78&S_CMP=HP (retrieved 30/07/2008).
- NAIK, S. 2004. Software requirements analysis. <http://www.devpapers.com/article/233> (retrieved 11/09/2007).
- NARDI, B., (ed.) 1996. Context and consciousness: activity theory and human-computer interaction. Cambridge, MA: MIT Press.
- NEALE, S. H. *et al.* 2007. Online biodiversity resources – principles for usability. *Biodiversity Informatics*, Vol. 4, pp. 27-36.
- PROTEANS SOFTWARE SOLUTIONS 2005. Best practices of requirements collection. In *Agile RUP for Product Development*. <http://www.proteans.com/CS/Web/blogs/agilerup/archive/2005/07/06/19.aspx> (retrieved 11/09/2007).
- RICHARD, V. 1997. Stakeholders, soft systems and technology: separation and mediation in the analysis of information system requirements. *Information Systems Journal*. 7 (1) , pp. 21–46.
- Slay, J. 2002. Human activity systems: a theoretical framework for designing learning for multicultural settings. *Educational Technology & Society*. 5 (1). http://www.ifets.info/journals/5_1/slay.html (retrieved 05/10/2008).
- SMITH, R. website. How does it all work: playing with technology. http://www.edtechleaders.org/chat_series/presentations/smith.ppt. (Retrieved 14/05/2008).
- SUDHAKAR 2005. Best practices of requirements collection. In *Agile RUP for Product Development*. <http://www.proteans.com/CS/Web/blogs/agilerup/archive/2005/07/06/19.aspx> (retrieved 14/05/2008)
- SUN, L., WILLIAMS, S., OUSMANOU, K. AND LUBEGA, J. 2003. Building personalised functions into dynamic content packaging to support individual learners. In: *Roy Technology Blueprint FAQs*. <http://www.jcapitaltech.com/faqs.php> (retrieved 11/09/2007).
- Technology Blueprint FAQs <http://www.jcapitaltech.com/faqs.php> (retrieved 11/09/2007).
- The Object Agency, Inc. 1995. Defining OO requirements analysis. <http://www.toa.com/pub/00DFOORA.PDF> (retrieved 11/09/2007).
- THE STANDISH GROUP 1994. *The CHAOS Report*.
- URQUHART, C. 2001. Bridging information requirements and information needs assessment: do scenarios and vignettes provide a link? *Information Research*, 6(2). <http://InformationR.net/ir/6-2/paper102.html> (retrieved 05/10/2008).

- VAT, K. H. 2005. Modeling human activity systems for collaborative project work: an IS development perspective. *Issues in Informing Science and Information Technology*. <http://proceedings.informingscience.org/InSITE2005/I05f65Vat.pdf> (retrieved 05/10/2008).
- WIEGERS, K. E. 2003. *Software requirements 2: practical techniques for gathering and managing requirements throughout the product development cycle*, 2nd ed. Redmond: Microsoft Press.
- WILLIAMS, D. AND KENNEDY, M 2006. A framework for improving the requirements engineering process effectiveness.
- WILLIAMS, S. (ed.) 2003. *Proceedings of the 2nd European Conference on e-Learning*, Glasgow Caledonian University Glasgow, Scotland 6-7 November 2003.

Knowledge Management Frameworks: A Review of Conceptual Foundations and a KMF for IT-based Organizations.

Kyoratungye Karemente, Jennifer Rose Aduwo, Emmanuel Mugejiera and Jude Lubega

In the new economic era, knowledge has become the primary source of wealth and consequently, the term knowledge economy or knowledge age. Rapid technological advancements and innovations have narrowed the gap between competing organizations such that the collective knowledge of employees is regarded as the key factor in producing innovative and competitive products or services. Organizations, since the early 1990s, have been forced to rethink the way they manage their intangible assets, which are in form of knowledge resources and therefore the need for knowledge management. Many organisations use knowledge management frameworks as a model that initiates and strengthens knowledge management activities in the context of achieving organisational excellence. However, different knowledge management frameworks do not fully address knowledge management activities across the organisation, such that each of them addresses certain knowledge management elements, while leaving others unattended to. The paper examined 21 knowledge management frameworks guided by three themes as knowledge management activities, knowledge management resources and knowledge management enablers (or influences) on knowledge management. A matrix was developed to capture the individual components advanced by each author with respect to knowledge management activities, resources and influences. Based on the matrices for activities, resources and influences, the individual components were harmonised and integrated in terms of relationships in the context of knowledge management.

The findings are that knowledge management activities are socially enacted activities that support individual and collective knowledge. The activities vary depending on which of the knowledge resources the organization aims at improving. Since each organization has a different focus, knowledge management activities take place in different contexts. These activities have been summarized as knowledge acquisition, creation, repository, sharing, use and evaluation. The organization should consciously choose which of these activities they intend to support in order to identify appropriate organizational variables and technology to enable them have effect. Based on findings, a new knowledge management framework has been proposed to guide practitioners and researchers when conducting knowledge management.

1. Introduction

The world is experiencing a new economic era in which knowledge has become the primary source of wealth [Savage, 1996] and this era has been termed the “knowledge age” or the “knowledge economy” [Sunassee and Sewry, 2002]. With the rapid technological advancements and innovations, the gap between competing organizations has been narrowed such that the trend now is to regard the collective knowledge of

the employees as the key factor in producing innovative and competitive products or services [Sunassee and Sewry, 2002]. Based on this development, Zack [1999] argues that organizations view knowledge as their most valuable and strategic resource, while Nonaka [1998] agrees by pointing out that in an economy, where the only certainty is uncertainty, the only sure source of lasting competitive advantage is knowledge. This change of focus since the early 1990s has forced organization to re-think the way they manage their affairs given that the emphasis is no longer on tangible assets but on intangible assets as well in form of knowledge resources and hence the need for knowledge management [Sunassee and Sewry, 2002].

Knowledge Management (KM) derives its importance from the support it provides to organizations to gain competitive advantage and effective working through sharing and re-using knowledge [Abdullah *et al.*, 2005]. In the global market, knowledge management initiatives are used to systematically control information and expertise to improve organizational responsiveness, innovation, competency and efficiency. The collaborative functions of groupware, for example, Lotus Notes encouraged real-time collaboration and information sharing through the concept of “places” as virtual rooms similar to internet chat-rooms where team members can work on a document or share tools [FTI, 2001]. According to Abdullah *et al.* [2005], there are many reasons why knowledge should be managed properly especially using collaborative technology. The reasons for knowledge management include information overload, technology advancement, increased professional specialization, competition, workforce mobility and turnover, and capitalization of organizational knowledge. In this paper, knowledge management is taken to mean the processes by which knowledge is acquired, created, communicated, shared, applied and effectively utilised and managed, in order to achieve an IT-based organisation’s existing and emerging needs including identification and exploitation of existing and acquired knowledge assets [Alavi and Leidner, 2001; Bhatt, 2001; Holm, 2001; Gupta *et al.*, 2000; Davenport and Prusak, 1998; Quintas *et al.*, 1997; Allee, 1997]. In this case, IT-based organizations have been taken to mean private or public organizations that apply IT systems to support and enhance their core operations including the processing of knowledge creation, storage/retrieval, transfer, and application [Alavi and Leidner, 2001].

Previous research on Knowledge Management Frameworks (KMFs) has enabled many organisations to conduct and implement knowledge management. These KMFs have served as foundations for planning and developing knowledge management systems (KMS) in IT-based organisations to achieve operational excellence. A KMS refers to a class of computer-based information system applied to managing organisational knowledge to support and improve the processes of knowledge creation, storage/retrieval, transfer and application. Most of KMFs discussed in this paper describe or prescribe how organizations manage their knowledge [Lai and Chu, 2000] but still show differences in many aspects including being incomprehensive in some ways [Holsapple and Joshi, 1999; Hsiangchu and Tsai-hsin, 2000]. Different frameworks do not fully address KM comprehensively to take care of organisational requirements [Calabrese, 2000], but each of them addresses specific aspects of KM elements with the implication

that other elements are left unattended to. However, with the advancement in technology and the specificity of many of the existing frameworks, there is need to review these frameworks in details and integrate them into a comprehensive model to facilitate future research and innovation in IT-based organisations conducting knowledge management.

This paper reviews and compares the existing KMFs proposed by various authors in the literature. A synthesis of the KMFs is undertaken to identify the commonalities and shortcomings. By combining these factors, the paper then proposes a new framework, which is believed to be more suitable for IT-based organizations. The paper is organized in the following manner. First, KM background is provided in the introduction followed by a discussion of the conceptual foundations of KM and its relation to information technology. Then a comparative analysis of the existing frameworks is presented, which leads to the development of the comprehensive KMF model. This is then followed by the presentation of the key distinguishing features of the new KMF and the conclusion.

2. Conceptual Foundations of Knowledge and Knowledge Management

2.1 Data, information and Knowledge

Data are at the lowest level of known facts, with little value on their own and have to be organized, analyzed, and interpreted in order to be of value. After the clean up of such data, it then becomes information. Accordingly, information can be described as assembled data into a message that is meaningful [Duffy, 1999]. Therefore, information is data or facts organised to describe a situation or condition [Wiig, 1993]. Information then becomes knowledge after it has been validated. It is the information with context, which provides the understanding and rationale associated with knowledge and cognitive experiences; its insight, judgment, and innovation. In general, knowledge can be experience, concepts, values, or beliefs that increase an individual's capability to take effective action [Alavi & Leidner, 1999; Allee, 1997]. This, therefore, implies that knowledge is a pre-requisite for decision making for many individuals and organizations.

2.2 Types of Knowledge

Previous research suggests the existence of different types of knowledge. According to Polanyi [1967], knowledge can be either *tacit* or *explicit*. Tacit Knowledge refers to the knowledge that has a personal quality that makes it hard to articulate or communicate. It can be said to be the knowing or the deeply rooted *know-how* that emerges from action in a particular context. The tacit dimension is based on experience, thinking, and feelings in a specific context, and is comprised of both cognitive and technical components. The cognitive component refers to an individual's mental models, maps, beliefs, paradigms, and viewpoints. The technical component refers to concrete know-how and skills that apply to a specific context [Popadiuk and Choo, 2006].

Explicit Knowledge refers to the codifiable component that can be disembodied and transmitted [Alavi and Leidner, 1999]. It is the *know-what* which can be extracted from the knowledge holder and shared with other individuals [Nonaka & Takeushi, 1995]. It

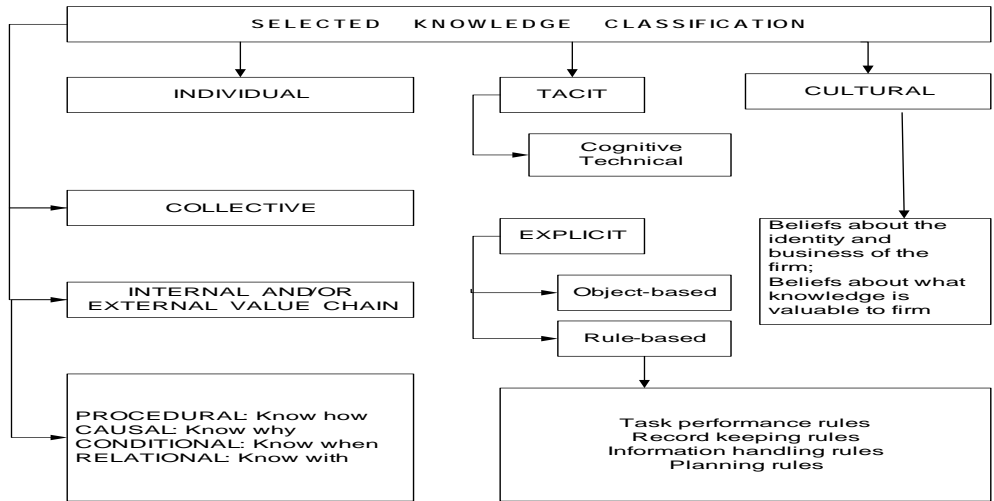
can be expressed in form of books, reports, data files, newsreels, audio cassettes, DVDs, CDs and other physical forms. The explicit dimension may also be classified as object based or rule-based. Knowledge is object-based when it is codified in words, numbers, formulas, or made tangible as equipment, documents, or models. It is rule-based when the knowledge is encoded as rules, routines, or standard operating procedures [Choo, 1998].

The rule-based knowledge is further broken down into four types of procedures [Cyert and March, 1992] as (a) task performance rules for accomplishing organizational tasks and facilitating the transfer of learning; (b) record-keeping rules on what records and how such records should be maintained by the organization; (c) information-handling rules that define the organization's communication system and (d) planning rules that guide the planning process and the allocation of resources among the activities of the organization.

Besides the tacit and explicit types of knowledge by Polanyi [1967], Choo [1998] introduced a third kind of knowledge - the cultural knowledge. This refers to the "assumptions and beliefs that are used to describe, and explain reality, as well as the conventions and expectations that are used to assign value and significance to new information" (p.112). Cultural knowledge is not codified but is diffused over the ties and relationships that connect a group. Although Nonaka and Takeuchi [1995] did not mention cultural knowledge, they attempted to distinguish between knowledge of the individual and the collective. Individual knowledge is created by and exists in the individual according to his/her beliefs, attitudes, opinions, and the factors that influence the personality formation. Social knowledge is created by and resides in the collective actions of a group, which forms their cultural knowledge heritage. It involves the norms that guide intra-group communication and coordination.

Based on the foregoing and previous researches, Figure 1 below shows the different classification of knowledge. Further classification was provided by Alavi and Leidner [2001] basing it on its usefulness (value chain) and Zack [1998] who proposed procedural, causal, conditional and relational attributes.

Figure 1: Categories of types of organizational knowledge adapted from Popadiuk and Choo (2006).



2.3 Knowledge Management

According to Bukowitz [1999], knowledge management is the process by which the organization generates wealth from its knowledge or intellectual capital. In a simplified way, knowledge management is the process through which organizations generate value from their intellectual and knowledge-based assets. Most often, generating value from such assets involves sharing them among employees, departments and even with other companies in an effort to devise best practices. Hayek (1945) pointed out that knowledge sharing was an old concept in organizations in which the most important asset is considered to be its ability to process information. Knowledge management has been defined by Kinney (1998) as “the process by which an organization creates, captures, acquires, and uses knowledge to support and improve the performance of the organization”. This definition complements that provided by Anthony (1965) who defined information handling as “...the process of collecting, manipulating and transmitting information, whatever its use is to be...”

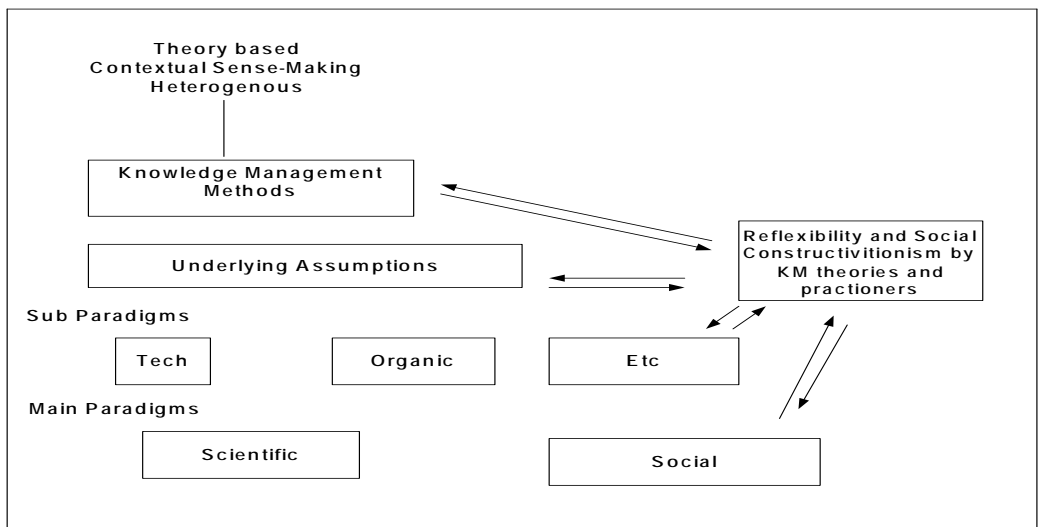
2.4 Knowledge Management Paradigms

Two main or meta-level paradigms to explicate knowledge management have been identified in the literature, namely, the scientific view and the social view [Hazlett and McAdam, 2005]. The scientific view of knowledge takes knowledge as “truth” [Alvesson & Willmott, 1996] and that knowledge is essentially a canonical body of facts and rational laws [Swann & Scarborough, 2001]. Gergen [1991] describes such indisputability as “scientists adding sanctity to ideology”. Researchers go through bodies of scientific facts, propose hypotheses, evaluate outcomes of deductions, and then build repositories of knowledge for organizational use.

Considering the limitations of solely relying on the scientific paradigm to interpret knowledge management, a different view or paradigm has been used [Demerest, 1997].

Such a paradigm is what is loosely referred to as the social paradigm of knowledge construction where Burgoyne, Pedlar, and Boydell [1994] state that the “philosophy of science has largely been replaced on the intellectual agenda by the history and sociology of knowledge which emphasizes cultural and historical processes rather than rationally superior knowledge.” This therefore implies that knowledge can be socially constructed rather than being seen as universal scientific truth, which Burgoyne *et al.* [1994] term “constructionist consciousness.” This concept agrees with Habermas’s view that knowledge constitutes human interest rather than being restricted to a functionalist science. The outcome of this debate is that knowledge management research can be undertaken using either the positivist approach or the interpretivist approach or both in combination but not any more on positivism alone. The implications of science and social paradigms for knowledge management research in relation to their sub paradigms, underlying epistemological assumptions and the applicable methods are shown in the Figure 2 below.

Figure 2: Reflexibility in Knowledge Theory and Praxis (adapted from Hazlett and McAdam (2005))



3. Knowledge Management Frameworks

In the 1990s, there was a strong inclination to research on Knowledge Management Frameworks (KMFs) to better understand knowledge management phenomenon. Many such frameworks have been developed and classified into two categories namely the descriptive and prescriptive frameworks [Holsapple and Joshi, 1999]. The descriptive frameworks attempt to explain the characteristics of the KM phenomenon while the prescriptive frameworks are geared to describing the methodologies to be followed when

conducting KM. In this paper, the attention will be on developing a broad framework that is both descriptive and prescriptive in nature with emphasis on content covering knowledge resources, knowledge manipulation activities, and influences dimensions on the conduct of knowledge management.

From the content angle, the knowledge resources dimension concerns itself with the characterization of an organization's resources, the dimension of knowledge manipulation activity identifies operations on these knowledge resources that can be executed in the organization's conduct of KM and the knowledge manipulation activities operate on the knowledge resources under the influence of various factors. The knowledge management influences dimension identifies the influencing or enabling factors that operate on the knowledge management activities and resources. Therefore, based on this analysis, Holsapple and Joshi (1999) contend that the understanding of KM phenomenon depends on:

- i. identifying the organizational knowledge resources that need management attention;
- ii. categorizing the activities that operate on the resources in the conduct of knowledge management, and;
- iii. outlining the factors which affect the conduct of knowledge management.

3.1 Review of Previous Knowledge Management Frameworks (KMFs)

Knowledge management (KM) frameworks provide the organizations with the central areas for consideration in KM efforts [Earl, 2001]. The frameworks can help such organizations to approach KM methodically and consciously [Okunoye, 2004]. Furthermore, they can help to identify a specific approach to KM, to define goals and strategies, to understand the various KM initiatives, and then to choose the best ones in the particular circumstances [Maier and Remus, 2001; Earl, 2001].

Several reviews have been done on existing frameworks but the authors of this paper have focused on framework reviews by Holsapple and Joshi [1999], Lai and Chu [2000] and Rubestein-Montano *et al.* [2001] which have been used to discuss the main features and assumptions in the existing frameworks. According to Okunoye [2004], most reviewers have generated a consensus on the need for a more generalized framework and further agree that the key components of the framework should be KM intellectual capital, KM processes and KM enablers. The authors of this paper share the same view.

3.2 Key Features of Previous Knowledge Management Frameworks (KMFs)

3.2.1 Knowledge Management Intellectual Capital

A debate has been raging on regarding the rapid changing nature of organisations' environment within which they operate especially in the developed economies. The debate goes on to discuss what constitutes a competitive advantage to an organization [Egbu, 2002]. Knowledge has been identified as being a source of competitive advantage for organizations and that it is fast overtaking capital and labour as the key economic resource in form of an intangible asset in advanced economies [Egbu, 2000; Edvinsson,

2000]. Hence, management of knowledge in organizations has to be done in an effective and efficient manner. The critical advantage of knowledge management is vested in the intellectual capital as a resource [Lai and Chu, 2000]. Different knowledge management frameworks have developed several resources dimensions which have been summarized into three types of intellectual capital namely human capital, structure capital, and customer capital [Holsapple & Joshi, 1999; Hahn and Subramani, 2000; and Okunoye, 2004].

The benefit and strategic importance of knowledge management provides leverage to an organisation to correctly identify which type of intellectual capital they can improve and apply to gain sustainable competitive advantage [Okunoye, 2004]. Bukowitz and Williams [1999] describe knowledge management as the process by which the organization derives wealth from its knowledge or intellectual capital. The components of intellectual capital form the knowledge management resources [Stewart, 1997], which are the organization's assets that can be used to achieve organizational goals.

The first knowledge resource is human capital, which refers to the employee capability to solve a problem and is the source of creativity. This is similar to the terms "employee knowledge," "employee competencies" and "professional intellect" proposed by Leonard-Barton [1995], Sveiby [1997] and Quinn, *et al.* [1996] separately. The framework by Leonard-Barton [1995] identified two types of knowledge resources - employee knowledge (human capital) and physical systems (structure capital) such as machinery and databases. In line with this thinking, the Petrash [1996] framework came up with four more types of intellectual capital: customers (referred to as customer capital), organizational processes, organizational structures, and organizational culture but the latter three relate to as organizational capital. Sveiby's framework is similar to Petrash but further incorporates customer capital within the external intellectual capital but excluding customers (e.g., suppliers).

The second intellectual capital is structural capital that relates to the organizing capability of an organization in order to meet the customers' needs. The organizing capability refers to organizational structure, processes, systems, patents, culture, documented experience and knowledge, and the capability to utilise knowledge through sharing and transferring [Stewart, 1997; Holsapple & Joshi, 1999; Mayo, 1998]. According to Lai and Chu [2000], this was considered to be similar to the terms "internal structures," "organizational capital" proposed by Sveiby [1997] and Petrash [1996] separately.

The third kind of intellectual capital is customer capital. It concerns the relationship between an organization and its stakeholders, such as a supplier or customer relationship, brands, and reputation [Stewart, 1997; Holsapple & Joshi, 1999]. This was called "external structure." by Sveiby [1997].

3.2.2 Knowledge Management Processes

Knowledge management processes are the sequential steps of conducting knowledge management in an organization. A number of KMFs have been proposed and some have been modified by various authors in as far as knowledge management processes (activities) are concerned. In addition, several reviews have been made that have

summarized or integrated the knowledge management processes [Holsapple & Joshi, 1999; Lai and Chu, 2000]. One such comprehensive review was by Lai and Chu [2000] who indicated in their review that all of the knowledge processes included in the previous frameworks were integrated based on the “content” of the processes rather than the “name” of the process and concluded that these processes can be classified into seven processes as initiation, generation, modeling, repository, distribution and transfer, use, and retrospect.

Knowledge management processes are socially enacted activities that support individual and collective knowledge and interaction [Alavi and Leidner, 2001]. These processes vary depending on which type of knowledge intellectual capital the organization aims at improving. Since each organization has a different focus, KM processes take place also in a different context.

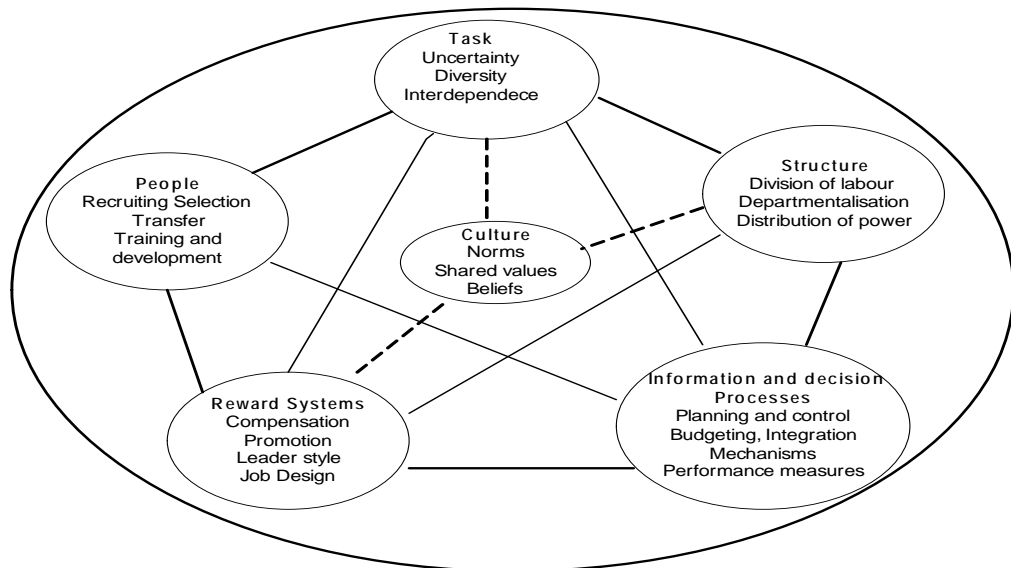
3.3.3 Knowledge Management Enablers

Knowledge management enablers (also called factors) have an impact (positive or negative) on the way knowledge management processes are manipulated. The outcome of knowledge management processes is affected by the enablers. Holsapple and Joshi [1999] identified five researchers who contributed to the discussion on knowledge enablers as Leonard-Barton [1995], Van der Spek Spijkervet [1997], Szulanski [1996], Wiig (1993), and Arthur Andersen and APQC (1996). These enablers have been summarized as *leadership, culture, measurement, technology, education, reward and incentive systems, and value and norms* [Holsapple and Joshi, 1999; Lai and Chu, 2000]. These enablers have been further classified into three groups a) *internal organisational variables*, b) *external organisational variables* and c) *facilitating variables* that cut across the previous two i.e. both internal and external such as information technology.

Organisational Variables

Okunoye [2004] used the conceptual framework on organizational issues developed by Galbraith [1977] to explain the need to identify and classify the internal enablers to the organization as organizational variables [Hurley and Green, 2005]. The original conceptual framework was modified by Okunoye [2004] to include the component of culture (internal culture) that related to the organization and this is presented in Figure 3. This chart therefore is a good representative of the organizational ‘internal’ variables that were modified by various researchers as enablers.

Figure 3. Okunoye [2004] Organisational Variables [adapted from Galbraith 1977]



Facilitating Variables have specifically been tagged to Information technology (IT), which has already made its mark in supporting the processes for knowledge creation, sharing, storage and use [Alavi and Leidner, 2001]. In this era of the internet, World Wide Web and ubiquitous computing, modern organizations no longer view IT as being only to support various organizational processes but also as a source of competitive advantage and organizational core capability [Holsapple and Joshi, 1999; Alavi and Leidner, 1999; Okunoye 2004]. IT has the ability to make the information and decision making processes easier. It also supports changes in the organizational structure and influences communication within and between organizations. Most of the contemporary organizations have been affected by IT in every aspect including the thinking of people and conduct of businesses at the workplace, if the organization is IT enabled [Lau *et al.*, 2001]).

External organizational Variables also known as environmental variables have been identified as enablers [Holsapple and Joshi, 2000; Lai & Chu, 2000; Okunoye 2004]. These are the factors outside the control of the organization and do arise from within the surrounding environment for the organization. Environmental variables have been taken to include governmental, economic, political, social, and educational factors [Holsapple and Joshi, 2000].

In summary, the knowledge management enablers (variables) can not be ignored when conducting knowledge management. Organisational change and transformation in the nature of work done by organization including knowledge management is better understood by looking at both the technological and the institutional (specifically environmental) changes that are reshaping economic and organizational activities [Orlikowski and Barley 2001]. This therefore provides a manifestation of the

interrelationship of the environment, organizational and technology variables including information technology.

4 Proposed Kmf for IT-based Organisations

Many organisations in conducting KM use Knowledge Management Frameworks (KMFs) as a basis for planning and developing knowledge management systems. All frameworks identified in this research describe how organizations manage their knowledge [Lai and Chu, 2000] but are different in several areas and are still lacking in some ways [Holsapple and Joshi, 1999; Hsiangchu and Tsai-hsin, 2000]. Different frameworks do not fully address KM across the full spectrum of organisational needs [Calabrese, 2000] especially IT-based organisations, but each of them addresses certain KM elements while leaving others unattended to. However, with the growing need to accelerate research in knowledge management, the authors have reviewed these frameworks in details and integrated them into a comprehensive KMF model to guide future research in IT-based organisations.

4.1 KMF Model Design

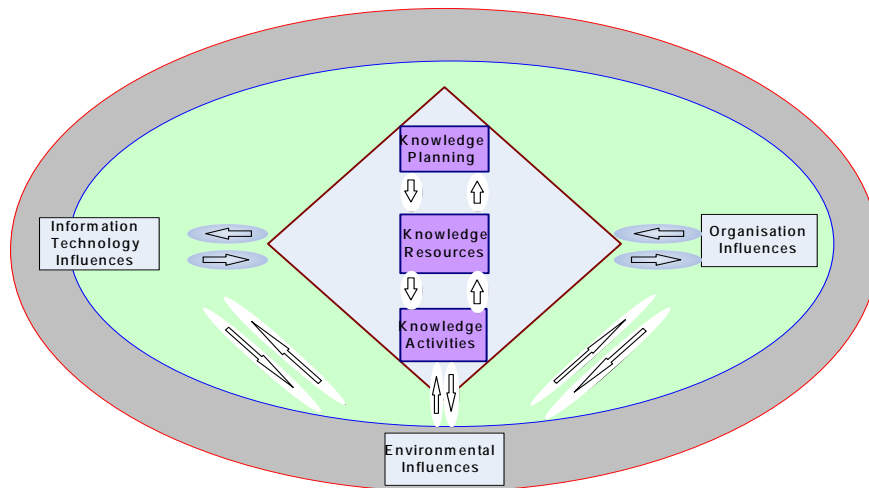
A review of literature on knowledge, knowledge management and knowledge management frameworks was undertaken; the purpose specifically was to identify commonalities and shortcomings in knowledge management frameworks so as to develop a generic knowledge management framework appropriate for IT-based organizations. A total of 21 knowledge management frameworks were reviewed from 1993 to 2004.

The review was guided by three themes as knowledge management activities, resources, enablers (or influences) adopted from Lai and Chu [2001] and other researchers. A matrix was developed to capture the respective components of each author with respect to knowledge management influences, resources and then activities. Based on these matrices for influences, resources and activities, the individual components were then harmonised in terms of relationships in the context of knowledge management. Therefore, some components were combined, where possible, in the sense that they refer to the same thing and consequently, a new list of components was derived for each matrix.

4.2 Structure of the Proposed KMF

Based on the literature survey and analysis, the authors have constructed a new comprehensive framework for IT-based Organisations, which addresses the shortcomings of the existing models. The proposed framework consists of two main distinguishing aspects or elements as 1) integrated Knowledge Management Influences Aspects encompassing environmental, information technology and organizational factors and 2) Knowledge Development Aspects that consist of knowledge management planning, resources and activities as shown in Figure 4. These two aspects interact with each other and within each aspect. The environmental factors relate to national infrastructure and culture in the surrounding environment while organizational factors concern corporate variables and information technology factors (over rapping with the environment) support the process of knowledge creation and its sustainability. These factors shape the other key knowledge development aspects of the framework that include knowledge planning, knowledge resources and knowledge activities (processes).

Figure 4: The Proposed Comprehensive Knowledge Management Framework for IT-based organisations



Based on the model, it is then possible to clearly identify the key components of the proposed KMF model as being:

- i) Knowledge Influences Aspects
 - a) Environmental Influences
 - b) Information Technology Influences
 - c) Organizational Influences
- ii) Knowledge Development Aspects
 - a) Knowledge planning,
 - b) Knowledge resources
 - c) Knowledge activities

4.3 Key Components of the KMF Model

4.3.1 Knowledge Influences Aspects

Whereas previous analyses on influences (called enablers or factors in previous KMFs) were based on five frameworks at the time [Lai & Chu, 2000], the authors' analysis has been based on 16 other frameworks that have made a contribution to influences including those after 2000. Although these influences had been further categorised into three groups as a) *environmental influences* b) *information technology influences* and c) *organisational influences*, the individual influences already identified had not been integrated into these category group, which the authors have done. Based on this approach, the authors have re-classified the influences as shown in Table 1 as identified by Holsapple and Joshi [1999], Lai & Chu [2000] and other researchers.

a) **Environment Influences** - The Environmental Influences are in the outer circle of the KMF model to represent governmental, economic, political, social, and educational factors. But these are further categorized into two ie the national culture and the national infrastructure. The *national culture aspects include* norms, values and beliefs while the national *infrastructure* aspects cover education, banking and cooperatives, transport and communication systems and other industry players [Holsapple and Joshi, 2000; Okunoye, 2004].

Table 1: Knowledge management influences

Authors	Environmental Influences		Information Technology	Organisational Influences								
	National Culture	National Infrastructure	Technology	Corporate Culture	Leadership	Corporate Infrastructure	Knowledge structure	Vision	Continuous Learning	Knowledge worker	Measurement	Reward & Incentive Systems
Andersen & APQC,1996			√	√	√						√	
Kai-Hin Chai, 1998			√	√								
Davenport et. al, 1998			√	√	√	√	√					
Hazlett & Gallaghrs '99			√	√		√						
Cai, Yan and Yi, 2000							√					
Skyrme,2000			√	√	√	√		√				
Storey and Barnett '00.			√		√				√			
Lee and Kim, 2001.			√							√		
Ndledla and Toit, 2001.				√	√							
Holsapple and Singh '01					√							
Jinxi and Jisheng,2001			√	√	√	√						
Lee and Kim, 2001.			√	√		√	√					
Egbu, 2002			√	√								
Charles H. Bixler, 2002			√		√	√			√			
Hasanahi, 2002 APQC.			√	√	√	√						
Moffett.et.al. 2003.			√	√	√							
Okunoye et.al. 2003	√	√	√	√	√	√			√		√	√
Van D.Spek /Spijkervet97			√	√	√						√	
Leonard –Barton, 1995			√	√	√				√			√
Szulanski, 1996												√
Holsapple & Jodhi,1999			√	√	√						√	√

b) **Information Technology (IT) Influences** - Information Technology has been also instrumental in enhancing communication and the interaction of individual, group, organizational, and inter-organisational knowledge [Nonaka and Takeuchi, 1995; Hedlund, 1994]. Although some elements of IT are part and partial of the internal organization like the local area network (LAN), computers and software that form the organisation’s information system, there is the other part of the IT that is external and hence organizations do not have direct control over those elements of IT such as the other networks, the internet and the communication system out side the LAN. So, it is crucial that IT, because of its strategic importance to all organizations and more especially those that are IT-based (intensive), should stand

out alone as a separate influence to knowledge management. Hence, information technology facilitates rapid search, access and retrieval of information, and enables collaboration and communication between organizational members. In essence, it plays a variety of roles to support an organization's KM processes [Alavi and Leidner, 2001; Lee and Hong, 2002; Wong, 2005].

c) Organisational Influences - Based on the summary provided in table 1, many of the influences relate to factors within the organization [Galbraith, 1977] and these include corporate culture, *leadership, corporate infrastructure, knowledge structure, vision, continuous learning, knowledge worker, measurement, reward and incentives, among others*. In view of their inter relationships as depicted in Figure 3, each of these influences has implications for KM efforts in organisations. It has been observed that particular attention be paid to organisational influences without which the success of KM becomes doubtful.

4.3.2 Knowledge Development Aspects

i) Knowledge Management Planning

Planning is one of the key principles of management that deals with setting objectives to achieve organizational goals [Vasquez *et al.*, 2000]. Although previous research on knowledge management frameworks has hinted on planning, it has not been emphasized as an important factor to start on before conducting a knowledge management activity. The authors argue that it should be the first step in the conduct of knowledge management and incorporated in the IT-based KMF. Holsapple and Joshi [1999] brought out the issue of planning in the focus dimension. The focus dimension identifies the primary intent of a framework, which in our view is in reference to planning. In the same vein, Arthur Anderson and APQC [1996], for example, identify the need to provide a basis for benchmarking the conduct of knowledge management within and between organizations and most likely at the beginning of a knowledge management project. Similarly, other researchers whose contribution was in reference to planning included Van der Spek Spijkervet [1997] by advocating for a conceptualize-reflect-act-retrospect cycle. In addition, Szulanski [1996] proposed the need to identify and examine the barriers that prevent transfer of best practices within an organization when conducting knowledge management. However, none of these previous frameworks placed knowledge management planning at the centre stage of their KMFs.

The authors contend and agree with previous researchers that knowledge management frameworks can help to identify a specific approach to knowledge management, to define goals and strategies, to understand the various KM initiatives, and then to choose the best ones in the particular circumstances [Maier and Remus, 2001; Earl, 2001] but this can only be achieved through a proper planning process. Knowledge management process therefore should begin with planning. This is more so given that knowledge management in its generality is a conceptual framework that encompasses all activities and perspectives required to plan for and gain an overview of the organizational knowledge assets and their conditions [Wigs, 1993]. This approach makes a knowledge management framework pinpoint and prioritize those knowledge areas that require

management attention; further it identifies the salient alternatives by suggesting methods for managing them, and conducts activities required to achieve desired results. Wigs therefore identified that planning was important in the management of knowledge [Vasquez *et al.*, 2000]. Maier and Remus, [2001] and Earl [2001] also recognized the importance of planning, which is reflected in their argument that knowledge management frameworks can help to identify a specific approach to knowledge management. It is against this background that a recommendation is being made that the starting point to conduct knowledge management should be the knowledge management planning function.

ii) Knowledge Management Resources

Previous authors have focused on knowledge management resources (also called intellectual capital in previous KMFs) as being composed of three components – human, structure and customer resource (capital). However, this categorization does not take into account the asset value brought about by IT. IT has generated many possibilities of advancing KM not only in terms of using it as a tool but also as a KM resource. Therefore, the authors propose a fourth knowledge management resource termed Collaborative Technological Capital (CTC) to cover concepts that do not properly fit in the earlier categorization.

Collaborative technology has been described as a tool that supports team members or other tools that share information and contribute to any knowledge management system [Hayden, 2003]. These technologies provide a platform that allows team members to collaborate on problems or projects, efficiently capture data and information, and transform it into knowledge. It is against this view that this new knowledge generated from collaborative technology becomes a knowledge resource or capital and hence the use of the term collaborative technological capital. Abdullah *et al.* [2005] supports the view that knowledge would be more useful if it could be shared and used among the organisation's staff that work together using collaborative technology at anytime, anyplace and anywhere. Hahn & Subramani [2000], on the other hand, argued that many organizations currently deploy corporate intranets so that important documents can be posted and accessed by other users browsing or searching through the site. The emphasis was that such documents do not follow a predefined structure but search and retrieval is achieved via search engines that locate documents using full-text search. The use of collaborative filtering technology that recommends documents is an alternative approach to locating documents relevant to a user's question or problem without structuring contents a priori. The collaborative filtering system records the browse and search behaviors of users and recommends documents based on the past behaviors of other users when they performed similar searches. This is an aspect of collaborative technological capital as a resource in knowledge management. The authors note that all the three researchers mentioned identified CTC but in the context of building knowledge management systems and not in the sense of being a component of the KMF as a resource.

Collaborative technological capital as a fourth resource in KM provides new grounds for reclassifying the contributions made by earlier authors. Leonard-Barton [1995] for example considered databases under structure but these can be also in collaborative technological capital. Sveiby [1997] looked at concepts, and computer and administrative system under structure capital but this too can be placed under collaborative technological capital. On the other hand, Hahn & Subramani [2000] and Abdullah *et al.* [2005] identified collaborative technological capital as a legitimate knowledge asset. It is against this view that collaborative technological capital as fourth knowledge resource is justified.

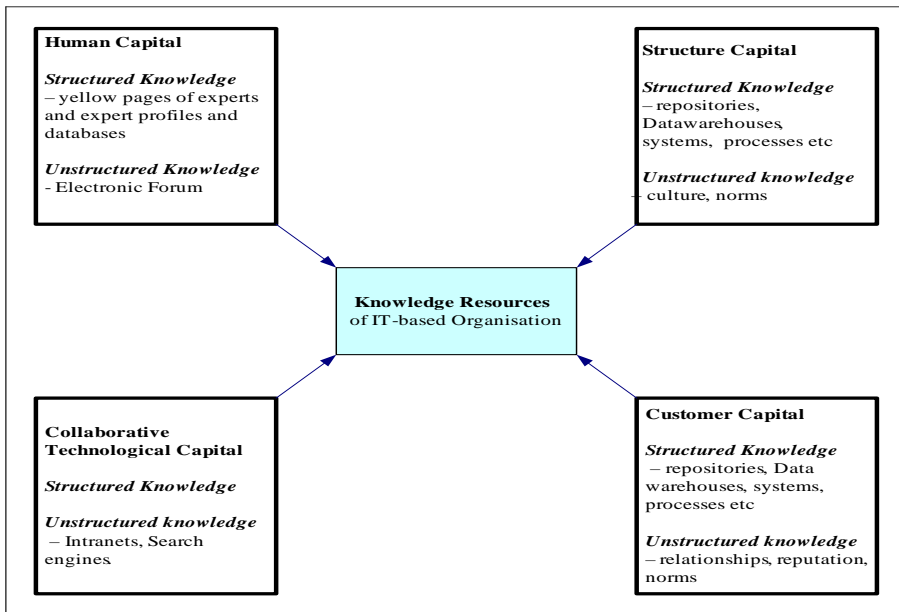
Table 2: Classification of Knowledge Management Resources

Author	Knowledge Resources			
	Human Capital	Structure Capital	Customer Capital	Collaborative Technological Capital
Leonard-Barton, 1995	Employee Knowledge	Physical systems (Machines)		Physical systems (Databases)
Petrash, 1996	Human Capital	Organizational capital/structures	Customer capital	
Sveiby, 1997	Employee competence	Internal structures (patents, concepts, models, & culture.)	External structures (customer & supplier relationships, brands)	Concepts, computer and administrative systems
Steward, 1997	Human Capability	Organizing Capability (structure, culture, processes, patents, experience)	Stakeholder Relationships (supplier, customer, brands, reputation)	Organizing Capability (systems, documented experience)
Klobas, 1997	Individuals	Recorded knowledge	External sources	
Hahn & Subramani, 2000	Individuals	organizational knowledge - repositories, Datawarehouses		Collaborative filtering (Intranets, Search Engine)
Abdullah <i>et al.</i> , 2005	Knowledge workers	Organizational capital (documents, repositories, processes, practices, norms)		collaborative technology capability any time/place/anywhere-email,s/engines

Knowledge management resources can have structured and unstructured segments for purposes of knowledge management support [Hahn and Subramani, 2000]. In this regard, the Figure 5 below represents the four components of knowledge management resources and the examples of such knowledge under each resource for IT-based organisations [Abdullah *et al.*, 2005]. The human capital resource, for example, has yellow pages of experts, expert profiles, and experts' database under structured knowledge and electronic forum under unstructured knowledge. The other resources also have examples that have relevance to IT-based organisations. The structure capital,

for instance, has repositories, data warehouses, systems, processes etc as structured knowledge while culture and norms fall under unstructured knowledge. The customer capital, on the other hand, has repositories, data warehouses, systems, processes under structured knowledge whereas relationships with stakeholders belong to unstructured knowledge. Finally, collaborative technological capital has intranets and search engines [Abdullah *et al.*, 2005; Hahn & Subramani, 2000] as unstructured knowledge. The re-classified knowledge resources can then be fitted into the proposed comprehensive knowledge management framework as a major pillar.

Figure 5: Knowledge Management Resources for IT-based organizations.



iii) Knowledge management activities

For the purpose of this paper but basing on previous frameworks, a new and simplified matrix has been developed as shown in Table 3 to serve as an analytical framework for the review and presentation of the activities more applicable to IT-based organizations. The knowledge activities were re-classified and summarized into six activities basing on content of each activity as analysed from previous frameworks. The new structure of activities is *knowledge acquisition, creation, repository, sharing, use and evaluation*. The authors took and considered these as the key management activities for the development of the generic comprehensive knowledge framework appropriate for IT-based organizations. From the 21 knowledge management frameworks reviewed, only 16 of them were focused on knowledge management activities. The organization should consciously choose which of these activities they intend to support in order to identify appropriate organizational variables and technology to enable them have effect.

Table 3: Knowledge Management Activities

Authors	Acquisition	Creation	Repository	Sharing	Use	Evaluation
1. Wiig, 1993		√		√	√	√
2. Nonaka, 1994	√	√				
3. Leonard-Barton,1995	√	√	√	√	√	
4. Arthur Anderson & APQC, 1996	√	√	√	√	√	
5. Choo, 1996	√	√			√	√
6. Szulanski, 1996	√				√	√
7. Van der spek and Spijkervet, 1997		√			√	√
8. Alavi, 1997	√	√		√	√	
9. Kai-Hin Chai,1998	√	√	√	√		√
10. M'Rad,2000)	√			√	√	√
11. Alavi and Leidner, 2001.		√	√	√	√	
12. Holsapple and Singh,2001	√	√			√	
13. Jinxi and Jisheng, 2001.	√	√			√	

A brief outline of each knowledge management activity identified in Table 3 is provided below:

Knowledge acquisition - Knowledge acquisition is the process of acquiring knowledge from both internal and external sources of the organisation. It is acquired through various ways for instance, internally through induction programme for new employees and externally through educational institutions and previous employers. The knowledge for example, customer service experience, which the new recruits bring to the organization, can be converted for the benefit of the organization [Koh *et al*, 2005]. The source of knowledge acquisition has no limits; it could be from superiors, customers, advertisements, magazines, newspapers and television. [Koh *et al*, 2005]. Information made available through ICT, manuals, memos and e-mails provides the supporting environment to enhance knowledge acquisition process.

Knowledge creation- Knowledge creation is the transformation of tacit knowledge to explicit knowledge and vice versa. The conversion of one kind of knowledge to another has been explained using the Spiral of Organizational Knowledge Creation Model proposed by Nonaka and Takeuchi [1995] as explained below:

Socialization: This is the creation of tacit knowledge from other tacit knowledge, through experiences shared by many members of the organisation. It rests on the transmission of tacit knowledge from one person to another without using formal communication but through observation, conversation, imitation and practice e.g. using teleconferencing technology.

Combination: This is the creation of explicit knowledge from an explicit knowledge source, for example through filtering, categorization or “re contextualising” of explicit knowledge, brainstorming for instance falls in this category. An example of combination is by using groupware technology.

Externalization: This emphasises converting tacit knowledge to explicit knowledge through codification e.g. by using e-mail and broadcasting technology.

Internalization: This involves converting explicit knowledge into tacit knowledge e.g. using visualization technology. The process of knowledge creation draws extensively from the existing knowledge base, for instance, transformation of explicit, tacit and cultural knowledge to new knowledge. When management tries to resolve an issue by finding a solution, it results in knowledge creation. Once a solution has been found and implemented successfully, the new knowledge can be made available organizationally by the management. [Koh *et al*, 2005].

Knowledge repository - Knowledge storage/repository is the ability of providing a centralised repository for knowledge storage, for example a public library and a database of related information about a particular subject. In the context of information technology, a knowledge repository is a machine-readable resource for the dissemination of information, either online or offline. A well-organized knowledge repository can save the organization money by decreasing the amount of employee time spent trying to find information for instance on company policies and procedures.

Knowledge sharing - Knowledge sharing is the means by which organization obtains access to its own and other organizations’ knowledge. Experience and research suggest that successful knowledge sharing involves extended learning processes rather than simple communication processes, as ideas related to development and innovation need to be made locally applicable with the adaptation being done by the ‘incumbent organisations’ for the ideas to be successfully implemented [Cummings, 2003]. When knowledge is shared in the organization to achieve an organizational goal, the knowledge is distributed. Sharing of knowledge takes place in two ways for example formal and informal. Formal sharing takes place through official channels like meetings, discussions, e-mail, web-postings and memos, while informal sharing takes place inside or outside the office, for instance, during breaks and time out. Deliberate management attempts can improve the knowledge sharing functions in the organisation. These measures could include community of practice, quality circles and buddy training [Koh *et al*, 2005].

Knowledge use - Knowledge use is the process of getting knowledge utilised for a particular purpose; this occurs, when knowledge is put into action for decision making or policy making. Knowledge utilization results in knowledge increase, by gaining expertise and insights. Employees learn through experience on how to deal with a particular type of enquiry efficiently. For instance, frequent use of the information helps employees to locate the information faster and become aware of the location in which the information resides. Knowledge is useless if it is not utilized. People do not just passively receive

knowledge; rather they actively interpret it to fit with their own situation and perspective [Nonaka, 1998]. Utilization of knowledge increases the expertise in a domain of action, and the user becomes an expert through repetitive practice.

Knowledge Evaluation - This is the process of appraising the knowledge available to the organisation say in its repository to assess whether it meets the organisation's objectives. It is also a process of determining whether new knowledge was created [Lai and Chu, 2000].

5. Differentiating Features of the Comprehensive KMF

The discussion of previous frameworks above provided the foundation for designing a comprehensive framework for IT-based organizations. This framework differs from the other previous KMFs in several ways:

- 1) It introduces and emphasizes the need to start the conduct of knowledge management with planning and continue interactively with the other components of the framework like knowledge resources, activities etc.
- 2) It acknowledges the impact of IT on KM as a knowledge management resource enabled through collaborative technologies. This therefore increases the knowledge resources from the three previously advanced to the current four namely human, customer, organization structure and collaborative technological capital, which now form part of the comprehensive KMF.
- 3) It harmonizes the knowledge management activities into six activities (acquisition, creation, repository, sharing, use and evaluation) that are considered as core to KM in IT-based organizations. These six activities summarize and integrate all the previous researchers' KM activities that have been advanced as the processes required and necessary in the conduct of KM in an organisation;
- 4) It harmonises all the individual influences previously advanced and classifies them into three groups ie environmental influences, information technology influences and organizational influences.

6. Conclusion

The framework provides a methodology for the conduct of knowledge management. It is intended to provide a basis for organizations to undertake the design of better policies, modification of actions and delivery of desired results. Accordingly, the framework presented can be used by any person intending to develop or evaluate knowledge management in any organization. The integration of previous frameworks into this comprehensive framework can aid the acceleration of research in knowledge management.

Previous research has shown that knowledge management provides a major competitive advantage and as such is a critical activity. This is so given that KM is the process of managing knowledge through a systematically and organizationally specified process for acquiring, organizing, sustaining, sharing, and renewing both tacit and

explicit knowledge of employees to enhance organizational performance and create value [Allee, 1997; Davenport and Prusak, 1998; Alavi and Leidner, 2001].

Knowledge management, as a managerial activity which develops, transfers, transmits, stores and applies knowledge for decision making in order to attain the organization's goals, requires careful planning before implementation. Furthermore, the activation of individual components within the framework should take cognisance of the fact that these components are inter-related, which is in line with Leavitt [1965] model of organisational change recommendations that makes a case for the interdependence of productive variables.

The framework, however, has not been evaluated to test its consistency and validity. It is suggested that further research should look into the evaluation of this framework.

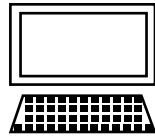
References

- ABDULLAH, R., SAHIBUDDIN, S., ALIAS, R. A, SELAMAT, M.H. (2005). A Framework For Knowledge Management System Implementation In Collaborative Environment For Higher Learning. *Journal of Knowledge Management Practice*.
- ALAVI, M., LEIDNER, D.E. (2001). Knowledge management and knowledge management systems: conceptual foundations and research issues. *MIS Quarterly*, Vol. 25 No.1, pp.107-36.
- ALAVI, M., LEIDNER, D.E. (1999). Knowledge management system: issues, challenges, and benefits. *Communications of the AIS*, Vol. 1 No.7, pp.1-37.
- ANTHONY, R.N. (1965). *Planning and Control Systems: A Framework for Analysis*. Boston: Harvard University Press, pp vi, 180.
- ALLEE, V. (1997). Twelve principles of knowledge management. *Training and Development*, Vol. 51 No. 11, pp. 71-4.
- BHATT, G. (2001). "Knowledge management in organizations: examining the interaction between technologies, techniques, and people". *Journal of Knowledge Management*, Vol. 5 No. 1, pp. 68-75.
- BURGOYNE, J., PEDLAR, M., & BOYDELL, T. (1994). *Towards the learning company*. London: McGraw Hill.
- BUKOWITZ, W. AND R. WILLIAMS (1999). *The knowledge management fieldbook*. Pearson Education Limited.
- CUMMINGS J. (2003). *Knowledge Sharing: A Review of the Literature*, The World Bank Operations Evaluation Department-OED, The World Bank Washington D.C
- DAVENPORT, T.H. AND PRUSAK, L. (1998), *Working Knowledge: How Organizations Manage What They Know*. Harvard Business School Press, Boston, MA.
- DUFFY, J.(1999). *Harvesting experience, reaping the benefits of knowledge*. Arma International.
- EDVINSSON, L. (2000). Some perspectives on intangibles and intellectual capital 2000. *Journal of Intellectual Capital*, Vol. 1, No.1, pp. 12-16
- EGBU, C.O. (2002). Information Technologies for Knowledge Management: Their Usage and Effectiveness. July 2002 PUBLISHED: August 2002 at <http://www.itcon.org/2002/8>
- GALBRAITH, J. (1977). *Organisational Design*. Reading, Mass: Addison Wesley, 1977.

- GUPTA, B., IYER, L.S. AND ARONSON, J.E. (2000). Knowledge management: practices and challenges. *Industrial Management & Data Systems*, Vol.100 No.1, p.17-21.
- FINANCIAL TIMES (WWW.FT.COM). (2001). FTIT June 6 2001 / Case studies. <http://specials.ft.com/ftit/june2001/FT3OM6GYJNC.html>. Access on May 16, 2007
- HAHN, J. AND SUBRAMANI, M.R (2000). A Framework of Knowledge Management Systems: Issues and Challenges for Theory and Practice. *21st International Conference on Information Systems, Brisbane, Australia*, December 2000.
- HAYEK, G. (1945). The use of Knowledge in Society. *American Economic Review*, XXX, No.4, September, 519 – 30; accessed 20th July 2004: www.virtualschool.edu/mon/Economics/HayekUseOfKnowledge.html
- HAZLETT, S.A., MCADAM, R AND GALLAGHER, S, (2005). Theory Building in Knowledge Management In Search of Paradigms. *Journal of Management Inquiry*, Vol. 14 No. 1, March 2005 31-42
- HAYDEN, T.W. (2003). The age of collaborative knowledge management system: A Technical View. *International Symposium on Collaborative Technologies and Systems 2003*.
- HEDLUND, G. (1994). A Model of Knowledge Management and The N-Form Corporation. *Strategic Management Journal*, (15), Summer 1994, pp. 73-90
- HOLM, J. (2001). Capturing the spirit of knowledge management. Paper presented at the *37 American Conference on Information Systems*, Boston, MA, August 3-5.
- HOLSAPPLE, C. AND JOSHI, K. (2000). An Investigation of Factors that Influence the Management of Knowledge in Organisation. *Journal of Strategic Information System*, (9:2-3), 2000, pp. 235-261
- HOLSAPPLE, C. AND JOSHI, K.(1999). Description and Analysis of Existing Knowledge Management Frameworks. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1999.
- HURLEY, T.A AND GREEN, C. W. (2005). Knowledge Management and The Nonprofit Industry: A Within And Between Approach. *Journal of Knowledge Management Practice*
- KLOBAS, J.E. (1997). *Information services for new millennium organizations: librarians and knowledge management*. in Raitt, D. (Ed.), *Libraries for the New Millennium*, Library Association Publishing, London, pp. 39-64.
- KOH, S.C.L GUNASEKARAN, A. AND ARUNACHALAM, S. (2005). “The application of knowledge management in call centres”. *Journal of Knowledge Management*, Volume 9 Number 4 2005 pp. 56-69.
- LAI, H. AND T'SAI-HSINCHU (2000). Knowledge Management: A Review of Theoretical Frameworks and Industrial Cases. *Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000*
- LAU, T., WONG, Y. CHAN, K., AND LAW, M. (2001). “Information Technology and The Work Environment – Does IT Change the Way People Interact at Work?”. *Human Systems Management*, (20:3), , pp. 267-280.
- LEAVITT, H. J. (1965). “Applying Organizational Change in Industry: Structural, Technological, and Humanistic Approaches”. In *Handbook of Organizations*, edited by James G. March. Chicago: Rand McNally.

- LEE, S.M., HONG, S. (2002). An enterprise-wide knowledge management system infrastructure. *Industrial Management & Data Systems*, Vol. 102 No.1, pp.17-25.
- NONAKA, I. AND TAKEUCHI, H.(1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford, UK: Oxford University Press, May 1995.
- NONAKA, I. (1998). *The Knowledge-Creating Company*. In *Harvard Business Review on Knowledge Management*. Harvard Business School Publishing, Boston.
- OKUNOYE, A. (2004). Context-Aware Framework of Knowledge Management *Cultural and Infrastructural Considerations*.
- OUINTAS, P., LEFRERE, P. AND JONES, G. (1997). Knowledge management: a strategic agenda. *Long Range Planning*, Vol. 30 No. 3, pp. 385-91.
- POPADIUK, S. AND CHOO C.W. (2006). Innovation and knowledge creation: How are these concepts related?. *International Journal of Information Management* 26 302–312
- RUBENSTEIN-MONTANO, B., LIEBOWITZ, J., BUCHWALTER, J., MCCAW, D., NEWMAN, B., REBECK, K. (2001). The Knowledge Management Methodology Team. “A Systems Thinking Framework for Knowledge Management”, *Decision Support Systems*, (31:1), pp. 5-16.
- SAVAGE, CHARLES M. (1996). *Fifth Generation Management: Co-Creating through Virtual Enterprising, Dynamic Teaming, and Knowledge Networking*. Boston: Butterworth-Heinemann.
- SUNASSEE, N. N AND SEWRY, D.A. (2002). A Theoretical Framework for Knowledge Management Implementation. *Proceedings of SAICSIT2002*, pp 235 – 245
- ORLIKOWSKI, W. AND BARLEY, S. (2001). Technology and Institutions: What Can Research On Information Technology And Research On Organisations Learn From Each Other. *MIS Quarterly*, (25:2), 2001, pp. 145-165.
- VASQUEZ, M. L., GOMEZ. C. F., LOPEZ, A.M. AND ROBLEDO, J. (2000). Knowledge Management Systems Assessment: A Conceptual Framework and A Methodological Proposal. *Proceeding of LAMOT 2000 Conference on Technology Management, Track 15 – Knowledge Management*
- WONG, K.Y. (2005). Critical success factors for implementing knowledge management in small and medium enterprises. *Industrial Management & Data Systems*, Vol. 104 No.9, pp.735-43.
- ZACK, M.H. (1999). Developing a Knowledge strategy. *California Management Review*, vol. 41, no. 3, 125-146

PART TWO



Information Systems

1

Managing University Research: Key Policy Issues

Anthony. J. Rodrigues

We examine the pursuit of research in a community of persons that constitute a University. The responsibilities of the different sectors of that community, as well the commitment required of the institution in order that research may flourish are discussed. Some generic principles and policy issues pertaining to university research are highlighted. Finally we outline a case on the politics of a research quality framework in a developed country. A contextual analysis of the case emphasizes the need for and importance of universities having balanced research policies.

1. Introduction

There have been numerous attempts at defining research which is not easy for ‘research is as elusive as the truth it seeks’ [Bonneau & Corry 1972]. Attempts have been made to decompose it into simple elements such as frontier research and reflective inquiry. Other attempts have been made to categorise it into:

- o Basic and applied ,
- o Pure and mission oriented, and
- o Fundamental and operational.

However, in all these versions a common spirit of research emerges, which, must stand for the advancement of knowledge through scholarly, scientific and creative activity.

2. Elements of Research

1. **Intellectual Curiosity** whether in response to a problem or simply for new horizons. For example to explain why certain data have the patterns they do; or to discover order in data seemingly without coherence, or to conduct inquiry beyond the boundaries at which data previously known had ended.
2. **Learning.** An advance in knowledge is rarely possible without an understanding of what was known before. This prior learning is needed so that the worker can distinguish between his own difficulty of understanding the subject and those problems in the subject that are without adequate explanation and exist for all others.
3. **Creative mind.** Seeking answers to problems requires a mind willing to apply lessons learned as the work progresses towards its logical or logistical

conclusion. This activity elicits a commitment which motivates the work with a heightened attention to its problems. It is this attention that is creative not the self that researches.

4. **Sustained work.** A high level of energy and uninterrupted time are needed to sustain research. This arduous process may be motivated by the quest for results, however, the intermediate results of the work being a process of trial and error appear to go round in circles often demotivating the worker to the point of exhaustion only for the cycle to start over and over again until there is clarity (order) and the loop is no more.

Research is then an activity pursued by the community of persons that constitute the university. It is useful to consider the responsibilities of the different sectors of that community, as well the commitment required of the institution in order that research may flourish.

3. General principles of a University's research policy:

- 3.1 Recognises that the international reputation of a university is largely based on its research reputation.
- 3.2 Remain committed to excellence in research, research training and research consultancy throughout the academic community.
- 3.3 Prides itself on its vibrant and positive research environment, frequently interdisciplinary in nature.
- 3.4 Covers the full range from investigator driven research carried out individually or in teams which adds to the international knowledge base, to research consultancies which more directly and rapidly benefit the community we live in. This depth and breadth in R&D provides quality personalised research training experience for higher degree by research students.
- 3.5 All staff appointed to carry out teaching, research and research training should continue to have the opportunity to do so.
- 3.6 Recognises the important nexus between undergraduate teaching and research.

4. Importance of Research to a University

4.1 The University should recognise that research forms a necessary and vital part of its function as a University. Through its policies and practices, it should seek to encourage the pursuit of excellence in the research that is undertaken. The University and its Schools should support areas which demonstrate or hold the promise of showing excellence.

4.2 The University should recognise the importance of research for assisting in:

- achieving the national and international goals of enhancing the quality of life and enhancing wealth creating potential;
- promoting an academic and intellectual ethos within the institution;

- ensuring relevance and vigour in the courses, and enthusiasm among the staff in the institution;
- attracting and retaining staff of the highest calibre;
- promoting and supporting excellence in undergraduate and postgraduate teaching, and providing a platform for future development of courses at these levels;
- promoting and developing excellence in postgraduate study;
- enhancing the reputation of the institution, its Schools and the staff therein;
- enhancing the level of external income to the University;
- promoting relations with local, national and international partners, including industry, business, commerce, governmental organisations, professional bodies, educational establishments, charities and the community in general.

5. Key Policy Issues

- 5.1 The University should attach considerable importance to research for the reasons given in section 4. All Schools should have some opportunity to undertake research, although the level of support for and expectation of Schools may vary selectively over time and between Schools.
- 5.2 Within the context of selective support, the focus of decision-making about the balance of research and other activities for a School, its disciplines and its staff should lie within the Schools. It is here that knowledge of available resources and expertise is most detailed and where strategic and operational objectives are set.
- 5.3 The University and its constituent Schools should consider that excellence in research, teaching, consultancy, professional practice, course development and management merit parity of esteem.
- 5.4 While recognizing that individual staff will be suited to different parts of this range of activities the University does not expect that all staff will be engaged in research as defined in section 6; although it does expect that all staff will be engaged in scholarship.
- 5.5 The University has the ultimate responsibility for taking an overview of its research portfolio and, through a Research and Consultancy Committee and the strategic planning process, should, where appropriate, negotiate with Schools on the balance of their activities. The priorities as they relate to research should be set out through the University Corporate Plan, which should be monitored through the Annual Operating Statement.
- 5.6 Schools will be encouraged by the University to be selective in the areas of research that they pursue, harnessing and developing the talents of the researchers available to them in the most effective and productive way. Schools will be expected to produce updated strategic plans annually; to monitor the activities contained therein; and to submit a copy of these plans to the Research and Consultancy

Committee annually. Schools should also make public their criteria for deciding how individual staff will balance their time between research, teaching, and other activities. All academic staff, including those on fractional contracts, should have the opportunity to be judged against these criteria.

- 5.7 The University should both encourage and, where possible, facilitate Schools and researchers to seek innovative opportunities and areas for research both within and between traditional disciplines.
- 5.8 The University should recognise that research requires resources including adequate time and space. With regard to the resources available and within the context of the overall priorities and selectivity policies agreed by it, the University should provide the Schools with enabling resources, and should monitor Schools in the use of these resources. In turn, Schools should be in the best position to advise on how best to distribute their available resources in order that the aims given here are realised to the best and most efficient effect.
- 5.9 The University should recognise that the aims given cannot be shown to be achieved unless quality research output appears in the public domain or, in the case of sensitive material, as confidential output. It is expected that such output will be the normal result of research.
- 5.10 The University should commit itself to providing effective, efficient and equitable support for researchers, via the infrastructure, on such matters as holding and publicising data on funding sources and research opportunities, advising on research bids, publicising the University's research successes, advising on research expenditure, administering the enrolment and progress of research students, publicising and co-ordinating internal and external policy matters related to research, and performing a central research co-ordination role in general.
- 5.11 Through School Research Committees and the Research & Consultancy Committee, the University should monitor the implementation of this policy. It will also review research performance, and the use of research resources through a Research Degrees Board, the progress of research students. Performance Indicators should be published against which progress can be measured, and an annual report on these matters be produced.
- 5.12 Ethics approval of research projects and monitoring of ethical practice in research will be carried out through Schools and the Research Ethics Committee, on behalf of the Research and Consultancy Committee. A report from the Research Ethics Committee on its activities will be submitted annually to Research and Consultancy Committee. A Research Ethics Code of Practice should provide further guidance

6. Definitions

- 6.1 The University should recognise that 'research' includes a number of activities which share the property of being innovative:

- basic research, which is experimental or theoretical and aims at acquiring new knowledge or offering new interpretations;
 - strategic research which is potentially applied but is in an area where the eventual applications are not clearly specifiable at the time;
 - applied research which is work undertaken to acquire new knowledge and is directly aimed at practical and applicable objectives;
 - near-market research which is innovative work aimed at generating or partially generating a specific product, artefact or idea for the commercial market;
 - creative work, particularly in the arts and allied subjects;
 - advanced pedagogic research of the type acceptable in national assessment exercises, including innovative research into teaching methodology and development of the curriculum.
- 6.2 It should be recognised that a continuum exists between these forms of research, and that all are equally commendable activities in pursuit of the University's research aims. However, it is also recognised that the availability of external funding for research may have the effect of giving preference to certain of these activities.
- 6.3 The University should also recognise that other related activities are either required or are equally desirable and valid in a School or staff portfolio even though these are excluded from the strict definition of research:
- **consultancy** mainly aimed at revenue generation through standard testing or standard application of methods;
 - **substantive consultancy** which uses research expertise (not simply routine methods or tests) on a contractual basis in order to achieve a specific contracted goal;
 - **scholarship*** whose purpose is to update staff in subject developments, which in any case is required of all staff;
 - necessary or voluntary professional practice which is non-innovative;
 - routine curriculum updating and routine updating of teaching methods, both of which are required of all teaching staff.

* Note: There is often a different meaning given to scholarship i.e. work intended to expand the boundaries of knowledge and understanding within and across disciplines by the analysis, synthesis and interpretation of ideas and information, making use of a rigorous and documented methodology and which results in publications of various kind.

7. University Research Policy and National Politics: A Cautionary Tale

Plans to introduce a national research quality assessment scheme for Australian universities were abandoned following the election of a Labor government in November 2007. The cost, however, was high as tens of millions of dollars were spent by the

government and universities preparing for the introduction of the controversial scheme in the following year 2008.

Labor's Acting Minister for Science and Innovation said the new government would abolish the "flawed research quality framework scheme". It would be replaced by "a new, streamlined, transparent, internationally verifiable system of research quality assessment" see [Maslen G. 2007].

The decision to assess research being conducted in Australia's universities and other public research organisations was announced by the former Conservative Prime Minister in 2004. An expert advisory group was established and headed by the late Sir Gareth Roberts, who had earlier led a review of Britain's research assessment exercise in 2003.

A report by the group in 2006, said that fundamental to the research quality framework model was the importance of review by peers and qualified end-users. Sir Gareth Roberts said "My experience in the United Kingdom clearly demonstrates that the only system which will enjoy the confidence and consent of the research community is one based on expert review. I am pleased that the Australian research quality framework (RQF) will be underpinned by this vital principle" see World University News 2008. The report endorsed RQF outcomes influencing future research block funding for universities and stated that the proposed model, once implemented, would establish greater transparency regarding the quality and impact of publicly-funded research.

With the scheme intended to begin in 2008, universities prepared to compete for the A\$550 million (US\$500 million) in additional research funding. The result was that hundreds of academics faced losing their jobs in a massive staff restructuring with those regarded by their bosses as less active in research were likely to be made redundant or forced to accept teaching-only positions. At the same time, top researchers were being poached from other universities in Australia and overseas with offers of \$250,000-plus a year salaries - double the usual amount a professor might earn. Entire research teams were lured from competing institutions in universities around Australia.

By the start of 2007, several universities had already conducted audits of the research strength of their academics, with staff ranked on a one to five scale and those scoring a one or two told they would not be included in the assessment exercise and could face teaching-only careers or retrenchment. Australia's largest university, Monash, was accused of 'culling' its academics in a bid to boost its research strengths by offering lecturers with poor research records voluntary separation packages. One critic claimed 300 staff had cleaned out their offices at the end of 2007 after accepting a \$10,000 bonus to quit, although this was more related to a sharp fall in IT enrolments.

In early 2007, the National Tertiary Education Union (NTEU), which represented a majority of the nation's academics, warned that there was widespread anxiety on campus about the impact of the scheme and that stress levels among academics were high and rising. In a survey of NTEU members, across the sector, the union's research coordinator, found many instances where academics had been 'leaned on' to leave or bring forward their retirements. The union lobbied the then Labor Opposition to abandon the scheme if it won office. With Vice-chancellors also opposed to its immediate introduction, the Labor Opposition agreed and, with the election landslide the RQF became a lost cause.

The newly elected Labor government said it would also address the inadequacies in current and proposed models of research citation. Labor's model would recognise the contribution of Australian researchers to Australia and the world.

The lessons learnt from this case are that the interpretation and implementation by some universities of the Research Quality Framework appeared to be in conflict with:

- the general principles cited above:
 - o 3.5 All staff appointed to carry out teaching, research and research training will continue to have the opportunity to do so;
 - o 3.6 Recognising the important nexus between undergraduate teaching and research;
- the generic policy issues raised above:
 - o 5.3 namely that excellence in research, teaching, consultancy, professional practice, course development and management merit parity of esteem; and
 - o 5.6 that inter- *alia* Schools should make public their criteria for deciding how individual staff will balance their time between research, teaching, and other activities.

The need for and importance of Universities having balanced research policies cannot be underestimated.

References

ACADEMIC POLICY STATEMENT Middlesex University 2006

BONNEAU AND CORRY, QUEST for the Optimum, Association of Universities and Colleges of Canada, 1972

MASLEN G. AUSTRALIA: Research quality scheme scrapped 2 December 2007. World University News August 2008

POLICY FOR RESEARCH, University of York 2000

RESEARCH AND DEVELOPMENT POLICY, University of Botswana RD02/02N

RESEARCH AS A UNIVERSITY FUNCTION, The University of Lethbridge Research policy 1992

RESEARCH POLICY IN CANADA, Davenport P. University of Western Ontario 2004

RESEARCH POLICY Oxford Brookes University 2003

2

Application of Grid Computing for on line Learning Resources

Sadeque Imam Shaikh

There is a big difference in education sector between first world and developing world. This difference is based on digital learning resources and computing power. Unfortunately all of these resources are geographically distributed all over the world. There is no doubt that ICT is playing a big role for introducing E-learning around the world. But if we really want to overcome these differences we need a revolutionary approach that support mutual use of geographically distributed computing and learning resources as an aggregated environment that will create new ways of flexibility, interoperability and extensibility. According to IBM there are millions of distributed computers on the web and most of the computing powers of those PCs are under utilized. Fortunately Grid is the technologies that can integrate all of these resources of knowledge's and produce super-computing power from those geographically distributed computers to access those knowledge's without sacrificing local autonomy. In this paper we will describe the ICT infrastructure for on line collaborative learning and then we will design service oriented Grid technology that will be able to support ICT infrastructure by generating supper computing power from distributed resources for sharing learning resources. We will also analyze the challenges related with technology, standard, security and performance of grid Technology for resource share and management based on literature review and cutting edge technologies of available industry standard software and toolkits. Finally we will propose recommendations for successful implementation of Grid technology with ICT for collaborative leaning revolution in the world. Methodology of this paper is based on study, analysis and literature review as well as empirical as we will do few experiments using grid tools to taste feasibility of this technology.

1. Introduction

E-learning is the key stream of Information Technology Revolution around the world as with the help of it learning materials can be delivered to the users even in remote place without having any physical presence of the teacher. Due to this revolutionary approach many techniques of E-learning have been developed and implemented by various organizations of the world like client-server, peer to peer or modern web service architecture [Vossen G.,Westerkamp, 2003].But the main disadvantages of E-learning is related with scalability, availability, distributed computing power and memory allocation .As a result E-learning is using in areas that do not highly require these requirements. Thus the power of E-learning cannot be explored properly for the welfare of man kind. But with the passage of time E-learning desperately requires sharing and further uses of knowledge resources, interoperability and different types of interactions among them and this is the place where the concept of grid technology is required. By recommending the use of grid computing as scalable, flexible, coordinated and secure

resource sharing among geographically distributed individuals and organizations [Foster, I and Kesselman, C. and Tuecke, S. 2001] based on E-learning, we will be able to define these matters. The middleware based on Grid Technology and recent developed web service [Newcomer, E. 2002] is capable to support developers an integrated architecture facilitating E-learning system development. Moreover to reuse the functionalities of grid services we have to implement the middleware as well as grid services. In this way, many organizations based on E-learning service and content providers will be able to take part for generating a very large scale integrated e-learning system.

ICT Infrastructure for supporting GRID

Grid Technology is completely based on Web services and web services completely depends on Internet. As a result Internet Backbone is the main source of Grid. As Grid technology will generate super computing power from distributed resources the internet line must be able to support those activities with maximum data rate. The Information and Communication Technology (ICT) of the world is divided in three parts. First the submarine cable that is the main backbone of the world which is the super high way of information and faster enough to support Grid as for example SEA ME WE-4 is the submarine cable that connects all most all countries of Asia, Middle East and Europe. This submarine cable provides at least 10 G.B Bandwidth to each of the countries which is sufficient enough to support Grid Technology. Second is the Fiber Optic Communication Backbone of respective countries. Every country has its own optical fiber backbone network. This should be sufficient enough to carry this 10 G.B bandwidth and should be capable enough to distribute 8 to 10 mbps to individual nodes by using wired or wireless broadband. This requirement is also necessary to access online application software's or online video streaming for E-Learning. Fortunately by January next Wimax will make a revolution for wireless broadband internet for rural area. As Grid with E-learning infrastructure can make an educational revolution for rural parts of the world all those countries must ensure internet availability with this ICT infrastructure to get maximum output.

Role of Grid Technology for Future E-learning System

E-learning is a revolutionary approach for spreading education around the world. The key elements of this technique are learner, author and administrator. Authors who may also be teachers are responsible for preparing content which is stored under the control of learning management system (LMS) and basically in a database [Vossen, G., P.Jaeschke, 2002]. Contents of E-Learning resources should be regularly updated and there should be facility to exchange contents among different resources or systems. Basically the Learning management system is maintained by an administrator. User or Learner can interface with this LMS any time from any place. These three key elements should be distributed in different systems. Different learners require different things. They have different ability, different target and different speed of learning. In this way

the learning groups are heterogeneous. For example a student of computing may want to run simulation software of robotics and another student may want to see the simulation result of chemical reactions. From this point of view to meet all the requirements of all learners E-learning environment must support integration of several resources and materials, the potential deviation from predetermined sequence of action [Casati, F., U.Dayal, Eds, 2002], personalization and adaptation and verifiability of work and accomplishment [Vossen, G., P.Oberweis, 2002]. Unfortunately most of the E-learning Systems are only trying to concentrate to reuse the resources for all rather than extending the size of the storage or computation power that can support simulation based learning technique. For example E-learning system [IMSContentPackagingSpecifications, 2005] is a traditional system that only concentrates on reusability rather than Web services let alone Grid technology. Very few E-learning systems are now trying to use the web service to enhance the integration of various applications that are very useful for E-learning. For example most of the Universities of USA and UK are offering Web service based architecture for E-learning to extend there quality of education all over the world using online. Now the fact is if we want to integrate the E-learning services of various universities of UK and USA definitely we have to require an intermediate platform that will search and select all of these E-learning services and will produce a giant product of knowledge that is the composition of all of these resources. The successful use of Web services for E-learning demands various types of application software's that will create user friendly environment between learners and E-learning resources to produce various types of services to learn quickly. For example information retrieval is an important feature of web based E-learning system because various types of learners may want to search for various information's from the millions of information's as a result search engine software has to be distributed to all learners or it has to be built on the E-learning Website that should be the ultimate choice for successful E-learning system [FU, 2004, FU.Y, Mostafa., 2004]. Moreover with the help of web service various types of simulation software can be installed with E-Learning systems specially for practical based education system. With the help of this simulation software learners will be able to do their lab based experiments residing in home and this may vanishes the criticism that E-learning may not effective for lab based education. Fortunately many organizations and universities have high quality learning resources with different simulation software's for different arenas. Unfortunately all of these valuable learning resources are distributed around the world .Next generation E-learning system must integrate all of these geographically distributed resources. But for this large volume of information and related application or simulation software's there is a necessity of large storage. On the other hand millions of learners can access or run these information's or simulation of application softwares.To manage these millions of activities definitely we need largest supper computing power which is not possible to build by the scientists at present or near future. From this point of view Grid is the technology that can combine storages as well as computing power of the processors from various computers, servers

of learning resources that are geographically distributed around the world and connected with web services. If it is possible with the help of Grid technology then it can be ensured that there will be no difference between on campus education and distance or E-learning rather than E-learning system based on Grid will dominate the class room or lab based education.

Grid Technology

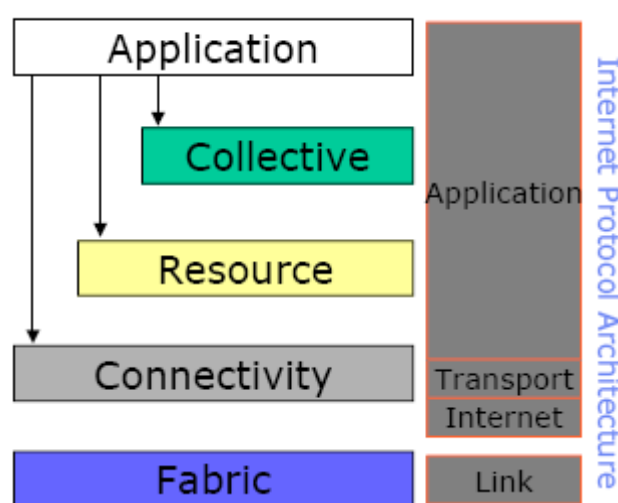
The concept of Grid has been taken from power grid of electricity where grid is the backbone to carry the main power that is generated from various sources and all users who have access to the power grid can share electricity without knowing from where they are getting the electricity. Like this in computing geographically distributed various types of resources like, CPU of computers, various types of servers or storages can be connected through internet with the help of web services to produce a super highway of information's, storages and computing power that can be shared by any user from any parts of the world. According to the IBM that most of the computers on the web are idle in most of the time. Grid technology can aggregate all of the powers from the CPUs to generate a super computing power that will help millions of users to access various types of services simultaneously. Grid computing unifies geographically distributed resources in such a way that it seems to be a large powerful computer. Grid computing extends this view to a large scale flexible, secure, coordinated resource sharing among dynamic collection of individuals, institutions and resources [Foster, I., C.Kesselman, S.Tuecke, 2001]. The next generation of scientific experiments and studies, popularly called as e-Science will be carried out by communities of researchers from different organizations that span national and international boundaries. However data grid service has to face two challenges of large datasets and multiple data repositories at distributed locations in data-intensive computing environments [Chervenak.A, Foster.I, Kesselman.C, Salisbury.C, and Tuecke.S 2000]

Web Services

Internet is the amazing invention for communication around the world but web service is the tool that makes internet as a service oriented technology for human beings. A web service is a stand alone software component that has a Unique Resource Identifier that works based on various types of standard protocols. Basically there are four types of protocols which are used by web services. Web service description language (WSDL) is used to specify the operations that are supported by web services [Christensen, E., F.Curbera, G.Meredith, S.Weerawarana, 2001]. The simple object access protocol (SOAP) is used to exchange structured data through web. Hyper text transfer protocol (HTTP) is used to help the function of (SOAP) protocol. Universal description discovery and Integration protocol (UDDI) is used to discover new service on the web [Bellwood, T., L. Clement, D. Ehnebuske, et al, 2002]

Integration of Grid Middleware for E-learning

Figure 1: Layered Grid Architecture



Basically grid technology is based on middleware that is used to provide all services of grid technology. Actually middleware acts as a protocol which is used for communication from source to destination and vice versa. This middleware is a combination of programs where communication technique has been defined. This middleware is divided in five individual layers which are responsible for individual tasks. The lowest layer is called fabric layer which will be responsible to provide uniform access to all E-learning resources like servers, storages that contain learning materials or softwares. It also ensures interoperability. This layer has to be installed locally in all learning resources to connect with Grid. Connectivity layer contains communication and security protocols to protect learning resources of grid. Cryptography algorithm is installed in this layer as a result authentication will be required for learner, teacher or for administrator. After authentication they will be able to access any resources. This grid security will also cooperate with local security technique of each resource. Resource layer is responsible to provide access to single learning resource and also look after the status of that resource. Collective layers coordinate global access for collection of resources. Calculation or information's are distributed in this layer. Finally application layer contains all Grid applications. The function of grid middleware has been already tested in various projects with Globus or Condor-G. Specially Globus toolkit with its Open Grid Service Architecture [Foster, I., C.Kesselman, J. Nick, S. Tuecke, 2002] has already been proved as a best and it should be the best for Grid on E-learning. To use the grid technology user just need Java enable internet browser.

4. Standard software and toolkits of Grid

A software toolkit is used for addressing key technical problems in the development of Grid enabled tools, services, and applications that actually contain a “bag of technologies”. This toolkit is also used for incremental development of Grid application and to implement standard Grid protocols and APIs that support open source license. This Globus Toolkit is responsible for authentication, scheduling, file transfer and resource description. Condor-G is also software that drives Globus toolkit to execute jobs of Grid or to solve any kind of computing related problem. Condor-G is the most successful software for Grid services and has been already implemented in various Grid related projects in Europe as it combines the strength of both Condor (Runs jobs within a single administrative domain) and Globus Toolkit (Runs jobs through many administrative domains).

4.1 Challenges of Grid Service

1. Poor ICT infrastructure to support Grid Technology especially in 3rd world countries who need it most.
2. Lack of awareness about the advantages of Grid
3. Lack of expert ICT professionals who may be capable of installing and maintaining Globus toolkit or Condor.
4. Grid Technology is not yet included as a major course in undergraduate and post graduate levels.
5. World class IT training providers like Microsoft, Cisco, Comptia have not yet started any professional certification course on Grid Technology.
6. Security may be a big challenge for Grid because attacks on Grid services allow attacker to get into any resources of the Grid.

4.2 Recommendations for Grid Service

1. Governments of the 3rd world countries have to implement a master plan of ICT that will support high speed internet access to support Grid Technology.
2. The main beneficiary of Grid technology based E-learning will be the developed countries as well as rural people. So First initiative has to be taken by them.
3. Ministry of education should come forward so that all schools, colleges and Universities can be registered under the same Grid platform for E-Learning.
4. Grid should be included as a major course in both under graduate and post graduate courses of computing or ICT.
5. Different ICT based organizations should send their ICT professionals for training to Europe or North America where Grid technology has been implemented.
6. It may be difficult to bring all E-learning resources of the world under same Grid platform but it is quite possible to bring the entire E-learning resources or Universities of SAARC region under same Grid Platform so that a cooperative learning and knowledge sharing environment can be established.

7. `Finally as security may be a major threat for Grid, organizations should take necessary initiatives to monitor the online security all the time.

5. Conclusion

Quality education is the first priority for sustainable socio-economic development and with the help of ICT E-learning is playing a great role for spreading online education around the world even in rural areas. Although ICT is reducing the difference between on campus education and distance education still there are few limitations of E-learning for Lab based education due to computation power. Fortunately Grid is the technology which can generate knowledge Grid with super computation power to access that knowledge by sharing distributed E-learning resources. It is the responsibility of scientists to concentrate on further research on Grid so that they can spread this light of the knowledge for the deprived people of rural areas of Asia and Africa rather than keeping it inside a small machine.

References

- BELLWOOD, T., L. CLEMENT, D. EHNEBUSKE, et al."UDDI Version 3.0", UDDI Spec Technical Committee Specification, 19 July 2002, <http://uddi.org/pubs/uddi-v3.00-published-20020719.htm>
- CASATI, F., U.DAYAL, Eds."Special Issue on Web Services", IEEE Bulletin of the Technical Committee on Data Engineering, 5(4)2002.
- CHERVENAK.A, FOSTER.I, KESSELMAN.C, SALISBURY.C, AND TUECKE.S (2000)"The data grid: Towards an architecture for the distributed management and analysis of large scientific datasets", Journal of Network and Computer Applications, vol. 23, no. 3, pp. 187–200.
- CHRISTENSEN, E., F.CURBERA, G.MEREDITH, S.WEERAWARANA."Web Service Description Language (WSDL) 1.1" W3C Note 15th March 2001, <http://www.w3.org/TR/2001/NOTE-wsdl-20010315>.
- FOSTER, I AND KESSELMAN, C.AND TUECKE, S. (2001)"The Anatomy of the Grid: Enabling Scalable Virtual Organizations", International journal of high performance Computing applications, 15(3).200-222.2001.
- FOSTER, I., C.KESSELMAN, J. NICK, S. TUECKE."The Physiology of Grid: Open Grid Service Architecture for Distributed Systems Integration", Proc. 4th Global Grid Forum (GGF4) Workshop 2002.
- FOSTER, I., C.KESSELMAN, S.TUECKE."The Anatomy of the Grid: Enabling Scalable Virtual Organizations,"Int. J of super Computer Application and High Performance Computing. 15(3) 2001, pp.200-222
- (FU, 2004) FU.Y, MOSTAFA.J, (2004) "Information Retrieval Web Services for Digital Libraries", Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries.
- "IMSContentPackagingSpecifications", (2005) Available <http://www.imsproject.org/content/packaging/index.html>
- NEWCOMER, E. (2002)"Understanding Web Services: XML, WSDL, SOAP, and UDDI", Addison-Wesley Professional; 1st edition (May 13, 2002).

The Globus Project, Introduction to Grid Computing and Globus Toolkit, www.globus.org.

VOSEN G., WESTERKAMP P. "E-Learning as Web Services (Extended Abstract)," Proc. 7th Int. Conf. on database engineering and application (IDEAS), Hong Kong, China, 2003, IEEE Computer Society Press.

VOSEN, G., PJAESCHKE. "Towards a Uniform and Flexible Data Model for Learning Objects", in Proc. 30th Annual Conf. of the Int. Bus. School Computing Assoc (IBSCA), Savannah, Georgia 2002, pp.99-129.

VOSEN, G., POBERWEIS, A. "Flexible Workflow Management as a Central E-Learning Support Paradigm", Proc. 1st European Conference on E-Learning (ECEL), Uxbridge, UK, 2002, pp.253-267.

3

Relationship between Information Systems Development Paradigms and Methods

Peter Nabende, Benjamin Ahimbisibwe, Jude T. Lubega

There are well established paradigms for information systems development but the methods used for Information Systems Development (ISD) have not been tied to most of these paradigms. Researchers have attempted to document the assumptions underlying different paradigms with the goal of making systems developers become aware of the assumptions and beliefs that they employ for a system development task. However, a number of Information Systems that have failed are as a result of lack of awareness by the information systems developers of some methods that should be used when dealing with an ISD problem. Information Systems researchers have not related ISD methods that can be used by practitioners to the identified ISD paradigms. In this paper, ISD methods have been classified under some of the major paradigms in form of a matrix. It is hoped that such classification will enable IS Developers easily identify methods they should use which in turn should lead to better quality of Information Systems, and a reduction in time and costs in Information Systems Development.

1. Introduction

Information is one of the most valuable assets of modern corporations. However, development of Information Systems (ISs) faces many problems. Some traditional problems include [Tolvanen 1998; Lycett et al. 2007]: an inadequate alignment of ISs with business needs, low productivity, and a large number of failures. In this paper, we attempt to relate some of the major IS development methods to major IS development paradigms with the aim of addressing some IS development problems.

The need to ground ISD methods in different paradigms has long been sought for. Before looking at ISD methods, Hirschheim and Klein [1988] attempted to document the relationship between systems development methodologies to paradigms, however, they found the process to be complicated because different methodologies such Soft Systems Methodology had properties of more than one paradigm. This should have been obvious to Hirschheim and Klein since the methodologies that they considered comprised of methods based on different assumptions. In the field of information systems development, Hirschheim [1989] proposed grounding existing ISD methods by analyzing them with regard to different paradigms. According to Hirschheim [1989], “a paradigm is a set of assumptions adopted by a professional community that allows its members to engage in commonly shared practices”. It is these assumptions that play a central role in guiding information systems development. Several researchers in the IS field later on discovered that although ISD methods were used extensively, they were actually not widely applied [Hardy et al. 1995; Wynekoop and Russo 1993]. To make it

easier to adopt suitable and specific ISD methods, Iivari et al. [1999] proposed a way of classifying ISD methodologies based on their features leading to comprehensible approaches that organizations could use to meet their business needs.

In this paper, we adapt the classification method used by Iivari et al. [1999] to relate a set of ISD methods to ISD paradigms. The paper is organized as follows: In section 2 we review the four traditional paradigms for ISD; in section 3 we review a set of well known ISD methods and summarize their characteristics in a table; in section 4, we map the ISD methods presented in section 3 to ISD paradigms to generate a relation as described in section 5. Section 6 concludes the paper and proposes future work.

2. Information Systems Development Paradigms

There exist different definitions for the term ‘paradigm’; however, the most popular is that of Kuhn [1962]: “A paradigm is a set of common beliefs and agreements shared between scientists about how problems should be understood and addressed”. This definition is clearly appealing to the aim of the work reported in this paper. Hirschheim and Klein [1989] also defined a paradigm as “the most fundamental set of assumptions adopted by a professional community that allows its members to share similar perceptions and engage in commonly shared practices”. Hirschheim [1989] proposed four paradigms for Information Systems Development which were initially used in organizational and social research [Burrell and Morgan 1979]. The paradigms (figure 1) include [Hirschheim, 1989]: functionalism, social relativism, radical structuralism, and neo-humanism. These paradigms are used in this paper as a starting point for classifying ISD methods. Table 1 gives the emphases and underlying assumptions of the paradigms.

Fig. 1. Information Systems Development paradigms (Adapted from [Burrell and Morgan 1979]).

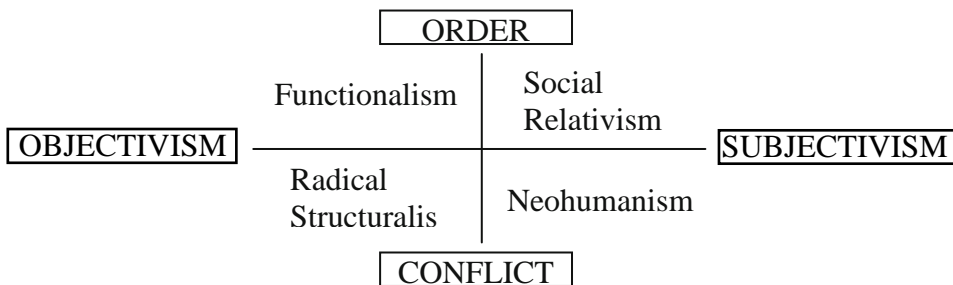


Table I: Features of ISD Paradigms (Adapted from [Hirschheim 1989])

Paradigms Features	Functionalism	Radical Structuralism	Social Relativism	Neohumanism
Emphasis	Explains status quo, social order, social integration, consensus, needs satisfaction and rational choice	Emphasizes the need to overthrow or transcend the limitations placed on existing social and organizational arrangements. Focus primarily on the structure and analysis of economic power relations	Seeks explanation within the realm of individual consciousness and subjectivity and within the framework of reference of the social actor as opposed to the observer of the action	Seeks radical change, emancipation and potentiality and stresses the role that different social and organizational forces play in understanding change. It focuses on all forms of barriers to emancipation (ideology, power, and psychological comparisons)
Assumptions	Epistemology: developer gains knowledge about the organization by searching for measureable cause-effect relationships Ontology: it is believed that there exists an independent, empirical organizational reality	Epistemology: dialectic inquiry in the specific form of a materialistic view of history and society Ontology: realism reflecting the belief in a pre-existing empirical reality	Epistemology: anti-positivism (that is, will and need to make sense of oneself and the situation) Ontology: Reality is not given, but socially constructed	Epistemology: Positivism for technical control, anti-positivism for mutual understanding and emancipation Ontology: realism for technical control, social construction for mutual understanding and emancipation

3. Information Systems Development Methods

Generally, a method is defined as a systematic way of doing something. In Information Systems literature, a development method is considered to be comprised of a set of procedures, techniques, tools, and documentation aids that help in developing an information system [Smolander et al. 1990]. Olle et al. [1991: 1-2] also define a development method as “a methodical approach to information systems development used by one or more persons to produce a specification or design product by performing a design process”. A method may be broken down into phases, and the phases into sub-phases that help system developers choose procedures, techniques, tools, etc. that are appropriate at each stage in a systems development project [Avison and Fitzgerald 2003: 20].

The necessity to classify ISD methods under ISD paradigms is due to the fact that in using a particular method, one is making an implicit or explicit assumption about the nature of the world and knowledge. As we have already seen, it is these assumptions that are associated with a given paradigm. Examples of ISD methods include Structured Analysis and Design [Yourdon 1989] and object-oriented methods of Booch [1991] and

Rumbaugh et al. [1991]. In the next subsections, we identify some of the ISD methods in existence and construct a table summarizing their features which are used for relating to ISD paradigms.

3.1 Structured Methods

Structured development methods arose out of the information systems community to help deal with various problems associated with the mismatch between information systems developed and the business requirements the systems were supposed to meet. The most common example is the Structured Systems Analysis and Design Methodology (SSADM) [Yourdon 1989]. These methodologies formalize the requirements elicitation process to reduce chances of misunderstanding the requirements and use best practice techniques to the analysis and design process.

3.2 Object-oriented Methods

Object-oriented methods find their origins from object-oriented programming [Rojas 1996]. Walker [1992] identified four main criteria that should be considered by any object-oriented method: abstraction, notation, and representational verisimilitude, validation, and reuse. Examples of well known object-oriented methods include: Object Oriented Development [Booch 1991], Hierarchical Object Oriented Development [Rumbaugh 1991], Object Modeling Technique [Rumbaugh 1991], Responsibility Driven Design [Wirfs 1990], Object Oriented Analysis [Coad and Yourdon 1990].

3.3 Rational Unified Process

The Rational Unified Process (RUP) is as a result of merging different object-oriented methods including the objectory process from Ivar Jacobson and those of Booch [1991] and Rumbaugh [1991]. The main idea behind the Rational Unified Process is to start small and iteratively build on that. The three main building blocks include: roles, work products, and tasks. More details about RUP can be found in the Rational Software White paper [Rational 2001].

3.4 Agile Development Methods

Agile development methods were initially aimed at developing software in a lighter, faster, and more people centric way [Beck et al. 2001]. Based on the goals identified, a number of principles were put forth concerning agile software development. Schuh [2004] gives examples of the best known of a growing number of software development methodologies that are based on the principles underpinning agile development: Adaptive Software Development (ASD); the Crystal methodologies; Dynamic Systems Development method (DSDM); Extreme Programming (XP); Feature-Driven Development (FDD); Lean Software Development.

3.5 Scenario Requirements Analysis Method (SCRAM)

The SCRAM is a requirements elicitation method which employs scenarios and prototypes to help users relate a design to their work or task context and there after develop requirements [Shin et al. 2003]. Scenarios relate to real world experiences and

can be expressed in different ways including using natural language, pictures, or other media [Gough et al. 1995; Carroll 2000]. The main goal of the method is to have a requirements specification in which user preferences for different design options are expressed [Sutcliffe 1997].

3.6 Summary of the characteristics of selected ISD methods

Table II shows a summary of the characteristics of some of the major ISD methods considered for classification under ISD paradigms.

4. Method For Relating ISD Methods To ISD Paradigms

The method we use for relating ISD methods to ISD paradigms is an adaptation of a procedure used by Iivari et al. [1999] for classifying ISD methodologies under ISD approaches. The basic idea of the method is to identify ISD paradigms of which an ISD method is an instance. We check whether the ISD method fits under an ISD paradigm by analyzing the correspondence of the features of the ISD method and paradigms through tables I and II respectively. If there is no correspondence that can be identified for a given ISD method, then we check if an existing paradigm can somehow be modified or generalized to assimilate the ISD method. The features of the new modified paradigm must correspond to the features of the ISD method before a relationship can be established between the method and paradigm. If at the end of modifying a given paradigm, there still exists no correspondence between the features of the method and the modified paradigm, a new paradigm has to be proposed associated with the ISD method. However, we do not attempt proposing new paradigms in this paper. Otherwise, the ISD method would first be developed to an ISD methodology and later to a specific ISD paradigm.

Table II. ISD Methods and their characteristics¹

Methods Characteristics	SSADM	Object-Oriented methodology	Agile Development Method	Soft Systems Methodology	SCRAM
Goal	To provide methods which help high quality (reliable and maintainable) software in a productive way	To provide a method which helps to ensure that the products are delivered to the user on time and within budget, that the products meet user requirements, that user request to modify the system and/or fix bugs are responded to in a timely fashion that increasingly sophisticated products are offered so as to keep competitive edge that the changes in standards and delivery technology are kept up and the project team feels motivated and successful	To provides methods for building applications in a very short amount of time; traditionally with compromises in usability, features and/or execution speed.	To provide learning methods to support debate on desirable and feasible changes	Requirements Acquisition, Modeling and Analysis [Zhu and Jin 2004]
Principles and Assumptions	Separation of the essential model from the implementation model; Careful documentation to make the development process visible; Graphical notations; Top-down partitionable transformation / process models to hide complexity; unambiguous minimum redundant graphic specification; Balancing of models; Design modules with high cohesion and weak coupling	Seamless analysis, design and Implementation; Encapsulation; Information (implementation) hiding	Active user Involvement; Teams must be empowered to make Decisions; Focus on Frequent Delivery; Fitness for Business is Criterion for Accepted Deliverables; Iterative and Incremental Development is Mandatory; All Changes During Development Must Be Reversible; Requirements are Baselined at High-Level; Testing is Integrated throughout the Lifecycle; Collaborative and Co-operative Approach	Use of notional system modules called 'human activity systems' to illuminate different philosophies which may be applied to any social system; An information system is a system to support the truly relevant human activity system	Active User involvement for acquisition of requirements.
Fundamental concepts	Essential model vs. Implementation model; Transformation; Data flow; Data Store; Terminator; Module; Cohesion; Coupling	Problem domain vs. Implementation domain; Object and Class; Encapsulation; Information hiding; inheritance; Polymorphism; Communication between objects		Philosophy; Human Activity Systems; Root definition; Relevant system	A scenario is considered as a continuum of the real world descriptions and stories to models and specifications

¹ Some of the characteristics associated with ISD methods in table II are adapted from work by Iivari et al. [1999]. References are also shown for other specific sources.

The classification method used by Iivari et al. [1999] follows the logic of a single hierarchy. First, the method checks whether any of the candidate ISD paradigms share a strict subset of the features with a new paradigm abstracted from an ISD methodology. If it does, then the new paradigm is inserted as a sub-paradigm. Next the method checks whether any of the candidate paradigms can be generalized or modified so that the resultant paradigms share a strict subset of features of the new paradigm formed from the ISD methodology. If so, the new paradigm formed from the ISD methodology is inserted in the resultant paradigm. For our case, we only associate ISD methods to ISD paradigms by looking at features common to both, and use a scoring scheme to indicate the level of relationship between a given ISD method and paradigm. We specifically find relationships with regard to ISD methods and not ISD methodologies since some of the ISD methods can be constituents of a larger ISD methodology and yet have completely different features. One obvious advantage that results from this is that establishing the degree of relationship between ISD methods and ISD paradigms simplifies establishing the degree of relationship between various ISD methodologies and ISD paradigms.

5. Results

From an analysis of the characteristics of the ISD methods shown in table II, and the assumptions and emphases of ISD paradigms shown in table I using the method introduced in section 4, we related ISD methods to paradigms as shown in table III.

Table III. Relationship between ISD methods and paradigms

Paradigms Methods	Functionalism	Radical Structuralism	Social Relativism	Neohumanism
SSADM	✓✓✓	✓✓✓	✓	✓
Object-Oriented Method (OOM)	✓	✓	✓✓	✓
Rapid Application Development (RAD)	✓	✓	✓	✓
Soft Systems Method (SSM)	✓	✓	✓✓	✓✓✓
Scenario Requirements Analysis Method (SCRAM)	✓	✓	✓✓✓	✓✓
Agile Method (Dynamic Systems Development Method - DSDM)	✓	✓✓	✓✓	✓✓

Key

✓✓✓ More fitting; ✓✓ Averagely fitting; ✓ Less fitting; **None** Not Applicable

Table III shows the level of relationship between ISD methods and ISD paradigms as specified by the Key. Given a new method and a new paradigm the same method used for obtaining the relationships can be used. An example to explain the results in table III follows. SSADM fits more under Functionalism and Radical Structuralism paradigms because most of its characteristics are found to fit within the assumptions and emphases of the two paradigms. SSADM, however is found to be less suitable for use under the Social relativism and Neohumanism paradigm, averagely fitting when used under the Social relativism paradigm, and less fitting when used under Functionalism and Radical Structuralism paradigm. This explanation can be used for the rest of the ISD methods.

5. Conclusion

We have determined the level of usability of some ISD methods under different ISD paradigms. Not all ISD methods and methods were analyzed in this paper. However, a similar approach can be used for all the established ISD paradigms and widely used ISD methods. The approach we adapted for obtaining the levels of relationships between ISD methods and paradigms takes into consideration emerging new ISD methods and the possibility of proposing new paradigms. Obtaining relationships between ISD methods and paradigms also serves to enrich on the information that is required to map ISD methodologies or approaches to ISD paradigms. As future work, we hope that a general framework for categorizing methods and paradigms can be developed.

Acknowledgments

Many thanks go to PhD students of 2006 intake at the Faculty of Computing and Information Technology, Makerere University whose company and comments helped in shaping this paper.

References

- AVISON, D.E. AND FITZGERALD, G. (2007). Where now for Development Methodologies?. *Commun. ACM*, 46(1): 78-82.
- BECK, K., BEEDLE, M., BENNEKUM V.A., COCKBURN, A., CUNNINGHAM, W., FOWLER, M., GREENING, J., GRENNINGS, J., HIGHSMITH, J., HUNT, A., JEFFRIES, R., KERN, J., MARICK, B., MARTIN, C.R., MELLOR, S., MELLOR, S., SCHWABER, K., SUTHERLAND, J., THOMAS, D. 2001. *Principles behind the Agile Manifesto*.
- BIGGS, P. 1994. A Survey of Object-Oriented Methods. University of Durham. Retrieved May 10th, 2007 from <http://students.cs.byu.edu/~pbiggs/survey.html>
- BOOCH G. 1991. *Object-Oriented Design with Applications*, Benjamin Cummings.
- BURRELL, G. AND MORGAN, G. (1979). *Sociological paradigms and Organizational Analysis*, London: Heineman.

- CARROLL, J.M. (2000). *Making Use: Scenario-Based Design of Human-Computer Interactions*. Cambridge MA: MIT Press. In SUTCLIFFE, A. 2003. Scenario-based Requirements Engineering. *Proceedings of the 11th IEEE International Requirements Engineering Conference*, 320-329.
- COAD, P. AND YOURDON, E. 1990. *Object-Oriented Analysis*. 2nd Edition, Prentice Hall.
- COCKBURN, A. 2001. *Writing Effective Use Cases*. Boston MA: Addison-Wesley. In SUTCLIFFE, A. 2003. Scenario-based Requirements Engineering. *Proceedings of the 11th IEEE International Requirements Engineering Conference*, 320-329.
- GOUGH, P.A., FODEMSKI, F.T., HIGGINS, S.A., AND RAY, S.J. (1995). Scenarios: An Industrial Case Study and Hypermedia Enhancements. *IEEE International Symposium on Requirements Engineering*, Los Alamitos CA, 10-17, IEEE Computer Society Press. In SUTCLIFFE, A. 2003. Scenario-based Requirements Engineering. *Proceedings of the 11th IEEE International Requirements Engineering Conference*, 320-329.
- HARDY, C.J., THOMPSON, J.B. AND EDWARDS, H.M. (1995). The use, limitations and customization of structured systems development methods in the United Kingdom. *Information and Software Technology*, 37 (9): 467-477.
- IIVARI, J., HIRSCHHEIM, R., AND KLEIN, K.H. (1999). Beyond Methodologies: Keeping up with Information Systems Development Approaches through Dynamic Classification. *Proceedings of the 32nd Hawaii International Conference on System Sciences*, IEEE.
- KUHN, T. 1962. *The Structure of Scientific Revolutions*. University of Chicago Press.
- LYCETT, M., MARCOS, E. AND STOREY, V. (2007) Model-driven Systems Development: An Introduction. *European Journal of Information Systems*, 16: 346-348.
- MACLEAN, A., YOUNG, R.M., BELLOTTI, V.M.E AND MORAN, T.P. (1991). Questions, options and criteria: elements of design space analysis. *Human-Computer Interaction* 6, 3 & 4, Special issue on design rationale edited by J.M. Carroll and T.P. Moran.
- OLLE, T.W., HAGELSTEIN, J., MACDONALD, I.G., ROLLAND, C., SOL., H.G., VAN ASSCHE, F, VERRIJN-STUART, A.A., (1991). *Information Systems Methodologies: A Framework for Understanding*, 2nd edition, Avon: Addison-Wesley Publishing Company, The Bath Press.
- RATIONAL SOFTWARE CORPORATION (RSC). (2000). Rational Unified Process: Best Practices for Software Development Teams. Retrieved June 1st, 2007 from http://www.augustana.ab.ca/~mohrj/courses/2000.winter/csc220/papers/rup_best_practices/rup_bestpractices.pdf
- ROJAS, T. (1996). Object-oriented Information Systems Development Methodologies by Means of Reverse Engineering. *Americas Conference on Information Systems*, Phoenix, Arizona, USA.
- RUMBAUGH, J., BLAHA, M., PREMERLANI, W., EDDY, F. AND LOENSEN, W. (1991). *Object-Oriented Modeling and Design*. New York: Prentice-Hall International Editions.
- SCHUH, P. 2004. *Integrating Agile Development in the Real World*. New York: Charles River Media.
- SHIN J.E., SUTCLIFFE, A.G. AND GREGORIADES, A. (2005). Scenario Advisor Tool for Requirements Engineering. *Requirements Engineering*, 10 (2): 132-145.
- SMOLANDER, K., TAHVANAINEN, V. P., LYYTINEN, K. (1990). How to Combine Tools and Methods in Practise - a Field Study. In STEINHOLTZ, B., SØLVBERG A. (eds) *Lecture Notes in Computer Science, Second Nordic Conference CAiSE'90*, Stockholm, Sweden, May, 195-211.

- SUTCLIFFE, A.G. AND RYAN, M. (1998). Experience with SCRAM, a Scenario Requirements Analysis Method. *In Proceedings of Third International Conference on Requirements Engineering*, Colorado Springs, CO, USA, 164-171.
- SUTCLIFFE, A.G., MAIDEN, N.A.M., MINOCHA, S. AND MANUEL, D. (1998). Supporting Scenario-based Requirements Engineering. *IEEE Transactions on Software Engineering*, 24: 1072-1088. In Sutcliffe, A. 2003. Scenario-based Requirements Engineering. *Proceedings of the 11th IEEE International Requirements Engineering Conference*, 320-329.
- SUTCLIFFE, A. (2003). Scenario-based Requirements Engineering. *Proceedings of the 11th IEEE International Requirements Engineering Conference*, 320-329.
- TOLVANEN, J-P. (1998). *Incremental Method Engineering with Modeling Tools: Theoretical Principles and Empirical Evidence*. Phd Thesis, Department of Computer Science and Information Systems, University of Jyväskylä, Jyväskylä.
- WALKER, I.J. (1992). Requirements of an Object-Oriented Design method. *Software Engineering Journal*, 7 (2):102-113.
- WIRFS-BROCK, R. (1990). *Responsibility Driven Design*. Prentice Hall.
- WYNEKOOP, J.L. AND RUSSO, N. (1993). Systems development methodologies: Unanswered questions and the research-practice gap, in DEGROSS, J.I., BOSTROM, R. AND ROBEY, D. (eds.), *Proceedings of the Fourteenth International Conference on Information Systems*, Orlando, FL, 181-190
- YOURDON, E. (1989). *Modern Structured Systems Analysis*. Englewood Cliffs, NJ: Prentice Hall.
- ZHU, H. AND JIN, L. (2004). Scenario Analysis in an Automated Tool for Requirements Engineering. *Requirements Engineering*, 5 (1): 2-22, London: Springer.

4

From the Ground Up: User Involvement in the Development of an Information System for use in HIV Voluntary Counselling and Testing

Kathy Lynch and Stacey Lynch

This research investigates end-user involvement in the development of a simple HIV Voluntary Counselling and Testing (VCT) data collection instrument and associated spreadsheets. The research involved end-user participation in developing a paper-based VCT data collection instrument, followed by the development of, and training in an associated electronic spreadsheet. The collection of data together with the electronic storage and retrieval of the data constitute, however simple, an information system. The findings show that involving the end-user in the development and use of something fundamental to their mission, not only gives them something tangible that can assist in developing strategies for identifying groups for VCT sessions, manage resources and data, and generate reports, but can also encourage the collection of useful and complete data, invoke a sense of achievement and positive prospects for the future. The study involved working with end users from six rural Ugandan NGOs who deliver VCT in rural communities in the Mukono district.

*Categories and Subject Descriptors: H.4 Information Systems Applications
General Terms: Information systems development*

1. Introduction

The killer disease of the 20 and 21 centuries is undoubtedly AIDS resulting from a HIV infection. The disease has spread rapidly across the globe; in 2006 sub-Saharan Africa was the most HIV infected region in the world with nearly 24.7 million people living with HIV – nearly two-thirds of the global burden [UNAIDSa n.d.]. In Africa as whole, the death rate in 2007 is reported to be 7 million [WikiAnswers n.d.]. These figures are more staggering when you look at the results alongside findings that only 12% of men and 10% of women in the general population had been tested for HIV and received their results [WHOa n.d.]. In addition, many of the men and women who seek HIV testing and counseling, are already in the advanced stages of the disease [Hogg, et al. 2006]. The WHO and UNAIDS proclaim that “early diagnosis presents an opportunity to provide people with HIV the information and tools to prevent HIV transmission to others” [WHOb n.d p16]. A voluntary counselling and testing (VCT) program has been formulated by the WHO and UNAIDS to encourage people to be pre-HIV test counselled, tested and post-HIV test counselled in an endeavour to prevent infection and transmission of HIV.

In Uganda, the occurrence of new cases of HIV is on the decline (15% in the early 1990's, 6.4% by 2004 [UNAID 2006]. Although the 6.4% figure is a great achievement over the 18 years, there is a significant difference between urban and rural prevalence; 10.1% and 5.7% respectively [UNAIDSa n.d]. It is also reported that HIV prevalence is also higher among women (7.5%) than men (5.0%) [UNAIDSb n.d]. These figures may however be misleading as data has been generated from only a small proportion relative to Uganda's population size with the rural/up-country regions being under-represented.

HIV counselling and testing is one of the factors claimed to have brought about the decline in HIV prevalence. (Other factors are the Ugandan government's advocacy of ABC -Abstinence, Be faithful, and Condom use). The acceptance of the delivery of HIV counselling and testing as a common and routine event in a community should increase an understanding of HIV and living with and preventing HIV transmission, as well as a probable increase in the uptake of HIV testing.

The Ugandan government in 1990 founded the AIDS Support Organisation (TASO). TASO is a critical link in HIV counselling and testing, and together with more than approximately 2000 of the non government organisations (NGOs) in Uganda are active in HIV work [Coutinho 2003] by informing, counselling and testing the population. The large-scale delivery of VCT however, is only one step towards addressing the HIV pandemic; scaling up best practice of VCT is another critical step [Coutinho 2003]. The findings presented in this paper hope to demonstrate one example of best practice in VCT data collection and management procedures.

This paper is organised as follows. Section 2 presents the context of the study - an overview of VCT – specifically in Uganda. Section 3 depicts the importance of the end-user in the design and development of information systems – even when they are primarily paper-based. Section 4 outlines the study design and results. The paper concludes with a discussion of the known benefits to-date of the information system, and its possible impact on surveillance of HIV in rural communities and deliverers of VCT.

2. HIV VCT

A greater knowledge of HIV status within a community (and/or nation) is critical to expanding access of HIV treatment, care and support in a timely manner as it offers people living with HIV an opportunity to receive information and tools to prevent HIV transmission to others [WHOa n.d]. The most successful method of gaining knowledge about HIV has been through Voluntary Counselling and Testing (VCT).

HIV VCT voluntary pre-test counselling, testing, and post-test counseling, it is “the process of providing counseling to an individual to enable him or her to make an informed choice about being tested for HIV..... VCT is an entry point for prevention and care” [Family Health International 2004 p5]. It is an efficient and cost-effective strategy in expanding access to prevention, treatment and care services. It facilitates behaviour change [Family Health International, 2004], reduces the stigma attached to those who live with HIV/AIDS [Coutinho 2003], and early VCT can lead to a delay

in HIV deaths [Family Health International 2004]. Knowing your HIV status (that is whether positive/reactive or negative) is important, as it assist in the way you behave and thus affecting the state of your life style and health. Undertaking VCT offers a personal and individual way to learn of ones HIV status, and a way to maintain a negative status or if HIV positive, a way of maintaining a quality of life as well as access to treatment, care and knowledge on how to prevent transmission to others.

It is disheartening to know, that even with over 2000 Ugandan NGOs who are active in HIV work, including working in VCT programs, there are still many Ugandans who do not know their HIV status [UNAIDSb n.d.]. Table I presents the results of a 2003-2005 study showing the percentages of all men, all women, HIV-positive men and HIV-positive women (aged 15–49 years) in Uganda who were ever tested for HIV and received the results (the number of participants was not reported in the report).

Table I. HIV status

Country Uganda: Date of survey 2004	
percentage of all men who knew their status	12.7
percentage of HIV-positive men who knew their status	23.5
percentage of all women who knew their status	10.8
percentage of HIV-positive women who knew their status	15.0

[source - Demographic and Health Surveys, MEASURE DHS, 2003–2005, cited in WHOa n.d. 2007 p35]

The report also showed that in most countries of sub-Saharan Africa, knowledge of status was higher among people living in urban areas than among those living in rural areas; indicating that more needs to be done for those who live in rural regions. Furthermore, even in settings where VCT is routinely offered, for example, programmes for prevention of mother-to-child transmission, the number of people who avail themselves of these services remains low in many countries. Stigma and discrimination are the two main factors that continue to stop people from having a HIV test [WHOb n.d.].

Health organizations worldwide are advising governments that the large-scale delivery of VCT is only one step towards addressing the HIV pandemic; scaling up best practice of VCT is another critical step [WHOa n.d. pg 48]. One way of doing this is in improving on the collection, reporting and dissemination of useful data regarding VCT. However, in practice, and in particular in rural areas where the majority of VCT is conducted by grass-root NGOs, formal - let alone valid and useful data, has rarely been collected with little analysis and dissemination. One of the issues in collecting the data is that many of the organisations conducting the VCT do not have an acceptable nor simple data collection instrument that they understand, nor a way to correlate the data into information that is useful to them. Often is the case ‘why and what to record in the note book?’ if there is no direct benefit. The study presented here, aimed to give

meaning to collecting data during VCT with the objective of benefiting those who collected it and their 'clients' through involving the end-user in the development of a simple data collection and analysis system.

3. End User Involvement In Information Systems Design

Information systems design is a complicated and time consuming process. It is considered to be the job of professionals. Regardless of how good the professional or practitioner is, many systems fail or are not used in the way that the system designer had in mind. Conversely, those information systems that are solely developed by the end-user are often missing technical and system features that would not only improve the robustness of this system, but also improve the technical efficiency of the system. It has long been argued that the participation of the end-user in the design of the information system tends to result in a more effective, efficient and accepted system. Collecting the user-requirements for a system is a challenging task in systems development [Pekkola et al 2006], "This includes information about the domain and context specific technical issues, and about multifaceted cultural, political, communicational, motivational, and personal issues" [ibib p21]. A development team that intricately involves both the end-user and system designers, though not a guarantee, has a greater chance of success of obtaining accurate information as stated by Pekkola et al [2006] than without the cooperation and collaboration between the two parties.

Systems development begins with an analysis of what is required of the system together with interaction with the end-user to obtain the data for the system. Though we are in the age of the 'paperless office' this is rarely the case, and many data items that are found in a system have initially be captured manually on a paper form. One of the disheartening trends in today's information systems development is that often the development of any associated paper-based data capturing form is produced with no more than a token gesture to end-user input (if at all). A way of giving meaning to a data collection instrument is to give a tangible reason for its collection. Involving the end-user in the design of the system –even a paper system, can improve the systems usability and acceptance. Without the end-user being involved in the design of the system, it can become less relevant, under-utilised or abused. Tushabe et al [2008] found that the functional usage of customized software in Uganda by Ugandan companies is approximately 42% [ibib p27]. An interesting figure but not a surprising one as the study also reports that only 12% of customized software in Ugandan organizations is locally developed [ibib p25]. The respondents indicated that one of the reasons for under-utilisation was that the end-users are "not playing a sufficient role within the software development process" [ibib p27]. Their study supports the hypothesis that if VCT counsellors were involved in the design of a system to assist them in the collection of data obtained during VCT, then they would collect more reliable data, as well as use the system more efficiency and effectively.

4. The Study

The study is situated in rural Uganda - one of the poorest countries in the world (GPD – per capita in 2007 was US\$1000, [CIA n.d.]), with little national infrastructure - however, it has made vast improvements within the last 10 years to alleviate the burden of poverty, disease, poor sanitation and limited utilities. HIV VCT has been regarded as a means to reduce the transmission of HIV. To be informed as to the state of HIV in Uganda (or any country for that matter), useful data needs to be collected, reported and used by those who are at the front line delivery of VCT.

This study hypothesised

1. VCT counsellors will collect more reliable and usable data if they are involved in the design of the data collection instrument.
2. The data generated could be used to inform planning.

Building on the hypotheses, the aim of the study is to develop a useful information system that could be used in rural Uganda for collecting VCT data.

Data obtained could then be used to:

1. Obtain accurate and informative data on HIV prevalence within villagers by gender, age, occupation and sexual behaviour.
2. Develop strategies to enable a more efficient and effective use of limited resources in the fight against the spread of HIV.

4.1 Study Design


The study is longitudinal in nature, using an ethnographic action research approach with its beginnings in investigating then discussing the user requirements for an information system to improve the efficiency and effectiveness of data collection during VCT.

The study was conducted in two stages:

Stage One consisted of end-user involvement in the design of a simple paper-based data collection form for use in the VCT process. From this form an electronic spreadsheet was developed, the NGO workers were trained in using the spreadsheet and in the importance of completed forms and participant's consent. The NGO involved in this stage was Voluntary Services Trust Team (VOLSET). VOLSET concentrates its VCT in rural villages in and around Lake Victoria, and in the Mukono sub-county. (A detailed description of the design of this stage can be found in Lynch, Lynch and Bazira, 2008.)

The VCT data collection form was printed on a single page, the questions were written in plain English, and there was adequate space to write the responses (see Figure 1).

Figure 1. VCT data collection form (version 1)



 VOLSET-VCT

Test Number: <input type="text"/>	
[format is test number / number of tests conducted that trip, eg. 04/45]	
Date: <input type="text"/>	Venue: <input type="text"/>
Name of Counselor: <input type="text"/>	Counselor Signature: <input type="text"/>
Client's Name: <input type="text"/>	Sex: <input type="text"/> Age: <input type="text"/>
Marital Status: <input type="text"/>	Occupation: <input type="text"/>
No. living biological Children: <input type="text"/>	
Physical address: <input type="text"/>	
Village: <input type="text"/> Parish: <input type="text"/> Sub-County: <input type="text"/> District: <input type="text"/>	
<input type="text"/>	
Postal Address / Tel number: (if applicable): <input type="text"/>	
No. Previous tests: <input type="text"/>	Date / Number of last test: <input type="text"/>
Estimated number of partners: <input type="text"/> 0 / 1-3 / 4-8 / >9	
How often does the client use condoms: <input type="text"/> na / never / usually / always	
Results: <input type="text"/> TRN / <input type="text"/> TRR	Test Number (repeated): <input type="text"/>
Name of Tester: <input type="text"/>	Signature: <input type="text"/>
Results entered into tracking spread sheet: <input type="text"/> NA / Yes	Signature: <input type="text"/>

The questions included— place of residence, demographic data (age, gender, occupation, fertility history), sexual behaviour data (marital status, number of partners, condom use), and date of last HIV test. Each of these variables were deemed necessary by the VCT counsellors as important in understanding their clientele, and therefore enabling them to assist them further if necessary. It is interesting to note that even though the participants were not aware of all (if any) of the detail contained in WHO or UNAIDS publications on guidelines for collecting VCT data, (such as WHOc n.d. p20-22), there were similarities between many of the variables identified and that recommended by WHO or UNAIDS. An Excel spreadsheet was constructed to accompany the data collection form. To reduce data entry errors caused by spelling or typing mistakes, the spreadsheet included drop-down lists from which to selected repeated data (such as gender, martial status, occupation, village). This technique also streamlined responses into categories for easy analysis.

Stage Two related to the refinement of the initial system (paper and electronic) by seventeen VCT counsellors from six different VCT providers who deliver VCT in rural parts Mukono sub-county (Uganda). In order to obtain feedback and improvements on the paper-based system, a focus group discussion was held during August 2008. At

the session, the participants were given a copy of the instrument. The conduct of the focus group was facilitated by a native English speaker with little Luganda, however a bi-lingual Ugandan research assistant was employed to take extensive notes of the discussion, and to translate if there was a need. After the session, the notes recorded were reviewed to extract repeating ideas, which were then coded. Analysis of the coded data was undertaken to highlight emerging themes. Ethics for the study was granted by the University of the Sunshine Coast (Australia) and the study was approved (and ethics approval sighted) by the Uganda National Council for Science and Technology.

4.2 Results

Stage One is reported in Lynch, Lynch and Bazira (2008), however, a summary is repeated here to present a full picture of the study. VOLSET started collecting data during VCT in August 2004. Exercise books were used to record participants details such as name, village, HIV result, and the data owner's signature. Between August 2004 and February 2006, 218 responses were collected, however less than 30% of responses were usable for analysis due to incomplete data. Between March 2006 and December 2007 over 1000 people had their data recorded using this refined collection method, as a result of participating in VOLSET's VCT program. Approximately 95% of the responses were usable; that is, the completion rate of the forms in regard to basic demographic data was high. Through discussion with VOLSET VCT counsellors it was discovered that this increase was due to the simplicity of the form, knowing what data to collect, the importance of complete data sets, and the usefulness of the information derived from the data.

Stage Two began in August 2008 with the refinement of the VCT data collection instrument by seventeen VCT counsellors from six organisations that deliver VCT to rural Ugandans in the Mukono sub-county. The Director of VOLSET informed the other participants the value of the summary information obtained from data collected using the VCT forms, and how this analysis has assisted his organisation

"we send it [the summary information] to organisations which have funded us for that programme or we also give it to the Government like the health centre in Ntenjeru and then Mukono District, Ministry of Health. And by doing so, our report works as an eye opener to the Government or to the local authority so that they can think about those isolated areas where we go HIV testing." [Festus Bazira]

Further comment from those who are VCT counsellors with VOLSET added that the reports assisted them in who to target for their VCT sessions, *"men or women, or villages that are not coming for HIV testing at all or VCT counselling."* Another counsellor indicated that it helped manage what they needed to bring to a VCT session, some sessions are conducted on an island that is over a 5 hour canoe ride *"It gives you an indication of the time required, how many testing kits you need to bring next time or how much ARVs are needed - things to bring along"*

Two new variables were highlighted as important information to collect and generated much debate and were discussed at great lengths during the focus group session. These were the inclusion of the participant's CD4 count, and their consent to their data being recorded.

CD4 count. The immune system contains different types of cells that help protect the body from infection. One of these specialized cells are called the CD4⁺ T-cells. HIV attacks these T-cells and uses them to make more copies of HIV, weakening the immune system, and making them unable to protect the body from illness and infection. The higher the CD4 count, the stronger the immune system. A CD4 count of 700 to 1000 in a regulated sample size is considered to be 'normal'. HIV infected people are considered to have a 'normal' CD4 count if their CD4 count is above 500, however, when a count is below 200 the patient is considered to have AIDS [AIDS n.d.]. Counsellors use the CD4 count to assist in educating HIV positive people in how to stay healthy in a hope of keeping their CD4 level above 200.

Some participants thought that including the CD4 count was important for everyone, others said it was only important if their HIV results returned a positive reaction. The discussion continued, with comments such as

"CD4 is necessary for one to know whether they are ready for ARVs or not. Therefore let this variable be on the form."

"These are people who are just being tested. They can't know their CD4 count. But if they have ever been tested and now they are coming for a second test or counselling, then it can be possible"

"CD4 is used by health workers. I don't think it is useful on this form. First time testing wouldn't know their CD4."

"Whoever tests [HIV] positive must be tested for CD4 count as well. Having a CD4 test is important. Therefore referral is important to tell them where to go for the CD4 count."

The decision from the discussion was that there should be a place for the CD4 count to be recorded. The decision was based on that once a HIV result returns a positive reaction, the person should be advised to have a CD4 test done. A CD4 test can be conducted at the same time as VCT, with a result imminent in approximately an hour (depending on the type of rapid test used). Providing the CD4 count on the form would be useful for future reference, as well as assisting in record keeping.

Consent. The WHO suggests that specific consent for HIV testing is required and is actually implied if general consent for medical care has been given (WHO n.d. p20-22). The focus group participants debated over the need to record that consent was given.

"there is need to improve on the form to provide for an area of consent from the client. The form does not show that the client consented to the testing or counselling."

"When we go to the Island, people come willingly. People come asking for VCT. However, though they come willingly we have to ask them to consent."

There was an extended discussion in regard to consent from a minor

"How do you get consent from a minor?"


"There were many implications when many adults never wanted their children to be tested. The practice however has changed. For the minors who can not

consent on their own, ...the caretakers can consent on their behalf. For children of 12 and above years, these can consent by themselves."

Even though, implied consent is given through allowing the counsellor to complete the form, it was decided to include a space to indicate that consent was obtained. The researcher questioned the need for a signature as this identified the participant. It was determined to make the recording of consent optional, and that a 'tick' indicating consent was sufficient.

Through active discussion and analysis of the notes taken during the discussion, the instrument was refined to include several new variables, these are listed in Table II together with the new form. These variables are consistent with the WHO guidelines for measuring national HIV prevalence [WHOd n.d. p20-22].

Table II Stage Two variables and form

<p>Stage Two added variables</p> <ul style="list-style-type: none"> • Number of children given birth to; number alive, pregnant? • Last testing organisation • VCT record card given to client? • Comments/recommendations (eg CD4) • Consent obtained for use of de-identified and aggregated data. 	
--	---

It is important to note that the version two of the form included a 'Private and Confidential' box at the top right of the form. This was as a result of discussion at the focus group; and is a reminder of the ethical conduct to be followed during HIV testing and VCT.

The spreadsheet was modified to suit the new VCT data collection form, both of which are in use across the organisations represented at the Stage two focus group session.

5. Conclusion and Future Work

Although this study was limited to one then six VCT organisations in the Mukono sub-county of Uganda, the results suggest that the involvement of the end user in the design of an information system – even if primarily paper-based, together with input

from information systems design professionals (who have an understanding of local needs), and a purpose for the data collection, are instrumental in the acceptance and use of information systems.

The one organisation that has been using the VCT data collection instrument and associated electronic spreadsheet since March 2006, has noticed a significant increase in the number of people attending VCT, repeat visits, and recognition that they are advancing in the fight against HIV, as stated by Bazira during the August 2008 focus group session in Mukono, *“our report works as an eye opener to the Government or to the local authority so that they can think about those isolated areas where we go HIV testing.”* Furthermore, those that record and analyse the data, are able to report on genuine statistics that they have an affinity with and understanding of, for example, Bazira continued, *“whenever we do the testing, the percentage is higher than expected; sometimes 15% people who are positive. And still many [more people] needs to be done [tested].”*

Once data are available and in a format that can be used and manipulated, information from the VCT data collection activities, can aid in developing strategies and plans for the best time and place to conduct VCT; comparison between sites (villages) in terms of when/who to visit, the number of counsellors and other resources required (eg brochures, VCT forms/cards, HIV rapid test kits).

Although the aim of the study was to develop a useful, simple and effective data collection instrument to assist VCT counsellors in understanding their client needs, develop strategies to assist them after VCT, and resource planning, the data gathered also shows information that can be used for inclusion into grant applications to up-scale and expand VCT coverage, and communicate the successes and challenges to the community (local, national and global).

Most of the 2000 NGOs that conduct VCT in Uganda need to report their work to health authorities within the country and to many organizations that financially support these NGOs, such as the Global Fund. The Global Fund and other large funding bodies to Uganda, have (and could do so again) suspended funding if monitoring and reporting is not done rigorously and based on empirical work.

“One of the major bottlenecks in the implementation of large country grants, such as Global Fund grants, is a weak monitoring and evaluation system, with inappropriate indicators, inadequate data collection systems, a lack of capacity to collect data at the local level, and an inadequate capacity for data analysis. As HIV services evolve, countries need support in revising and strengthening monitoring alongside the implementation of programmes.” [WHOa n.d. p49].

The development and acceptance of a user directed VCT data collection system such as the one presented in this paper can enable an empirical base for HIV VCT coordination, planning, training and procurement in a resource-limited region such as Uganda. Each of these can directly lead to the up-scaling of VCT in an attempt to reduce the transmission and increase the prevention of HIV.

Furthermore, what we found interesting, is that even though the WHO has developed guidelines for the collection of VCT data, the participants in this study had little knowledge of these guidelines, let alone implemented by many of the NGOs that operate in rural, deep village or up-country villages in Uganda. The VCT data collection form developed during this study is inadvertently very similar to the WHO guidelines for addressing, up-scaling and expanding VCT into rural regions. The WHO guidelines were designed not only to collect information in regard to HIV prevalence, but at the same time suggest that data are collected on social, behavioural and biomedical factors. Data on some of these factors can be extracted from the form refined by the VCT counsellors who participated in the study presented in this paper.

This study has shown that involving those who use and need a system in the development of the system, results not only in a system that is used, but one that gathers acceptance further a field amongst peers and colleagues. This is due to an affinity with the wider-user group, an appreciation of the users, listening to their concerns and needs, as well as given them ownership of the system. End-user involvement in systems design and development is a recommended and accepted practice in systems development [Pekkola et al 2006]- however, it is not undertaken as often as required. This view is supported by Tushabe et al [Tushabe et al 2008], where they report that software applications (in the case of this study, an information system) need to be developed for Ugandans by Ugandans for the application to be successful.

Acknowledgements

The authors would like to acknowledge the support given to the project by the University of the Sunshine Coast, VOLSET, and the Ugandan VCT counsellors who participated in the focus group session.

References

- AIDS n.d. CD4 count. Retrieved April 2009 <http://aids.about.com/od/technicalquestions/f/cd4.htm>
- CIA n.d. Retrieved August 2008 at <https://www.cia.gov/library/publications/the-world-factbook/geos/ug.html>
- COUTINHO, A. (2003). *A situation analysis of HIV/AIDS in Uganda and the role of VCT*. Retrieved August 2008 http://www.aidsuganda.org/pdf/role_of_VCT.pdf
- FAMILY HEALTH INTERNATIONAL (2004). *Monitoring HVI/AIDS Programs: Participant Guide*. A USAID Resources for the presentation, care and treatment.
- HOGG, R., MAY, M., PHILLIPS, A., COSTAGLIOLA, D., STERNE, J., SABIN, C., DE WOLF, F., LEDERGERBER, B., D'ARMINIO MONFORTE, A., JUSTICE, A., GILL, J., FUSCO, G., STASZEWSKI, S., ROCKSTROH, J., CHÈNE, G and EGGER, M. (2006). Rates of disease progression according to initial HAART regimen: A collaborative analysis of 12 prospective cohort studies. *Journal of Infectious Diseases* 194: 612-622
- LYNCH, K., LYNCH, S and BAZIRA, F. (2008). Facilitating Grass-root NGO's Adoption of Information Systems and Information Communication Technologies. *International Journal of Computing and ICT Research* Special Issue (2008) (pp35-41). Michigan State University Press.

- PEKKOLA, S., KAARILAHTI, N., and POHJOLA, P. (2006) Towards formalised end-user participation in information systems development process: bridging the gap between participatory design and ISD methodologies. In *Proceedings of the ninth conference on Participatory design: Expanding boundaries in design - Volume 1*. 1-5 August 2006. Trento Italy
- TUSHABE, F., BARYAMUREEBA, V., BAGYENDA, P., OGWANG, C. and JEHOPIO, P. (2008) The Status of Software Usability in Uganda. In *Proceedings Fourth Annual International Conference on Computing and ICT Research*, 3-5 August 2008. Kampala Uganda.
- UNAID 2006. ABCs of HIV prevention. Retrieved August 2008 http://www.usaid.gov/our_work/global_health/aids/News/abcfactsheet.html
- UNAIDSa n.d. AIDS epidemic update. Retrieved August 2008 <http://www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2006/default.asp>
- UNAIDSb n.d. Report on the global AIDS epidemic. Retrieved October 2008 <http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/2008/>
- UNAIDSc n.d. UNGASS country progress report Uganda, January 2006 to December 2007. Retrieved October 2008 <http://www.unaids.org/en/KnowledgeCentre/HIVData/CountryProgress/2007CountryProgressAllCountries>.
- WHOa n.d. Towards Universal Access: Scaling up priority HIV/AIDS interventions in the health sector. Progress Report, April 2007. Retrieved October 2008 http://www.who.int/hiv/mediacentre/universal_access_progress_report_en.pdf
- WHO b n.d. Policy statement on HIV testing. Retrieved October 2008 <http://www.who.int/hiv/pub/vct/statement/en/index.html>
- WHO c n.d. HIV Testing and Counselling Online Toolkit. Retrieved October 2008 http://whqlibdoc.who.int/publications/2005/924159327X_eng.pdf
- WHO d n.d. Guidelines for measuring population health. Retrieved October 2008 <http://www.who.int/hiv/pub/surveillance/guidelinesmeasuringpopulation.pdf> .
- WikiAnswers n.d. Retrieved October 2008 http://wiki.answers.com/Q/How_many_deaths_or_related_illnesses_have_occurred_due_to_aids_and_hiv

5

Translating Transliterations

Jörg Tiedemann and Peter Nabende

Translating new entity names is important for improving performance in Natural Language Processing (NLP) applications such as Machine Translation (MT) and Cross Language Information Retrieval (CLIR). Usually, transliteration is used to obtain phonetic equivalents in a target language for a given source language word. However, transliteration across different writing systems often results in different representations for a given source language entity name. In this paper, we address the problem of automatically translating transliterated entity names that originally come from a different writing system. These entity names are often spelled differently in languages using the same writing system. We train and evaluate various models based on finite state technology and Statistical Machine Translation (SMT) for a character-based translation of the transliterated entity names. In particular, we evaluate the models for translation of Russian person names between Dutch and English, and between English and French. From our experiments, the SMT models perform best with consistent improvements compared to a baseline method of copying strings.

1. Introduction

Transliteration is the process of representing words from one language using the approximate phonetic or spelling equivalents of another language [Arbabi et al. 1994]. Models for automatic (machine) transliteration are useful in the handling of unseen terms for various NLP tasks. The most frequent use of machine transliteration is in the representation of names in languages with different writing systems or alphabets, for example English and Russian. In tasks such as MT and CLIR, the lack of a comprehensive bilingual dictionary including the entries for all entity names makes machine transliteration necessary [Kashani et al. 2007]. The main motivation for integrating a machine transliteration module in NLP applications is to handle unseen terms in a proper way so that performance of that application is improved. Several researchers have used a variety of approaches to machine transliteration that involve either modeling a direct mapping between two orthographies or considering the phonetic representation for transforming strings into each other or a combination of both [Oh et al. 2006].

A task that is not addressed in the literature (at least to our knowledge) is that of translating transliterated names. With this, we refer to the translation of names that have been transliterated from a different writing system in a third language. Consider the examples in Table I where Russian names have been transliterated into: English, French, German, and Dutch. As we can see, names are spelled very differently even though all the four languages use more or less the same writing system (the Roman alphabet). Such spelling variations for the same source name in target languages arise due to language specific differences, for example in the way of encoding pronunciations [Hsu et al. 2007].

Table I. Russian names in four European Languages

English	French	German	Dutch
Alexander Pushkin	Alexandre Pouchkine	Alexander Puschkin	Aleksandr Poesjkin
Nikita Khrushchev	Nikita Chruschtschow	Nikita Khrouchtchev	Nikita Chroesjtsjov
Yuri Andropov	Iouri Andropov	Juri Andropow	Joeri Andropov
Leonid Brezhnev	L'eonid Brejnev	Leonid Breschnew	Leonid Brezjnev

The problems usually addressed by transliteration between different writing systems apply here as well: The different spelling variants try to match the underlying phonetic description which is usually not known to a cross-lingual application. A dedicated module for the translation of (unknown) transliterated entity names is expected to help a system (for example MT) in the same way a transliteration module improves performance across writing systems.

In our experiments we look at the translation of Russian proper names between English, Dutch, and French using two different models that map the orthography directly without considering (or modeling) the underlying phonetic representation. In particular, we look at task-specific weighted finite state transducers and character-based statistical machine translation models trained on names extracted from Wikipedia².

In section two, we review related work; in section three, we introduce the models used for the translation task; in section four, we describe the experiments and show results with a discussion; we finally conclude in section five.

2. Related Work

In this paper we adapt both machine translation and transliteration models for automatically translating transliterations. We briefly review recent work associated with the two classes of models.

For machine transliteration, four types of models have been proposed by Oh et al. [2006]: grapheme-based transliteration models, phoneme-based transliteration models, hybrid transliteration models, and correspondence-based transliteration models. Classification of these models is based on the units that are used for transliteration: graphemes only, phonemes only, or both graphemes and phonemes. Different methods and techniques have been used under these models, typical classifications including: statistical methods, rule-based methods or a combination of both. Based on these methods, significant research has been dedicated towards developing techniques for machine transliteration in languages that use different writing systems. Research has also been done with regard to handling of transliterations; however, most of it is associated with measuring similarity between transliterations across different writing systems [Hsu et al. 2007; Lam et al. 2007].

Statistical Machine Translation (SMT) models have also been adapted for machine transliteration. Matthews [2007] for example uses Moses, a state of the art Phrase-based Statistical Machine Translation system (PSMT) on transliterating proper names between

² Wikipedia is a free online encyclopaedia that can be accessed via <http://www.wikipedia>

English and Chinese, and English and Arabic. Matthews [2007] shows that a machine transliteration system can be built from an existing PSMT system whose performance is comparable to state-of-the-art systems designed specifically to transliterate.

Although a lot of work has been done for Machine transliteration, little has been done with regard to translating transliterated entity names whose origin is in a different writing system. We adapt two approaches used in previous work for the translation task in this paper: Finite state automatons [Graehl 1997] and PSMT models [Matthews 2007].

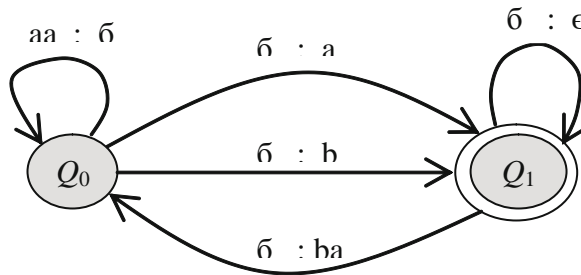
3. Machine Translation Models

In this section, we describe the methods and techniques we have used for translating transliterated names. The first method uses Weighted Finite State Transducers (WFSTs) and the second is based on Phrase-based Statistical Machine Translation (PSMT) models.

3.1 Weighted Finite State Transducers

A Finite State Transducer (FST) is an automaton that transforms one string into another. It can be seen as a network of states with transitions between them which are labeled with input and output symbols. Starting at some state and walking through the automaton to some end state, the FST can transform an input string (by matching the input labels) to an output string (by printing corresponding output labels). Figure 1 is an example of an FST where each arc is labeled by an input and output string separated by a colon while the nodes represent states (two states in this example)

Fig. 1. Example of a Finite State Transducer (Adapted from [Mohri 1997])



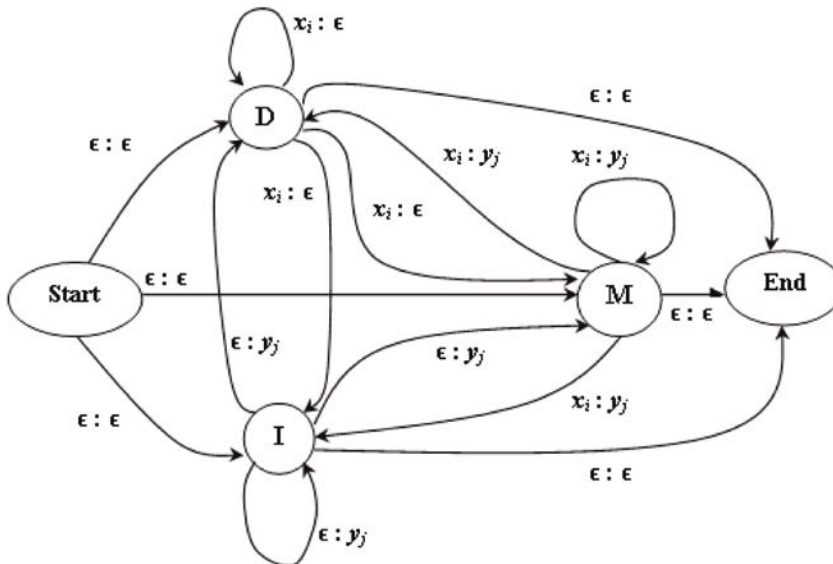
Weighted Finite State Transducers (WFSTs) are automatons in which each transition in addition to its usual input label is augmented with an output label from a possibly different alphabet, and carries some weight element [Mohri 2004]. Assuming that the weights form proper probability distributions, we can compute the probability of certain transformations and compare them with other competing transformations. In this way the most likely transformation can be chosen according to the model.

For the translation task in this paper, we first model the transformation between transliterated names as a sequence of edit operations on the character level (Figure 2).

For this we define a state **M** for substituting a source language character x_i with a target language character y_j , a state **D** for inserting a source language character x_i (by defining the output label to be the empty string ϵ) and a state **I** for inserting a target language character y_j (by setting the input label to be the empty string ϵ). Furthermore, we add a start state and an end state that can be reached from any other state by matching the empty string ϵ and producing the empty string ϵ . The general structure of the FST illustrating the nodes associated with the edit operations is shown in Figure 2.

We have chosen the structure described above with respect to related work on similarity estimation using pair HMMs [Mackay and Kondrak 2005; Wieling et al. 2007] that take on a similar structure. It is, however, not necessarily an advantage to model insertion, deletion and substitution with different states; we define some variants of the transducer in figure 2. Basically, we add transitions performing arbitrary substitutions and source/target character insertions at each state and let the training procedure decide how to make use of the explanatory power of the model chosen. We then use various numbers of states to see if additional power leads to improved results.

Fig. 2. Edit distance WFST model



A final modification we applied is related to the splitting of strings into characters. In the models above we simply use transformations of single characters without considering any context. However, there are often contextual dependencies between adjacent characters and various character n-grams correspond to some degree to an underlying phonetic representation. For example, the Dutch spelling ‘*oe*’ usually refers to an /*u*/ sound. Because we did not involve any sophisticated phonetic transcription we simply change our splitting strategy in the following way: We define a set of vowels in each language and split strings into adjacent vowel, non-vowel or white-space characters. Here we assume that especially vowel combinations correspond to certain sounds but also that

combinations of consonants are often used to transliterate certain phonetic elements (for example “**sjk**” in the Dutch {**Poesjkin**} as opposed to “**shk**” in corresponding English {**Pushkin**}). We also do not take care of character ambiguities and simply define fixed disjoint sets of vowels and consonants.

3.2 Phrase-based Statistical Machine Translation

Phrase-based statistical machine translation (PSMT) is the current state of the art in data-driven machine translation. It is based on the well-known IBM models trained on large parallel corpora but using bilingual phrase-tables instead of word link probabilities and fertility parameters. In PSMT several components are usually combined in a log-linear model (translation models, reverse translation model, word and phrase penalties, language models, distortion parameters, etc) with weights optimized using minimum error rate training. Various tools are available for training such a model and “decoding” (translating) input strings according to the model.

PSMT can be used on the character level instead of the word level [Matthews 2007]. The entire procedure is directly applicable in the same way as standard word-level PSMT systems are organized. Instead of aligning words we have to align characters in parallel data, which will be translated transliterations in our case instead of translated sentences. Phrases now refer to character N-grams instead of word N-grams and language models are also defined over character sequences instead of word sequences.

The advantage of PSMT over the transducer model described earlier is that the extracted phrase tables (character N-grams) now cover a lot of contextual dependencies found in the data. In this way we hope to find better transformations by translating sequences of characters instead of single characters. Furthermore, we do not have to model insertions and deletions explicitly but leave it to the translation table to change the lengths of translated strings. Another advantage is the explicit inclusion of a target language model to weight the possible outcomes of the system. In the transducer model this is not easily possible as we include deletion operations. A language model would always prefer shorter strings and therefore force the system to over-use the deletion operations when transforming strings.

For actual translation, we use standard SMT decoders with monotonic decoding (because we do not expect any character movement on the target side in translations that should correspond to the same underlying phonetic representation).

4. Experiments

4.1 Data sets

Our data set is extracted from an English Wikipedia database dump from 2008/07/24. We used simple patterns to identify Russian names looking at the structured information in Wikipedia info-boxes. Basically we looked at entries that match the pattern (Russian|Russia|Soviet) in categories such as “citizenship”, “nationality” and “place of birth”. Translations of these names are taken from links to other languages which are also given for every Wikipedia page. In this way we collect all names potentially from Russian origin and their correspondences in other languages. We save all name

pairs for the language pairs we are interested in, performing some extra normalization. For example, the German Wikipedia page referring to “Nikita Khrushchev” includes his middle name (“Nikita Sergejewitsch Chruschtschow”). In order to fix this problem we use the following heuristics for names with different numbers of space separated elements: Firstly, we remove abbreviated middle names (such as “H. W.” in “George H. W. Bush”). Secondly, only the first and the last element are considered if there is still a difference in the number of elements (or only the last element is taken if one of the names only contains one element).

Another normalization that has to be done is to switch the order of first and family names. In some languages it is very common to use the family name followed by a comma and the first names (“Clinton, William Jefferson” instead of “William Jefferson Clinton”). Here we simply change the order by swapping the strings preceding and following a comma for all names in the database. Note that we swap the strings first and then check for matching number of name elements. As a final pre-processing step we convert all names to lower case versions to reduce the number of parameters in our models.

Unfortunately, the data set extracted in this way is quite small which is probably due to the requirement of having an info-box (with matching contents). Nevertheless, we obtained 199 pairs of names for English-Dutch and 372 pairs for English-French. We did not manually check them and, therefore, our database includes names which are not typically Russian (such as Marc Chagall, born in the Russian empire as son of a Jewish family). However, we assume that there are only very few of these exceptions.

From each of our data sets we removed 50 name pairs to form our two test sets for both language pairs. Each of the two test sets is used for evaluating all the models tested. The remaining pairs are used for training and/or tuning model parameters.

4.2 Evaluation

There are different ways that can be used for evaluating the translations generated. One obvious way is to calculate the accuracy on a test set of unseen names, i.e., computing the proportion of correctly translated names in that set. However, accuracy is a very strict measure with respect to character-based translation where one single mismatch is counted in the same way as a completely dissimilar pair of strings. Furthermore, for many transliterated names, several alternatives may be acceptable in a language (for example, “Chrushchev” instead of “Khrushchev”) but only one reference is given in our data. Therefore, other measures for string similarity should be considered. We have chosen to use the Longest Common Subsequence Ratio (LCSR) as our main evaluation measure. It is defined for a pair of strings as the ratio of the length of the longest common subsequence of characters and the length of the longer strings. It produces values between 0 and 1 where 1 is a perfect match between the two strings.

4.3 Training WFSTs

Parameters of WFSTs can be trained from data using a forward-backward algorithm. The training set simply contains pairs of correct matching transliterations from the two given languages and they need not to be aligned. The forward-backward algorithm iteratively maximizes the probability of observing this training data set by adjusting the

internal model parameters in a hill-climbing manner. The algorithm converges to a local maximum depending on the initial model chosen.

In our settings we run the training procedure with a uniform initial model and five other randomly chosen initial models. In this way we reduce the likelihood to end up in a suboptimal model at least to some extent. For training, we use Carmel, a free toolkit for manipulating finite state automata [Graehl 1997]. The training procedure is implemented in the Carmel system and also the procedures for obtaining the most likely strings given some input sequence and model parameters.

As discussed earlier, there are several models that can be applied to our translation task. In particular, the transducer structure in terms of states and their possible emissions can be varied. The first model refers to the edit distance model using separate states for substitutions (and matching); insertion (modeled as inserting target language characters) and deletion (modeled as inserting source language characters). In this model we introduce some kind of model bias by restricting the type of emissions to be of a certain kind at each state. We can also remove this bias by allowing all possible types of emissions (including insertions on the source and target language side) from any state in the WFST. The idea here is to let the training procedure decide how to make use of the hidden layer of states without defining the function of each state. This is basically a test to see if the forward-backward procedure is capable of learning some underlying structure which is not given to the system when training its parameters. Of course, we still have to define the number of states to be used in the WFST before training its parameters. In our experiments, we applied WFSTs with one up to five states (excluding start and end state) and a fully connected graph with uniform initial settings. Furthermore, we also ran the training procedure with five additional randomly chosen initial parameters.

Finally, we can also modify the input and output alphabets by changing the way of splitting strings into symbol sequences. Previously, we simply used character sequences for training and testing. Now we also like to test the technique discussed in section 3 on WFSTs; i.e. splitting words into sequences of vowel/non-vowel N-grams. After splitting our data in this way, we can apply the same training procedures as for the preceding WFSTs.

4.4 Training PSMTs

For training and decoding the PSMT models we used the publicly available toolkit Moses [Hoang et al. 2007] with its connected tools GIZA++ [Och and Ney 2003] and IRSTLM [Frederico et al. 2008]. Adjusting the system to our transliteration task basically refers to manipulating the data sets used for training, tuning and decoding. This means that names have to be split on the character level and spaces have to be converted to some other kind of separation marker (we used ‘_’ for this purpose which otherwise does not appear in our data). Furthermore, we selected monotonic decoding for obvious reasons and left other parameters unchanged. Hence, the model uses standard settings for the alignment with GIZA++ (character alignment in our case), standard heuristics for the extraction and scoring of phrase alignments (character N-grams with a maximum length of 7 characters) and standard settings for the minimum error rate training (MERT) used for tuning. The language model is a 5-gram model estimated from the target language

side of our training data using the standard smoothing technique implemented in the IRSTLM toolkit (Witten-Bell smoothing).

There are various fixed parameters that can be tuned in the PSMT models as described above. Among others, we could change the maximum size of phrases to be considered, various phrase extraction techniques can be used and language model parameters can be modified. In our experiments we did not tune these training specific parameters. Instead, we concentrated on modifying the models in the following ways: firstly, we changed the training data in such a way that the set for tuning is part of the training set instead of keeping a separate set for tuning. In our basic setting we remove 50 additional name pairs from the training set to be used for tuning the SMT model parameters. In another setting we simply used them for training as well. Here, we were interested in seeing how increasing the training set influences the performance before training (especially with our tiny training set). Furthermore, we also like to know if tuning on parts of the training set may still lead to improvements on the test set.

Secondly, we changed the pre-processing step from character splitting to vowel/non-vowel splitting as described in the previous sections for WFSTs. Here, we do not expect a similar effect on the results as we expect for the WFSTs using this technique. This is because contextual information is already integrated in the phrase-based SMT model to a large extent and important character combinations already appear in the extracted phrase table with appropriate scores.

A last modification we investigated is to apply a larger language model. It is well-known that SMT models produce better results in general when increasing the language model. However, the transliteration task is different and common character combinations in the target language may not necessarily be as common in transliterated names. Hence, we like to test the impact of adding data from a larger set of target language strings to estimate the character language model for our task.

4.5 Results

Let us first have a look at the baseline for our task. A common technique in machine translation for handling unknown words is to leave them untouched and to copy them to the target output. For names (usually a large portion of the unknown words) this is certainly a good strategy if alphabets at least very similar. Hence, the baseline for our task refers to this strategy of copying the strings even for transliterated names. Table II shows the scores for the baseline in terms of LCSR and accuracy³ (acc.) on our test sets of English-Dutch and French-English, each with 50 names.

Table II. Results for baseline approach of copying strings

baseline	LCSR	acc.
Dutch-English	0.88	0.32
French-English	0.89	0.26

³ We complement LCSR scores with accuracy scores in order to show the magnitude of completely correct translations as well

As we can see in Table II, the LCSR scores for both English-Dutch and French-English are quite high already, which means that English and Dutch, or French and English spellings of Russian names are not very far from each other. Even the accuracy is rather high considering the strict nature of this measure. Table III shows the translation results using our WFST models.

As we can see in Table III, the WFST models do not perform very well. None of the Dutch-English WFSTs actually improves the baseline scores, neither in LCSR nor in accuracy. The translation performed seems to actually harm the system especially when looking at accuracy scores. The baseline of leaving names unchanged should be preferred instead. For French-English we can observe a slight improvement of the edit distance WFST. In accuracy, the vowel/non-vowel model also performs better than the baseline. Furthermore, for both language pairs, we can see that the edit distance WFST does not have a clear advantage over a single-state WFST. There is only a slight gain in accuracy for English to Dutch and French to English translation, but otherwise the scores are the same. From our experiments we can also conclude that the training procedure is not capable in learning a hidden underlying structure from the data. However, looking at the size of our training data, this was not to be expected either. Looking at the large differences in scores for various numbers of states it seems that the algorithm easily gets stuck in suboptimal maxima. Finally, the second splitting strategy using vowel and non-vowel sequences does not improve the performance either. On the contrary, it actually hurts the model which is a bit surprising. One reason might be the increased sparseness of our data set including larger sets of input and output symbols, which now contain character N-grams. The only improvements compared to the character-based WFST can be seen in the accuracy for English to Dutch and French to English translations. The score for English-Dutch, however, is still below the baseline.

Table III. WFST Translation results

WFST (characters)	Dutch-English		English-Dutch		French-English		English-French	
	LCSR	acc.	LCSR	acc.	LCSR	acc.	LCSR	acc.
edit distance	0.88	0.22	0.87	0.20	0.90	0.28	0.89	0.24
1 state	0.88	0.22	0.87	0.18	0.90	0.26	0.89	0.24
2 states	0.79	0.00	0.88	0.18	0.79	0.02	0.88	0.14
3 states	0.81	0.12	0.80	0.04	0.85	0.14	0.81	0.0
4 states	0.81	0.06	0.85	0.22	0.78	0.02	0.81	0.02
5 states	0.78	0.02	0.78	0.02	0.79	0.04	0.83	0.04
vowel/non- vowel	0.83	0.20	0.84	0.28	0.88	0.30	0.87	0.20

Let us now look at the results from the PSMT system in Table IV. Here we can see a clear improvement of the translation quality as measured in LCSR scores. Except for the non-tuned models with large language models, all LCSR scores are above the baseline and even accuracy scores exceed the baseline in various cases (but by far not all of them). The importance of training data can be seen in the figures for Dutch and English where the tuning set is included in the otherwise very small training data set. For those experiments, we obtain the highest scores in both translation directions. For English-French, for which we have a larger training set available, we do not see a similar behavior. A separate development set seems to be preferable. Also, the impact of tuning is mixed and it is not clear how MERT is affected by a setting where the development set is not kept apart from training.

The second modification of our training data, the split of characters into vowel/non-vowel sequences, performs quite well, especially for English to Dutch. However, a clear advantage of this technique over the standard pre-processing technique cannot be seen.

The final test refers to the inclusion of a larger language model. Here, we included the English, French and Dutch Europarl data [Koehn 2005] for estimating character-based language models. We can clearly see that the additional data sets harm the translation process and only after tuning do the scores get back to the level of other models using the small language models from the parallel training data. Looking at the weights after tuning we can also see that the language model weights are very low when using the large data set. This seems to suggest that the overall influence of a language model on translation quality is rather low in our case.

Table IV. PSMT Translation results

Moses (characters)	Dutch-English		English-Dutch		French-English		English-French	
	LCSR	acc.	LCSR	acc.	LCSR	acc.	LCSR	acc.
without tuning	0.89	0.24	0.90	0.28	0.91	0.22	0.88	0.14
tuned	0.92	0.30	0.90	0.28	0.93	0.46	0.91	0.28
{tune} \subset {train} without tuning	0.91	0.40	0.92	0.32	0.91	0.28	0.89	0.18
	0.93	0.34	0.91	0.40	0.91	0.38	0.90	0.28
vowel/non- vowel without tuning	0.90	0.28	0.91	0.48	0.93	0.46	0.90	0.28
	0.89	0.32	0.92	0.44	0.93	0.44	0.91	0.36

large LM								
without tuning	0.82	0.06	0.82	0.06	0.76	0.02	0.84	0.02
tuned	0.91	0.26	0.92	0.44	0.90	0.22	0.90	0.22

Table V. Examples from the test set Dutch-English showing some typical problems of translating transliterations with the models

Dutch input	Correct English	WFST English	Moses English
andrej tarkovski	andrei tarkovsky	andrey tarkovski	andrey tarkovsky
anna koernikova	anna kournikova	anna koernikova	anna kurnikova
Aleksandr solzjenitsyn	aleksandr solzhenitsyn	aleksandr solzyenitsyn	alexandr solzhenitsyn
anton tsjehov	anton chekhov	anton tsyekhov	anton chechov
andrej sacharov	andrei sakharov	andrey sakharov	andrei sakharov
dmitri sjostakovitsj	dmitri shostakovich	dmitri syostakovitsy	dmitri sjostakovich
leonid brezjnef	leonid brezhnev	leonid brezynev	leonid bruzhnev

In Table V some examples of translations from the Dutch/English test set are shown. In these examples we can see typical problems especially of the WFST model. In particular, we can see the problem of consistent character substitutions without considering local context. For example, “i” is consistently translated into “i” in English and “j” into “y” using the WFST. In the PSMT model, contextual dependencies are better covered due to character n-grams in the translation table. However, there are still ambiguities causing problems in examples like “tsjehov”→“chechov” (instead of “chekhov”).

5. Conclusions

In this paper we have looked at the problem of translating transliterated names. Spellings of names originally coming from a different writing system vary substantially even for related languages. We investigated two types of models for translating such names between Dutch-English and English-French. We applied various types of weighted finite state transducers and phrase-based statistical machine translation models to our task. The models were trained on name pairs of Russian origin extracted from Wikipedia. In our experiments the SMT approach performed best, consistently beating the baselines. The results are encouraging especially when considering the tiny training set that was available to us. The results also show that specialized models like the ones we have tested in this paper for translating transliterations may help to handle unknown words and hence improve performance in NLP applications such as MT, CLIE, and CLIR. As future work, we would like to determine whether improvements can be obtained by investigating extended structures of the WFST models for example while incorporating contextual information represented in the model.

References

- ARBABI, M., FISCHTHAL, S.M., CHENG, V.C., AND BART, E. (1994). Algorithms for Arabic name transliteration. *IBM J. Res. Develop.*, 38(2).
- FREDERICO, M., BERTOLDI, N., AND CETTELO, M. (2008). IRSTLM Language Modeling Toolkit, Version 5.10.00. FBK-irst, Trento, Italy.
- GRAEHL, J. 1997. Carmel finite state toolkit. <http://www.isi.edu/licensed-sw/carmel>.
- HOANG, H., BIRCH, A., CALLISON, C., ZENS, R., AACHEN, R., CONSTANTIN, A., FEDERICO, M., BERTOLDI, N., DYER, C., COWAN, B., SHEN, W., MORAN, C., AND BOJAR, O. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 177-180.
- HSU, C-C., CHEN, C-H., SHIH, T-T., AND CHEN, C-K. (2007). Measuring similarity between transliterations against noisy data. *ACM Transactions on Asian Language Information Processing (TALIP)*, 6(1), Article no 5.
- KASHANI, M.M., JOANIS, E., KUHN, R., FOSTER, G., AND POPOWICH, F. (2007). Machine Transliteration of names in Arabic text. In *Proceedings of Association for Computational Linguistics (ACL) Second Workshop on Statistical Machine Translation (WMT07)*, ACL, 1-13.
- KOEHN, P. 2005. Europarl: A Multilingual Corpus for Evaluation of Machine Translation. In *Proceedings of MT Summit*, Phuket, Thailand.
- LAM, W., CHAN, S-K, AND HUANG, R. (2007). Named Entity Translation Matching and Learning: with application for mining unseen translations. *ACM Transactions on Information Systems*, 25(1).
- MACKAY, W. AND KONDRAK, G. (2005). Computing Word Similarity and Identifying Cognates with pair Hidden Markov Models. In *Proceedings of the ninth Conference on Computational Natural Language Learning (CoNLL)*, Ann Arbor Michigan, 40-47.
- MATTHEWS, D. (2007). *Machine Transliteration of Proper Names*. Masters Thesis, School of Informatics, University of Edinburgh.
- MOHRI, M. 2004. Weighted Finite State Transducer Algorithms An Overview. AT&T Research, Shannon Laboratory.
- MOHRI, M. 1997. Finite State Transducers in Language and Speech Processing. *Computational Linguistics*, 23(2), 269-312.
- OCH, J.F. AND NEY, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1), 19-51.
- OH, J-H, CHOI, K-S., AND ISAHARA, H. (2006). A Comparison of different Machine Transliteration Models. *Journal of Artificial Intelligence Research*, 17(2006), 119-151.

6

Decision Support in the Operating Theatre – Usability Aspects

John Kizito

Anesthesiology deals with such a complex social system that it can spawn over an infinite number of states. Diagnesia, a prototype built to offer decision support to anesthetists continuously estimates the likelihood and unlikelihood of diagnoses during surgery, by applying arguments for and against the different diagnoses, and presents the most probable diagnoses to the anesthetist. In this paper, we present the usability aspects and/or design decisions pertaining the prototype

1. Introduction

When patients go to the operating theatre for surgery, they need some sort of anesthesia in order not to feel conscious pain. This paper will focus on this type of anesthesia, referred to as general anesthesia. Anesthesia suppresses some vital functions of the patient. Kizito [2008] and Ballast [1992] discuss several effects of this. These include loss of water, blood loss, vasodilatation, suppression of autonomic nervous function, and so on. Since the patient is unconscious, unaware of pain, and immobile during the period he/she is anesthetized, he/she is unable to express him/herself. We thus have a specialist in the theatre responsible for the health of the patient during this time –the Anesthetist.

The specialist is able to monitor the state of the patient by use of monitoring devices. These devices continuously display the patient's physiological data, which when coupled with registrations of certain events, medical history, and drug administration can be used to estimate the state/health of the patient. Such data may include, but is not limited to, blood pressure, heart rate, respiratory rate, oxygen saturation, and anesthetic concentration of gas mixture. When any of these values goes out of the normal range, the Anesthetist has to take appropriate action.

In [Kizito, 2008], we discussed in detail the states of the patient. We categorized them into three; familiar (to the anesthetist), urgent, and diagnosing states. In the familiar state, the Anesthetist knows the typical treatment to give the patient. The urgent state is unfamiliar but nevertheless, the Anesthetist needs to give some treatment even without knowing the cause of the problem. The diagnosing state is also unfamiliar but not urgent and thus can not be diagnosed like the familiar case. In this paper, we present approaches towards providing information in all these three dimensions.

Diagnesia, a prototype designed to offer decision support to the Anesthetist, continuously estimates the likelihood and unlikelihood of each of the diagnoses in its list, and presents such relevant information at such an abstract level that, this information

coupled with his expert knowledge, the Anesthetists decision making process can be facilitated [Kizito, 2008; Pott and Feber, 2005]. Physicians often refer to their clinical decision making process as more art than science, and suggest that while computers might be programmed to deal with the scientific, analytical aspect of their work, they will never be able to capture the “art” of a skilled clinician [Pople, 1992]. The decision support system can thus not do without the experts’ knowledge. This paper discusses the use of Information and Communication Technology (ICT) techniques to relay this kind of information and also discusses how such information is presented to the specialist. We shall also discuss the design decisions made in accordance with usability.

Using some defined indicators as inputs, Diagnesia builds a set of rules that are used to continuously estimate likelihood and unlikelihood of diagnoses in its set by computing corresponding evidence probabilities. Every diagnosis has a set of indicators and counter indicators, each with a certain strength/weight. The indicators are used to estimate the evidence probabilities [Kizito, 2008]. At any one point, the outcome of this computation is a paired set of probable diagnoses and corresponding estimated probabilities. This paper will not focus on how these probabilities are computed but rather on how the findings are relayed to the relevant personnel.

2. Usability Design Decisions

In this section, we look at a number of aspects considered when designing Diagnesia’s user interface. The user interface is an important aspect of the system. It is through this interface that the system keeps the user informed about what is going on through appropriate feedback within reasonable time. Diagnesia is designed in such a way that the information displayed can easily refresh every so often by setting a value of the refresh timer.

The system should speak the users’ language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. It should follow real-world conventions, making information appear in a natural and logical order. The terms and/or abbreviations used to describe diagnoses, arguments, and the like, are those that anesthetists are familiar with, taken from literature of anesthesiology [Aitkenhead et al., 2001].

The user interface contains two major screens. An input screen (mainly for simulation and/or testing purposes) and the output screen (for Anesthetists use). This paper will focus on the output of the system thus the output screen. It is also very important that the information displayed on the interface is actually relevant to the anesthetist. The next sub-sections describe the approach used to ensure relevancy of the information displayed by the output screen of the decision support system (DSS).

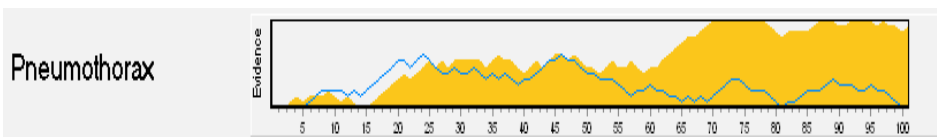
2.1 Most probable diagnoses (the familiar state)

Previous research has attempted the use of approximation strategies in offering decision support to diagnosis [ten Teije and van Harmelen, 1996, 1997]. Diagnesia follows a similar approach by using estimated probability as a measure of the most likely diagnosis. The likelihood and unlikelihood of a diagnosis is estimated as a ratio of the total score

of all arguments (expected to contribute to the evidence) that are true to that of all those expected to make a contribution for the evidence in question (see Kizito [2008] for details).

When designing the way Diagnesia displays both the likelihood and unlikelihood, our intention is to make it easy for one to distinguish between the two. Although the likelihood may be superimposed on the screen more than the unlikelihood, the unlikelihood is also very important and should not be neglected. In collaboration with a usability class in the Department of Artificial Intelligence of the University of Groningen, we adopted a logarithmic scale, with a measure of time and evidence, as shown in Figure 1. This allows us to put more emphasis on the current state (far right) of the patient as well as keeping track of the historical states (towards the left). The orange shaded area shows the likelihood of the diagnosis at a given time, whereas the blue (later changed to green) line shows the unlikelihood.

Fig. 1. Indication of likelihood and unlikelihood of diagnosis *Pneumothorax*



2.2 The urgent state

In [Kizito, 2008], we discussed some states of the patient. We noted that the state of the patient may be unfamiliar to the anesthetist and, in such cases, there is not enough time to investigate the cause of the problem. The anesthetist needs to give a treatment in order to bring back the patients state to normal. In this state, it is important for the anesthetist to have an idea of what category the problem is and the extent to which it is life threatening. Is it a *Cardiovascular*, *Respiratory*, or *Anesthesia depth* problem? If we have a problem in more that one of these categories, where do we have more urgency?

Approximation strategies are informed by particular properties of the domain knowledge [ten Teije and van Harmelen, 1997]. The work of Bravata et al. [2004] shows specificity as one of such properties that affect diagnostic closure when using approximation techniques. We thus attempt to provide some information in a more specific and precise manner. We consequently use three sets of icons to reflect the patient's state along these three dimensions: the heart (for indicating problems with the cardiovascular system), lungs (for the respiratory system) and eye (for the depth of anesthesia). By moving the icon for a particular category to some direction, it should indicate more life threatening situations where as the withdrawal to the opposite direction should indicate more stable state. The icons may be in one of four positions; normal (not life threatening), low, intermediate, and high (life threatening).

Taking an example of the Anesthesia category, it is interesting to note that the system could easily come up with a scenario in which the patient has a low sleeping depth and at the same time a low awakening. This can be obtained in case the two arguments *High MAC* (argument for *sleeping depth*) and *Low amplitude pulse oximeter* (argument for *awakening*)

happen to be true at the same time. It is a contradiction because one is not expected to have an increasing depth of anesthesia and at the same time waking up! However, since it is a possible scenario according to the rules built into the system, a decision of handling this case has to be made. Considering the fact that it is more dangerous for a patient to wake up without the knowledge of the anesthetist than sleeping a bit more, it was decided that the awaking alarm overrides the one of sleeping depth. So, if such a scenario ever arose, the awakening icon will move to reflect the state of the patient along the dimension. We shall discuss more about these icons in 2.5.

2.3 Non Measurable (Observable) Variables

Diagnosis uses a number of variables to test the truth of the corresponding arguments. The likelihood of a diagnosis is then estimated depending on the number of arguments for the diagnosis that evaluate to *true* and their corresponding weights. In some instances, it is hardly possible to reach conclusive diagnostic closure as the system may require additional observation [McIlraith and Reiter, 1992]. For example the system cannot tell whether or not the patient is sweating. These arguments are important and may be needed to confirm certain diagnoses. At the moment, we are unable to directly measure these variables. Perhaps it is possible to investigate the possibilities of measuring them but in the scope of our study the cost of this investigation may not be worthwhile. Some of them probably can be computed from some other information measured from the monitoring devices. Nevertheless, the system is designed in such a way that it is not hard to add more arguments for disorders.

Such arguments have further been categorized into three groups: one group can be observed, checked on the user interface, and considered in probability computations; another group cannot be considered when estimating probabilities but can be highlighted in the status bar of the user interface when there is suspicion that it should be observed; the last group is not considered at all by the DSS.

Checkable on the user interface: In this category, a checkbox has been provided on the input screen. For instance if the anesthetist realizes that the patient is sweating, he/she simply needs to check the box provided for sweating and the DSS will take it into account when estimating the probabilities of diagnoses. This method is not so user friendly and was limited to only two arguments (Mottled skin and Sweating) since one has to remember to uncheck such checkboxes when the observed action stops.

Alert message in status bar: As the number of observable arguments increased, the method described above could not be used anymore. Another mechanism was devised. A status bar was added at the bottom of the output screen of the system. When the diagnosis indicated by an argument is supported by any other measurable argument with a probability greater than 0 (zero) and qualifies to be among the probable diagnoses, then we can advise the anesthetist to check the observable arguments. We do this by putting some text in the status bar explaining *what to observe and for what diagnosis*. This text will disappear as long as the diagnosis in question loses the support. The text should slowly scroll to the left when it is too much to be accommodated in the available space of the status bar. A sample display of this feature will be presented in 2.5.

Other. The method described above can only work if the diagnosis indicated by the observable argument has at least one other measurable argument. Otherwise it will never get a chance of displaying its “help” information since it cannot raise any probability greater than zero. Such cases, where the only argument known for a diagnosis is an observable one, are not implemented in the system. For example, the system cannot know that there is a *widening of the QRS complex on the ECG* and since this is the only argument we have for *hyponatraemia*, the diagnosis (*hyponatraemia*) is not in the set of 36 catered for in Diagnesia.

On addition to the diagnosis affected by the *Other* category above, Diagnesia does not cater for *hemorrhage*. This is also due to the unease of measuring the disorder. Hemorrhage is a copious discharge of blood from the blood vessels. According to a textbook of anesthesia [Aitkenhead et al., 2001], blood loss can be estimated by weighing swabs, measuring the volume of blood in suction bottles and assessing the clinical response to fluid therapy. Estimation is always difficult where large volumes of irrigation fluid have been used.

2.4 Unavailable Measurements

Diagnesia is designed to be integrated with the current systems in order to automatically read the input values. However in normal cases, where there is no need for extra monitoring, the standard set of monitoring variables may be measured and are available. However, a variable may not be read because of a failure in the integration software, disconnection or damage in cables or any other computer and/or connection failures. In such cases, we think it is wise to assume that the variable in question is simply not being measured.

Because of this uncertainty of which variable is available for measurement, the system is designed in such a way that every variable may or may not be available at a certain point in time. When a variable is unavailable for measurement (irrespective of the cause), the DSS drops all arguments related to the variables from the set of arguments to be considered. We simply assume that we know nothing about it. In fact, we do not know if it is *normal* or *out of range*. If a diagnosis indicated by such a variable is supported by some probability high enough to qualify it among the probably diagnoses (despite the absence of at least one of its arguments), the DSS will show nothing about the corresponding argument for the unavailable variable.

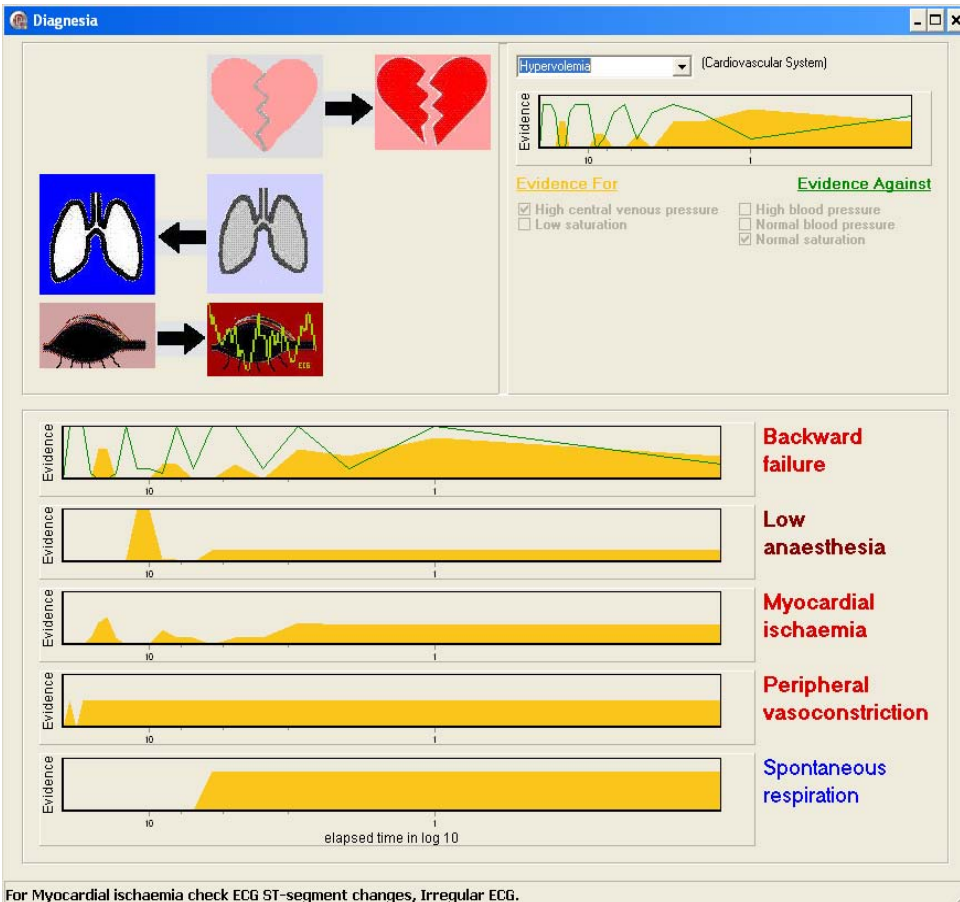
This method affects the algorithm for the estimation of probabilities since the weights are assigned and thought about with the assumption that we have all the arguments contributing to the probability of the diagnosis. It should also be noted that this might make a certain diagnosis become *single argument*, which have their own challenges (not discussed in this paper) since they easily give rise to 100% evidence.

2.5 The User Interface

Figure 2 shows a complete snapshot of a sample output screen of Diagnesia. In the top left corner are the icons used to reflect the patient’s state along the three dimensions: the heart, lungs, and eye as mentioned earlier in 2. The further the icon moves to the right the more critical the state of the patient. The previous state of the icon is

shown by a faded version of the icon and an arrow to show the trend. In order to provide more information along with the movement of the icon, the criticality of the situation has been coded with visual effects: the heart “breaks”, the lungs turn black, and the eye either opens (patient is waking up – not in picture) or one of two typical electroencephalogram (EEG) curves for deep anesthesia is superimposed on the closed eye icon. Clicking on any of the icons displays the production rules used to determine the state of the icon group in question

Fig. 2. Diagnesia user interface



In the lower section of the screen, the system displays up to five most likely diagnoses (that might further explain the state of the patient). These diagnoses are initially arranged in order of the difference likelihood – unlikelihood however, when the system re-computes the probabilities and has to replace one or more of the currently displayed five diagnoses, the one with the smallest difference will be replaced first without re-ordering the diagnoses. This prevents movement of the graphs considering that one may be observing a particular graph. In addition, the name of the new entrant to the list is displayed in bold font to reflect the change.

In the top right corner of the screen is highlighted the first likely diagnosis however, the anesthetist may decide to instead observe another diagnosis by selecting its name from the drop-down list above the graph. This could be because the anesthetist chooses to observe the pattern of the likelihood for a certain diagnosis, which may not be displayed among the probable diagnoses. This can be relevant when the anesthetist expects or suspects something to happen because of his/her knowledge about the patient's health condition, past experience, action taken by the surgeon, or any other factors that may cause the anesthetist's suspicion. When this happens, the system does not override the selected diagnosis with any other until the anesthetist decides to (by selecting another one from the list). Furthermore the category of the selected diagnosis is shown besides the select box. For *Hypervolemia* (in Figure 2), we have *Cardiovascular System*.

Yes, *Diagnosis* can suggest diagnoses with corresponding probabilities, but *if I really want to find out how it came to such a conclusion?* Anesthetists want to be aware of the situation as clear as possible. It is very important that the information supplied by the DSS does not conflict with the strategies of the anesthetist. It should therefore be possible for the system to display summarized and clear information that explains how it comes to the suggested conclusion. Consequently, we have the arguments for and arguments against that the system has used to approximate the current probabilities of the selected diagnosis. A checked box against the (counter) argument implies that the (counter) argument was found to be true.

Color-coding: We group all the diagnoses in four different categories namely cardiovascular, respiratory, anesthesia, and others. Each of these categories is assigned a color-code. That is to say, red for cardiovascular, blue for respiratory, brown for anesthesia, and black (default) for others. The names of the diagnoses against their corresponding graphs are printed in the respective color using this scheme. In Figure 2, we have a lot of red diagnoses thus a possible reason for the broken heart. On addition, there are two more colors that are coded on this interface. Orange is used in the graphs to display the likelihood where as green is for the unlikelihood. This corresponds to the headings *Evidence For* and *Evidence Against*. The colors used were specifically chosen with reasons; red for the color of blood; blue for air (respiratory system); brown being the color of anesthetic bottles; orange for being bright; green is used to show *OK* (or *GO* for instance on traffic lights). The use of a shaded orange graph with a green line is to make a clear distinction between the likelihood and unlikelihood.

Lastly we have a status bar that gives extra information to the anesthetist. As mentioned earlier, the probabilities are estimated using data generated from monitoring devices however, some information that could be used to verify a certain diagnosis can only be observed by the anesthetist (see 2.3). We display such information in the status bar at the bottom of the screen. In Figure 2, we advise the anesthetist to check *ECG ST-segment changes* and *Arrhythmia (Irregular ECG)*, which are arguments for Myocardial ischaemia although its probability has been computed based on the values of Forward failure and Backward failure. This interface is a result of consultations with the usability class in the department of Artificial Intelligence, specialists from the University Hospital,

and computer scientists from the department of Mathematics and Computing Science, University of Groningen. With all such aspects put into consideration, the designed system should be very usable and with ease.

3. Discussion And Conclusion

This paper discusses the usability aspects of Diagnesia, a DSS designed to enhance the decision making process of the anesthetists by improving their situation awareness. We discussed a number of aspects that include using probability theory [Miller et al., 1982; Weinstein and Fineberg, 1980] to estimate the likelihood and unlikelihood of disorders in the operating theatre. We present such information to the anesthetist using a graph with a logarithmic scale in order to superimpose the current state of the patient over the historical data. On addition, we present the patients state using icons in order to acquire information about cases that are not familiar to the anesthetist. We also attempt to take into consideration disorders that may be brought about by arguments that we can not measure (perhaps can be observed) as well as those whose values may be unavailable for whatsoever reason.

We have also discusses some challenges still faced with this body of work. Disorders whose arguments can only be observed by the anesthetist are still a challenge. We also need to find a better way of resolving the misinformation brought about by missing or unavailable measurements. Despite the fact that adoption of cognitive models into artificial intelligence systems requires a substantial amount of time [Pauker and Kassirer, 1980], there is more to be done in this body of work with adequate testing in mind.

Acknowledgments

Special thanks go to the following people for the tremendous work and efforts that contributed to this paper:

- Constanze Pott, University of Groningen
- Gerard R. Renaldel de Lavalette, University of Groningen
- Albert Ballast, University Medical Center Groningen

References

- AITKENHEAD, A. R., ROWBOTHAM, D. J., AND SMITH, G. (2001). chapter 40, pages 501–523. Fourth Edition.
- BALLAST, A. (1992). Warning Systems in Anesthesia. PhD thesis, University of Groningen.
- BRAVATA, D. M., SUNDARAM, V., MCDONALD, K. M., SMITH, W. M., SZETO, H., SCHLEINITZ, M. D., AND OWENS, D. K. (2004). Evaluating detection and diagnostic decision support systems for bioterrorism response. In *Emerging Infectious Diseases*, volume 10.
- KIZITO, J. (2008). Diagnesia: A prototype of a decision support system for anesthetists. In *Proceedings of the Third International Conference on Broadband Communications, Information Technology & Biomedical Applications (BroadCom' 08)*, pages 12–19, Pretoria, South Africa.
- MCLLRAITH, S. AND REITER, R. (1992). On tests for hypothetical reasoning.

- MILLER, R. A., POPLE, H. E. J., AND MYERS, J. D. (1982). Internist-i, an experimental computer-based diagnostic consultant for general internal medicine. *The New England Journal of Medicine*, 307:468–76.
- PAUKER, S. G. AND KASSIRER, J. P. (1980). The threshold approach to clinical decision-making. *The New England Journal of Medicine*, 302:1109–17.
- POPLE, H. E. (1992). chapter Heuristic Methods for Imposing Structure on Ill-Structured Problems: The Structuring of Medical Diagnostics. Westview Press, Boulder, Colorado.
- POTT, C. AND FEBER, J. L. (2005). Using cognitive modeling for requirements engineering in anesthesiology. In RE '05: Proceedings of the 13th IEEE International Conference on Requirements Engineering, pages 491–492, Washington, DC, USA. IEEE Computer Society.
- TEN TEIJE, A. AND VAN HARMELEN, F. (1996). Computing approximate diagnosis by using approximate entailment. In Fifth Internat. Conf. on Principles of Knowledge Presentation and Reasoning, pages 256–265.
- TEN TEIJE, A. AND VAN HARMELEN, F. (1997). Exploiting domain knowledge for approximate diagnosis. In Proceedings of IJCAI'97, pages 434–439.
- WEINSTEIN, M. C. AND FINEBERG, H. V. (1980). Clinical decision analysis.

7

Implementing Successful E-health Implementations within Developing Countries

Stella Ouma, Marlien E. Herselman and Van Greunen

The use of Information Communication Technologies (ICT) within healthcare can make significant changes in the daily operations of hospitals particularly in the developing world. A technology assessment of five hospitals based within Nyanza Province in Kenya was conducted to find out how hospitals are embracing the use of ICT. Both primary and secondary data were collected to be used in the study. A qualitative study was used through the application of a multiple case study to investigate five randomly selected hospitals. Structured interviews, open ended questionnaires and observations were used as methods to collect data from the various hospitals. In order to collect relevant data the participants were divided into three categories. Managers (n=5), hospital staff members (n=31) and patients (n=24). Therefore a total of (n=60), participated in the study. The findings revealed that just like in the majority of the developing nations, there are very few computers and e-health solutions that are currently used in the hospitals as a result of various challenges in Kenya. Consequently, this paper analyses the challenges and provide a way forward for developing nations when implementing e-health solutions.

1. Introduction

The developed countries have embraced the use of information communications technologies (ICT) within the hospitals and health clinics. A few examples of the use of ICT include computerisation of medical records, electronic scheduling for appointments, use of the Internet for the purposes of communication and the use of magnetic cards [Tomasi, Facchini and Maia, 2004:867].

In the case of developing countries, Ojo et al. [2007:3] state that the majority of African countries are grappling with major socioeconomic development challenges. Samake and Mbarika [2007] concur by stating that the challenges include wars, diseases and poverty, which affect the provision of medical care for both the rich and the poor.

As much as this is the case, Richardson [2006] argues that healthcare providers and governments have no choice but to meet healthcare demands for future citizens and the application of e-health is therefore fundamental.

Hence to overcome the challenges that Africa is facing, there is a need to improve information and communication exchange in the healthcare industry in order to accelerate knowledge diffusion and increase access to information [Ojo et al., 2007].

Of more concern are the rural areas within the developing countries. The authors in IOM [2001], state that the quality of healthcare is essential to all. The factors that are of significance in the provision of quality healthcare include timeliness, patient centeredness, efficiency, effectiveness, equity and safety. These factors however do not exist in the provision of healthcare in some of the rural communities. According to Chang, Bakken, Brown, Houston, Kreps, Kukafka, Safran and Stavri [2004:449] the rural inhabitants actually receive fewer healthcare services compared to the problems that they have. Hence, they can be referred to as undeserved populations. This is because rural hospitals are faced with many challenges that lead to a negative effect on the quality of service that is provided in the rural communities [Ruxwana, 2007:103]. Information Technology therefore, can dramatically revolutionise the delivery of healthcare thereby making it safer, more efficient and more effective [Thielst, 2007].

According to ABARD [2005:4], the use of ICT can enhance the quality of services in rural areas by providing a better and more cost-effective approach to service delivery. This can be done by implementing various e-health solutions within rural hospitals.

Based on these conclusions on the use of ICT, a technology assessment was conducted in the rural areas of Kenya, particularly in Nyanza Province to find out how the hospitals are using ICT to improve service delivery within the rural hospitals.

Nyanza province is one of the eight provinces in Kenya. The population comprises 2,291,069 females and 2,107,326 males with a total of 4,398,395 as per the 1999 population census [Nyanza Provincial Development Plan, 2007]. More over, the Province has a total of 35 hospitals, 20 sub district hospitals, 127 health centres and 240 dispensaries. In terms of human resource, it has 156 doctors and 2271 nurses [Ministry of Health, 2006].

According to the Central Bureau of Statistics, Kenya [2006], 65 per cent of the population in Nyanza province are living below the poverty line. The province also has the highest prevalence of human immunodeficiency virus [HIV] infection in the country. Therefore, there is a need to address the quality of services, increase efficiency and reduce costs in rural areas in order to improve healthcare services.

2. E-health Solutions

E-health can be described as any electronic exchange of health information within the healthcare industry by use of different stakeholders [Kwankam, 2004; Denise, 2003]. There are various e-health solutions that can be used in rural areas to improve the quality, efficiency and to reduce costs within the rural hospitals. Key among them includes electronic health records, hospital information systems, telemedicine and the internet.

2.1 Electronic health records (EHR)

Initially known as computer patient records [Katchakis and Tsinakis, 2006], the concept of electronic health records has revolutionised to what it is today from the 1960's [Hanson, 2006:105].

EHR can be relied on to act as a back up in cases of emergencies and when patients change locations unlike the case of paper based records given the fact that they are easily accessible [Blair, 2007; Novak,2005].

Essential functions of EHR include shared health records, support for external information requests, provision of security and message transfer of health records [Edwards, 2007].

The benefits of using EHR include improved quality of healthcare [Miller and West, 2007], reduced medical errors and reduced costs [America Academic of physicians, 2007], access to medical record information [Gans, Kralewski, Hammons and Dowd, 2005] and time savings [Thakkar and Davis, 2006].

2.2 Hospital Information Systems (HIS)

Clinical managers and health planners rely on information in order to make decisions regarding effective functioning of health facilities, allocate resources and also to make strategic policies [AbourZahr and Boerma, 2005]. Therefore, HIS consist of different softwares that are integrated in order to capture data in specific sections of the hospital [Garrido, Raymond, Jamieson, Liang and Wiesenthal [2004:21-22]], handle the workflow of daily medical services and also assist in managing financial, administrative and clinical data [Yang, Yeh and Wang, 2006:174]. The benefits of using HIS include improved quality, better communication, staff efficiency, reduced transcription costs and increased patient safety and increased revenues[Garrido et al., 2004].

2.3 Telemedicine

Telemedicine can be referred to as the provision of medical services from a distance [Wooton, Craig and Patterson, 2006:1]. This includes diagnosis, treatment and prevention of diseases

The types of telemedicine can be categorised as real-time or pre-recorded telemedicine. Information is sent and received by the participants almost immediately in the case of real time telemedicine while in the case of pre-recorded telemedicine, information is captured and then transmitted later for subsequent reply [Anthony et al., 2005:288-293].

Examples of pre-recorded telemedicine include tele-electrography, tele-obsterics and tele-radiology [Mea, 2006:43-45]. Examples of real time telemedicine include tele-consultation, tele-pathology and tele-dermatology [Wooton, Craig and Patterson 2006:52-60].

2.4 The Internet

The role of the Internet in healthcare cannot be ignored. The Internet provides a platform where various stakeholders of e-health are able to achieve various goals. Key among the solutions that the Internet provides in healthcare includes:

- **Business operations:** There are various types of e-business that are conducted by organisations and individuals online within the healthcare industry. These business models include business to business to consumer models [Olatokun and Ajiferuke, 2006], business to business and business to consumer models [European

Commission Enterprise Directorate General, 2004. Some examples of business models within the healthcare industry include virtual doctor visits, online medical suppliers and automated systems [Tan, 2005].

- **Research:** The Internet has provided a platform for conducting a lot of research in healthcare. This has led to the use of online experiments, randomised trials and surveys [Couper, 2007]. Additionally the numbers of publications on healthcare issues have increased tremendously [Curry, 2007]. Professional medical education: Medical learners can now share a lot of digital information from the various digital libraries on the Internet which have been reviewed by various researchers [Ruiz, Mintzer and Leipzig, 2006:209-210]. Additionally various professionals in medical fields can now access various web pages in order to take continuous medical education [Tan, 2005] hence improving current standards of healthcare by use of the availed technologies on the Internet [Dario et al., 2004].
- **Consumer Informatics:** Patients are using the Internet to get information, interact with their physicians and order pharmaceutical products online [Podichetty and Biscup, 2003]. Hence the patients are now taking charge of their health status by staying informed with issues regarding to their health [Lorence, 2006].

3. Research Motivation

Currently the selected rural areas are using the traditional ways to access health and those who are currently using ICT do not know the power that exist behind it, in improving healthcare services.

The study seeks to establish the current structures in place in terms of technologies, equipments and communication media available within the rural areas. More importantly it examines challenges that exist which may not favour the implementation of e-health solutions within the developing world.

The contribution of the paper will be to advice on should be done since majority of the hospitals in the developing do not know how to go about implementing e-health solution with the resources available. The main research question was to find out how ICT can be applied in rural hospitals to support E-health solutions in the developing world.

4. Background of the Hospitals

The background of the rural hospitals that were investigated are discussed below:

4.1 Background of the rural hospitals

This section looks at the background of all the selected hospitals. The hospitals include Homabay district hospitals, Nyamira district hospital, Bondo district hospital, Kisumu district hospital and Chulaimbo health centre.

4.1.1 Homabay district hospital

Homabay district hospital is located in Asego division in Homabay district. The hospital was established by the colonial government in Kenya in 1958. According to the district health records (2007/2008), the hospital is a level-IV service delivery hospital and

serves a total of 96,936 outpatients and 10,046 inpatients, a total of 106,982 patients a year. It is the most sought after hospital in the district in terms of use of resources and human resources for effective service delivery. This is confirmed by Homabay district health plan (2007/2008), which states that the utilisation rate of the hospital is 107 per cent. This shows that they serve more patients than the expected figure. In the human resources section, it has around 211 employees. The hospital serves the entire district since it is also a level-III referral hospital and has specialised staff (district health records, 2007/2008).

4.1.2 Nyamira district hospital

Nyamira district hospital is located in Nyamira division at Kebiringo trading centre. Opposite the hospitals are various kiosks and shops, with a busy road in between. Nyamira guest house is located next to the hospital. The hospital was established in 1979 as a government project. According to Nyamira district health records (2006) the hospital serves between 160 and 260 outpatients at the hospital every day with a total of 143,236 a year. The utilisation rate of the hospital, according to Nyamira district health plan (2007/2008), is 225 per cent, which is very high. Major diseases manifested by patients at this hospital are malaria, diseases of the respiratory system and pneumonia.

4.1.3 Bondo district hospital

Bondo district hospital was started by the colonial government as a dispensary and it operated as such until 1970 when the late Jaramogi Oginga Odinga raised funds for the construction of the wards. After the completion of the wards in 1972, the hospital was upgraded to a health centre, finally becoming a district hospital in 1998. Bondo district hospital is located in the Maranda division of Bondo. It serves around 44,837 patients per year and has a utilisation rate of 285 per cent (Bondo district health plan, 2007/2008) with only 94 professional staff members to deal with the workload. The major diseases that affect the patients are malaria, HIV and pulmonary TB (Bondo district health records, 2007).

4.1.4 Kisumu district hospital

The hospital was started in 1942 by the then Ministry of Social Services. The hospital is located in Winam division in Kisumu town along Ang'awa road opposite Kenyatta sports ground in Kisumu. The hospital serves a population of 142,400 patients with an utilisation rate of 164.1 per cent (Kisumu district health plan, 2007/2008) and around 337 staff members. The hospital is a level IV hospital and a third referral level within the district. The complaints that the patients in the hospital suffer from include malaria, HIV, pneumonia and pulmonary tuberculosis (Kisumu district health plan, 2007/2008).

4.1.5 Chulaimbo health centre

Chulaimbo health centre is located along the Kisumu-Busia road. The hospital was started on 20 January 1976 by the government with the idea of facilitating training for healthcare workers in proper rural health service delivery. The facility serves a

population of 75,678 people (projected from 1999 census). It is a first level referral health facility. The government of Kenya health facilities refer patients to Chulaimbo from Maseno division are Siriba dispensary and Chulaimbo including Ekwanda and Ipari health centres in Vihiga district and Nyahera Kisumu. Chulaimbo has become a hub for outpatients; attendance has risen from 32,952 patients (2004) to 46,464 in the year 2005 (Chulaimbo PRHTC work plan (2006/2007)). The utilisation rate is 684.5 per cent (Kisumu district health plan, 2007/2008).

5. Methodology

In this section, the study design, data collection instruments and participants have been discussed.

5.1 Study design

A qualitative study was used. This was done through a multiple case study approach. Through random sampling, five rural hospitals were identified to be used in this multiple case study. These hospitals include Bondo district hospital, Nyamira district hospital, Kisumu district hospital, Homabay district hospital and Chulaimabo health centre.

5.2 Data collection instruments

The data collection instruments used included observations, interviews and questionnaires.

5.3 Observations

This involves watching a particular phenomenon within a certain period in order to understand a particular outcome. Lategan and Lues [2006:21], state that observation refers to the physical act of keeping certain variables and making notes of what was observed. Therefore the participants were observed in this research including the technologies of ICT that is available in each of the selected hospitals.

5.3.1 Interviews (Semi structured)

Semi structured interviews were used to collect data from the selected hospitals. The interviews were used mainly to access collect data from the participants concerning the existing technologies in place. These interviewees therefore included various e-health stakeholders. These included management, doctors, nurses and patients at selected hospitals in Nyanza province in Kenya.

Additionally to confirm the reliability and validity of the collected information, all interviews were captured using a tape recorder and later transcribed to Microsoft Excel application for further analysis.

5.3.2 Questionnaires

The questionnaires focused on the background or history of ICT within the rural hospitals, the access level of ICT and the current condition of the ICT infrastructure in place that support e-health solutions in the various hospitals.

Open ended questionnaires were. The questionnaires were categorized into three. The first questionnaire was for managers, another set was for the staff and finally the last category of questionnaires was particularly for patients in order to collect relevant data and to back up the in.

The English language was used to conduct the interviews and also on the questionnaires as Nyanza province uses different languages.

5.4 Participants

Purposive sampling was used to select the managers, doctors, nurses and patients as a sample that is truly representative of the users of the total population, who use the hospitals. The participants were categorized into three namely managers, staff members and patients. The managers that were selected were the medical superintendents who are in charge of the hospital operations. The hospital staffs included doctors, nurses and clinical officers. A total of 12 participants took part from each hospital. An overall of 60 participants therefore took part in this study as shown in table 1, below.

Table 1: Expected participants in the study in the selected hospitals and data collection techniques that were used.

	Observations	Still Photos	Tape Recording	Interviews	Questionnaires
Bondo Hospital	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients
Kisii Hospital	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients
Maseno Hospital	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients
Kisumu Hospital	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients
Homabay Hospital	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients	1 Manager 6 Staff 5 Patients
Total	60 participants from 5 different hospitals				

6. Results

From the findings this section has been categorised into three parts. In the first part, we look into the existing technologies within the hospitals, the perception of the participants in regard to benefits of using ICT's is discussed in the second part and finally barriers to e-health implementations are discussed in the last section.

6.1 Available ICT technologies within the hospitals

The first sub question sought to investigate the current technologies in place which could be used to support e-health. The methods that were used to answer this question are:

- a. Questionnaires
- b. Interviews

A questionnaire was posed to managers, patients and staff members of the rural hospitals. Majority of the patients seemed not to be aware of any ICT technologies in place. However, the managers through the questionnaires were quick to point out the technologies in placed and the managers contributions were backed up by the staff members through the interviews which were conducted.

Table 2: ICT infrastructure and e-health technologies in place within the hospitals

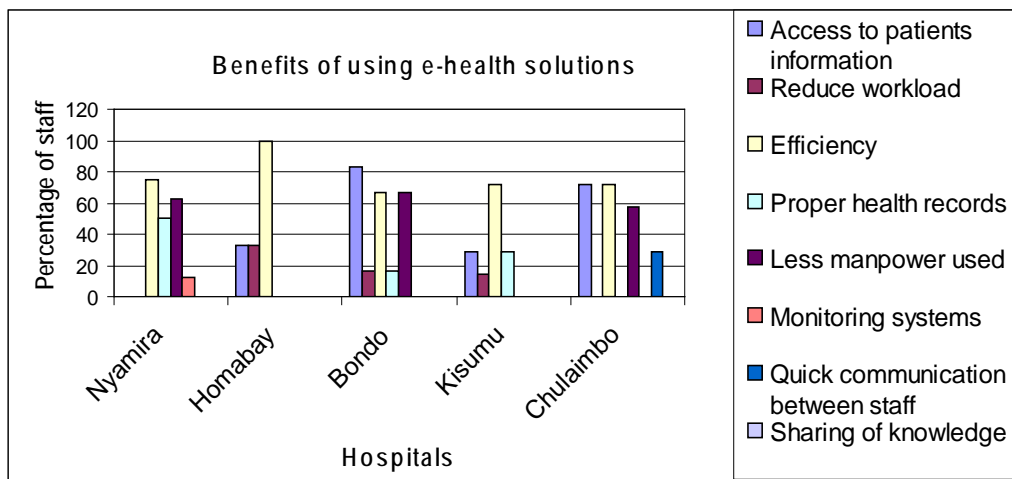
Participants	ICT infrastructures and e-health solutions
Respondent 1 (Homabay district hospital)	Has got ten computers, one photocopier and one printer. In addition it has an internet connection and telephone services are also available.
Respondent 2 (Bondo district Hospital)	Only administrators have got access to mobile phones. The other staff members do not have access to telephones. There is one computer and a printer at the records office. No internet connectivity is available.
Respondent 3 (Nyamira district hospital)	Has got six computers for use, two printers and a network that connects two computers. Additionally it has information systems operating in the billing department, at the pharmacy and at the chest office.
Respondent 4 (Chulaimbo health centre)	Only administrators have access to telephones. The rest of the staff members use their personal mobile phones for communication. In addition the hospital has got a computer and a printer for administrative work.
Respondent 5 (Kisumu district hospital)	Telephone services are available in addition to two computers, a fax machine, and a printer.

In summary, the few technologies in place among the hospitals included, computers, telephones, a network and an internet connection. In addition information systems are used within two hospitals. Homabay district hospitals have information systems at the pharmacy and at the billing department while the Nyamira hospital has information systems at the radiology office, in the pharmacy and at the billing department.

6.2 Benefits of using ICT technologies as depicted by the participants

Through the interviews conducted, the 60 participants were able to point out their understanding of how they would benefit if ICT technologies were to be implemented. The participants managed to point out the following benefits as shown in figure 1:

Figure 1: Participants perception on benefits of using e-health solutions



From the diagram above the participants identified the following benefits:

- Increased efficiency
- Access to patient's information
- Reduced workload
- Proper health records
- Available monitoring systems
- Quick communication between staff
- Sharing of knowledge

Additionally, the various staff members and the management totalling to 36, pointed out that the use of ICT within the hospitals would improve the quality of services in the hospital in addition to reducing costs.

6.3 Existing barriers to E-health implementation

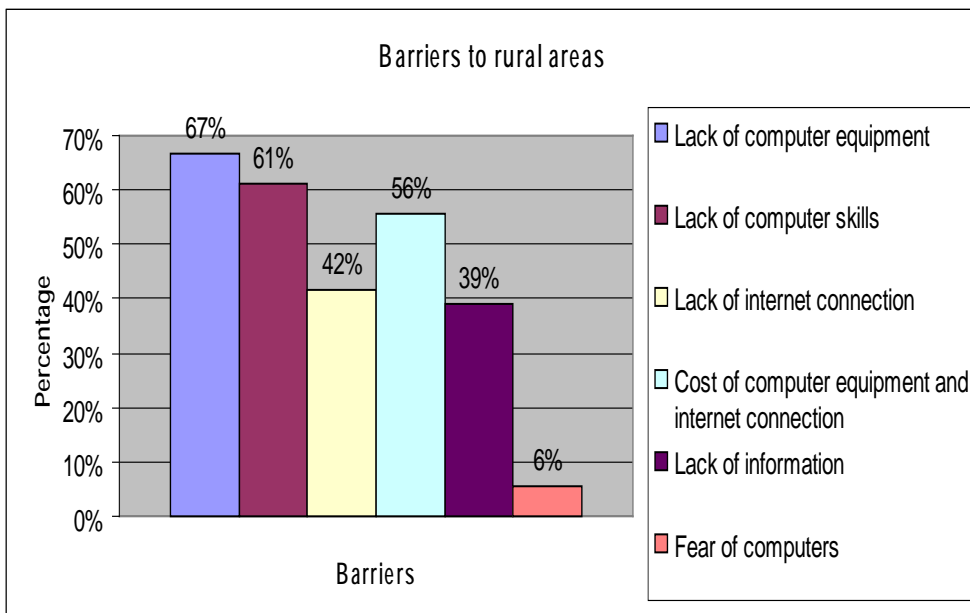
Through the questionnaires, we sought to find out why e-health benefits could not be realised within the hospitals. The managers and staffs (36) were given nine options from

which they would identify what they consider as barriers to e-health implementation within their respective hospitals. The options included the following.

- Option 1: Lack of computer equipment
- Option 2: Lack of computer skills
- Option 3: Lack of Internet connection
- Option 4: Out-dated and unreliable equipment
- Option 5: Lack of broadband connection
- Option 6: Working style not suited to the use of computers
- Option 7: Cost of computer equipment and Internet connection
- Option 8: Fear of computers
- Option 9: Lack of information

Figure 2 below shows the findings:

Figure 2: Participants perception on barriers that affect the use of ICT within the rural areas



As shown in figure two the barriers that the participants mainly selected included lack of computer equipment, lack of computer skills, cost of computer equipment and internet connection, lack of internet connection, lack of information and fear of computers respectively.

6.4 Findings from observations

Observations that were carried out ascertained that a lot of the processes within the hospitals are still done manually.

Patients have to buy their own files and prescription books. The files are then kept

at the records office after they are attended to. However when making the next visit, patients make a long queues as they wait to pick their files before they can be attended to. In some hospitals however, patients go home with their health records that are stored in form of books. In the process some of the patients loose the books.

It was also observed that some patients cannot be attended to since the procedures demand that they must come with books hence they are first sent to go and buy a patient's book where the patient's prescription will be recorded in addition to a file, despite insisting that they have no money.

Additionally, the matrons were manually drawing the worksheet that contains the duty schedule for the nurses. In addition outdated schedules are pinned on the wall in the wards. The pharmacy is using an information system to monitor the use of drugs.

In some hospitals patients have specific days on which they come to the hospital. For example, the hospital has days when referred patients come to the hospital or clinic days for mothers to bring the children to the hospital. This is done in order to share the resources available (i.e. the records office has few employees and so can only work with few patients at a time, the doctors are also few).

It was also confirmed that the only hospitals that use information systems in the pharmacy and at the billing points are the Nyamira and Homabay district hospitals. Apart from these, other operations are still conducted manually just like the other hospitals.

7. Discussions

From the findings, a lot of the functions within the hospital are manually done since there are very few ICT's in place, therefore a lot remains to be done. Additionally the use of ICT technologies in place currently are characterised by the following as a result of existing challenges:

- i. **Unavailability** - Currently these hospitals do not have enough technologies in place. Bondo district hospital and Chulaimbo hospital do not have telephone landlines in place as majority of staff use personal mobile phones. Each of these hospitals has only one computer. Only Homabay hospital has an internet connection in the matron's office. This is due to the availability of limited funds to put these technologies in place.
- ii. **Unreliability** - Additionally the telephones in place experience very frequent break downs forcing members of staff to move from point to point in order to communicate leading too a lot of time wasteges. This situation was most prevalent at Kisumu district hospital. Additionally, the internet service at Homabay district hospital was not functioning at the time of conducting the research. ICT's need to be put in place in such a way that it is reliable, to avoid making unnecessary errors and to promote efficiency.
- iii. **Inaccessibility** - Moreover majority of the staff members do not have access to the ICT technologies in place. Especially accessibility to the internet connectivity and accessibility to the available computers. This is due to the inadequacy in the number of the facilities in place.

- iv. **Lack of skills** – Majority of the staff members are not trained or equipped with basic computer operations skills. Therefore they may not be able to embrace the use of ICT technologies.

Unless these challenges are addressed, the hospitals will continue to implement e-health solutions that are not accessible, unavailable, unsustainable and unreliable as is the case currently from the research findings in Nyanza province, Kenya.

Moreover, these research findings support a similar research done in Eastern Cape Province, South Africa by Ruxwana(2007) and the findings show that they face similar challenges.

This calls for a way to deal with these challenges which have crippled implementations of ICT solutions within the developing nations.

Recommendations

From the findings, it is depicted that the developed countries are grappling with quite a number of challenges that are acting as barriers to the implementation of e-health solutions. Some of the major barriers include lack of computer infrastructure and lack of skills. Hence, it is imperative that e-health solutions are applied over a period of time to allow for a smooth integration and transition considering the fact that not all the challenges can be addressed at once by the developing nations but may be implemented in stages. By doing these, the governments within the developed nations will be able to plan and even measure the output even as the implementation process proceeds in the various stages. The proposed stages are as outlined in the following subsections.

- **Phase one (Initial stage)**

This stage requires various stakeholders in e-health to work together towards the implementation of e-health.

During this phase the governments and hospital administrators should be in the process of purchasing various ICT equipments, needed within the hospitals. This should include computers and its accessories and networking equipments. More importantly all the hospitals should have telephones to improve quality of service within the hospitals.

In addition to purchasing the ICT equipments, software development should begin at this stage. The government may work with various researchers to develop templates of open source softwares for use within the dispensaries, health centres, sub-district hospitals and the district hospitals, provincial hospitals and the national hospitals. This will act as a cheaper option compared to purchasing of on shelf software which requires licensing for a particular number of machines thereby limiting its use within the rural hospitals.

Moreover, the various staff members (doctors and nurses) should be trained during these phase in order to support e-health implementations.

- **Phase two (Deployment stage)**

Local area networks and wide area networks should be set up at this stage to facilitate the communication between various hospitals. In addition the various software templates for the health information systems that were developed should be customised to be used at the different hospitals. Telemedicine equipment should also be purchased at this stage. In addition, policies governing the use of e-health solutions should be set up. If possible there should be cross-sector linkages between various government ministries to support the implementation of e-health.

- **Phase three (Implementation stage)**

Since the hospitals shall have had ICT structures in place the use of health information systems, telemedicine and the Internet can then be implemented in each hospital to improve the quality of services provided to the patients.

- **Phase four (Evaluation stage)**

The use of various e-health solutions in place should be reviewed at this stage in order to determine the way forward for the hospitals. In accordance to the findings, changes should be made where possible. A way forward should also be proposed after analysis of the findings.

Note

The initial, deployment and implementation stage can overlap according to the time frame given based on the availability of resources at hand available within the various countries.

8. Conclusions

ICT's can improve the quality of services in rural hospitals, reduce costs and improve the efficiency. However, there are a few technologies that exist within the rural areas, these technologies are not sufficient for the implementation of e-health solutions. For e-health solutions to be implemented successfully, the key is to first deal with the barriers or challenges that exist within the developing nations.

Only when the challenges have been resolved, will the rural areas be able to enjoy the enormous benefits that e-health solutions provide. Therefore the challenges should be addressed in a step by step process in accordance to availability of resources at hand within the given periods in order to implement the use of e-health solutions within developed nations.

9. References

ABARD NEWS. 2005. *Why e-health?* e4, Nov. [Online]

Available at: http://66.102.9.104/search?q=cache:6_JXgS2sGMMJ:www.a-bard.org/Repository/ABARD%2520Newsletter%25204%2520-%2520November05.pdf+ABARD+News&hl=en&ct=clnk&cd=3

[Accessed: 17 May 2007].

- ABOUZAHAR, C & BOERMA, T. (2005). Health information systems: the foundations of public health. *Bull World Health Organ*, 83 (8), Geneva, Aug.
- AMERICA ACADEMY OF FAMILY PHYSICIANS. (2007). *Potential Benefits of EHR*, [Online]. Available at: <http://www.centerforhit.org/x1117.xml> [Accessed: 05/07/2007].
- ANTHONY, S. C., BENSIK, M., ARMFIELD, N., STILLMAN, J. & CAFFERY, L. 2005. Telemedicine and rural healthcare applications. *Journal of Postgraduate Medicine*, 51(4), Oct-Dec: 286-293
- BABBIE, E. (2005). *The basics of social research*. Thomson Wadsworth.
- BLAIR, R. (2007). Disaster-Proof Patients. *Health Management Technology*, 28(2), Feb.
- BONDO DISTRICT HEALTH RECORD (2006). Bondo district health plan, *Ministry of Health*. Republic of Kenya.
- BONDO DISTRICT DEVELOPMENT PLAN. 2002-2008. Effective management for sustainable economic growth and poverty reduction. Kenya: *Ministry of Finance and Planning*.
- BONDO DISTRICT HEALTH PLAN 2007/2008. *Ministry of Health*. Republic of Kenya.
- CENTRAL BUREAU OF STATISTICS, 2006. *Kenya facts and figure*. [Online] Available at: <http://www.cbs.go.ke/downloads/pdf/Kenyafacts2006.pdf?SQMSESSID=101561fb04e4511cde553e35a3da421e&SQMSESSID=5c6af6c69dda837b8417fcd03e7045b0> [Accessed: 06-03-2007].
- CHANG, B., L., BAKKEN, S., BROWN, S., S., HOUSTON, T., K., KREPS, G., L., KUKAFKA, R., SAFRAN, C. and STAVRI, P., Z. (2004). Bridging the digital divide: Reaching vulnerable populations. *Journal of the American Medical Informatics Association*, 11(6), pp: 448-457, Nov / Dec 2004.
- CHULAIMBO HEALTH PLAN 2002-2007. *Ministry of Health*. Republic of Kenya
- COOMBES, H. (2001). *Research using IT*. Basingstoke: Palgrave, 2001.
- COUPER, M. P. (2007). Issues of representation in e-health research (with a focus on web surveys). *American Journal of Preventive Medicine*, 32(5S).
- CURRY, S. J. (2007). E-health research and healthcare delivery: Beyond intervention effectiveness. *American Journal of Preventive Medicine*, 32(5).
- DARIO, C., DUNBAR, A., FELICIANI, F., GARCIA-BARBERO M., GIOVANNETTIL S., GRASCHEW, G., GÜELL, A., HORSCH, A., JENSSEN M., KLEINEBREIL, L., LATIFI, R., LLEO, M. M., MANCINI, P., MOHR, M. T. J., GARCÍA, P. O. , PEDERSEN, S. , PÉREZ-SASTRE, J. M. & REY, A. (2004). Opportunities and challenges of e-health and telemedicine via satellite. *Eur J Med Res Supplement. Proceedings of ESRIN-Symposium*, July 5, Frascati, Italy.
- DENISE, 2003. *The case for e-health*. [Online] Available at :http://www.eipa.nl/Publications/Summaries/03/2003_E_01.pdf [Accessed: 23-02-2007].

EDWARDS, J. (2007). Electronic Health Records: Essential IT Functions and Supporting Infrastructure. *Gartner Publication*, G00145016, Jan.

EUROPEAN COMMISSION, ENTERPRISE DIRECTORATE GENERAL. (2004). *Electronic Business in Health and Social Care Service Sector*. The quantitative picture Diffusion of ICT and e-business in 2003/04. [Online]

Available at: http://www.ebusinesswatch.org/resources/health_generic/SR10-I_Health.pdf

[Accessed: 05/06/2007]

GANS, D., KRALEWSKI, J., HAMMONS, T. & DOWD, B. (2005). Medical Groups': Adoption of Electronic Health Records and Information Systems. *Health Affairs*, 24(5), Sep/Oct.

GARRIDO, T., RAYMOND, B., JAMIESON, L., LIANG L. & WIESENTHAL, A. 2004. Making the business case for hospital information systems. *Journal of Healthcare Finance*, 31(2): 21–22.

HANSON, C. W. 2006. *Healthcare informatics*. New York: McGraw Hill.

HOMABAY DISTRICT HEALTH RECORD (2007/2008). Homabay district health plan Republic of Kenya: *Ministry of Health*.

HOMABAY DISTRICT DEVELOPMENT PLAN 2007/2008. *Effective management for sustainable economic growth and poverty reduction*. Kenya: Ministry of Finance and Planning.

HOMABAY DISTRICT HEALTH PLAN 2007/2008. Republic of Kenya: *Ministry of Health*.

INSTITUTE OF MEDICINE (IOM), 2001. *Crossing the quality chasm: A new health system for the 21st century*. [Online] Available at: www.iom.edu

[Accessed on: 05/06/2007]

KISUMU DISTRICT HEALTH RECORD (2007/2008). Kisumu district health plan. Republic of Kenya: *Ministry of Health*.

KISUMU WEST DISTRICT DEVELOPMENT PLAN. 2002, 2008. *Effective management for sustainable economic growth and poverty reduction*. Kenya: *Ministry of Finance and Planning*.

KISUMU DISTRICT HEALTH PLAN 2007/2008. Republic of Kenya: *Ministry of Health*.

KURTINAIYTYTE, L. (2007). *E-health: The use of ICT in developing healthcare system*. Multiple case study of European countries: Denmark and Lithuania. Published dissertation in International Marketing, University of Halmstad..

KWANKAM, S. Y. 2004. What E-Health can offer? *Bull World Health Organ*, 82(10), Oct.

KRUGER, S. J. & WELMAN, J. C. (2001). *Research methodology*. Cape Town: Oxford University Press.

LUES, L. & LATEGAN, L., L. (2006). *Research ABC*. Sun Press

LEEDY, P.D. AND ORMROD, J.E. (2005). *Practical research: planning and design*. Columbus: Prentice hall.

LORENCE D. (2006). Web-based consumer health information: Public access, digital division, and remainders. *Medscape General Medicine*, 8(2):e4.

MEA, V. D. (2006). Pre-recorded telemedicine. In *Introduction to telemedicine*. 2nd ed. WOOTON, R., CRAIG, J. & PATTERSON, V. (Eds.). London: RSM Press.

- MILLER, R. H. & WEST, C. E. (2007). Market watch: The Value of Electronic Health Records In Community Health Center. *Health Affairs*, 26(1). Jan. /Feb.
- MINISTRY OF HEALTH (KENYA). (2006). *Kenya health facilities by district, type and running agency as by 2005-2006*. [Online]
- Available at:<http://www.health.go.ke/Healthfacilities/Health%20facilities%20by%20District.xls>
- [Accessed on: 12/12/07]
- NOVAK, K. (2005). *Reducing costs through Electronic Health Records and Services*. Benefits and Compensation Digest. [online].
- Available at: www.ifebp.org
- [Accessed: 05/06/2007].
- NYAMIRA DISTRICT DEVELOPMENT PLAN 2002, 2008. *Effective management for sustainable economic growth and poverty reduction*. Kenya: Ministry of Finance and Planning.
- NYAMIRA DISTRICT HEALTH PLAN 2007/2008. Republic of Kenya: *Ministry of Health*.
- NYANZA PROVINCIAL DEVELOPMENT PLAN, 2007(KENYA). *Ministry of planning and development*.
- OLATOKUN, W. M. & AJIFERUKE I. (2006). *E-commerce challenges and policy considerations in Nigeria. Encyclopaedia of e-Commerce ,e-Government and Mobile Commerce*. Volume 1. London: Idea Group Reference.
- OLIVIER, M. S. (2004). *Information technology research: A practical guide for computer science and informatics*. Pretoria: Van Schaik
- OJO, O., OLUDAY, O., OLUGBARA, O., SUNDAY, M., ADIGUN, O. J., EMUOYIBOFARHE, O. & XULU, S. S. 2007. An architectural framework for rural e-healthcare information infrastructure with web service-enabled middleware support. In *Proceedings of HELI NA 2007, conference on scientific at a glance* , January 8–12, 2007. Bamako, Mali[Online].
- Available at: <http://www.sim.hcuge.ch/helina/6.pdf>
- [Accessed on: March 15th 2007].
- PODICHETTY, V. K. & BISCUP, R. S. (2003). E-Health, a new approach in healthcare practice. *Internet Journal of Allied Health Sciences and Practice*, 1(2).
- RICHARDSON, R. J. 2006. *The role of e-health*. [Online] Available at: http://www.medetel.lu/download/2006/parallel_sessions/abstract/0407/Richardson.doc?PHPSESSID=cc6aa23c71108db22338556e858f183e.
- Accessed: [05/03/2007].
- RUIZ, J. G., MINTZER, M. J. & LEIPZIG, R. M. (2006). The impact of e-learning in medical education. *Academic Medicine*, 81(3).
- RUXWANA, L. K. (2007). *Technology assessment of rural hospitals in the Eastern Cape province*. Dissertation in Information Technology, Tshwane University of Technology.
- SAMAKE, K. B. & MBARIKA, V. W. (2007). *HELINA 2007 scientific programme*. [Online]
- Available at:<http://www.sim.hcuge.ch/helina/22.pdf>
- [Accessed: 05/06/2007]

- TAN, J. 2005. *E-Healthcare information systems; an introduction for students and professionals*. Jossey bass.
- THAKKAR, M. & DAVIS, D. C. 2006. *Risks, Barriers, and Benefits of EHR Systems: A Comparative Study Based on Size of Hospital, Perspectives in Health Information Management*. [Online].
Available at:
http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_031769.pdf
[Accessed: 20/05/07].
- THIELST, C. B. (2007). The New Frontier of Electronic, Personal, and Virtual Health Records. *Journal of Healthcare Management*, 52(2), Mar. /Apr.
- THOMAS, A. B. 2004. *Research skills for management studies*. London: Routledge.
- TOMASI, E., FACCHINI, L. A. & MAIA, M. S. (2004). Health information technology in primary healthcare in developing countries. *Bulletin of the World Health Organization*, Nov. 82(11).
- WOOTON, R., CRAIG J. & PATTERSON, V. (eds). (2006). *Introduction to telemedicine*. London: The Royal Society and Medicine Press.
- WUTOH, R., BOREN, S. A., & BALAS, E. A. (2004). E-Learning: A Review of Internet-Based Continuing Medical Education. *Journal of Continuing Education in the Health Professions*, 24: 20-30.

8

The Evaluation of Information Visualisation Techniques Using Eye Tracking

André P. Calitz, Marco C. Pretorius, Darelle Van Greunen

The general increase in the use of information visualisation techniques has highlighted the need for methodologies to evaluate the user interface of software system utilising these techniques. Usability evaluation techniques have evolved over several years in order to assess the user interface of systems with regard to efficiency, interaction flexibility, interaction robustness and quality of use. The evaluation of the user's thought processes when using software systems is difficult to access with traditional usability techniques. Eye movement data can supplement the data obtained through usability testing by providing more specific information about the strategies that users apply. In this paper, the authors investigate how eye tracking data can supplement the usability evaluation data of information visualisation techniques, by applying eye tracking in a usability evaluation case study of an information visualisation tool. The results of a pilot study have previously been reported and these results are compared specifically to the main case study. The results of the main case study are reported in this article and illustrate that eye tracking does provide additional value to the usability evaluation of information visualisation techniques.

1. Introduction

The aim of information visualisation techniques is to present data using methods that accurately communicate information, and needs minimal effort for comprehension and interpretation [Lee et al. 2003]. The availability and ability of information visualisation tools to generate data, in order for users to make informed decisions, has increased. However, the ability to access and analyse information to make informed decisions is on the decline [Spotfire 2001]. There is a growing need for usable solutions to facilitate and support the visualisation of data and the associated decision-making process. It is very important that these tools be evaluated to determine that these solutions are indeed usable.

Usability testing is the process of obtaining information from users about a software product's usability by observing them using the product [Benel and Ottens 1991]. Testing usability means making sure that users can use a software product quickly and easily to accomplish required tasks. The role of the user's visual attention can provide additional information on usability testing, but is difficult to assess with traditional usability methods such as click analysis, questionnaires or asking users what was important and what they focused on. Eye tracking has become a capable tool to answer research questions relating to where the user's visual attention is on the screen.

Research has shown that the user's eyes do not wander randomly and that people look at specific areas of interest on the screen [Sibert and Jacob 2000]. Information can be obtained where the user's attention is at any point in time and provide an indication to how the user perceived the information viewed. In usability engineering, eye tracking assists software designers to evaluate the usability of screen layouts. The assessment of the user's visual attention cannot be measured by means of think-aloud protocols or questionnaires, utilised by traditional usability studies. Incorporating eye tracking with software usability evaluation can provide additional knowledge that is not obtained from traditional usability testing methods [Karn et al. 1999].

Information visualisation tools allows for the exploration of large amounts of data and information in order to gain a better understanding and to discover hidden relationships and patterns within data. Traditional usability evaluation techniques have been used for the evaluation of information visualisation tools. This paper investigates whether the addition of eye tracking in formal usability evaluations will provide additional insights that are not available using standard usability methods. The results of a previous pilot study have been reported at the SAICSIT conference [Pretorius et al. 2005]. This study will report the results from the main case study.

2. Information Visualisation Techniques

Information visualisation (IV) is defined as *“the use of computer-supported interactive, visual representations of data to amplify cognition, where cognition is the acquisition or use of knowledge”* [Card et al. 1999]. IV enables users to discover, make decisions and provide explanations about patterns, groups of items, or individual items. The purpose of IV is therefore to gain insight and understanding. IV links the human mind and the modern computer, two very powerful information processing systems [Gershon and Eick 1997]. It is a process that transforms data, information and knowledge into a visual form by utilising people's natural strengths for rapid visual pattern recognition.

Text-based interfaces require cognitive effort to understand their information content. IV seeks to present information visually, in essence to offload cognitive work to the human visual perception system [Andrews 2002]. Interpreting network data sets for example Internet traffic patterns, network throughput, usage patterns and application delay, is made much easier by visual representations of this complex data.

A variety of IV techniques exist [Card et al. 1999]. The usability of these IV techniques is however largely untested. Usability studies on IV techniques have been researched in the past [Freitas et al. 2002; Stasko et al. 2000]. This study proposes that eye tracking will add in the usability of these IV techniques. Investigating users' strategies in using IV techniques can give us an understanding of their perception of these IV techniques. Plaisant [2004] states that IV is making steady gains and that new evaluation procedures should be explored. This study is a step towards new evaluation methods for IV techniques.

3. Usability and Eye Tracking

According to Jacob Nielsen [Nielsen 1993] usability is a quality attribute that assesses how easy user interfaces are to use. ISO 9241 defines usability as the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments [9241-11 1998].

Usability testing involves measuring the performance of users on tasks with regard to the ease of use, the task completion time, and the user's perception of the experience of the software application [Preece et al. 2002]. Usability testing can be conducted within a usability laboratory or by means of field observations. The usability evaluation for this research was conducted by means of a formal usability evaluation in the Nelson Mandela Metropolitan University (NMMU) usability laboratory. Formal usability testing is an empirical method that requires the design of a formal usability experiment that is carried out under controlled conditions in a usability laboratory.

Most traditional evaluation methods fail to provide useful results on usability evaluation of interfaces for IV [Freitas et al. 2002]. Therefore, usability techniques that will provide useful results on the usability evaluation of IV techniques need to be explored. Eye tracking is proposed as an additional usability method with traditional usability evaluations that will provide meaningful and useful results.

The integration of eye tracking with software usability evaluation can provide insights into Human-Computer Interaction that are not available from traditional usability testing methods [Karn et al. 1999]. Eye tracking contributes to design knowledge and evaluation by providing detailed capture of user interaction behaviour [Tzanidou et al. 2005]. Eye tracking can be defined as a technique to determine eye movement and eye-fixation patterns of a person [Namahn 2001]. The human eye moves by alternating between saccades and fixations. A saccade is the quick movement of the eye in order to move focus from one area to the next. A fixation is the time spent looking at the new area on the screen. An eye tracker follows the eye around during its saccades and tracks the location of the fixation points. Eye tracking studies have been used in diagnosing the effectiveness of designs with point of interest detection (fixation) and information transmission via eye movement (scan path) as two main indicators [Yoneki 2006].

Software designers can gain useful information about natural human movements, by tracking eye movements [Newman 2001]. Eye tracking is based on the fact that a record of a person's eye movements, while doing a task, provides information about the nature, sequence and timing of the cognitive operations that took place [Rudmann et al. 2003]. Eye tracking data can expose response biases of participants resulting from an artificial testing environment. This would be undetected in traditional usability testing techniques and therefore eye tracking data results in a higher validity of usability data [Schiessl et al. 2003].

Traditional usability methods provide data on a descriptive level only, while the additional application of eye tracking offers insight in the origins of a problem [Schiessl et al. 2003]. Perception and comprehension are just some of the stages where problems can occur. Eye tracking allows detailed analysis of these stages.

Goldberg et al. [2002] states that variables that are derived from eye tracking methods can provide insight into users' decision making while searching and navigating interfaces. Scan paths, the time spent looking at various areas of interest on the screen and the use of visual attention are some of the benefits that eye tracking can add to the usability testing of software and specifically IV techniques.

4. Case Study: Appvis 1.0

A case study approach was followed which involved the design of a usability experiment with eye tracking in order to evaluate the usability of an IV tool. The IV tool selected for the case study was AppVis 1.0, a prototype system developed by a Masters student in 2004 [Rademan 2004]. A pilot study preceded the case study. Preliminary results showed that eye tracking data combined with usability evaluation data did add value to the evaluation of Network Management Tools [Pretorius et al. 2005]. The case study results will be discussed next.

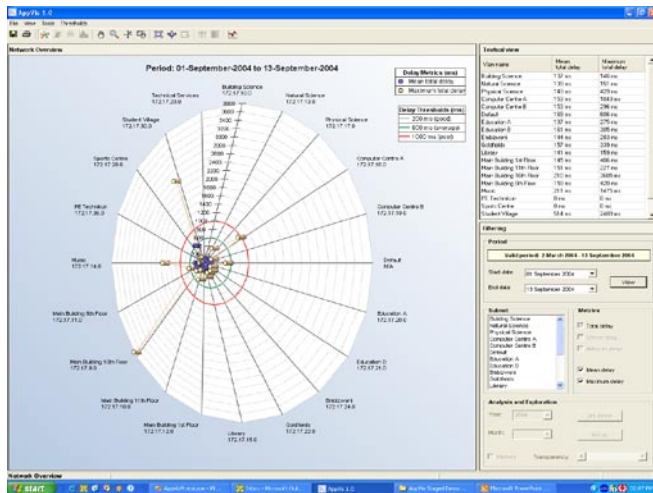
4.1 AppVis 1.0 and the IV Techniques used

AppVis 1.0 is an IV tool that allows network managers at the Nelson Mandela Metropolitan University (NMMU) analyse and explore application performance on the NMMU networks. Application performance management entails the comprehension of how a network application performs from a user perspective [Rademan 2004]. NMMU has an extensive network infrastructure that supports several application services. This prototype system, AppVis 1.0, uses novel IV techniques to visualise the application delay performance of the Integrated Tertiary Software (ITS) application implemented at NMMU. AppVis 1.0 contains various different types of IV techniques discussed in the next subsection. These IV techniques will be evaluated in this study.

4.1.1 Network Overview

The star metaphor is the IV technique used to display a network overview of application total delay. The visualisation technique plots both the mean and maximum application total delay for each Virtual Local Area Network (VLAN), across an associated axis emanating from the origin of the graph for user specified periods. Application delay metrics that are plotted further from the origin indicate higher delay measures for a VLAN while those plotted close to the origin indicate low delay measures for a VLAN. Threshold bands are incorporated into the visualisation technique as concentric rings that categorise the delay measures plotted for each VLAN on the graph.

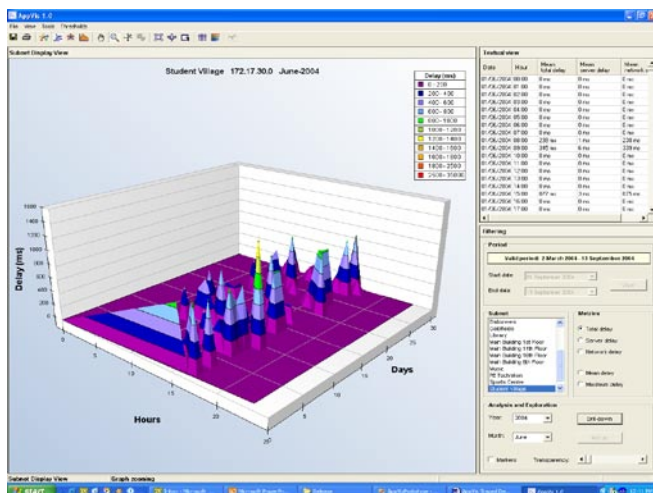
Figure 1: Network Overview (star metaphor)



4.1.2 Subnet View

The landscape metaphor is the IV technique used to display the subnet view of application delay for an individual VLAN. The visualisation technique indicates periods of peak application delay across a monthly period by making use of colour-coded surface levels on the graph. The application delay is displayed in milliseconds (ms) for each hour of every day for a selected month. The X-axis represents the hours of a day, the Y-axis represents the delay in ms and the Z-axis represents the day within the month.

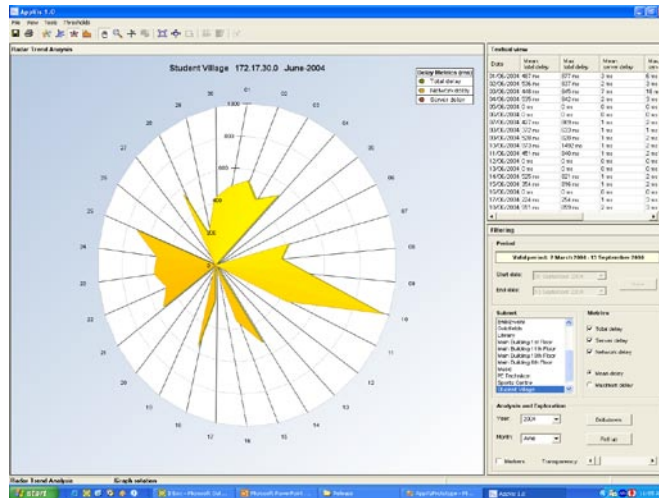
Figure 2: Subnet View (landscape metaphor)



4.1.3 Radar Trend Analysis

The radar graph is the IV technique used to display application delay trends across time. This graph displays application server, network and total delay of a VLAN one above the other across a series of axes in a clockwise fashion at higher levels of aggregation based on time. Application delay can be displayed across the days of a selected month or across the months of a selected year.

Figure 3: Radar Trend Analysis (radar graph)



4.2 User Profile

AppVis 1.0 is classified as a Network Management tool and for use by network managers at NMMU. The main criterion for the users of this product was to have a sound knowledge in the domain of network performance management and to have previous knowledge of IV techniques. A background questionnaire was used to screen the participants for this evaluation. The background questionnaire was sent to all Network Managers employed at the NMMU. This questionnaire reflected the possible participant's NM tool experience, computer experience, IV technique experience, age and gender. Questions regarding the participants' eye sight were also asked in the questionnaire. Eight participants (all the network managers at NMMU) were tested in this study. Three of the participants wore glasses or contact lenses and it did not affect the accuracy of the eye movement data. Their right eye was calibrated for the eye tracking purposes. The participant profile is included in Table 1. One participant was colour-blind. Seven participants had previous experience with IV techniques.

Table 1: Profile of the participant population

Total	Gender		Age		
	Male	Female	20-29	35-49	
8	7	1	3	5	
Total	Years Experience			NM tools Used	
	<1	1-3	4+	1-3	4+
8	1	2	5	3	5

4.3 Evaluation Metrics

4.3.1 Usability Metrics

Several usability techniques and metrics [Andrews 2002; Dumash and Redish 1999; Faulkner 2000; Rosson and Carol 2002; Rubin 2002] and eye tracking techniques and metrics [Benel and Ottens 1991; Cowen et al. 2001; Gao 2001; Goldberg and Kotval 1999; Goldberg et al. 2002; Renshaw et al. 2003; Xu 2000] were investigated for this study. The usability and eye tracking metrics identified to use in the case study is depicted in Table 2.

The task list for this case study was designed in such a way as to allow these metrics to be captured. Data was collected and calculated by means of: video recordings; eye tracking video and data files; post-test questionnaires; and monitoring of tasks. The post-test questionnaire used a five-point Likert scale and was used to gather feedback on the participants' perception of the user interface and system. The questionnaire was a combination based on the Computer System Usability Questionnaire and metrics obtained by Freitas et al. [2002] in the usability of IV techniques. The questionnaire results are not discussed in this paper.

4.3.2 Test Procedure

The participants were welcomed and briefed about the experiment, which was followed by an explanation of the equipment to be used. It was explained that only the eye, voice and stimulus display would be recorded. The participant was required to complete an informed consent form. Participants were encouraged to share their thoughts and opinions of a task after it was completed. A training task was given to the participants where they were briefed about the system goals and objectives as participants were not familiar with AppVis 1.0.

After the training tasks, the participants were given time to become comfortable in front of the PC before the eye tracking calibration commenced. A 9-point calibration with corner correction was used at all times. The participants were asked to keep their head as still as possible during the experiment as to minimise inaccuracy caused by head movements. Participants were offered the opportunity to stand up and relax half-way through the experiment. After every three or four tasks, depending on the task length, the accuracy of the eye movements was checked. If the accuracy would appear to be incorrect, the participant's eye would be recalibrated. Data recording commenced with the test administrator reading the task, and ended with the participant either answering or completing the task. The duration of the experiment was between 40 minutes and

one hour. Following the tasks, a post-test questionnaire was administered. Finally an experiment debriefing was conducted.

Table 2: Usability and eye tracking metrics

Usability metrics	
Effectiveness	Task completion rate. The percentage of tasks each participant completed successfully in the task list, including the percentage of tasks completed per participant and per task.
	Number/percentage of tasks completed with and without assistance. Number or percentage of participants who completed a task without assistance and also those who completed with assistance.
	Error rate recovery. Monitoring the number of errors made by the participant, as well as the total errors from which the user could not recover.
Efficiency	Task completion time. Measuring the total time participants spent performing the assigned tasks.
	Real-time events. Monitoring and filtering events such as the click of a mouse, the push of a key or the participant making a comment.
Satisfaction	Post-test questionnaire. A five-point Lickert scale was used to collect participant responses relating to the overall perception of the system.
	Interview. Likes, dislikes and issues encountered.
Eye Tracking Measures	
Number of fixations	The number of fixations is negatively correlated with search efficiency. Large numbers of fixations point to less efficient search perhaps resulting from poor display element arrangements.
Fixation duration	Relatively long fixation duration is an indication of the complexity and difficulty of a display.
Number of fixations on each AOI	This metric was an indication of the importance of a system element. For example, instructions being read or not. (Area of Interest (AOI)).
Number of gazes on each AOI	The eyes are drawn to informative areas. This metric also reflects the importance of a system element.
Scan path	The strategy that a participant used to complete a task was obtained from this metric. This gives an indication of the efficiency of the arrangement of elements in the user interface.
Time to the 1st fixation on the target AOI	This metric is useful for tasks where a precise search target exist.

5. Results

The following section discusses the task(s) for each IV technique used in the AppVis 1.0 together with the traditional usability results obtained and the eye tracking results obtained. Each section contains a conclusion of the results found for the specific IV technique.

5.1 Network Overview

5.1.1 Network Overview: Task 1

Participants were required to determine which VLAN had the highest maximum total delay, using the Network Overview (star metaphor) IV graph. This information could be extracted from the graph or from the textual view.

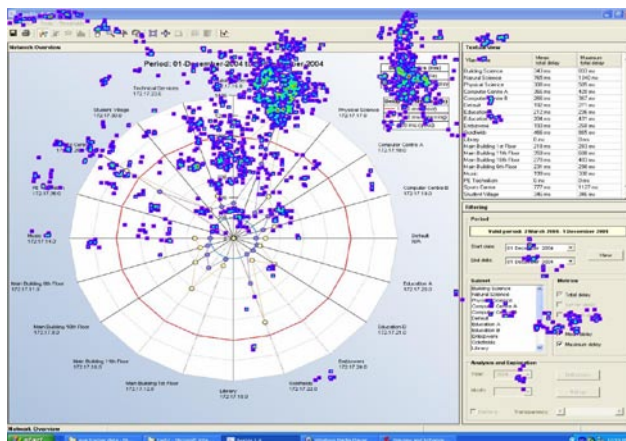
Usability Results: This task had a 100 % completion rate and all participants gave the correct answers. This task had a mean completion time of 13 s (StdDev 2.62s), with a maximum time of 17s and a minimum of 10s.

Eye Tracking Results: The eye tracking data showed that all the participants used the graph to obtain their answer with the number of fixations, fixation percentage and gaze percentage favouring the graphical AOI. The graphical AOI had a mean gaze percentage of 72.60 % (StdDev 13.23 %). Figure 4 shows the heat map of all the participant fixations. The “hotter” areas, green and red in this example, indicate the areas mostly viewed. In this case the correct answer is mostly viewed. A large concentration of the fixations is also on the legend, indicating the participants’ strategies. Participants first looked at the colour of the delay on the legend, before identifying it on the graph.

5.1.2 Network Overview: Conclusions

All the participants obtained the correct answers from the graph, in short enough time, fixating in the correct areas. This all indicated that the graph was easy to use and provided the relevant data and information to network managers.

Figure 4: Heat map for Network Overview Task 1



5.2 Subnet View

5.2.1 Subnet View: Task 1

Participants were required to filter the subnet view to display only the server delay occurring across the specific VLAN. Once the Subnet View (landscape metaphor) graph was filtered the participants needed to determine the highest server delay.

Usability Results: The mean completion rate for this task was 100 % with all participants filtering the graph correctly. A mean time of 10.14 s was needed to filter the graph. However, the mean task completion time was 39.86 s (StdDev 11.75 s) with the participants taking particularly long to answer the question. It is important to note that not one participant gave the correct answer. One participant made a comment that it is difficult to read from the graph. Only one participant tried to rotate the graph, but had difficulty doing so. The participant expected to rotate by dragging the graph and did not know that the rotate option must be selected from the toolbar first.

Figure 5: Fixation path for Subnet View Task 1

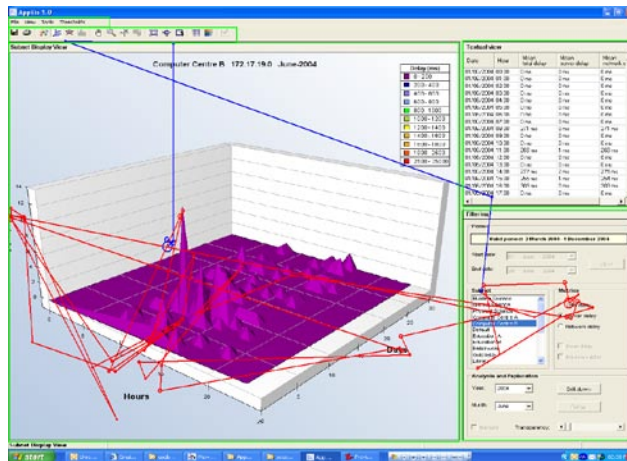
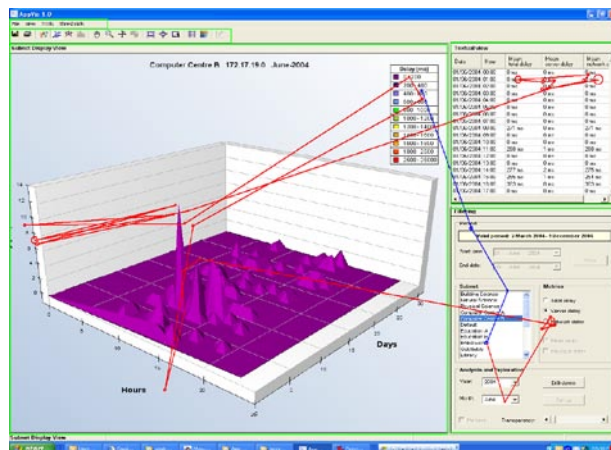
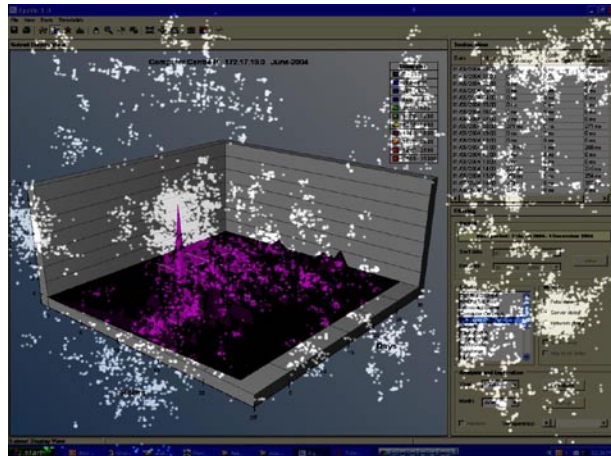


Figure 6: Another fixation path for Subnet View Task 1



Eye Tracking Results: The mean time to the first fixation on the filtering AOI, where the graph was filtered, was only 3.15 s (StdDev 0.85 s), showing that participants expected to filter the graph here. A mean of 34.33 fixations (StdDev 7.17) were recorded after the graph was filtered, indicating the difficulty participants had in finding the answer. The graphical AOI had the most fixation and gaze time with seven participants using the graph to obtain the answer. This indicates that the participants struggled to obtain the answer from the graph. Figure 5 and 6 shows the fixation paths of various participants. It can clearly be seen that the participants looked at the peak of the graph and tried to match it up with the corresponding label on the Y-axis. The correct answer was between 12 and 14 ms. It can be seen from Figure 5 and 6, that not one fixation was made on the correct area between 12 and 14 ms. The participants fixated mainly in the area between 6 and 10 ms. Figure 7 shows the focus map of all participant fixations for this task, with highlighted areas indicating dense distributions of the fixations. The Y-axis to the left is “highlighted” between 6 and 10 ms, with almost no light between 12 and 14 ms. The eye tracking results indicated that this graph was difficult to interpret.

Figure 7: Focus map for Subnet View Task 1



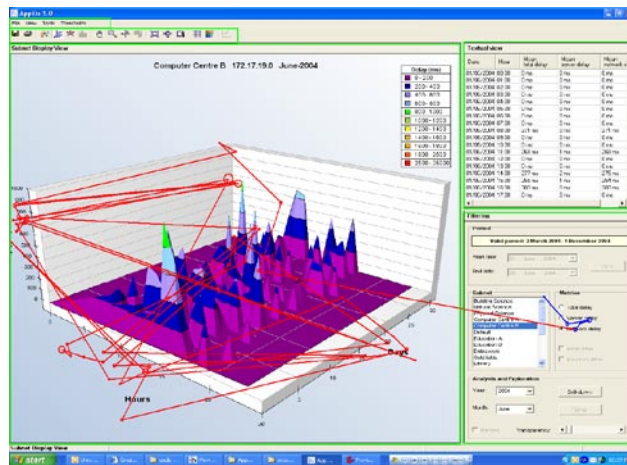
5.2.2 Subnet View: Task 2

Similar to the previous task, this task required the participants to filter the subnet view to display only the network delay occurring across the specific VLAN. Once the graph was filtered the participants needed to determine the highest network delay.

Usability Results: The mean completion rate for this task was 100 % with all participants filtering the graph correctly. A mean time of seven seconds was needed to filter the graph, a 3.14 s improvement on the previous task. The mean task completion was also improved by 10.43 s. Once again, the participants took a long time to answer the question. This time, three participants answered the question correctly. A participant also commented on the difficulty of reading the graph. One participant clicked on the textual view, but gave the incorrect answer as the textual view did not sort correctly.

Eye Tracking Results: Once again, participants fixated on the filtering AOI relatively fast, in order to filter the graph. Less fixations (26.17) were needed to give the answer after the graph was filtered. Participants found their answers quicker than in the previous task. However, this time only two of the participants used the graph, while the other used the textual AOI. The graphical AOI still had the most fixation and gaze time, but this time the participants who used the textual AOI, first used the graph and then proceeded to the textual AOI. This indicates that the participants struggled with the graphical AOI and proceeded to the textual AOI, where they found their answers. Figure 8 illustrates the fixation path of a participant who used the graph to obtain the answer. Again, it can clearly be seen that the participant fixated in the incorrect areas (between 600 and 900 ms), with no fixations on the correct areas (between 900 and 1000 ms).

Figure 8: Fixation path for Subnet View Task 2



5.2.3 Subnet View: Conclusions

The usability results of Task 1 indicated that participants struggled to read information from the Subnet View, but the eye tracking results showed that participants did not understand the 3-Dimensional graph, because they fixated on the wrong areas of the graph, resulting in providing incorrect answers. The eye tracking results of Task 2 established the results by showing that several participants attempted to read information from the graph, but because they struggled, proceeded to search in the textual AOI to obtain their answers. The use of a 3-Dimensional graph is questioned by the authors and the data could possibly be displayed more effectively using 2-Dimensional graphs.

5.3 Radar Trend Analysis

5.3.1 Radar Trend Analysis: Task 1

This task was completed successfully by filtering the radar trend analysis to display the maximum network, server and total delay for 2004. The participants were then asked to identify the maximum total and network delays.

Usability Results: This task had a 100 % task completion rate, meaning all participants filtered the graph correctly. Seven participants identified the maximum total delay

correctly, but only five identified the maximum network delay correctly. The mean time for the total delay answer was 6.86 s (StdDev 3.29 s), while the mean time for the network delay answer 14.57 s (StdDev 7.85 s).

Eye Tracking Results: The eye tracking data showed positive results for the filtering task with a mean time of only 2.35 s (StdDev 1.64 s) till the first fixation on the filtering AOI. This shows that participants expected to find the information there.

The first answer was also obtained relatively easily, with a mean of 7.86 fixations (StdDev 3.72) before the answer was given. The mean average fixation duration of 0.312 (StdDev 0.030) indicated that participants did not find it difficult to extract the information from the graph. The eye movement data showed that only one participant used the textual AOI to find the answer.

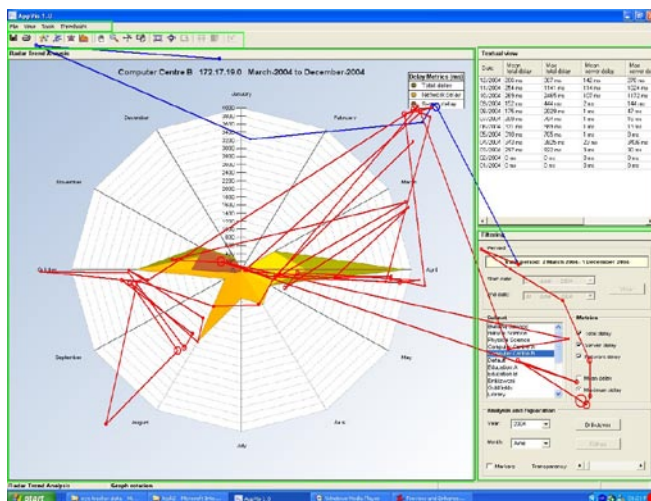
The second answer proved to be more difficult. The usability results already showed that fewer participants answered correctly. The answer was much more difficult to obtain with a mean number of 18 fixations (StdDev 3.42) before the answer was given. This indicates a less efficient search for this answer. Two participants used the textual AOI to obtain their answer this time.

Figure 9 shows the fixation path of a participant obtaining the wrong answer for the network delay. The fixations on the peaks of August and October can clearly be seen, but the participant incorrectly identified August as the answer (correct answer was October). The participants had difficulty to distinguish between peaks at the lower end of the graph.

5.3.2 Radar Trend Analysis: Conclusions

The usability results showed that participants found the first answer easily but struggled more with the second answer. Eye tracking data indicated that participants had difficulty to distinguish between peaks at the lower end of the graph. It is suggested that the value legend on the graph should be allowed to rotate, so that values can be compared easier, or the legend should be extended to the lower end of the graph as well.

Figure 9: Fixation path for Radar Trend Analysis Task 1



6. Conclusions

The increased use of IV techniques has necessitated the use of other evaluation techniques with traditional usability evaluation methods, for the usability evaluation of IV techniques. Eye tracking was identified as a valuable technique to provide the added value. Not only will it provide an insight into the strategies that users employ when using IV techniques, but it also provides us with such measures like fixation and gaze times in relevant areas of interfaces, reflecting the importance of these areas. The number and duration of fixations also gives us insights into the difficulty of extracting information from the IV techniques used.

Eye tracking data was obtained in a case study on a network management tool using various IV techniques. The results indicated that participants found information easy and interacted effectively with the star metaphor used in the network overview. The radar graph used in the radar trend analysis was found to be relatively easy to work with but would be easier to understand if a value legend was available for the bottom half of the graph. Eye tracking data also showed that participants had difficulty extracting information from the subnet view using the landscape metaphor (a 3-Dimensional graph). Not one participant gave the correct answer for the first task and several participants retreated to the textual view for the second task.

The combined usability and eye tracking data showed that the star metaphor IV technique was the easiest to use. The radar graph IV technique had some problems, because the participants could not identify the highest peaks towards the bottom half of the graph. This was due to the value legend not being available in the bottom half of the graph. If this is rectified this IV technique could prove to be useful. The 3-Dimensional landscape metaphor was the IV technique that participants disliked and had difficulty using. Freitas et al. [2002] states that most traditional evaluation methods fail to provide useful results on usability evaluation of interfaces for IV. In this study, eye tracking gave further insight, showing that participants had difficulty with the interpretation of the 3-Dimensional graph because they fixated on the wrong areas of the graph. The use of 2-Dimensional graphs would be a more suitable solution.

The integration of eye tracking with usability evaluation can provide insights into Human-Computer Interaction that are not available from traditional usability testing methods [Karn et al. 1999]. This paper showed that by adding eye tracking evaluation methods to traditional usability evaluation methods, value will be added to the evaluation of IV techniques. The eye tracking data added to the understanding of the users' data interpretation strategies as it allowed the scan paths to be analysed. The combined method offered the opportunity to measure user actions, but to also record eye movements which were critically important when dealing with IV techniques.

IV is making sound growth and new evaluation procedures should be investigated Plaisant [2004]. A case study has been conducted to combine usability evaluation methods with eye tracking methods and the proposed combination of these methodologies has been successfully implemented. Analyses and results of the case study indicate that eye tracking does give added value to the usability evaluation of IV techniques.

7. Future Research

Future research may include inter-disciplinary studies on the preferences of participants of textual view versus graphical views of interfaces. Several results of the colour-blind participant showed that in some instances, colours could not be distinguished. This resulted in the participant identifying incorrect objects and answers on the graphs. Future research could include evaluating the effectiveness of colours used on the graph by colour-blind participants. Future research will have more participants completing a longer task list. Other IV systems where participants have already obtained experience would also be tested.

Acknowledgments

The authors would like to thank the Telkom Centre of Excellence Programme and the Department of Computer Science and Information Systems at the NMMU for making this research possible.

References

- 9241-11, I. *Guidance on Usability* 1998. Available from: <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16883>.
- ANDREWS, K. (2002). Information Visualisation. Tutorial Notes. Graz University of Technology Infflegasse 16c A-8010 Graz. <http://www2.iicm.edu/ivis/ivis.pdf> , Date accessed: 05 September 2005.
- BENEL, D. AND OTTENS, D. (1991). Use of an Eye Tracking System in the Usability Laboratory. *Proceedings of the Human Factors Society 35th Annual Meeting*.
- CARD, S., MACKINLAY, J. AND SHNEIDERMAN, B. (1999). *Readings in Information Visualisation: Using Vision to Think*. Morgan Kaufmann Publishers, Inc.
- COWEN, L., BALL, L. AND DELIN, J. (2001). An Eye Movement Analysis of Webpage Usability. *In Proceedings of the annual HCI 2002, 317-335*.
- DUMASH, J. AND REDISH, J. (1999). A Practical Guide to Usability Testing. *Intellect Books*.
- FAULKNER, X. 2000. Usability Engineering. *Macmillan Press*.
- FREITAS, C., LUZZARDI, P., CAVA, R., WINCKLER, M., PIMENTA, M. AND NEDEL, L. (2002). Evaluating Usability of Information Visualization Techniques. *Proceedings of 5th Symposium on Human Factors in Computer Systems*.
- GAO, T. 2001. Processing Numeric Table Data in Electronic Formats: Eye Movement Patterns and Searching Efficiency. Masters Paper. *UNC School of Information and Library Science*.
- GERSHON, N. AND EICK, S. (1997). Information Visualisation, *Proceedings of 1997 IEEE Computer Graphics and Applications*, 29-31.
- GOLDBERG, J. AND KOTVAL, X. (1999). Computer Interface Evaluation Using Eye Movements: Methods and Constructs. *International Journal of Industrial Ergonomics, Volume 24, 631-645*.
- GOLDBERG, J., STIMSON, M., LEWENSTEIN, M., SCOTT, N. AND WICHANSKY, A. (2002). Eye Tracking in Web Search Tasks: Design Implications. *In Proceedings of the symposium on ETRA 2002: Eye tracking research & applications symposium, 51-58*.

- KARN, S., ELLIS, S. AND JULIANO, C. (1999). The Hunt for Usability: Tracking Eye Movements. *CHI 99 Extended Abstracts*. 173. NY: ACM Press.
- LEE, M., REILLY, R. AND BUTAVICIUS, M. (2003). An Empirical Evaluation of Chernoff Faces, Star Glyphs, and Spatial Visualizations for Binary Data. *Proceeding of the Australian Symposium on Information Visualisation*, 1-10. Sydney: Australian Computer Society Inc.
- NAMAHN. 2001. Using Eye Tracking for Usability Testing. www.namahn.com/resources/documents/note-eyetracking.pdf , Date accessed: 09 September 2005.
- NEWMAN, J. 2001. Eye Tracking on the Internet. *Cornell University, Human Computer Interaction Lab*. 13 Dec. 2001.
- NIELSEN, J. 1993. Jacob Nielsen's Alertbox: Usability 101. <http://www.useit.com/alertbox/20030825.html>, Date accessed: 19 May 2004.
- PLAISANT, C. 2004. The Challenge of Information Visualization Evaluation. *Proc. IEEE Proceedings of Advanced Visual Interfaces 2004*, 109-116.
- PREECE, ROGERS AND SHARP. 2002. Interaction Design: beyond human-computer interaction. *John Wiley & Sons, Inc.*
- RADEMAN, J. 2004. The Visualisation of Application Delay Metrics for a Customer Network. Masters Dissertation. Dept of CS/IS, *Nelson Mandela Metropolitan University*.
- RENSHAW, J.A., FINLAY, J.E., TYFA, D. AND WARD, R.D. 2003. Designing for Visual Influence: an Eye Tracking Study of the Usability of Graphical Management Information. *INTERACT2003, Ninth IFIP TC13 International Conference on Human-Computer Interaction, Zurich, Switzerland, September 2003*.
- ROSSON, M. AND CAROL, J. (2002). Usability Engineering: Scenario-Based Development of Human-Computer Interaction. Morgan Kaufmann Publishers.
- RUBIN, J. (1994): Handbook of Usability Testing. *John Wiley & Sons, Inc.*
- SCHIESSL, M., DUDA, S., THÖLKE, A. AND FISCHER, R. (2003). Eye tracking and its application in usability and media research. www.cs.ucl.ac.uk/staff/J.McCarthy/pdf/library/eyetrack/EyeTracking-ResearchApplications.pdf Date accessed: 07 April 2004.
- SIBERT, L. AND JACOB, R. (2000): Evaluation of Eye Gaze Interaction. *Proceedings of the CHI 2000, ACM, New York*, 281-288.
- SPOTFIRE. 2001. Spotfire DecisionSite: Making High Quality Decisions at e-Business Speed. <http://www.spotfire.com> , Date accessed: 07 April 2004.
- STASKO, J., CATRAMBONE, R., GUZDIAL, M. AND MCDONALD, K. 2000. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies* 2000;53 663-694.
- XU, A. 2000. Eye-tracking study on identifying and analyzing user behavior - eye movement, eye fixation duration and patterns - when processing numeric table data in electronic PDF format. Masters Paper. UNC School of Information and Library Science.
- PRETORIUS, M. C., CALITZ, A. P. AND V. GREUNEN, D. (2005). The Added Value of Eye Tracking in the Usability Evaluation of a Network Management Tool. *Proceedings of the 2005 annual research conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT), White River, South Africa*.

- RUDMANN, D. S., MCCONKIE, G. W. AND ZHENG, X. S. (2003). Eyetracking in Cognitive State Detection for HCI. *ICMI'03, November 5–7, Vancouver, British Columbia, Canada*.
- TZANIDOU, E., PETRE, M., MINOCHA, S. AND GRAYSON, A. (2005). Combining eye tracking and conventional techniques for indications of user-adaptability. *INTERACT 2005 : human-computer interaction, Rome, 12-16 September 2005*. vol. 3585, 753-766.
- YONEKI, E. (2006). Sentient Future Competition: Ambient Intelligence by Collaborative Eye Tracking. *European Workshop on Wireless Sensor Networks (EWSN)*.

9

A Model for the Adaptation of Contact Centre Computer User Interfaces

Bronwin Jason and André Calitz

Maintaining effective customer contact is important because customer interaction is seen as a valuable asset for building lasting and profitable customer relationships. Contact centres, are the primary interaction point between a company and its customers. One of the important challenges in today's contact centre solutions is to increase the speed at which contact centre agents (CCAs) retrieve information to answer customer queries. CCAs however have different capabilities, experience and expertise. Studies have provided empirical support that user performance can be increased when the computer user interface (UI) characteristics match the user skill level. As a result, software systems have to become more individualised and cater for different users.

Users are normally classified as either experts or novices, and in some cases somewhere in-between. There is evidence to support the fact that novice and expert users behave differently when using a specific UI. Expertise and skill affects the way users interact with software. It is thus envisaged that an adaptive user interface (AUI) which dynamically changes from a novice UI to an expert UI could possibly improve users' performance. This research focuses on applying this concept of AUIs to the domain of contact centres and investigates whether an AUI for CCAs could improve the CCA's performance. A model was proposed for AUIs which was applied to the domain of contact centres. Implementation was done on the proposed model as proof of concept and the model was then evaluated. Results indicate a difference in interaction data for novice and expert CCAs and this shows that the proposed model can be used to provide an AUI for both novice and expert CCAs. Most users approved the use of AUIs and the research showed that AUIs can improve the speed at which CCAs provide call resolution.

1. Introduction

Contact centres are the primary interaction point between a company and its customers and are rapidly expanding in terms of both workforce and economic infrastructure [Kool and Mandelbaum 2001; Mandelbaum 2004]. The contact centre personnel managing the calls logged by customers are referred to as contact centre agents (CCAs) [Steel 2003]. CCAs differ in their ability to respond to customer queries and do not interact with the computer user interface (UI) in a similar manner. CCA's have different capabilities, experience and expertise.

A field study conducted by the authors at three South African contact centres in 2006 revealed that newly appointed CCA's often find the complexity of contact centre UIs overwhelming. CCA's current computer UIs are static in nature and are the same for each CCA. CCA's often have to adapt to the contact centre UIs and this adaptation

process decreases the CCA's performance. Studies have provided empirical support that user performance can be increased when the computer user interface (UI) characteristics match the user skill level [Jettmar 2000]. As a result, software systems have to become more individualised and cater for the different users.

The above emphasises the importance of adaptive user interfaces (AUIs). AUIs which has widely been recognised as a being a central component of Intelligent User Interfaces (IUIs), have been proposed as a promising attempt to overcome the above mentioned problems of human-computer interaction complexity [Hook 2000]. Instead of CCA's adapting to the UI of contact centres, AUIs involves the UI adapting to the CCA. It is thus envisaged that by providing a more personalised UI according to the CCA's expertise level, an improved customer experience could be achieved.

An AUI model has been developed for contact centre computer UIs (Jason and Calitz 2008). The aim of this research is to determine if existing AUI models can be combined with an IUI model for CCAs and if such interfaces would improve productivity and usability. Section 2 of this paper discusses the problem domain of this research study, namely Differences in Expertise, IUIs and AUIs. The proposed AUI model is introduced in Section 3. Classification of user expertise is discussed in Section 4. Evaluation of the model is presented in Section 5 and Section 6 focuses on current and future research related to AUIs.

2. Related Work

2.1 Contact Centres

Contact centres are defined as an operation that uses personnel and various multimedia customer-contact channels in sophisticated ways to deliver a variety of services to customers [Kool and Mandelbaum 2001; Mandelbaum 2004]. The various areas of operations for contact centres are a help desk or a service desk. A help desk can be defined as a single point of contact for customer problem resolution [Frants 2006]. Service desks are defined as a central point of contact between the customer and all IT related areas whereby customers can use multiple channels for requesting services [Frants 2006; Microsoft 2002]. The main difference thus between a help desk and service desk is that a service desk enables the usage of multiple channels to service customers.

Even though contact centres utilise state-of-the-art technology, all contact centres rely on people to act as the intermediaries between the information in the database and the servicing of the customers [Parbhoo 2002]. These people employed by contact centres are the CCAs whose responsibility it is to receive the calls placed by customers and either redirect those calls or personally assist the customer, as previously mentioned. These CCAs, who are trained and skilled in customer service, have different levels of expertise [Australian National Audit Office 1996]. The next section discusses the differences in expertise levels.

2.2 Differences in Expertise

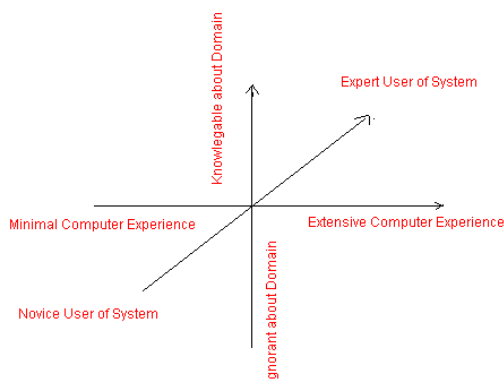
Prumper [1991] states that there are usually two overlapping criteria used to differentiate between novice and expert users. These two criteria are knowledge and the time spent

working with a particular system [Prumper 1991]. Nielsen [1993] supports Prumper [1991] and declares three main dimensions along which users' experience differs (Figure 1).

These three dimensions are [Wu 2000; Nielsen 1993]:

1. Experience with the system;
2. Experience with computers in general; and
3. Experience with the task domain.

Figure 1. User Cube [Nielsen 1993].



The most common dimension used when discussing user expertise is the user's experience with the specific UI. Users are normally classified as either experts or novices, or somewhere in-between. There is evidence to support the fact that novice and expert users behave differently [Hudson *et al* 2007].

Novice and expert users have different methods of selecting the correct menu item. An expert user typically knows the menu item they want to select and is able to memorise the location of menu items. Contrasting to expert users, novices do not tend to know what menu item they want to select or its location and tend to have to search for it. This means that expert users generally tend to make faster menu selections. This also suggests that features that differentiate searching behaviour from other types of motion may help to differentiate novice and expert use. When searching menus, features that approximate searching include the number of submenus open and how often the cursor "dwells" over a menu item [Hudson *et al* 2007].

Novice users are generally concerned with how to do things instead of how fast they can do it. Novice and expert users' behaviour usually differ dramatically at the physical level of interaction [Buxton *et al.* 1993]. An example of this difference is that novice users use the mouse to select from a menu whereas expert users press an accelerator key. Novice performance builds the skills that transition to expert performance when the basic action of the novice and expert are the same [Buxton *et al.* 1993].

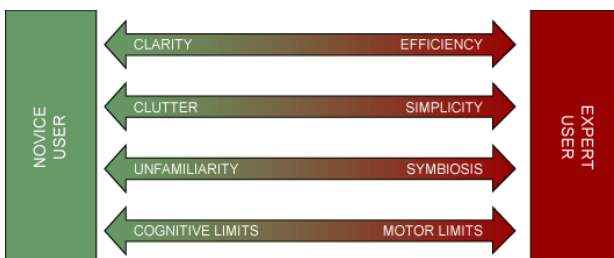
Expert users don't only know more but they know "differently". They have the ability to take large amounts of information and see it as connected units. Expert users are goal orientated. When using an interface, they quickly deduce goals and actions

to achieve those goals. They want a highly efficient interface and would thus like the number of interactions to be reduced [Wu 2000].

Padilla [2003] states that the efficiency of using an application is limited by the motor load for expert users and by the cognitive load for novice users. The UI for novice users needs to thus be rich in explanatory elements to assist with this cognitive understanding of the screen. This explanatory value comes at an expense for expert users (Figure 2) [Padilla 2003].

As previously mentioned, novice users become skilled with time. This statement suggests that novice CCAs can become expert CCAs over a certain amount of time. An AUI that shifts from a UI for novices to a UI for experts thus seems promising. Before we can look at AUIs, a brief look into IUIs needs to be done. The next section further discusses IUIs.

Figure 2. The Spectrum of user's needs [Padilla 2003].



2.1 Intelligent User Interfaces

In the early days of personal computers the UIs consisted of text based interfaces. Graphical User Interfaces (GUIs) have since replaced these text-based interfaces but despite the advancement, GUIs are still hard to learn for inexperienced users. IUIs are a way of alleviating this problem.

IUIs can be considered the next wave of interfaces. *IUIs are human-computer interfaces that aim to improve the efficiency, effectiveness, and naturalness of human-machine interaction by representing, reasoning, and acting on models of the user, domain, task, discourse, and media (e.g. graphics, natural language, gesture)* [Maybury and Wahlster 1998].

IUIs currently constitute a major direction of Human Computer Interaction (HCI) research, towards the provision of high-quality user-computer interaction. They are considered especially important when the aim is towards supporting heterogeneous user groups with variable and diverse needs, abilities and preferences since they facilitate a more 'natural' interaction, i.e. effective and efficient user-computer interaction [Maybury and Wahlster 1998].

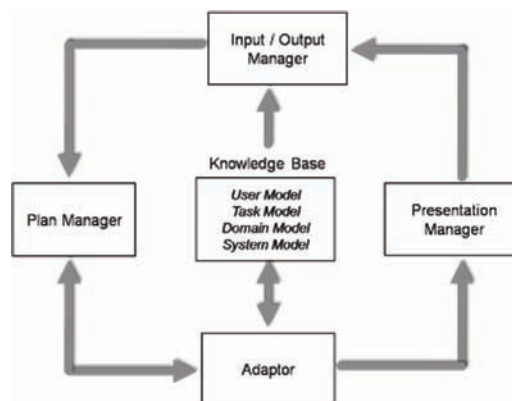
A suitable IUI architecture depicting the main functions of an IUI is provided in Figure 3. This specific architecture was proposed by Tyler *et al.* [Tyler *et al.* 1991]. The main components of this architecture are the Input/Output Manager, Plan Manager, Presentation Manager, Adaptor and Knowledge Base [Tyler *et al.* 1991].

The Input/Output provides the user with multimodal means of input. The Knowledge

Base is the key component of the IUI as it is a repository that is used to adapt the user interface (UI) to the user's needs. The Plan Manager assists the user with achieving high-level goals by using knowledge of the user's current goals and plans. The Plan Manager has the ability of detecting errors and correcting them, interpreting ambiguous requests and the ability to help users map high-level goals into low-level application commands. The Adaptor ensures that the UI is modified to best fulfill the user's needs and thereby assists the user with task completion. The Presentation Manager determines the most suitable modality and modality techniques to display to the user [Tyler *et al.* 1991].

According to the above mentioned IUI architecture, the Adaptor component and the Presentation Manager are responsible for adapting the UI to the user's needs. The adaptor component is responsible for maintaining the user model so that the Presentation Manager can best decide how to adapt the UI accordingly. It can therefore be concluded that adaptivity is recognised as being a central component of an IUI. The next section further investigates adaptive user interfaces (AUIs).

Figure 3. Tyler *et al.*'s Intelligent Interface Architecture (Extract) [Tyler *et al.* 1991].



2.2 Adaptive User Interfaces

A clear distinction needs to be made between an adaptable and an adaptive system. A system is called adaptable if the user is provided with tools to customise the UI. This is an attractive objective to provide the user with facilities for tailoring the system according to his personal tasks and needs. This kind of individualisation gives control over the adaptation to the user. A system is called adaptive if it is able to change its own characteristics automatically according to the user's needs [Jameson 2002].

An AUI can be defined as a software artifact that dynamically changes the appearance, function or content of the UI in response to its user interaction experience [Langley 1999]. It improves interaction with the user by constructing a user model based on partial experience with that user [Akoumianakis 2000; Langley 1999; Keeble and Macredie 2000]. Thus an adaptive interface does not exist in isolation and is designed to interact with human users. Furthermore, the interface is only adaptive if the interaction with the user is improved. Improvements should also rather result from generalisation over past experiences and carry over to new interactions [Langley 1999].

The above definition of AUI will seem familiar to some as it takes the same form as common definitions for machine learning. The main differences are that the user plays the role of the environment in which learning occurs, the user model takes the place of the learned knowledge base, and interaction with the user serves as the performance task on which learning should lead to improvement. In this view, AUIs constitute a special class of learning systems that are designed to aid humans, in contrast to early work on machine learning, which aimed to develop knowledge-based systems that would replace domain experts [Langley 1999]. A further investigation was done into various AUI models in order to select the most suitable one. Oppermann (1994) suggests that all AUIs should consist of the following criteria [Oppermann 1994]:

- 1) *Afferential component of Adaptivity*: This component handles the kind of data to be recorded. Data can both be implicitly or explicitly collected and then recorded. The Knowledge Base stores the recorded data.
- 2) *Inferential component of Adaptivity*: This component consists of the adaptive system analysing the gathered data to identify from the user behavior possible adaptation indicators, and hence inferences are made. The inferential component can be based on rules and heuristics that are represented by various models (user/system/task, etc.).
- 3) *Efferential Component of Adaptivity*: This component specifies how the system should be adapted based on data obtained in the Knowledge Base. The Knowledge Base and in particular the user model is thus the core component need for personalisation.

The above mentioned components were used as a basis to investigate the criteria of various AUI architectures/models. The criteria specified was found to be evident in all AUI models but in some cases not explicitly defined. An example of an incomplete AUI model will be discussed in the following section.

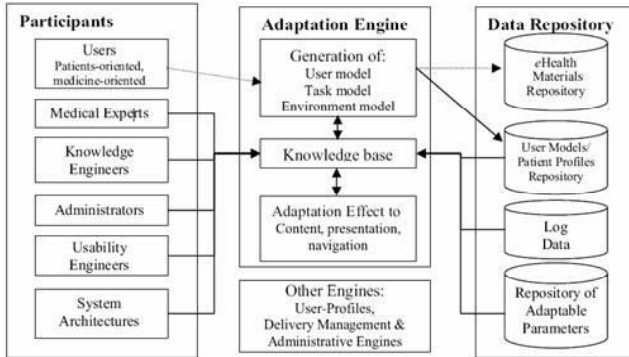
2.4.1 The E-Health AUI Model

Pechenizkiy *et al.* [2005] proposed a framework for an AUI for eHealth systems (Figure 4). The components of this model and how they relate to the criteria will now be discussed. As seen in Figure 4 the arrows emphasise information flows crucial to the AUI process. There are three major groups of framework components: Participants, Data Repositories and Different Engines.

The Participants are the various users of the system. The Data Repository is a repository of all information needed. The Adaptation Engine serves as the core component for adaptation and consists of the Knowledge Base, a Model (user, task, and environment) Generator and an Adaptation Effect provider. The inferential component of adaptivity can be made up from rules found in the User Model but it is not explicitly defined. The Adaptation Effect provides various kinds of adaptations such as adaptation to content, to presentation and to navigation. This component corresponds to the efferential component of adaptivity. The Model Generator generates the user /task / environment model by either implicitly or explicitly acquiring user information.

The acquiring of information evident in this component, therefore, corresponds to the afferential component of adaptivity but the afferential component of adaptivity is also not explicitly defined [Pechenizkiy *et al.* 2005].

Figure 4. General Framework of Adaptive eHealth System [Pechenizkiy *et al.* 2005].



The AUI for eHealth systems cannot therefore be viewed as a complete model and was designed specifically for eHealth systems and not CCs. It does however specify the basic architecture of an AUI model and so could be used to produce an AUI model which supports CCAs. An AUI model was proposed that satisfies all criteria of AUIs. The next section describes the design of the proposed AUI model and the extent to which it can be adapted to suit the domain of CCs.

3. Proposed Model

3.1 Model Design

The AUI model proposed in Figure 5 is a general architecture and can be applied to any system. This section discusses the various components of the proposed model (Figure 5) and the interaction between these components.

The components of the AUI model are:

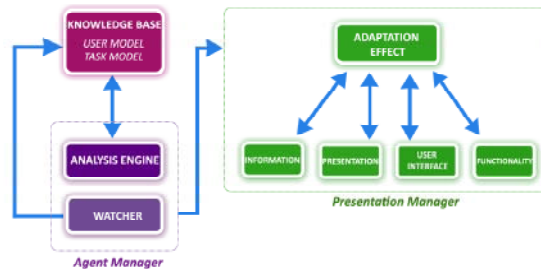
- 1) *Knowledge Base*: The Knowledge Base serves as the core component for AUIs acting as a repository by making use of various models. Among all the possible models pertained in the Knowledge Base, the field of user model research gains the most attention. The reason for this is that adaptivity requires the system to have a certain amount of knowledge about the user and user modeling fills this void. The Knowledge Base is not however limited to the models provided by Figure 5.
- 2) *Analysis Engine*: The Analysis Engine uses the user model and other models in the Knowledge Base to derive new user information. The analysis engine can update the user model based on new information learned about the user or it can initiate an event such as suggesting something to the user. It also responds to queries from the application. The Inferential component of Adaptivity is thus satisfied by the

Analysis Engine. The Analysis Engine also has the ability to make inferences.

- 3) *Watcher*: The Watcher collects information about the user either implicitly by observing the user's behaviour while he/she interacts with the system or explicitly and updates the Knowledge Base. The Afferential component of Adaptivity is thus satisfied by the Watcher.
- 4) *Adaptation Effect*: The Adaptation Effect provides various kinds of adaptations such as adaptation to information, presentation, user interface and navigation. The Efferential component of Adaptivity is thus satisfied by the Adaptation Effect. The Adaptation Effect decides how to adapt the UI to the user's behaviour based on data obtained from the Knowledge Base.

The proposed model can be specialised to the domain of contact centres by incorporating contact centre information in the Knowledge Base. The other components can then be configured to operate within the new domain.

Figure 5. Proposed AUI Model.



All components of adaptivity as proposed by Opperman (1994) are thus satisfied by the proposed AUI model. The proposed AUI model also fits into Tyler *et al.*'s Intelligent Interface Architecture (Figure 5) and this will now be discussed.

The Knowledge Base for both Tyler *et al.* and the proposed AUI model remains the same. The Adaptor component of Tyler *et al.*'s Intelligent Interface Architecture as discussed in Section 2.3, receives input from the Plan Manager, updates the user model and ensures that the UI is modified according to the user's needs. As seen in Figure 5, the Analysis Engine and the Watcher serve as the Adaptor component of Tyler *et al.*'s Intelligent Interface Architecture. These components collectively provide the same function if not more, as that provided by the Adaptor. The Presentation Manager component of Tyler *et al.*'s Intelligent Interface Architecture is discussed in Section 2.3 This component receives input from the Adaptor component and uses knowledge obtained to decide how the UI can be adapted to the user's needs. As seen in Figure 5, The Adaptation Effect serves as the Presentation Manager of Tyler *et al.*'s Intelligent Interface Architecture. The next section will discuss the implementation of the proposed model as a proof of concept.

3.2 Model Implementation

Nelson Mandela Metropolitan University (NMMU) ICT Service Desk was used as a case study for the purpose of this research. The main purpose of the NMMU ICT Service Desk is to provide solutions to computer related problems and to log problems for the users (staff and students) on all campuses. At present the NMMU ICT Service Desk uses FrontRange Solutions HEAT Product Suite [FrontRange.com]. The call logging steps for the NMMU's current Service Desk software are:

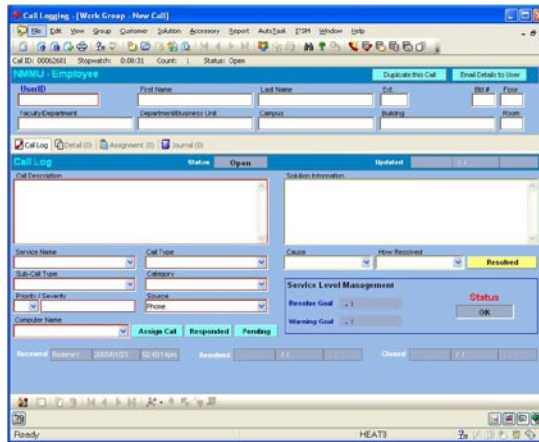
1. Search for the customer;
2. Provide Call Description details;
3. Assign the call to a technician/CCA; and
4. If the call was assigned to the CCA receiving call, provide the call's Solution Description.

Problems were found in the existing UIs of the NMMU ICT Service Desk. As seen in Figure 6, the Call Logging UI appears cluttered and requires the CCAs to recall what they have previously done rather than recognise. It is thus not an easy to use interface, especially for newly appointed CCAs. The implemented prototype should allow the user to successfully log a call placed by a customer. Implementation of the Knowledge Base was done using SQL Server 2005 combined with an extract of the existing HEAT database. The User and Task Models were implemented in XML and currently reside within the database.

The User Model consists of data for each Informative Moment and the potentially predictive features associated with that Informative Moment. Informative Moments and predictive features will be discussed in the next section.

The Watcher and Adaptation Effect component was implemented in C#. As the user performs a task, the watcher collects the interaction data and updates the User and Task Models pertained in the Knowledge Base. The Watcher updates the Task Model by specifying which steps of a task are complete and incomplete and what the current step is. Current task and step information is displayed to the user in a section allocated to the delivery of task based information (top section of Figure 7 and 8) similarly to that implemented by Singh and Wesson (2007) [Singh and Wesson 2007]. The Watcher updates the User Model by updating the values for the Informative Moments for each task (textbox entry or drop down list selection) the user performs. The Adaptation Effect implemented provides adaptation to the UI. UIs were designed and implemented for expert users and novice users, hereafter known as the expert UI and the novice UI

Figure 6. NMMU ICT Service Desk Call Logging UI (Screenshot)..

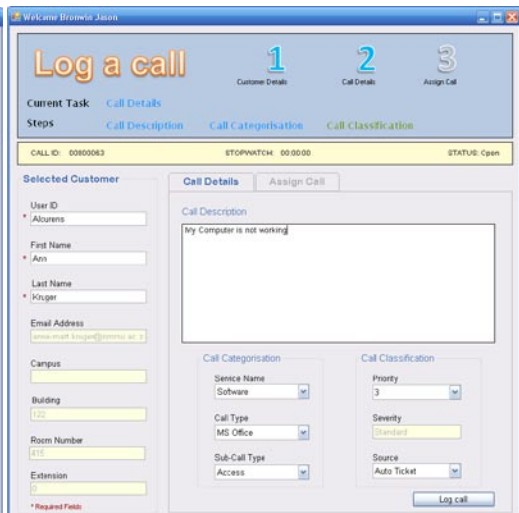
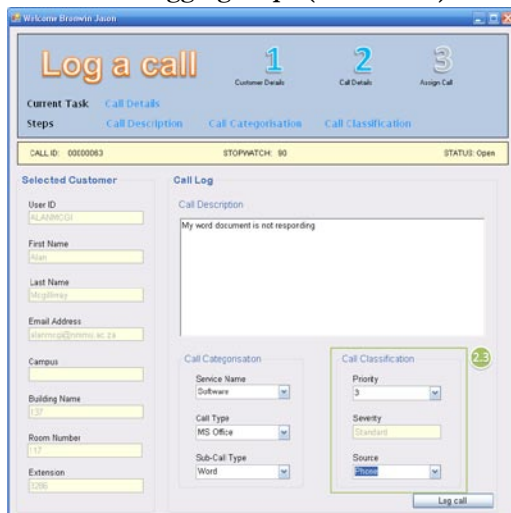


The novice UI (Figure 7) displayed a separate screen for each of the call logging steps guiding them step-by-step through the call logging process. It was previously mentioned that novice users are more concerned with how to do things instead of what to do and this suggests that novice users need to be guided through their tasks. A step indicator indicating the next step appeared as the user completed a task and the user is constricted in the order in which they complete a task. More warning messages and error messages are given to novice users as opposed to expert users and these messages are also more descriptive.

The expert UI (Figure 8) displayed only one screen for logging a call and they are not constricted to do tasks in a step-by-step manner, as the novice users were. They were given less guidance (no next step indicator) and more freedom when performing tasks.

Figure 7. Novice UI for Step 2 (provide call description details)of Call Logging Steps (Screenshot).

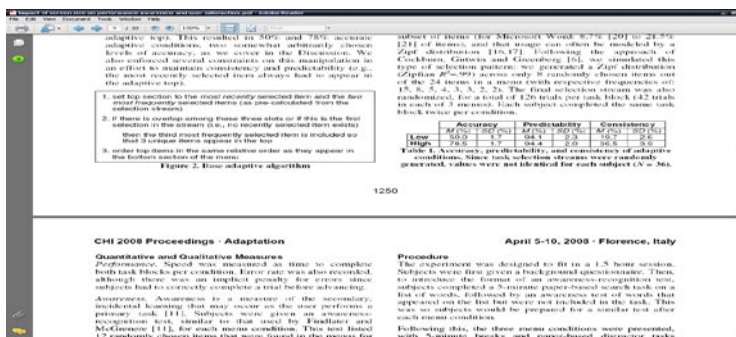
Figure 8. Expert UI (Screenshot).



3.1 Adaptive Lists

The selection lists for the expert UI was implemented to be adaptive. The research study applied the same technique to selection lists that previous research applied to adaptive menus [Hudson *et al.* 2007, Findlater and McGrenere 2008]. The research study used a Base adaptive algorithm (Figure 9) as used by (Findlater and McGrenere, 2008) which dynamically changes the selection list based on the user's previous list selections made. This algorithm incorporates both recently and frequently used items and is commonly used in Microsoft Office 2003's adaptive menus. The novice UIs contained static instead of adaptive lists, as to remain consistent with user expectations and to avoid confusion among novice users.

Figure 9. Base adaptive algorithm [Findlater and McGrenere 2008].



The Based adaptive algorithm (Figure 9) dynamically determines which three items should appear in the top portion of the selection lists. The research study moves the three unique items to the top section instead of copying them.

The research study does not utilise split menus to differentiate the top three unique items and thus provides differentiation by painting the top section a different font and selection colour than the rest of the items in the list. It is envisaged that based on the CCA's performance using the novice UI, he/she will dynamically be provided the expert UI. The system can thus predict which interface the CCA is most suited to use, based on certain potentially predictive features. These potentially predictive features will be further discussed in the next section.

4. Classification

4.1 Training Data

Hudson *et al.* [2007] define Informative Moments as user actions which can be readily isolated, are indicative of the phenomena they wish to study, model or predict, and can be easily and accurately labeled. They gathered data obtained from menu operations and used that as Informative Moments [Hudson *et al.* 2007]. Our Informative Moments consist of the various drop down list selections as well as text entered in a Textbox.

Textbox entries classified as Informative Moments are when the user enters text for the Call and Solution Description. Drop down list selections classified as Informative Moments are when the user selected a Customer, Service Name, Call Type, Priority, Source, Campus, Contact, Cause and Resolved option. For each Informative Moment, data for a number of possibly predictive features is captured. Hudson *et al.* [2007] developed potential features that could be predictive of a user's skill level. These features are not based on a task model but rather on low-level mouse and menu data which could be used in any application as they are not application specific [Hudson *et al.* 2007]. Our tasks constitute as mostly selecting items from a list of items, similarly to selecting from a menu and thus work done by Hudson *et al.* [2007] could prove to be useful.

The following potentially predictive features were selected for a drop down list selection from Hudson *et al.*'s. [2007] as most appropriate for this research project:

1. *Total Time (seconds)*: Total Elapsed time within the action (starting when the list opened and ending when it closed). This is a summative value of all the selection times for the drop down list selection.
2. *Y Mouse Velocity (pixels/second)*: Average velocity of the mouse during a list operation in the Y direction.
3. *Y Mouse Acceleration (change in velocity/second)*: Average unsigned acceleration of the mouse during a list operation in the Y direction.
4. *Dwell Time (seconds)*: Time spent dwelling (not moving) during the interaction sequence.
5. *Average dwell time (seconds/count)*: Time spent dwelling divided by the number of menu items visited.
5. *Nr of Items visited (count)*: Total number of list items that were visited or passed through during list action.
7. *KLM Diff (seconds)*: Difference between KLM predicted time and selection time for the action. The Keystroke Level Model (KLM) involves constructing a detailed, task specific model of expert behavior [Hudson et al. 2007]. To obtain the KLM predicted time, we used a modified version of the model created by Hudson et al. [2007]. The model we used is presented in Figure 10. This model is useful as it is presumed that expert users would perform at speeds closer to the predicted time than novices would.
8. *KLM Ratio (dimensionless)*: KLM predicted time divided by the actual time for the action.
9. *Selection Time (seconds)*: Elapsed time within the action (starting when the list opened and ending when it closed).

Figure 10. KLM for List Selection.

$$\begin{aligned}
 & (LM + T) + C \\
 & = (1.35 + 1.1) + 0.2 \\
 & = 2.45
 \end{aligned}$$

where:

LM is the *Look & Mental Operator*,
 T is the *Travel* to move to mouse to the targets,
 and
 C is the mouse *Click* when selecting a item

5. Evaluation

5.1 Participants

A convenience population of 15 postgraduate participants was selected from the Computer Science & Information Systems Department at NMMU as novice users. A background questionnaire indicated that the above mentioned participants had no previous experience with service desk software and in particular, they had no previous experience with the HEAT Product Suite. All participants possessed a degree and had moderate to high experience using computers. Irrespective of their extensive computer experience, because they had possessed no contact centre domain knowledge and had no previous experience with service desk software, they were classified as novice users (Section 2.2). Most participants were male (n=12), while only three participants were female and their ages ranged from 21-28 years. A conference paper [Jason and Calitz 2008] which was previously written by the authors only had novice user data available; however additional testing has since been done whereby the performance data of expert users were collected.

A population of five participants was selected as expert users. Four of these participants were CCAs at the NMMU ICT Service Desk and 1 participant was a technician. Four of the participants possessed a degree and 1 completed numerous IT courses. They all had moderate to high experience using computers. The CCAs (N=4) used the Call logging section of HEAT daily for 1-6 years. The technician has used certain sections of HEAT's Call Logging facility and uses other facilities of HEAT daily. These participants were thus considered expert users as they possessed an extensive computer experience, had a good amount of domain knowledge and had used HEAT Call Logging facility before (Section 2.2). Most participants were female (N=4) and their ages ranged from 24-48 years.

5.2 Tasks

Both expert and novice UIs required participants to log two calls of medium complexity, chosen to mimic real-world activities. The participants were not given instructions as to how to log these calls. For each Informative Moment encountered, data for each predictive feature was captured. Novice users were only given the novice UI and thus

only logged two calls. They were then asked to complete a post test questionnaire evaluating the novice UI.

Expert users were first given the novice UIs, with which they performed the same two tasks the novice users were given. They were then asked to complete a post test questionnaire evaluating the novice UIs. After completing the post test questionnaire, they were given the expert UI and two calls to log. They were then asked to complete a post test questionnaire evaluating the expert UI.

Participants performed the tasks using the Tobii T60 eye tracker as it was envisaged that eye gazing could be used as another variable which distinguishes novices from experts.

5.3 Quantitative Results

The data was statistically analysed using Statistica. Some interesting mean results were found for the Dwell Time, Total Time, Y Mouse Velocity and KLM Difference predictive features and these will now be discussed.

5.3.1 Total Time

The overall mean results for the Total Time predictive feature is shown graphically in Figure 12. There was a significant difference in speed between the novice and expert users. Results indicate that the Novice users were slower than the Expert users. In particular, expert users' selection times for the Service Name, Call Type, Campus, Contact and Cause Informative Moments appear to be considerably faster than novice users'. This is not surprising as novice users do not generally know which service name and call types a particular call should belong to. Novice users also tend to not know whom to assign calls to and hence selection times for the assignee's campus and users ID (Campus and Contact) are lower for novice users. The Cause Informative Moment is directly related to Service Name, and hence similarly to the selection time of Service Name, the selection time of Cause for novice users is much slower than that of expert users.

5.3.2 Dwell Time

The overall mean results for the Dwell Time predictive feature is shown graphically in Figure 12. There was also a significant difference in speed between the novice and expert users for dwell time. Results indicate that the Novice users spent more time not moving during the interaction sequence than the Expert users. Similarly to Total Time predictive feature results above, it is not surprising that most time was spent dwelling during the Service Name, Call Type, Campus, Contact and Cause Informative Moments.

Figure 11. Mean results for Total Time predictive feature.

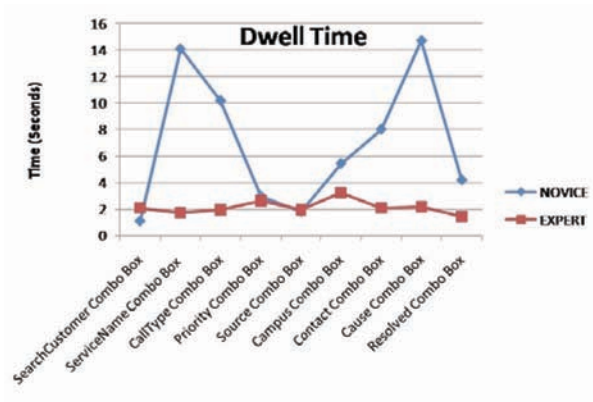
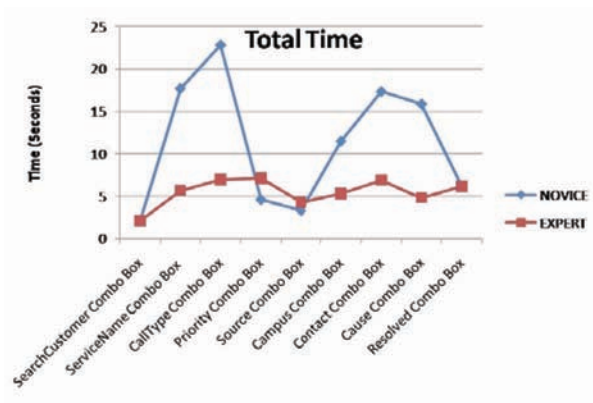


Figure 12. Mean results for Dwell Time predictive feature.



5.3.3 Y Mouse Velocity

The overall mean results for the Y Mouse Velocity predictive feature is shown graphically in Figure 13. Contrary to the Total and Dwell Time predictive features, results don't indicate any significant difference between the novice and expert users for Y Mouse Velocity. This could however be explained by the fact that all participants, labeled novice and expert, has extensive computer experience and thus mouse usage experience; as opposed to Total and Dwell Time which requires the user to have domain knowledge.

5.3.4 KLM Difference

The overall KLM difference mean results are graphically represented in Figure 14. With the exception of the Priority Informative Moment, the expert users have a lower KLM difference for the Informative Moments than the novice users. A low KLM difference is indicative of expert behaviour. These results are thus interesting as it further confirms our selection of users as novices and experts.

Figure 13. Mean results for Y Mouse Velocity predictive feature.

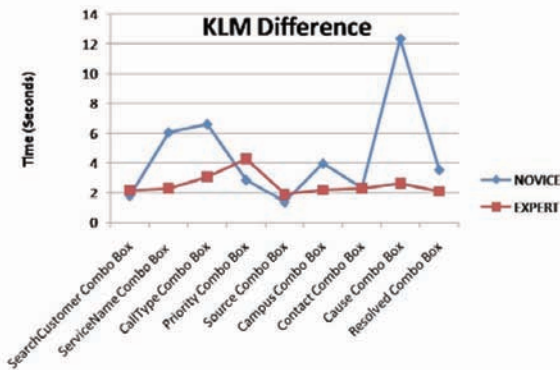
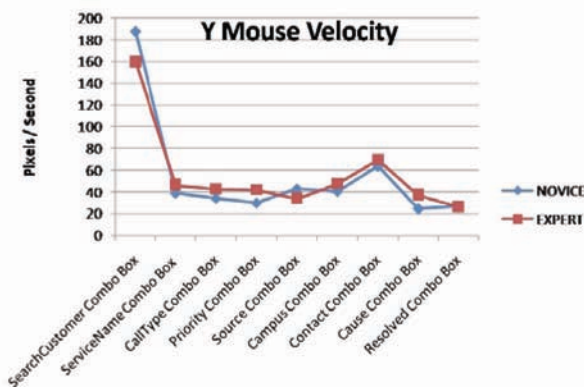


Figure 14. Mean results for KLM Difference predictive feature.



5.4 Qualitative Results

Questions and further comments were solicited after the evaluation by means of a user satisfaction questionnaire. The questionnaire was intended to determine the user’s general impressions of the system and its functionality. Participants were required to respond to questions (Table 1) by providing an appropriate rating on a 5-point Likert scale, where 1 represents strongly disagree and 5 represents strongly agree. Results were captured and statistical functions were applied in order to determine the mean and the standard deviation for each question. The results of these responses are summarised in Figure 15. Figure 15 presents the question number which corresponds to the question number presented in Table 1.

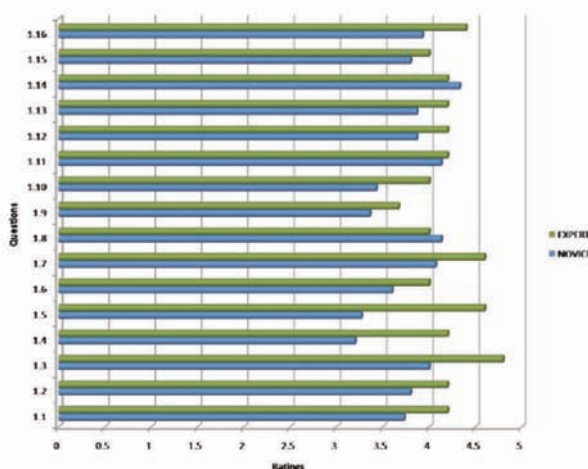
Both novice and expert users found the UI to be pleasant, but novice users could not complete the tasks quickly as opposed to expert users. This could be because of their unfamiliarity with call centre terminology and thus they did not know which selections to make from the various drop-down lists. Expert users found the UIs more simple than novice users and also found it somewhat easier to learn the system. The standard deviation for novice users is relatively large indicating there is not sufficient consensus among the participants whereas the standard deviation for expert users are relatively smaller as there was more of a consensus among the participants.

Expert users were additionally asked a few open ended questions regarding the Expert UI. All expert users found the Expert UI easy to use and in particular, found the adapted lists beneficial. The CCAs preferred the adaptive list to the static ones found in the novice UIs. All but one participant found no disadvantages of using adapted lists. The one disadvantage found was the use of colour for the most frequently used items. Some useful benefits mentioned of adapted lists for contact centres were quicker call logging, easier to log common problems and quicker to find information. Expert users also found the transition from a novice UI to an expert UI beneficial to CCAs as they felt that the transition would help novice user's get familiar with the interface.

Table 1. Questions asked in Post-test questionnaire

Nr	Question
1.1	Overall, I am Satisfied with how easy it is to use the system
1.2	It was simple to use this system
1.3	I can effectively complete the tasks using this system
1.4	I am able to complete the tasks quickly using the system
1.5	I can efficiently complete the tasks using the system
1.6	I feel comfortable using the system
1.7	It was easy to learn to use this system
1.8	I believe I became productive quickly using this system
1.9	The system gives error messages that clearly tell me how to fix problems
1.10	Whenever I make a mistake using the system, I recover easily and quickly
1.11	The information (such as on screen messages and help) provided with this system is clear
1.12	The information provided for the system is easy to understand
1.13	The information is effective in helping me complete the tasks and scenarios
1.14	The interface of this system is pleasant
1.15	I like using the interface of this system
1.16	Overall, I am Satisfied with this system

Figure 15. Quantitative Results from Post-Test Questionnaire



6. Conclusion And Future Work

An AUI model was proposed that is suited to the domain of contact centres. The prototype of the model has been implemented and evaluated. Novice users performed tasks using the novice UI whereas expert users performed tasks using both the novice UI and expert UI. Initial statistical results indicate that the selected predictive features could be used to differentiate novice from expert CCAs. These significant differences between novice and expert users could thus be used to successfully shift the UI from the novice UI to the expert UI. Current research is to investigate whether eye tracking could be used as an additional predictive feature which indicates CCA's expertise. Future work would involve getting more expert users to obtain possibly better results.

The envisaged benefits of this research would be increased performance of CCAs. This research helps novices become skilled/expert users and thus a secondary benefit includes training CCAs.

7. Acknowledgments

Acknowledgements are due to the Telkom Centre of Excellence Programme and the Department of Computer Science and Information Systems at the Nelson Mandela Metropolitan University for making this research possible.

8. References

- AKOUMIANAKIS, D., SAVIDIS, A. AND STEPHANIDIS, C. (2000). Encapsulating intelligent interactive behaviour in unified user interface artifacts. *Interacting with Computers*. Crete, Greece, Elsevier, pp. 383-408.
- AUSTRALIAN NATIONAL AUDIT OFFICE. (1996). Telephone Call Centres. [Online] Available: <http://www.anao.gov.au/WebSite.nsf/Publications/4A256AE90015F69B4A2569120006FCD1>.
- BUXTON W.A.S., KURTENBACH, G. P AND SELLEN A.J. (1993). An empirical evaluation of some articulatory and cognitive aspects of "marking menus". *Human Computer Interaction*. 8 (1), pp. 1-23.
- FINDLATER, L. AND MCGRENERE, J. 2008. Impact of Screen Size on Performance, Awareness, and User Satisfaction with Adaptive Graphical User Interfaces. CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. Florence, Italy, ACM.
- FRANTS, M.G. (2006). The Evolution of the IT Help Desk to the Service Desk. CompuCom. Available: <http://www.compucom.com/wwwimages/Corp/Images/CompuComServiceDeskWhitePaper0106.pdf>.
- FRONTRANGESOLUTIONS. [Online] Available: <http://www.frontrange.com/productssolutions/subcategory.aspx?id=35&ccid=6>.
- JAMESON, A. (2002). Adaptive Interfaces and Agents. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*. Lawrence Erlbaum Associates, Inc. pp. 305 – 330.

- JETTMAR, E. (2000). Adaptive Interfaces: Effects on User Performance. [Online]. Available: <http://www.dangerisland.com/true/papers/AdaptationJettmar.html>.
- HOOK, K. (2000). Steps to take before intelligent user interfaces become real. *Interacting with Computers*, 12, pp. 409-426.
- HUDSON, S.E., HURST, A. AND MANKOFF, J. (2007). Dynamic detection of novice vs. skilled use without a task model. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. KEEBLE, R.J. AND MACREDIE, R.D. 2000. Assistant agents for the world wide web intelligent interface design challenges. *Interacting with Computers*. Elsevier, pp. 357-381.
- JASON, B.A. AND CALITZ, A.P. (2008). An Adaptive User Interface Model for Contact Centres. *Proceedings 11th Annual SATNAC Conference*. Wild Coast Sun, South Africa. KOOLE, G. AND MANDELBAUM, A. 2001. Queuing Models of Call Centers An Introduction. [Online]. Available: <http://www.cs.vu.nl/obp/callcenters>.
- LANGLEY, P. 1999. User Modeling in Adaptive Interfaces. *Proceedings of the Seventh International Conference on User Modeling*. Springer-Verlag, pp. 357-370.
- MANDELBAUM, A.. (2004). *Call Centres: Research Bibliography with Abstracts*. Israel.
- MAYBURY, M. AND WAHLSTER, W. (1998). Intelligent User Interfaces: An Introduction. *Readings in Intelligent User Interfaces*. San Francisco, Kaufmann, M. pp. 1-13.
- MICROSOFT. (2002). Service Management Functions: Service Desk [Online]. Available: <http://www.microsoft.com/technet/solutionaccelerators/cits/mo/smf/smfsvcdk.msp>.
- NIELSEN, J. (1993). *Usability Engineering*. Academic Press.
- OPPERMANN, R. (1994). *Adaptive User Support: Ergonomic Design of Manually and Automatically Adaptable Software*, New Jersey, USA, Lawrence Erlbaum Associates Inc.
- PADILLA, M. (2003). Strike a balance: User's expertise on interface design. IBM white paper.
- PARBHOO, S. (2002). CCMA National Call Centre. [Online] Available: <http://www.ccma.org.za/DisplayNews2.asp?ID=8>.
- PECHENIZKIY, M., PUURONEN, S. AND VASILYEVA, E. (2005) Towards the Framework of Adaptive User Interfaces for eHealth. *CBMS '05: Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems*. Washington, DC, USA, IEEE Computer Society, pp. 139-144.
- PRUMPER, J., FRESE, M., ZAPF, D. AND BRODBECK, F. C. (1991). Errors in computerized office work: differences between novice and expert users. *SIGCHI Bull.*, 23, pp. 63-66.
- SINGH, A. AND WESSON, J.L. (2007). Designing an Intelligent User Interfaces for Contact Centres. *Proceedings 10th Annual SATNAC Conference*. Sugar Beach Resort, Mauritius.
- STEEL, A. Understanding and Enhancing Call Centre Computer- Human-Human Interaction. (2003). *CHI '03 extended abstracts on Human factors in computing systems*.
- TYLER, S.W., SCHLOSSBERG, A.R., GARGAN, R., COOK, L.K AND SULLIVAN. J.W. (1991). An intelligent interface architecture for adaptive interaction. In *Intelligent User Interfaces*, SULLIVAN, J.W. and TYLER, S.W., Eds. ACM Press, New York, NY, pp. 95-109.
- WU, J. 2000. Accommodating both Experts and Novices in One Interface. [Online]. Available: <http://www.otal.umd.edu/UUGuide/jingwu/>.

10

A Flexible Biomedical Ontology Selection Tool

Gilbert Maiga

The wide adoption and reuse of existing biomedical ontologies available in various libraries is limited by the lack of suitable tools with metrics for their evaluation by both naïve users and expert ontologists. Existing evaluation tools perform technical evaluation of the structure of an ontology as presented by its design and knowledge representation. These find little use in evaluating the processes and representation of granularity presented by biomedical ontology. In this paper we present an evaluation tool as part of a flexible framework that enables users to select a suitable biomedical ontology for use in building applications that integrate clinical and biological data. Requirements for such a tool were elicited in a descriptive survey using questionnaires, and a prototype developed. The tool also enables ontology modelers to iteratively elicit new requirements for improving upon existing biomedical ontologies, leading to new ones that are able to integrate data across structure, processes and granularity. By facilitating biomedical ontology evaluation, the tool contributes towards reuse of existing biomedical ontologies. This helps to avoid the need for costly time consuming tasks of developing entirely new biomedical ontologies. The utility of this tool was demonstrated in experiments that evaluated the infectious disease ontology. The results were validated using a questionnaire based human assessment.

1. Introduction

The wide adoption and reuse of existing biomedical ontologies available in various libraries is limited by the lack of suitable tools and lack of knowledge about properties users require to select a suitable ontology for a task [Alani and Brewster, 2006; Kalfoglou and Schorlmer, 2006]. Existing evaluation tools are known to differ by scope, evaluation type, purpose, inputs, process, metrics used, outputs and the level of user involvement. These perform technical evaluations of the structure of an ontology as presented by its design and knowledge representation [Alani& Brewster, 2006; Fernandez *et al.*, 2006; Hartmann *et al.*, 2004; Lozano-Tello & Gomez-Perez, 2004; Tartir *et al.*, 2005]. These tools therefore find little use in evaluating the processes and representation of granularity presented by biomedical ontology.

In this paper we present an evaluation tool as part of a flexible framework that enables users to select a suitable biomedical ontology for use in building applications that integrate clinical and biological data. Requirements for such a tool were elicited in a descriptive survey using questionnaires, and a prototype developed. The tool also enables ontology modelers to iteratively elicit new requirements for improving upon existing biomedical ontologies, leading to new ones that are able to integrate data across structure, processes and granularity.

The utility of the tool was demonstrated in experiments that evaluated the infectious disease ontology. The results were validated using a questionnaire based human assessment. The tool and human assessment results had high positive correlation (>0.5) when matched by scope, granularity and biomedical integration metric scores for the different use case scenarios. This is an indication of the success of the tool used, and the validity of our approach. By facilitating biomedical ontology evaluation, the tool contributes towards reuse of existing biomedical ontologies. This helps to avoid the need for costly time consuming task of developing entirely new biomedical ontologies.

The rest of this paper is organized as follows. Section 2 discusses current approaches to ontology evaluation. Section 3 explores existing ontology evaluation tools and their metrics, while section 4 explains how the tool was derived. Section 5 explains the prototype tool and its utility in evaluating the infectious disease ontology. Conclusions are given in section 6.

2. Current Approaches to Ontology Evaluation

There is no unifying definition of ontology evaluation [Gangemi *et al.*, 2005]. Evaluation determines the quality and adequacy of an ontology for use in a specific context for a specific goal [Fernandez *et al.*, 2006]. It is a technical judgment of the contents of an ontology with respect to requirements specifications, competency questions, or a meta ontology as a frame of reference [Gangemi *et al.*, 2005; Gomez-Perez, 2004; Guarino and Welty, 2002]. Ontology evaluation is important in order for them to become widely adopted and reused by industry and the wider web community [Alani and Brewster, 2006; Kalfoglou and Schorlmer, 2006]. There is also lack of knowledge about properties users require to select a suitable ontology for a task [Alani and Brewster, 2006]. Existing approaches to ontology evaluation use various contexts and conduct evaluation at different levels of complexity. A taxonomy of evaluation approaches based on type and purpose that adopts levels [aspects] of vocabulary, taxonomy, semantic relations, application, syntax, structure and design is provided by Brank *et al.* [2005] and given in table 1.

- 1) Gold standard approaches that compare an ontology to a manually built golden standard ontology or other representation of the problem domain for which an appropriate ontology is needed [Gomez_Perez, 1994uarino, 1998]. Lexical or conceptual levels of evaluation may be used. Lexical comparison uses similarity between the lexicons of two ontologies. At the conceptual level, taxonomic structure and relations are used for comparison [Sabou, *et al.*, 2006].
- 2) Task based approaches that use the ontology in an application and evaluate the quality of results [Porzel and Malaka 2004]. This approach has problems in that: (i) it is difficult to assess the quality of the supported task; (ii) it is difficult to create a neutral experimental environment where there are no other factors influencing the application performance. [Sabou, *et al.*, 2006]
- 3) Data or corpus driven approaches. These evaluate the congruence of an ontology with a given corpus to determine how appropriate it is for the representation of knowledge of the domain represented by the texts [Brewster *et al.* 2004].

- 4) Assessment by humans to show how well the ontology meets a set of predefined criteria, standards, requirements as in OntoMetric [Lozano-Tello and Gomez-Perez 2004] and the peer-review based approach [Supekar 2005].

Table 1. A level based Taxonomy of Ontology Evaluation approaches [Brank *et al.*, 2005]

Evaluation level	Evaluation approach			
	Golden standard	Application based	Data or corpus driven	Human assessment
Lexical, vocabulary	X = applied	X	X	X
Hierarchy	X	X	X	X
Semantic relations	X	X	X	X
Content application	X	X	Not applied	X
Syntactic	X	Not applied	X	X
Structure, architecture, design	Not applied	Not applied	Not applied	X

Table 1 points to existing approaches being mainly for conducting technical evaluation of the structure of an ontology domain as represented by its taxonomy, using different levels. It indicates lack of a unifying approach for evaluating ontologies, creating an obstacle for their reuse [Alani and Brewster, 2006]. There is also lack of knowledge about properties users require to select a suitable ontology for a task [ibid]. The approaches in table 1 evaluate for the ontology structure (taxonomy) and not for biomedical processes and representation of granularity. With these approaches, evaluating biomedical ontologies remains challenging as they seek to integrate clinical and biological data across structure, processes and granularity so as to achieve interoperability between sources [Kumar *et al.*, 2006; Yugyung *et al.*, 2006].

3. Ontology Evaluation Tools and Metrics

The need for tools to evaluate ontologies before their reuse is justified by the expense, time and effort required to build new ones [Fernandez *et al.*, 2006]. Existing evaluation tools include AKTiveRank [Alani& Brewster, 2006], OntoQA [Tartir *et al.*, 2005], OntoMetric [Lozano-Tello & Gomez-Perez, 2004], ODEVal and OntoManager [Hartmann *et al.*, 2004]. They differ by scope, evaluation type, purpose, inputs, process, metrics used, outputs and the level of user involvement (table 2) as explained here.

3.1 OntoQA

The OntoQA tool measures the quality of an ontology using schema and instance metrics. The schema metrics of OntoQA address the design of the ontology schema while instance metrics measure the size and distribution of instance data [Tartir *et al.*, 2005]. Schema metrics are; relationships diversity (RD) and schema deepness (SD). RD is the ratio of the number of non-inheritance relationships, divided by the total number

of relationships defined in the schema. It indicates the diversity of relationships in an ontology. SD is the average number of subclasses per class. It describes the distribution of classes across different levels of the ontology inheritance tree.

The instance metrics are of two types, class and relationship metrics. Class metrics are: class utilization (CU), cohesion (Coh), class instance distribution (CID), class connectivity (CC) and class importance (CI). CU is the ratio of the number of populated classes divided by the total number of classes defined in the ontology schema. It is used to indicate which of two populated ontologies has a richer KB. Coh of an ontology is the number of connected components of the graph representing the KB. It represents the number of connected components in the KB. CID is the standard deviation in the number of instances per class. It provides an indication on how instances are spread across the classes of the schema. CC indicates the centrality of a class. Formally CC is the total number of relationships instances the class as with instances of other classes. Class Importance is the number of instances that belong to the inheritance sub tree compared to the total number of class instances in the KB. It indicates where developers should focus on getting data if the intention is to get a consistent coverage of all classes in the schema.

OntoQA's relationship instance metrics are relational richness (RR) and importance (RI). RR indicates how the relationships defined for each class in the schema are being used at the instances level. It is number of relationships that are being used by instances that belong to an ontology compared to the number of relationships that are defined at the schema level. Relationship Importance measures the percentage of instances of a relationship with respect to the total number of relationship instances in the KB. This metric is important in that it helps in identifying which schema relationships were in focus when the instances were extracted and inform the user of the suitability for their intended use.

OntoQA enables knowledge engineers and researchers to find and analyze useful ontologies on the semantic web. It takes as input a crawled populated ontology or a set of user supplied search terms and ranks them according to some schema or instance level metrics related to various aspects of an ontology. A user is able to tune ontology ranking features to suit their needs. The tool has been validated by comparing its result to other approaches and to expert users [Tartir *et al.*, 2005].

3.2 AKTiveRank

AKTiveRank [Alani& Brewster,2006] is another tool proposed for ranking ontologies using metrics of the Class Match Measure (CMM), the Density measure (DM), Semantic similarity (SS) and Betweenness (BM) measures. The CMM assesses the coverage of an ontology for a given search term. The DM estimates information-content and level of knowledge detail of classes. The SS calculates how close the classes that match the search terms are in an ontology. The BM determines classes that are central to an ontology. AKTiveRank uses as input search terms provided by a knowledge engineer and ontologies from a search engine (e.g. Swoogle). It ranks the ontologies using a number of classical metrics and compares the results with a questionnaire based human

study. The results show that AKTiveRank has great utility although there is potential for improvement [ibid]. Its shortcoming is that it is that search terms can only be matched with ontology classes, and not with properties or comments [Alani & Brewster, 2006].

3.3 OntoMetric

OntoMetric [Lozano-Tello & Gomez-Perez, 2004] is a multi criteria decision making method that helps knowledge engineers to determine the suitability of a particular ontology for a project. A user selects an ontology using dimensions that specify the: 1) ontology content; 2) implementation language; 3) development methodology; 4) software used to build ontology; 5) costs of using the ontology in the system [Hartmann *et al.*, 2004]. OntoMetric is based on the analytic hierarchy process [Saaty, 1977], a selection process with four steps: 1) decide on selection criteria; 2) rate the relative importance of these criteria using pair-wise comparisons; 3) rate each potential choice relative to each other choice on the basis of each selection criterion, via pair wise comparisons of the choices; 4) combine ratings in steps 2 and 3 to get an overall rating for each potential choice. OntoMetric therefore offers the flexibility to select the hierarchy for the decision criteria to be used in evaluations. It however offers no specific support for evaluating integrated systems operating in dynamic environments like biomedicine.

3.4 ODEval

ODEval [Hartmann *et al.*, 2004] is used by ontology designers to evaluate knowledge representation for ontologies implemented in Semantic Web languages before they are used in applications. It detects possible knowledge representation inconsistency, incompleteness and redundancy in concepts for each considered language. ODEval has a high level of user involvement. It uses unpublished ontologies as input and generates ontology descriptions as output. No metrics are specified for ranking ontologies.

3.5 OntoManager

OntoManager is used by administrators, domain experts and business analysts to determine the truthfulness of an ontology with respect to its problem domain. It supports semi-automatic ontology improvement in response to the users' needs analysis. Its architecture has four functions of Monitor, Analyze, Plan and Execute. These monitor user interactions, analyze collected data, plan actions required for the changes discovered, and execute the changes to update the underlying ontology based application. It supports optimization of an ontology according to the users' needs. It is easy for end users to apply and use. However, evaluation quality is low. It is best applied in domains where information on ontology usage is available to identify relevant concepts of an ontology. The limitation on usage information does not allow to evaluate an ontology in general [Hartmann *et al.*, 2004] (Table 2)

Table 2. Comparison of some Ontology Evaluation tools

Criteria	OntoManager	OntoQA	AKTiveRank	OntoMetric
Users	Domain experts, admins., analysts	Knowledge engineers	Knowledge engineers [KE]	Knowledge engineers
Purpose	Fit ontology to requirements	Ontology ranking	Ontology ranking	Fit ontology to domain task
Evaluation context	Structure & user evaluation	Structure & knowledge base	Structure evaluation	Structure, cost, methodology, language, software
Input	Ontology in use	Search terms	Search terms	Ontology in use
Metric or criteria	User needs	Schema and instance metrics	Class density, semantic similarity, centrality measure	Multi-level tree [MTC] of 160 characteristics
Output	Ontology fitting requirements	Ranked ontology	Ranked ontology	Ontology fit to project
Validation type used	Not defined	Expert users	Questionnaire assessment	Not defined

Table 2 reveals that none of the current evaluation tools can be used across scope, in all evaluation contexts, inputs, metrics and outputs. However, as evident from the table, while the tools have been used to conduct structural and user evaluations, they do not define metrics useful for the dynamic processes and granularity presented by biomedical ontology. This lack of concern about dealing with heterogeneity including granularity has also been pointed out by Sabou, *et al.* [2006]. Furthermore, emphasis has been placed on defining structural metric on the basis of concepts while ignoring relations between concepts, a major limitation since relations provide valuable information in the search for the right ontologies, at the correct level of granularity [ibid].

4. An Evaluation tool for Biomedical Ontologies

In this paper we present a tool for evaluation of biomedical ontologies. Tool development is guided by:

- i. Existing literature on ontology evaluation frameworks [Maiga and Williams, 2008].
- ii. General requirements for ontology evaluation tools as given in the literature, revealed (table 2).
- iii. Specific requirements for biomedical ontology evaluation tools derived by a descriptive survey.

4.1. Specific Requirements for a biomedical Ontology Evaluation Tool

Literature review and document analysis were used to describe the theoretical requirements for a tool to help users assess and select a biomedical ontology suitable to their task. A questionnaire based descriptive survey was then used to validate the theoretical requirements with the proposed beneficiaries (biologists and medical doctors) of the tool. This objective guided the framing of the following key data collection questions on biomedical data integration model and evaluation tools namely: What is the scope (users and use cases) of a biomedical ontology evaluation tool? What general properties should the tool have? How should biomedical ontologies be compared using the tool? What are its inputs processes and outputs? How should such a tool be built?

The survey tested for agreement by respondents to characteristics of the proposed evaluation tool. Structured interviews and the questionnaire were pretested on twenty health care workers. Questionnaires were used to collect data and clarify requirements for the meta-model and tool. The survey then used questionnaires in which 630 randomly selected biomedical workers (580 medical doctors and 50 biologists) in Uganda were used as the study population. Filled questionnaires were returned by 404 medical doctors and 46 biologists. In the survey, these potential users of a biomedical integration system were asked for their level of agreement with proposed characteristics of the tool. The statistical package for social sciences (SPSS) was used to determine the level (%) of user agreement with the proposed characteristics of the tool. The results of the field study are presented in tables 3.

4.2. The Results

Table 3 provide user perspectives on the requirements of a biomedical ontology evaluation tool.

Table 3: Requirements for the Evaluation Tool

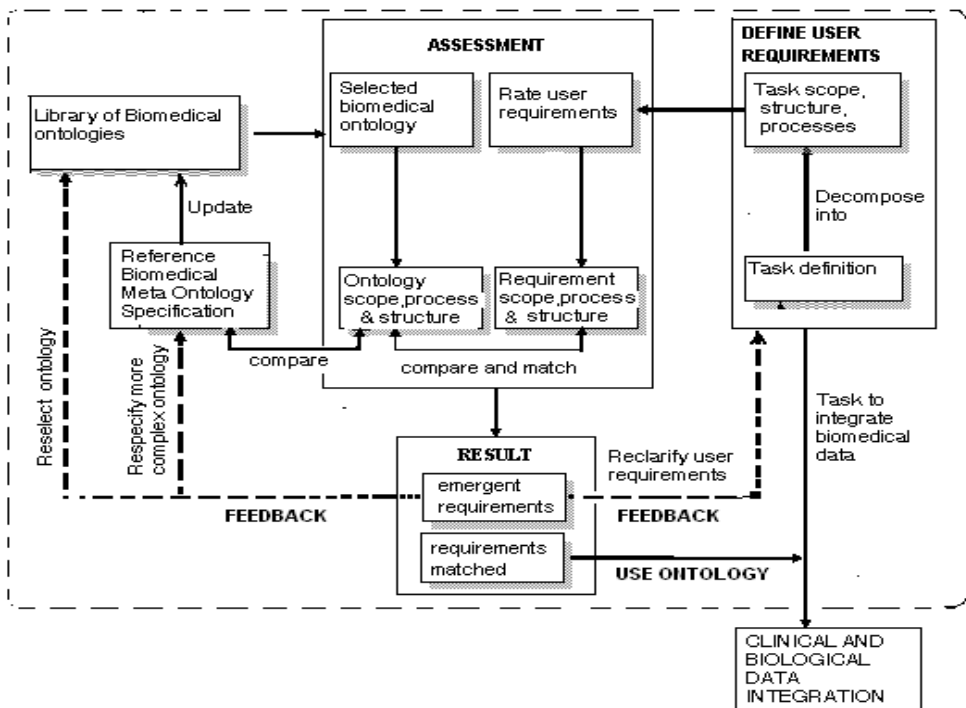
Tool requirement	Molecular Biologists		Medical doctors	
	No. of responses	Agreement Level [%]	No. of responses	Agreement Level [%]
Ontology Visualization	46	91	404	70
Determine scope	46	85	404	73
Generate task requirements	46	90	404	89
Match requirements to ontology	46	68	404	72
Provide user feedback	46	90	404	91
Ease of use	46	94	404	95

The table reveals a high level of user preference for the following features and processes in a tool to evaluate a biomedical ontology: visualize an ontology, determine its scope, generate requirements for a task and match them to an integration model, provide user feedback for re-specifying requirements and be easy to use. The requirements inform the principles adopted in the tool design.

4.3. A flexible User centered Evaluation Framework

The utility of using requirements in table 3 was demonstrated by a prototype tool developed as part of a flexible user centered framework for evaluating biomedical ontologies [Maiga and Williams, 2008]. The framework (figure. 1) provides different users with the flexibility to iteratively search through an ontology library using multiple criteria. This is more likely to result into the selection of a suitable ontology for a given task, or re-specification of new requirements for an ontology to fit the task. In the framework user needs motivate the assessment criteria in a formative evaluation, providing systematic feedback to designers and implementers in order to influence the process of developing new biomedical ontologies. Feedbacks in the framework provide a way to re-specify user requirements when extending, pruning and improve existing ontologies. This helps to avoid the huge effort of starting or building entirely new ontologies [Alani and Brewster 2006].

Figure 1. User Centered Framework to Biomedical Ontology Evaluation [Maiga & Williams, 2008].



4.4. Overview of the Evaluation Tool

The prototype tool described in this paper assumes a mixed context for assessing a biomedical ontology, applied in a flexible evaluation framework described in fig.2. It enables a user to select a biomedical ontology that supports development of a suitable application for their integration task. Scope, granular density and biomedical ontology structural integration are used as metrics for selecting an ontology.

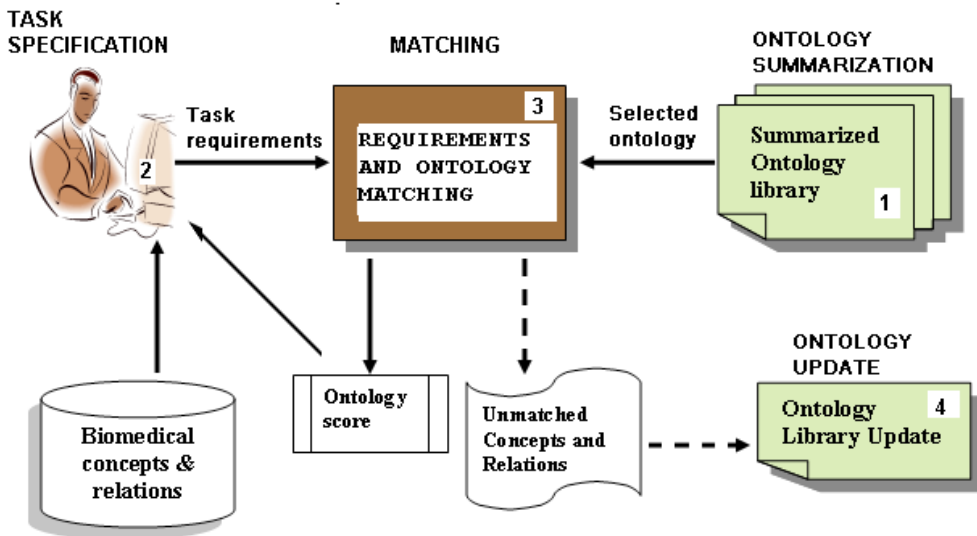
4.5. Tool Architecture and Design

The design of the evaluation tool is guided by:

- i. Requirements for a biomedical evaluation tool [table 3]
- ii. Literature on the design of other tools for evaluating ontologies.

Existing ontology evaluation tools differ in their design [Alani& Brewster, 2006; Tartir *et al.*, 2005; Fernandez *et al.*, 2006]. They vary in the type and format of data inputs, processing and outputs generated as presented [table 1]. These differences mean that users have to learn to work with new tools, data formats and outputs. The prototype tool design follows a process [activity] centered architecture guided by both requirements [table 3] and current knowledge on the design of ontology evaluation tools. The design therefore represents key requirements for such a tool as indicated in tables 1 and 3.

Fig. 2. Architecture of the Evaluation tool.



Tool design incrementally builds on existing tools in order to enable effective evaluation, promote user familiarity and system up-take by encouraging compatibility with previous data formats [Boardman, 2004]. An incremental approach that builds on existing tools design is advocated and used here because:

- 1) It builds on existing knowledge and technology, promotes user familiarity and compatibility with previous data formats, leading to a tool that is easy to use, a requirement for the evaluation tools.

- 2) It helps to achieve design goals that are implementable within the limited time and man-power resources in which research is conducted [ibid].

An incremental approach therefore enabled this study to build on previous design expertise manifested in already available tools. The tool is designed around the following core processes or activities of: 1) ontology summarization; 2) task determination; 3) matching and update, shown in the design.

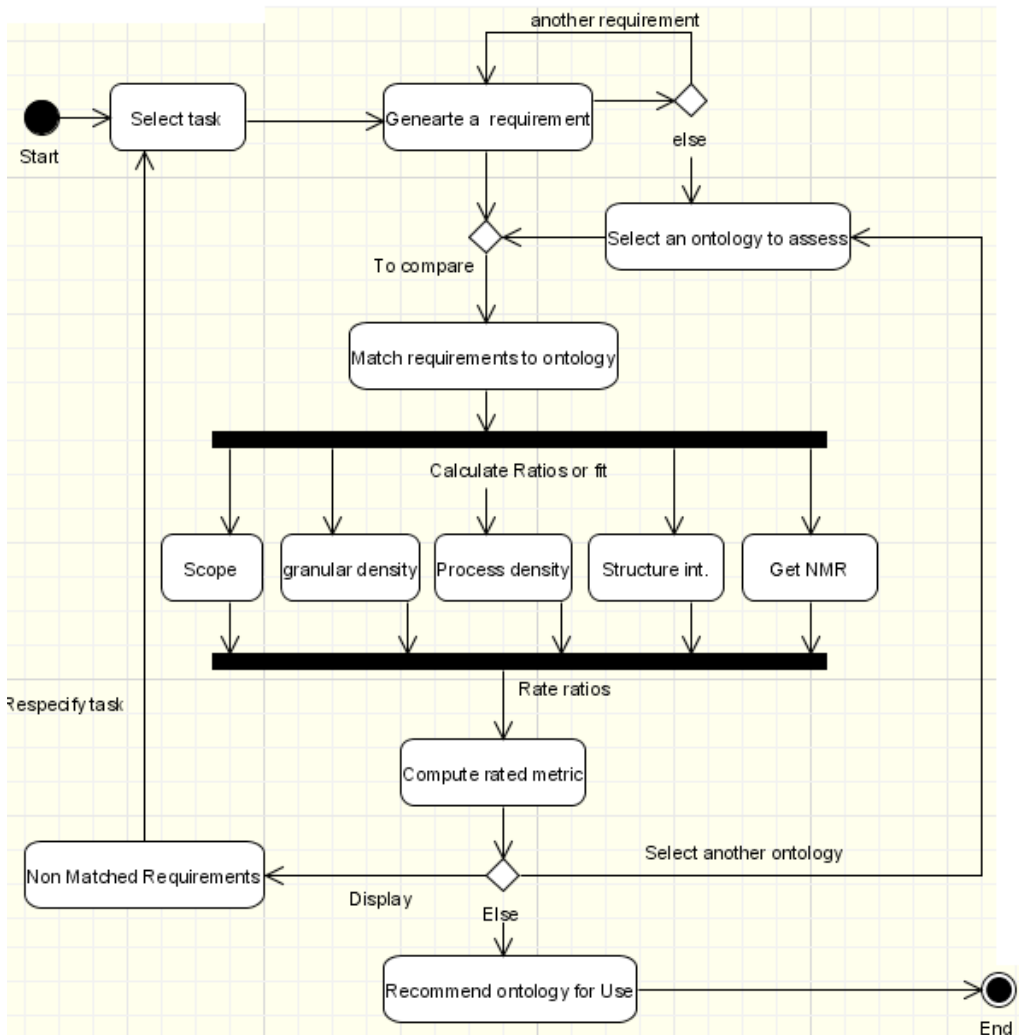
Ontology summarization abstracts the top level classes and relations in an ontology. Concepts and relations representing the thematic categories of an ontology are extracted to enable users understand and select an ontology [Noy, 2004; Zhang *et al.*, 2007]. Summary concepts are used to build an ontology database to facilitate quick comprehension prior to its assessment and selection by non expert users.

The task specification enables a user to select or specify a biomedical data integration activity from a database. A task is a statement of the purpose for integrating biomedical data e.g. “a need to relate gene profile data to a disorder in the population.” A task is decomposed into a set of requirements or search terms using objects and relations. The tool computes metrics for rating a biomedical ontology using a matching algorithm. User requirements are compared to an ontology selected from a library. The tool calculates and returns metrics and a description of any unmatched terms

4.6. The Tool's Matching Algorithm

The tool computes metrics for rating a biomedical ontology using a matching algorithm. User requirements are compared to an ontology selected from a library. The comparison process calculates and returns a metric and a description of any unmatched queries. On the basis of the returned result, a user can make a decision on whether to use the evaluated ontology, redefine the task or select another ontology. The UML activity diagram in Figure 3 is used to present the logic of an algorithm.

Figure 3. A UML Activity Diagram of the tool's Matching Algorithm



4.7. The Evaluation Metrics

In this paper, we describe three types of measures used to rate a biomedical ontology. These are scope, granular density and biomedical ontology integration.

4.7.1. Granular Density

Granularity, articulated as collectives linked by relations assumes a perspective that aggregates individuals into collections [Rector and Rodgers, 2005]. A related perspective articulates granularity based on categories of non-scale dependency (NSD), relationships between levels, set theory and mereology [Keet, 2006]. NSD granularity provides levels ordered through primitive relations like structural *part_of*, the spatial *contained_in* and the subsumption *is_a*. Logically, set theory also supports the *is_a* relation while mereology supports the *part_of* relation. Assuming multiple perspectives of granularity [as in this

paper] that combine NSD granularity, relations between perspectives, set theory and mereology makes it possible to seamlessly shift from one granular perspective to another [ibid].

In a biomedical ontology, granularity is articulated using relations between classes like *part_of* or *contained_in*. The *part_of* relation is both an intra and trans ontological as it can be used to model relationships within and between biological and clinical classes respectively. The ratio of relations [relational richness] has been used as a structure based metric for comparing ontologies [Tatir *et al.*, 2005]. The perspective of granularity as collectives linked by relations is adopted here and used to define a metric for assessing the ability of the biomedical meta-ontology to represent granularity. Intra and trans-ontological relations in the meta ontology are used to define the metric. Trans ontological relations like *part_of*, *has_part* and *contained-in* model collectives. The density of trans ontological relations [granular density] in a biomedical ontology is an indicator of its ability to model collectives. A metric for measuring the level to which an ontology models granularity can therefore be expressed by the ratio of trans ontological relations that model collectives to all trans ontological relations. This ratio or the granular density [G_D] is given by:

Let: R_{io} represent an intra ontological relation in the ontology; R_{ic} represent an intra ontological relation modeling collectives; R_{to} represent a trans ontological relation in the ontology, and R_{co} represent a trans ontological relation modeling collectives, then the Granular density G_D is given by the equation

$$G_D = \frac{[\sum Ric + \sum Rco]}{[\sum Rio + \sum Rto]} \frac{[\sum Ric + \sum Rco]}{[\sum Rio + \sum Rto]} \dots\dots\dots$$

..... (Equation 1)

The metric provides a way to select between various ontologies for the task of integrating biomedical data across levels of granularity that are common in biomedicine. Good representation of granularity in a model enables tracking of entities and attributes across levels, leading to building of better data integration systems [Kumar *et al.*, 2006].

4.7.2. Biomedical Ontology Structure Integration

Clinical and biological classes and their relations are part of biomedical ontology structure. Between these classes, trans domain relations [e.g. *participates_in*] are identified. The term semantic density has been used to describe the level of connectedness of an ontology using relations [D'Aquin *et al.*, 2007]. The more trans ontological relations that are defined, the greater is the degree of connectedness in the ontology. The density of trans ontological relations between entities in an ontology is here used as an indicator of the level of integration [connectedness] and overlap between clinical and biological ontologies. A measure of the level of integration between clinical and biological data in an ontology can therefore be expressed as the proportion of trans ontological relations to all relations as defined in equation 2.

Let : R_{io} represent an intra ontological relation in the ontology; R_{to} represent a trans ontological relation in the ontology, then biomedical ontology structure integration ($[BOS_{int}]$) is given by

$$BOS_{int} = \frac{\sum R_{to}}{[\sum R_{io} + \sum R_{to}]} \frac{\sum R_{to}}{[\sum R_{io} + \sum R_{to}]} \dots\dots\dots$$

(Equation 2)

This metric provides a user with way to compare biomedical ontologies and select between various alternatives, a suitable one for a biomedical data integration task.

4.7.3 Scope

Scope is a measure of how well an ontology's classes represent requirements for an integration task. For the tool, it is expressed as the number of objects [% search terms] in the requirements that are present or found in the ontology model. These scores are basis for a user to select the ontology, re-specify requirements or select another ontology to rate. The result also displays any unmatched requirements [search terms].

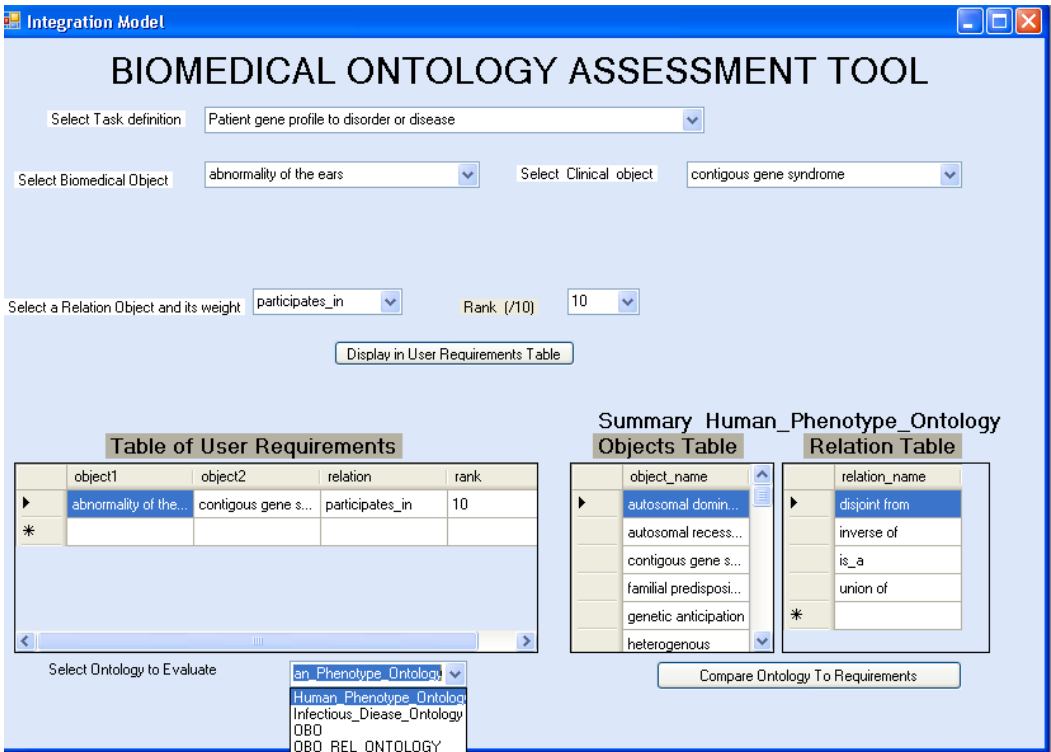
5. The Prototype

The tool was implemented using visual studio 2005 integrated development environment, in C# and SqlExpress2005 database management system (DBMS). The database summarized the human phenotypic ontology (HPO) and the infectious disease ontology (IDO), selected from the open biomedical ontologies (OBO) library. The OBO library is a collection of standardized ontologies (60 plus) being reformed or developed according to a set of principles that represents biomedical reality in a way that enables them to interoperate in the task of integrating biomedical data [Smith *et al*, 2007]. The emphasis on standardization informed selection of biomedical ontologies from the OBO library.

5.1. The Tool's User and Results Interfaces

The tool's user interface (figure 4.) has text boxes with drop down menus for selecting: 1) a biomedical integration task; 2) a biomedical object; 3) a clinical object; 4) a relation for the objects; 5) the ontology to be assessed. The interface also has panes for displaying requirements and the selected ontology. A user selects an integration task and its relevant biomedical objects and relations. The objects and relations constitute requirements for the task and are ranked on a scale of one to ten before being displayed. A use also selects an ontology, displays and compares it to the requirements. The biomedical objects and relations are used as the search terms against which an ontology selected from the database is assessed. The tool returns metric scores and a description of any unmatched search terms (figure 4).

Figure 4. The Tool's User Interface



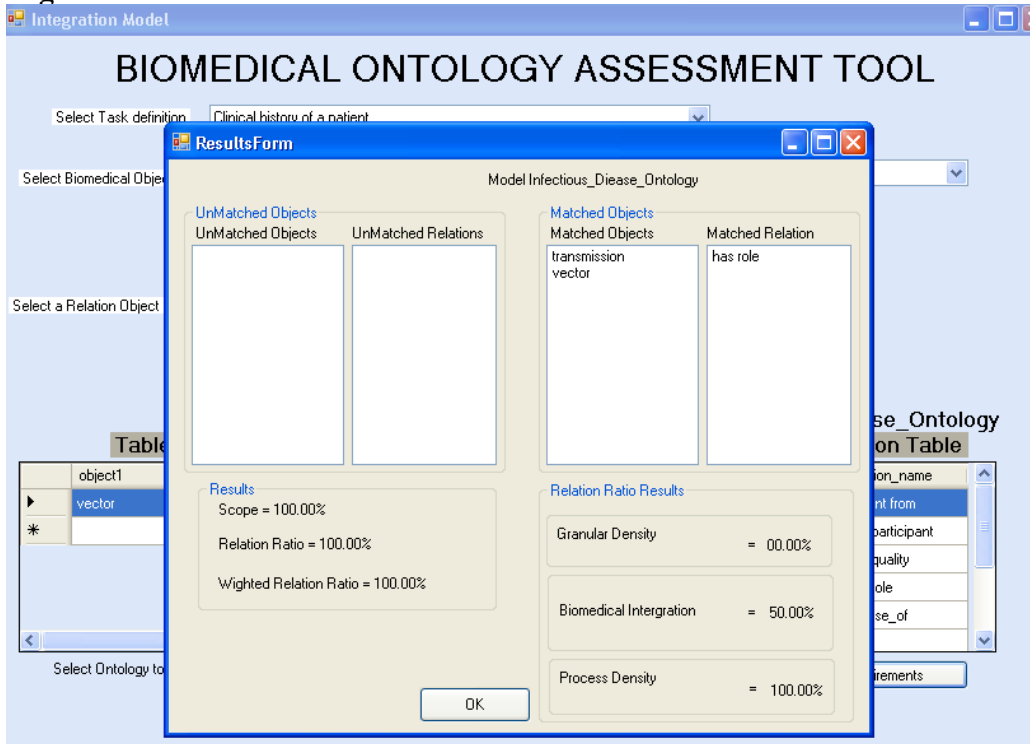
5.2. Evaluating the infectious disease Ontology

The tool was used to assess the infectious disease ontology [IDO], a set of interoperable ontologies that provide coverage of the infectious disease domain. IDO defines general entities relevant to biomedical and clinical aspects of infectious diseases. IDO is being built as a coordinated effort using best practices of the OBO consortium.

Experiments using the tool compared the ontology for their ability support different use case scenarios of biomedical data integration applications. For each use case, similar steps followed to assess an ontology are: 1) select a task; 2) generate requirements as a set of search terms; 3) rate the relative importance of requirements for the task; 4) display requirements; 5) select and visualize an ontology to be assessed; 6) compare search terms to selected ontology and generate any unmatched requirements. These may be used to re-specify a task, recommend ontology update or for use to build applications for the user's integration task or process.

Use case scenarios differed by the integration task selected, the requirements generated as a set of search terms and by the relative importance of these requirements or search terms for the users task. The output of using the tool to assess the IDO ontology for the different use case scenarios as illustrated in fig. 5 are presented in table 4. The results are displayed as scores of scope, granular density and biomedical integration.

Figure 5 . The Tool's Results Interface



5.3. Tool Validation

Previous attempts to validate ontology evaluation tools have included: 1) comparing the results of assessing an ontology in experiments using the tool to those by other approaches as controls [Tartir *et al.*, 2005]; 2) comparing the results of assessing an ontology using the tool to human assessment by expert users, ontology consumers and domain experts [Tartir *et al.*, 2005; Alani& Brewster, 2006; Cross and Pal, 2006],

The first approach is ideal for mature ontologies that have previously been evaluated by other tools using similar or related metrics, with data available for comparison. For IDO, no such data was available. The results of assessing the IDO with the tool (TA) are therefore compared to those by a questionnaire based human assessment (HA) study using a Pearson correlation [r] for significance. A random selection of 32 biomedical workers (18 medical doctors and 14 biologists) was used for this test. The questionnaire had a screen shot of the IDO model, two tables with six use case scenarios for applying the model and instructions on how to fill in the tables. For each scenario (search terms and relation), a question was asked about the models (IDO) ability to represent scope, biomedical relation and granularity. The answers to these questions were used to calculate the scope, granular density and biomedical integration ratios. The tool results (TA) of assessing IDO and corresponding ones from a questionnaire based human assessment [HA] for the different use case scenarios (expressed as percentages) are presented in table 4. The corresponding Pearson correlation coefficients are given in table 5.

Table 4 Tool Assessment (TA) and Human Assessment (HA) of different use case scenarios

		Scope		Granularity		Biomed	
		TA	HA	TA	HA	TA	HA
1	Vector <i>has_role</i> in transmission mode of disease	100	100	00	0	50	0
2	y-linked <i>has_participant</i> metabolism abnormality incubation period <i>part of</i> progression rate	50	50	100	50	67	50
3	y-linked <i>has_participant</i> metabolism abnormality infectious disease course <i>has_quality</i> phenotypic variability	25	15	50	0	67.7	50
4	Host has participant dormancy period	100	75	100	0	50	45
5	Host <i>has_quality</i> familial predisposition	50	55	00	0	50	0
6	Vector <i>is_a</i> transmission mode of disease	100	100	100	100	50	50

5.8. The Pearson Correlation

Table 5. Pearson Correlation [r] Comparison of Tool Scores against Human Assessments

Metric scored	Pearson correlation
Scope	0.945
Granular Density	0.608
Biomedical Ontology Integration	0.576

The r values from the tool and corresponding ones from the questionnaire based human assessment are not matched exactly. They however show a moderate to strong positive correlation between the two sets [$r = 0$ means zero correlation; $r = 1$ means strong positive correlation; $r = -1$ means strong negative correlation]. The scope ration is more highly correlated with biomedical ontology integration being the least.

The tool and human assessment had exact matches of scope, granularity and biomedical integration for the use case scenario of the *is_a* relationship. The tool and human assessment also had exact or near exact matches of scope and granularity for the use case scenarios of the *has_role*, *has_part* and *has_quality* relations, an indication of the success of the tool used, and the validity of our approach.

6. Discussion and Conclusions

The evaluation tool presented in this paper provides users with the flexibility to select a suitable ontology for use in building applications that integrate clinical and biological data. Given the large numbers of biomedical ontologies available in libraries, this tool can save time for both naïve users and expert ontologists by enabling them to quickly assess and select an ontology from a large collection (of summarized ontologies) before further evaluation of its taxonomy.

The tool can also enable ontology modelers to quickly gather new requirements for improving on existing biomedical ontologies, leading to new biomedical ontologies that are better able to integrate data across structure, function, processes and granularity. This is likely to contribute towards solving the problem of the persistent lack of a generic all purpose methodology for integrating biomedical data, as earlier recognized by Grenon *et.al.* [2004], leading to their re-use and adoption.

The evaluation process has some limitations. Human assessment of the infectious disease ontology presented some difficulties that are associated with ontology evaluation by human subject. Respondents were only exposed to screen shots of the IDO without sufficient accompanying documentation about it, which may have affected the quality of the response and choices made on the questionnaire. The respondents [medical doctors and biologists] are not ontologists and reported difficulty when interpreting and reasoning about the different use case scenarios of an ontology, even with detailed and clear instructions.

7. Acknowledgements

We wish to thank Makerere University's school of Graduate studies and the Faculty of Computing and Information technology for sponsoring this work. Special thanks to Dr Mathias Brochhausen and staff of the Institute for Formal Ontology and Medical Information Science, Saarland University, Germany, for hosting us and making the facilities of the institute for formal Ontology and medical information Science [IFOMIS] available to us when they were required.

References

- ALANI, H. & BREWSTER, C. (2006). Metrics for Ranking Ontologies. In *Proceedings of the 4th International Conference on Evaluation of Ontologies for the Web*. EON Workshop Edinburgh International Conference Center. Edinburgh UK. May 22nd 2006.
- BOARDMAN, R. (2004). Improving Tool Support for Personal Information Management. PhD Thesis.
- BRANK, J., GROBENLINK, M. AND MLADENIC, D. (2005). A Survey of Ontology Evaluation Techniques. *At the conference on Data Mining and Data Warehouses* SiKDD.
- BREWSTER, C., ALANI, H., DASMAHAPATRA, S., AND WILKA, Y. (2004). Data Driven Ontology Evaluation. *Proceedings of the Language Resources and Evaluation Conference LREC 2004*, Lisbon, Portugal.
- CROSS, V. AND PAL, A. ONTOCAT: (2006). An Ontology Consumer Analysis Tool and Its Use on Product Services Categorization Standards. Available: <http://protege.stanford.edu/conference>

- D'AQUIN, M. SABOU, M. DZBOR, C. BALDASSARRE, L. GRIDINOC, S. ANGELETOU, & E. MOTTA. (2007). WATSON: A Gateway for the next Generation Semantic Web Applications. *ESWC 2007*
- FERNANDEZ, M., CANTADOR, I., & P CASTELLS. (2006). CORE: A Tool for Collaborative Ontology Reuse and Evaluation.
- GANGEMI, A., CATENACCI, C., CIARAMITA, M. & LEHMANN, J. (2005). A theoretical framework for ontology evaluation and validation. *In Proceedings of SWAP 2005*.
- Gomez-Perez, A. (1994). Some ideas and Examples to Evaluate Ontologies. Technical Report KSL-94-65. Knowledge System Laboratory. Stanford University. Also in Proceedings of the 11 the Conference on Artificial Intelligence for Applications, CAIA94.
- GUARINO, N. & WELTY, C. (2002). Evaluating Ontological Decisions with OntoClean. *Communications of the ACM*. 45, No. 2.
- GÓMEZ-PÉREZ, A. (2004). Ontology Evaluation. *Handbook on Ontologies*. 251-274.
- GRENON, P., SMITH, B., AND GOLGBERG, L. (2004). Biodynamic Ontology: Applying BFO in the Biomedical Domain. *From D. M. Pisanelli ed., Ontologies in Medicine, Amsterdam: IOS Press*, 20-38
- HARTMANN, J., SPYNS, P., GIBOIN, A., MAYNARD, D., CUEL, R., SUÁREZ-FIGUEROA, M. C. & SURE, Y. (2005). Methods for ontology evaluation. EU-IST Network of Excellence NoE IST-2004-507482 KWEB.
- KALFOGLOU, Y. & SCHORLMER, M. (2006). Issues with evaluating and using publicly available ontologies. *In Proceedings 4th International EON Workshop, Evaluating Ontologies for the Web*, Edinburgh, Scotland.
- KEET, C.M. (2006). A Taxonomy of types of granularity. *IEEE International Conference on Granular Computing GrC*.
- KUMAR, A., LINAYIP, Y., SMITH, B. & GRENON, P. (2006). Bridging the gap between medical and bioinformatics: An ontological case study in colon carcinoma. *Computers in Biology and Medicine* 36, 694–711
- LAMBRIX, P. & TAN, H. (2006). Sambo. A system for aligning and merging biomedical ontologies *Web Semantics: Science, Services and Agents on the World Wide Web*, 4 196 – 206
- LOZANO-TELLO, A., GOMEZ-PEREZ, A. AND E. SOSA 2004. Selection of Ontologies for the Semantic Web.
- MAIGA, G. & WILLIAMS, D. (2008). A Flexible Approach for User Evaluation of Biomedical Ontologies. *International Journal of Computing and ICT Research*. ISSN 1996-1065. Volume 2 Dec. 2008, 62-74; Available: <http://www.ijcir.org/volume2-number2/index.php>
- NOY F. & MUSEN, M.A. (2002). Evaluating Ontology-Mapping Tools: Requirements and Experience. *Available: http://seer.nj.nec.com*. Accessed 6 may 2006.
- PORZEL., R. AND R. MALAKA. (2004). A task-based approach for ontology evaluation. Proc. of the 16th European Conf. on Artificial Intelligence (ECAI 2004), Valencia, Spain, August 2004.
- RECTOR. A., AND ROGERS, J. (2005). Granularity Scale & Collectivity: When size does and doesn't matter.
- SAATY, T. (1977). "A Scaling Method for Priorities in Hierarchical Structures". *Journal of Mathematical Psychology*, Vol.15, (1977) 234-281.

- SABOU, M., LOPEZ, V, MOTTA, E. AND UREN, V.(2006). Ontology Selection: Evaluation on the Real Semantic Web, *Fourth International Evaluation of Ontologies for the Web Workshop (EON2006)*, Edinburgh, UK, May (2006).
- SMITH, B., CEUSTERS, W., KLAGGES, B., KÖHLER, J. KUMAR, A.,LOMAX, J., MUNGALL, C., NEUHAUS, F, RECTOR, AL., ROSSE, C., (2005). Relations in Biomedical Ontologies. *Genome Biology*.
- SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W., GOLDBERG, JL., EILBECK, K., IRELAND, A., MUNGALL, C.J., LEONTIS, N., ROCCA-SERRA, P, RUTTENBERG, A., SANSONE, S., SCHEUERMANN, R.H., SHAH, N., WHETZEL, P.L., & LEWIS, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, Vol 25 (11).
- SUPEKAR, K. (2005). A peer-review approach for ontology evaluation. *Proc.8th Intl. Protégé Conference*, Madrid, Spain, July 18–21, 2005.
- TARTIR,S., ARPINAR B., MOORE, M., SHETH, A.AND ALEMAN-MEZA, B. (2005).OntoQA: Metric-Based Ontology Quality Analysis. *In IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, Houston, TX, USA, IEEE Computer Society (2005) 45 -53
- ZHANG, H., CHENG, G., AND QU YUZHONG.2007. Ontology summarization based on rdf sentence graph. In *proceedings of the 16th international conference on World Wide Web*. Banff, Alberta, Canada.

11

Web Content Filtration according to context of use: Case Study of Accessibility Guidelines

Rebema Baguma, Jude T. Lubega, Roger G. Stone and Th.P. van der Weide

In this paper, we propose an approach for filtering Web based content according to context of use based on user roles and other use case scenarios. The purpose of the filtration approach is to make such Web content easier to use for the target audience. Context of use is important for the usability of Web based content particularly that used by different groups of people with different roles, interests and skill base. We use the Web content accessibility guidelines (WCAG) to demonstrate how the approach can be used to improve usability of Web based content. Other than accessibility guidelines, the approach is relevant to ANY Web based content for any subject if intended for multiple classes of users.

1. Introduction

The Web has become a universal interface for various applications on the Internet, intranets and extranets [Murugesan and Ramathan, 2001]. However, Web users have different needs and roles and their skills and cognitive abilities also vary widely [Murugesan and Ramathan, 2001; Donnelly and Maggenis, 2003]. Therefore, one-size does not fit all its users and it is important to customize the Web interface to suit individual users' or a group of users' needs.

Nevertheless research on Web customization/personalization has so far been focused on obtaining and analyzing profiles of anonymous users particularly for e-commerce applications [Murugesan and Ramathan, 2001; Shwabe and Rosssi, 1999] leaving out cases with defined audience groups. In this paper, we discuss an approach for filtering Web based content according to context of use based on user roles or other use case scenarios. Context of use is important for the usability of Web based content particularly that used by different groups of people with different interests and skill base [Donnelly and Maggenis, 2003]. The Research-Based Web Design and Usability guidelines [Koyani et al. 2005] classify context of use in relation to Web usability as a user interface issue. The guidelines advise that among other user interface issues, the context within which users will be visiting a website should be considered.

We use the example of the Web based Web Content accessibility guidelines by the World Wide Web (W3C) [W3C, 1999] to demonstrate how the approach can be used to improve usability of Web based content. The example shows how to derive contexts of use for a given piece of Web based content as well as how to customize the content and user interface according to the chosen contexts of use.

Other than accessibility guidelines, the approach is also useful for ANY Web based content for any subject if intended for multiple classes of users.

The rest of the paper is organized as follows: Web content filtration approach according to context of use, an example using the Web Content Accessibility Guidelines (WCAG), conclusion and future work, references.

2. An approach for Filtering Web Content according to Context of Use

The approach for filtering Web content according to context of use proposes organization and presentation of Web based content according to contexts of use based on audience roles and other use case scenarios. The purpose of the approach is to make such Web based content easier to use for the target audience depending on their roles and tasks. Instead of one view for all users, the Web content can be organized into contexts/views according to user roles, tasks and or other use case scenarios. Customized views make it easier for a user to find required content. Figure 1 shows the architecture of the approach.

Figure 1 The architecture of the approach for filtering Web based content according to context of use

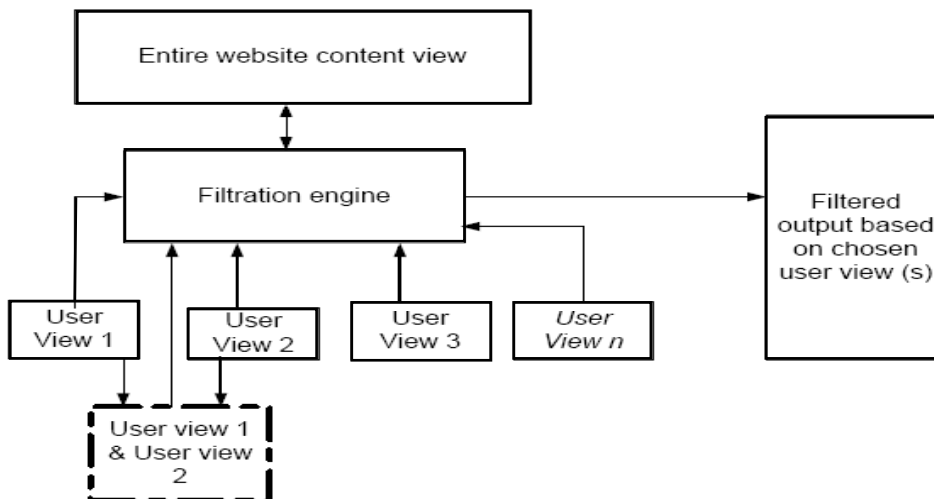


Figure 1 shows that instead of presenting the Web content in one view for all, it can be organized and presented according to different user contexts presented as user views (user view 1, 2, 3 & n). To filter the Web content according to user view (s), you apply a filtration technique represented by the filtration engine. Based on the selected view (s), the filtration engine sieves the entire content and outputs customized content. Therefore the filtration engine acts as a middle man between the entire content stored on the Web server and users' contexts of use expressed at the client end. It receives contexts of

use (user views) requests from users, filters the entire content and displays only the content requested. This helps a user to find quickly the content he/she wants, create a one to one experience, save time and increase user traffic. According to Murugesan and Ramanathan [2001], Web personalization can increase customer loyalty and attract a broader audience.

However this approach is rigid given that contexts of use have to be predetermined. This is difficult if not impossible given the dynamic nature of the Web audience characteristics such as preferences, skills and interests. As a solution to this potential shortcoming, the approach retains a view of the entire content in the customized interface. Any user with a context of use not yet represented in the interface will use the entire content view to look for the content of interest. Note that we use “not yet” because the interface is expected to evolve as more audience groups evolve and new user contexts get discovered.

In the next section, we demonstrate application of the approach on the Web based Web Content accessibility guidelines (WCAG). The purpose of the demonstration is to illustrate how the approach can be used practically to derive contexts of use for a given piece of Web based content as well as how to customize the content and user interface according to the chosen contexts of use.

3. An Example: Filtering Web Accessibility Guidelines According To Context Of Use

3.1 Introduction

Accessibility guidelines are aimed at all people with a role and responsibility in the procurement and development of IT products and services [Koyani et al. 2005]. These include: developers, IT procurers/managers and evaluators. The Web Content Accessibility Guidelines (WCAG) [W3C, 1999; W3C, 2008] provides a general set of guidelines that can be used to make Web based systems accessible to People with Disabilities (PWDs). Presently, WCAG is the most comprehensive and widely adopted accessibility guidelines [Donnelly and Maggenis, 2003].

But despite the wide recognition of the WCAG guidelines, they have been criticized for being overlong, unreadable and difficult to find required advice for much of its potential audience [Donnelly and Maggenis, 2003; Lazar et al. 2005; Clarke, 2006; Pickard, 2006]. Hence organizing and presenting the guidelines according to contexts of use could make the guidelines easier to use for the potential audience.

Kelly and colleagues [2005] noted that the vast majority of individuals working to make websites accessible, both developers and managers are not experts in accessibility or access technologies and may never be. More so, accessibility is just one of the aspects of developing websites they are seeking to understand. Therefore accessibility guidelines should be organized and presented in a way that makes it easy to find required advice for the different groups of users.

The next sub section discusses contexts of use by which the Web Content accessibility guidelines can be organized and presented namely; Web page component, type of disability, level of use and structure of the document. The choice was based on the roles

and interests of the audience groups defined by Donnelly and Magennis [2003] and general use case scenarios such as the logical structure in which content is presented.

3.2 Contexts of use for Web Accessibility Guidelines

3.2.1 Web Page Component:

Most of the Web accessibility guidelines including WCAG offer quantifiable rules but Web developers often fail to implement them effectively [Bigham and Ladner, 2007]. One of the challenges reported is that available accessibility guidelines are difficult to integrate into existing developer workflows and rarely offer specific suggestions that are developer oriented [Bigham and Ladner, 2007]. The Web page component context of use organizes and presents accessibility guidelines according to one of the concepts at the heart of Web development -that is components of a Web page. A Web page is a piece of information that can be displayed by the Web browser, a software used for accessing the Web [Yuen and Lau, 2003]. It is a resource that holds information on the Web. A web page can be as simple as displaying a single message or as complex as a communicator for fully functional databases of large organizations shared on the internet [Yuen and Lau, 2003]. Information in a Web page can be text, images, audio, video or a combination of two or more. A web page as an entity consists of a set of components or elements that make it a functional resource. These include: content, navigation and user interface.

Content: The content of a document refers to information conveyed to the user through natural language, images, sound, movies or animations [Thatcher et al. 2002]. Content is accessible if it can be viewed and accessed by the majority of users including PWDs [Thatcher et al. 2002]. Viewing is concerned with the format of the content such as text or audio while access is about the structure in which the content is presented that is how a document is logically organized for example by chapter, with an introduction and table of contents [Thatcher et al. 2002]. Structure is important for Web accessibility because assistive technologies for PWDs such as screen readers render Web content to the user in a linear form. Content is regarded the most critical element on a Web page than navigation, functionality and interactivity [Koyani et al. 2005]. It is unreasonable to spend resources on providing right content if that content cannot be accessed or viewed by the target audience.

Navigation: Web navigation is the method of getting around a given page, or moving within the website and on to other Web pages [Thatcher et al. 2002]. Common Web navigation tools include navigation menu, links, headings, which ideally can be activated by using either a mouse and or the keyboard. A Web page's navigation system is accessible if PWDs are able to perform navigational tasks successfully. Of course for some PWDs this may require the use of their standard support technology (e.g. screen readers for the visually impaired).

User Interface: The user interface of a Web page refers to the objects that the end user perceives and interacts with [Thatcher et al. 2002]. This covers the way in which

navigational objects are represented, which interface objects activate navigation, the way in which multimedia interface objects are synchronized, which interface transformations take place and the presentation of tasks that require users to input information such as a survey form [Shwabe and Rossi, 1999]. An accessible user interface is one where all the perceptible and interactive tasks of a Web page can be understood, perceived and utilized by PWDs.

Usability guidelines such as the Research based Web Design and Usability guidelines [Koyani et al. 2005] define users' usability expectations to be related to navigation, content and organization. One of the guidelines advises Web developers to focus on meeting user expectations on content, navigation and organization. Therefore structuring accessibility guidelines according to content, navigation and user interface also increases the likelihood of developing Web based systems that are more usable to PWDs. This is desirable given that despite the significant efforts for a more accessible and usable Web for PWDs, usability of Web based systems even for those that have passed accessibility tests is still a major challenge for PWDs [Vanderheiden, 2000].

Baguma and Lubega [2008] provide an example classification of Web design considerations for improved Web accessibility according to Web page components, a justification as to why each consideration is necessary for the accessibility of each Web page component, relationships between the accessibility of the three components and how such relationships can be exploited to improve the usability of Web accessibility guidelines.

Other than Web page components, other context of use used for filtering Web accessibility guidelines was type of disability, level of use and structure of the document. These are discussed next.

3.2.2 *Type of Disability*

There are different disabilities that can affect a person's use of the Web. These include: visual impairments (low vision, color blindness, blindness), hearing impairments (deafness, hard of hearing), physical disabilities (motor disabilities), speech disabilities, cognitive and neurological disabilities and ageing related disabilities [W3C, 2005].

Different disabilities cause different barriers on the Web but as noted by W3C [2005], sometimes different disabilities require similar accommodations. For instance, someone who is blind and another with unstable hand movement both require full keyboard support.

In addition, different Web projects may have varying levels of accessibility requirements for different types of disabilities for example a website of an educational institution without audio content would be less concerned about advice on hearing impairments but more concerned about visual, cognitive and physical disabilities. Another website selling music would be most concerned about hearing impairments. However, WCAG 1.0 does not categorize and present guidelines according to type of disabilities. WCAG 2.0, mentions against each guideline, the disability (ies) it benefits but this is only done at guideline level. One cannot get at once all the guidelines about a

particular disability. To get such information, one has to go through the entire guidelines document.

The content filtration approach shows how to organize and present Web accessibility guidelines according to types of disabilities that affect a person's use of the Web. The classification is based on the definitions of the different types of disabilities given in [W3C, 2005]. A complete categorization of WCAG 1.0 guidelines according to types of disabilities is available on request from the authors.

In addition to type of disability, the example application also shows how to filter Web accessibility guidelines according to level of use. This is discussed next.

3.2.3 Level of Use

In a study carried out by Colwell and Petrie [1999] on WCAG 1.0, participants had problems obtaining the guidance needed, could not navigate the guidelines effectively and were always not clear whether they were accessing guidelines or techniques. The implication of such navigation problems is the risk of missing important advice hence producing inaccessible systems.

Web development has more stakeholders than just developers (Donnelly and Maggenis [2003]).

According to a study conducted by Donnelly and Magennis in 2003 [2003], users of Web accessibility guidelines want information that is tailored to their roles and responsibilities. People involved in IT procurement need assistance in drafting the request for tender/proposal and assessing compliance with the agreed contract. IT project Managers need an accurate overview of accessibility problems and the implications of compliance to avoid them. Developers require detailed technical guidance and illustrative examples with clear functional requirements.

With the organization and presentation of the Web content accessibility guidelines (WCAG) both WCAG 1.0 and WCAG 2.0, the non technical audience cannot easily find appropriate advice while the technical audience (developers) lacks guidance on how to humanize accessibility aspects of their systems. Consequently developers focus more on passing automated accessibility tests, which are technically oriented and less on usability of their systems to PWDs [Asakawa, 2005]. Moreover, passing automated accessibility tests does not mean a website is accessible [Asakawa, 2005, Kelly et al. 2005]. An image may pass the test because it has an alternative text description but the image can still be inaccessible if the alternative text does not make sense to the user.

The level of use context organizes and presents accessibility guidelines according to level of use with two sub contexts namely: understanding the guidelines and the techniques. Although WCAG 2.0 is split into 2 main documents- that is understanding WCAG and techniques for WCAG, each of these is a separate document which means users have to move back and forth between the main guidelines page and the documents. Instead of two separate documents in addition to the main guidelines page, the level of use context presents both views in a single document.

The Understanding guidelines view is aimed at both developers and non technical users. The non technical need to gain an accurate overview of accessibility problems

and the implications of compliance, while developers need an over all awareness to inform the integration of accessibility into early stages of Web development as well as determination of design implications.

The understanding guidelines view provides developers with a layer of advice to use during early stages of Web development. By the time coding starts, accessibility should be an integral part of the requirements so far determined and remain part of subsequent iterations. This way, accessibility will not be an isolated property but a desirable property integrated with functional and other non functional properties. Such designs once implemented will produce websites that are both technically and humanly accessible. As Mline et al. [2005] put it, a holistic approach to accessible design relies on a better understanding of users' needs on the part of designers, developers and managers of digital resources.

The techniques view provides developers with practical guidance and illustrative examples on how to code accessible Web pages. This view is important for reference. It makes it easy for developers to find guidance on a given technique for a particular problem anytime.

3.2.4 Structure of the document

Other than the three main contexts, the content filtration approach for the Web Content accessibility guidelines includes a fourth context- that is structure which organizes and presents the guidelines according to the logical structure in which content is presented. It has three sub views namely: preliminary which keeps preliminary content such as abstract and table of contents 'hidden' and only displays it when needed, guideline which filters the guidelines on a per guideline basis and other information which filters appendices, references and acknowledgement. The structure context of use improves usability of the accessibility guidelines in that instead of having peripheral content such as table of contents, abstract, references and acknowledgement displayed with the main guidelines content, this information can be displayed only when it is needed. The guideline sub view makes it easy to refer to specific guidelines while using the guidelines.

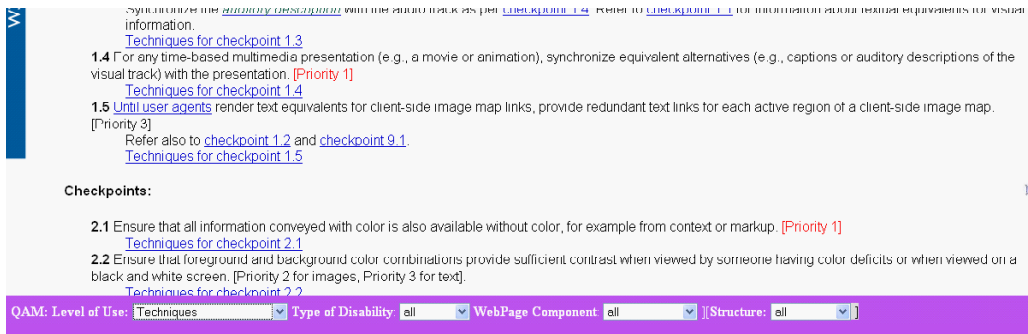
Other than accessibility guidelines, the approach is also useful for ANY Web based content for any subject if intended for multiple classes of users.

3.2.4 A Prototype System

A prototype system for filtering accessibility guidelines according to context of use and other use case scenarios was developed using the WCAG 1.0 guidelines. We chose WCAG 1.0 rather than the latest version, WCAG 2.0 because until 11th December 2008, WCAG 1.0 was still the stable and referenceable version of WCAG as was stated on the WCAG 2.0 overview page by 20th November 2008. The source code of WCAG 1.0 website was extracted and recoded to add CSS - Cascading Style Sheets style tags, java script code and more XHTML tags for customizing the organization and presentation of the guidelines according to the different contexts. In the customized user interface, the user sees a large frame containing the guidelines document and a small navigation frame. The navigation frame offers a four-dimensional viewing opportunity based on

the four contexts. Thus the user may select “techniques” (rather than “understanding”) from the “level of use” menu. Each menu choice causes a small client-side script to run which dynamically selects a suitable style sheet to filter the display in the large viewing frame. A choice in one context may be combined with choices in other contexts. For example the “techniques” choice could be combined with “visual” (rather than “hearing”, “physical” or “cognitive”) from the “disability” menu to give a display of techniques for visual impairment. Figure 2 shows a snapshot of the prototype system.

Figure 2: A snapshot of the prototype system



The CSS approach has a more optimal design for the static Web based content such as descriptive guidelines compared to the database driven design used by some accessibility guidelines like the Irish IT accessibility guidelines [10]. Content filtering using CSS is much lighter and easier to implement than a database driven approach. Besides with the CSS approach, it is easier to switch between the full content view to customized views suiting different audience roles. This is possible because of the unified interface. The filtration approach can be applied to other documents of related nature particularly those covering various subjects for a diverse audience.

Conclusion And Future Work

In this paper we have discussed and demonstrated an approach for filtering Web based content according to context of use based on user roles or other use case scenarios. Context of use is important for usability of Web based systems particularly those used by different groups of users for different needs. Other than the Web accessibility guidelines used to demonstrate how the approach can be used practically, the approach is useful for ANY Web based content for any subject if intended for multiple classes of users.

References

ASAKAWA C. (2005). What's the Web is Like If You Can't See It? International Cross Disciplinary Workshop on Web Accessibility (W4A), 23rd-26th May 2006, Edinburgh, UK

- BAGUMA, R. AND LUBEGA, T.J. (2008). A web design framework for Improved accessibility for people with disabilities (WDFAD), W4A 2008 134-140.
- BIGHAM AND LADNER, (2007). Accessmonkey: A Collaborative Scripting Framework for Web Users and Developers, W4A 2007 Technical Paper May 07–08, 2007, Banff, Canada. Co Located with the 16th International World Wide Web Conference
- CLARKE, J. (2006). To hell with WCAG, retrieved from <http://learningtheworld.eu/2006/to-hell-with-joe-clark/> on 10th January 2008.
- COLWELL AND PETRIE, H. (1999). Evaluation of Guidelines for Designing Accessible Web Content, INTERACT'99 Workshop: Making Designers Aware of Existing Guidelines for Accessibility August, 1999
- DONELLY AND MAGGENIS (2003). Making Accessibility Guidelines Usable, N. Carbonell, C. Stephanidis (Eds.): User Interfaces for All, LNCS 2615, pp. 56–67, 2003.
- KELLY ET AL. (2005). Forcing Standardization or Accommodating Diversity? A Framework for Applying the WCAG in the Real World, W4A at WWW2005, 10th May 2005, Chiba, Japan
- KOYANI ET AL. (2005). Research based Web design and Usability Guidelines retrieved on 10th Jan. 2009 from http://www.opsi.gov.uk/acts/acts1995/ukpga_19950050_en_1
- LAZAR, ET AL. (2005). Improving Web accessibility: A study of Webmaster perceptions, *Computers in Human Behaviour*, 20, 2005, 269288
- MLINE ET AL. (2005). Are guidelines enough? An introduction to designing Web sites accessible to older people, *IBM Systems Journal*, Vol 44, No 3, 2005.
- MURUGESAN, S AND RAMANATHAN, A. (2001). Web Personalisation- An overview, J. Liu et al. (Eds.): AMT 2001. LNCS 2252.pp.65-76, 2001.
- PICKARD (2006). The WCAG 2.0: What a Whopper, retrieved from <http://learningtheworld.eu/2006/to-hell-with-joe-clark/> on 5th Jan. 2008
- SHWABE, D. AND ROSSI, G. (1999). Designing Personalized Web applications, WWW10, 2001, Hong Kong.
- SHWABE AND ROSSI, (1999). Developing Hypermedia Applications using OOHDM retrieved from on 11th January 2008.
- SLOAN ET AL. (2000a). Accessible Accessibility, CUU '00 Arlington VA USA
- SLOAN ET AL. (2000b). Contextual Web Accessibility -Maximizing the benefit of Accessibility Guidelines

THATCHER ET AL. (2002). Constructing Accessible Web Sites, web professional to web professional an Apress imprint.

VANDERHEIDEN, (2000). Fundamental Principles and Priority Setting for Universal Usability, CULl '00 Arlington VA USA

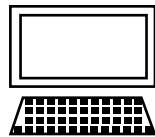
W3C, (1999). Web Content Accessibility Guidelines 1.0 retrieved from <http://www.w3.org/TR/WCAG10/> on 19th Jan. 2008

W3C, (2008). Web Content Accessibility Guidelines (WCAG) 2.0 retrieved from <http://www.w3.org/TR/WCAG20/> on 18th Jan. 2008

W3C, (2005). How People with Disabilities Use the Web, 2005, retrieved on 20th April 2007 <http://www.w3.org/WAI/EO/Drafts/PWD-Use-Web/Overview.html-use>

YUEN, P.K. AND LAU, V. (2003). Practical Web Technologies, Addison-Wesley

PART THREE



ICT Sustainable Development

1

SIM or Application Layer? An Implementation-Level Analysis on the use of Mobile Phones for ICT Development

Hannah Thinyane

In recent years, mobile phones have started to become popular in their use as a platform for ICT based development projects. This paper provides an implementation level analysis on the use of SIM card programming and application layer programming. This is a particularly important consideration when the program is to be run in developing nations due to the older handsets that are frequently used. The paper describes the ramifications that each layer would have on the application that is created, in particular in the context of developing nations. It then describes a case study of a development project where we have implemented two such applications, illustrating the principles described in the paper.

1. Introduction

In 2003, the United States Agency for International Development (USAID) defined two categories of Information and Communication Technologies (ICTs): traditional ICTs and modern ICTs [Ruseten and Ramirez 2003]. USAID define traditional ICTs as technologies that are engrained in most people's daily lives such as television, radio, and fixed line telephones. Modern ICTs are defined as computers and communication systems between computers such as the Internet, desktop computers, laptops, mobile phones and PDAs. There is often great debate amongst researchers about the compatibility of modern ICTs in developing communities [Pade 2007]. Polikanov and Abramova [2003] identify three distinct schools of thought: pessimists who argue that progress in ICTs will only broaden the digital divide; optimists who propose ICTs as the remedy to all problems faced by developing nations; and realists who maintain that societies will adapt technologies to serve them best.

From a realists perspective, Pade notes that it is assumed that traditional technologies are more suitable in meeting the needs of rural communities as they already face challenges such as poor education and illiteracy [2007]. McNamara also warns against overlooking traditional ICTs which in most cases are more appropriate, affordable, and adaptable to local community needs than modern ICTs [2003]. If the only pre-requisite of a technology to be seen as traditional is that it is engrained in most people's daily lives, we argue however, that due to the proliferation of mobile phones in developing nations, they can now be seen as a viable technology. A report compiled by the Wireless World Forum states that 66% of all South Africans own one or more mobile phone contracts [2006]. Another study compiled by the Consultative Group to Assist the Poor

(CGAP) and United Nations Foundation reports that in early 2006, the mobile phone became the first ICT to have more users in developing countries than in developed ones [Ivatury and Pickens 2006]. Donner [2004] reports that each day, thousands of people in developing nations purchase mobile telephones. On top of this, mobile phones have been used as a platform for development initiatives in a wide range of areas: the fishing industry in India [Reuben 2006]; a telemedicine network for rural clinics in Palestine [Zatari 2007]; online banking in Africa [Dybwad 2005]; and micro-entrepreneurs in Rwanda [Donner 2004]. These figures speak of the proliferation and acceptance of mobile phones within developing nations.

For applications to be available to mobile phone users, they can either be written as web-based applications that require a constant connection to the Internet, or they can be written to be run on the mobile phone. An issue facing many developing nations is a lack of physical infrastructure, including network infrastructure [Dutta 1997]. Although mobile phone reception covers fair portions of developing countries, it can at times be flaky, causing web-based applications to drop their connections. It is for this reason that this research investigates applications that reside on the mobile phone itself, rather than require constant connection to the internet.

Primarily there are two ways that these mobile applications can be run: on the SIM card itself, or as an application that runs on the handset. The major difference between the two techniques is that the latter relies on more capabilities on the handset itself than the former. As in all technologies, there is a roll down in the capabilities supported by handsets, with technologies that were considered innovative (for example the camera) slowly being available on even the most basic handsets. Also, with improvements in memory, processor, screen size, and power consumption, new mobile phone technologies are continually emerging on the market for much the same price as their predecessors. Surprisingly, this roll down of technology is slower than would be expected. In March 2008, Japan became only the second country (behind South Korea) to take delivery of no second generation mobile phones [Agence France-Presse 2008]. This illustrates that even in developed nations, consideration must be made for the capabilities of the mobile phone handsets.

Over the past three years, we have been involved in an ongoing ICT for development research project in a rural area in the former Transkei in South Africa. The research reported in this paper arose after visiting the area for a year and seeing how mobile phones were so ingrained into the lives of the people who lived in the area (see **Figure 1**). The original objective of the project was to develop and field-test the prototypes of a simple, cost-effective and robust, integrated e-business/telecommunication platform, to deploy in marginalized and semi-marginalized communities in South Africa where a large proportion (42.5%) of the South African population live. The project evolved however to include several other sub-projects, one of which is to investigate the use of mobile phones as a platform for ICT for development.

Figure 1: Mobile phone charging station at a local school

This paper presents an implementation-level analysis on the use of mobile phones as a platform for ICT for development. It compares and contrasts developing programs on SIM cards and developing application layer programs, given the particular context and constraints in place in developing countries. In particular this paper references development on Java Card [Sun Microsystems 2008a] platforms and Java Mobile (J2ME) [Sun Microsystems 2008b] platforms. Although they are not the only SIM based and application platforms available for mobile phones, they are the environments that the author has most experience with.

2. Implementation Techniques

As mentioned in the introduction to this paper, there are two levels that programs can be written on that reside on the mobile phone: SIM card level; and application level. SIM card level programs are executed solely on the SIM card and therefore make no demands on the mobile handset capabilities. This makes them particularly suited to older mobile phone handsets. However, as they are limited to the size of a SIM card, they are forced to have a very small footprint and cannot make use of any of the peripheral devices included in the mobile handset. Application level applications on the other hand can take full advantage of any peripheral devices, but as handsets can differ so vastly, may not be suited for as large a variety in mobile phones as SIM based applications. This section will now present a detailed description of the two levels of programming, with respect to the context of creating applications for low end mobile phones.

2.1. SIM Based Programming

Java Card technology allows the development of applications for earlier mobile phone models by targeting development at the SIM card level, making it unnecessary to have support at the application level. The Java Card environment shares the same architecture as the standard Java environment. Due to the limited resources on current SIM cards, Java Card does not support the following Java features: it doesn't support all primitive types [Oestreicher 1999]; and does not allow the dynamic download of classes. Instead all classes are packaged using a converted to create one executable file. This in turn reduces the size by pre-linking for the execution on the card as far as possible [Oestreicher 1999].

Java Card uses a Kilobyte Virtual Machine (KVM) which provides a basis for the Connected, Limited Device Configuration (CLDC). This configuration defines small, mobile devices with memory ranging from 160 to 512 Kbytes [Helal 2002a]. It is called the Kilobyte VM in reference to the small footprint of the platform [Sun Microsystems 2000]. Java Card has been used for many different purposes: to implement electronic money for e-commerce purposes [Insik and Ingook 2001]; storing presence information for mobile phone address books [Moyo 2007]; and as a store for sensitive data in the context of near field communication [Madlmayr et al. 2007].

2.2. Application Layer Programming

Although Java Micro Edition (J2ME) was released in June 1999 [Helal 2002a], it wasn't until September 2000 that Java was commercially deployed on mobile phones [Lawton 2002]. By April 2001, over three million Java handsets were sold worldwide [Helal 2002b]. Although not the only application platform available for mobile phones, it has been pushed as the de facto standard on mid-range and high-end mobile phones for the past eight years [Lawton 2002].

Like Java Card, J2ME makes use of a KVM with very similar functionality. A difference is that J2ME allows for user data to be stored in a simple, secure manner by making use of persistent storage, such as record stores. The deployment and installation of Java Mobile applications is straightforward, as it simply makes use of JAR and JAD files which automatically prompt the user to confirm the installation [Sun Microsystems 2000].

3. Case Study

Dwesa / Cwebe is a coastal region located in the previous homeland of the Transkei in the Eastern Cape, South Africa. It has an estimated population of 15000 people living in 2000 households. The inhabitants of Dwesa/Cwebe are traditionally subsistence farmers who depend on their crops for their livelihood [Palmer et al. 2002]. Figure 2 (a) shows a typical Dwesa / Cwebe homestead. The region features a coastal nature conservation park which is owned by the community around a reserve and a hotel. The region has a high potential for eco and cultural tourism due to the rich cultural heritage and the marine conservation project undertaken at the nature reserve. We consider the Dwesa / Cwebe region to be ideal to take advantage of the global upsurge in eco-tourism activities.

The Dwesa / Cwebe region comprises of two distinct communities. The Dwesa people on the southern side of the Mbashe River, and the Cwebe people on the northern side. This delineation is not only geographical, but extends to the philosophical outlook on the underlying world views of the two communities [Palmer et al. 2002]. This introduces an interesting cultural dynamic, with the Dwesa community being more educated and open to change, and the Cwebe community representing a more traditional and static culture.

Figure 2 (a) Dwesa / Cwebe region; and (b) Ntubeni School in Dwesa region

From a political viewpoint, Dwesa / Cwebe is only the second successful land restitution claim case in South Africa, and the first in the Eastern Cape. We believe that it is therefore an ideal location to use as a testbed for the implementation of ICTs in marginalized communities. Like most marginalized communities, Dwesa / Cwebe suffer from major infrastructure problems including limited electricity availability and connectivity, minimal telecommunication infrastructure, poor quality of the transport infrastructure, but we feel more importantly, sub-standard education facilities. The schools that do exist are also under-funded, under-equipped, and under-staffed. Figure 2 (b) is an example of such a school in the Dwesa region, which even lacks basics such as a roof.

In a recent study of 265 residents in the Dwesa area, it was found that just under half (46.5%) of the residents either own or have direct access to a mobile phone (where direct access refers to access to a mobile phone in their household). Of these, 73% charge their mobile phones at the local school or shop, and 12.5% use either a solar inverter or a motor battery. The remaining 14.5% charge their phones either at the nearest town with electricity (45 minute drive by public transport), at a clinic, or anywhere they can find.

3.1. Siyakhula Living Laboratory

The Siyakhula Living Laboratory (LL) is a joint venture between the Telkom Centres of Excellence (CoEs) at Rhodes University and the University of Fort Hare respectively. The original objective of the project was to develop and field-test the prototypes of a simple, cost-effective and robust, integrated e-business / telecommunication platform, to deploy in marginalized and semi-marginalized communities in South Africa where a large proportion (42.5%) of the South African population live. The project has evolved into an experiment on the adaptation of the Internet to rural areas in South Africa, based on the deployment of ICTs in schools, which together realize a distributed access network [Thinyane et al. 2008]

The current deployment of ICT infrastructure in Dwesa has been concentrated around schools which provide a centralized location, accessible to many villages. This allows us to piggy back on existing societal structures which facilitate community

acceptance, buy-in, and eventual ownership of the infrastructure. Another important benefit of locating computer laboratories in schools is that they are some of the few places in Dwesa that are connected to the main power grid (some local stores and health clinics situated on the main road also have power).

In the introduction to this case study, it was mentioned that the Dwesa / Cwebe region has a high potential for eco and cultural tourism. It is this potential that led to us partnering and training with the local craft initiatives and entrepreneurs. An interesting point about the Dwesa nature reserve is that during the summer holidays, the reserve is full of holiday-makers. In other times of the year, the place is empty. We proposed that with the extra market available online, the entrepreneurs and craft people in the area would be able to sell their goods to a greater market. With this in mind we created an online portal where residents could advertise their goods. In reality, we developed two different front-ends to the portal, one a SIM-based application, and the other a J2ME application. The following two sections describe the applications.

3.2. Mobil-e-Com

Mobile-e-Com [Slay et al. 2007] is a SIM-based application we developed to support second economy entrepreneurship. It was designed as a solution to what we saw as a major gap where current e-commerce solutions for development do not exploit the proliferation of mobile phones among the communities in developing countries. Mwabu and Thorbecke [2001] and Pade [2007] recognize the importance of rural based growth on the route to development. It was with this in mind that Mobil-e-Com was created, specifically to harness the currently dormant entrepreneurship potential of the second economy in rural areas.

Mobil-e-Com uses mobile phones to facilitate e-commerce transactions between sellers and buyers via mobile phone. The framework allows sellers to upload details, and if the phone supports it, photos of their goods for sale. Using their phone they can browse and update their goods already for sale. Buyers can browse via an online web-portal and select items they wish to purchase by adding them to a virtual shopping basket. When buyers check out their purchases, the framework forwards the buyers contact details to the seller. It is then up to the seller to contact the buyer to arrange details such as delivery and payment. An add-on to the Mobil-e-Com application is the optional micro-payment infrastructure that can be used to pay for their goods.

As mentioned earlier, Mobil-e-Com was developed to be sufficiently deployable on the SIM card of entry-level second generation mobile phones. Mobil-e-Com could have easily been developed as a J2ME application, but this would necessitate Java support and capabilities on the handsets, reducing its accessibility from the older generation of mobile phones.

Excluding greater support for older mobile phones, there are four other benefits to deploying the solution on a SIM card: independence from the handset; alliance with mobile operator's network; minimal resource utilisation; and ease of deployment. By deploying Mobil-e-Com on the SIM card, it makes it independent of the handset, ensuring operational consistency. In effect, this ensures that the program will work in

an identical fashion across different handsets. This is important when considering the effects caused by a poorly designed (or poorly displayed) user interface. Assuming the user interface undergoes user testing before it is released, this guarantees the same user experience across mobile handsets. By providing a SIM based solution, the application is implemented with the architectural framework of the mobile operator's network. This allows for optimizations on data transport protocols used, and allows access to operators' servers (messaging, authentication, etc). Mobile phone resource utilization is also minimal, as the application runs entirely from the SIM. This is important, particularly for older phones, as the resources on the mobile phone handsets are low to start with. Finally, when using SIM based applications, deployment is undertaken using Over The Air (OTA) messaging infrastructure on mobile networks [Gemplus 2006]. This is particularly useful for rural, marginalised and isolated communities. Using OTA, applications, new updates and services can be transferred to mobile phones or SIM cards by the network operator. This allows applications to be downloaded onto a SM card without physically having to connect to it. By having an OTA backend, the operator's servers send requests through an OTA gateway to the mobile phone. The mobile phone user then pulls then accepts the request and downloads the application.

Mobil-e-Com was developed using GemXplore Developer [2007], a SIM application development toolkit. For this prototype, there was no network operator buy-in so testing was undertaken using a simulator. EasyOTA [Gemplus 2006] was used to test the OTA transfer and communication between the SIM card and the (simulated) network operator.

3.3. Java-e-Com

The Java e-Commerce platform, Java-e-Com was developed using J2ME for newer mobile phones. In contrast to Mobil-e-Com, this application was designed to be run on the mobile phone itself (rather than the SIM card). It was however designed for the same conditions as Mobil-e-Com. The remainder of this section highlights the benefits that developing on the application layer brings to Java-e-Com, paying particular attention to grounding it in the same physical environment as Mobil-e-Com.

As with SIM based applications, a J2ME application provides portability and interoperability to the user. One application can be developed and executed on different types and varieties of handsets. Also, a link to a J2ME application can be transferred to the handset via OTA and then be downloaded from the Internet and installed on the phone.

The primary benefit of an application layer program is in strict contrast to the SIM based application: it can have greater access to the mobile phone functions and peripheral devices such as cameras that come as part of the mobile handset. Another important benefit of an application layer program is that the J2ME provides a more complete implementation of the J2SE, and therefore supports richer applications than the Java Card alternatives. This provides programmers with more leeway to create graphically rich and attractive applications. This can be a very useful extension, particularly when considering that large parts of developing countries suffer from semi-literacy and

illiteracy [Terry 2003]. This supports work such as that undertaken by Indrani, Aman and Kentaro [2006], where graphically rich interfaces were created to support illiterate and semi-literate users.

4. Conclusion

This paper has compared SIM based programming with application layer programming within the context of ICT for development. It has highlighted critical considerations that need to be made when programming for these conditions, such as lack of network infrastructure, but more importantly older mobile phone handsets. It has shown how applications can be created particularly for these environments using SIM based programming that make very little demands on the physical phone handset. To properly illustrate the difference this has to programming for other contexts, it paralleled the discussion of SIM based programming with a discussion of a J2ME application.

References

- AGENCE FRANCE-PRESSE (2008) *Japan marks funeral for second-generation phones* in Asia One Digital. Published on Wednesday 12 March, 2008. (Online) <http://www.asiaone.com/Digital/News/Story/A1Story20080312-53982.html>
- DONNER, J. (2004) *Microentrepreneurs and Mobiles: An Exploration of the Uses of Mobile Phones by Small Business Owners in Rwanda*. Information Technologies and International Development. 2(1): p. 1-21.
- DUTTA, A. (1997) *The physical infrastructure for electronic commerce in developing nations: historical trends and the impact of privatization* International Journal of Electronic Commerce 2(1): p. 61-83.
- DYBWAD, B. (2005) *Cellphone proliferation in Africa creates basis for low-cost mobile banking*. [Online report] 29 August 2005 [cited 22 March 2007]; Available from.
- GEMPLUS (2006) *GemConnect - OTA*. [cited 2008 September 29]; Available from: <http://www.gemplus.com/techno/ota/>.
- GEMPLUS (2007) *GemXplore Developer*. 14 February 2007 [cited 2008 September 29]; Available from: http://www.gemplus.com/products/gemxplore_developer/.
- HELAL, S. (2002a) *Pervasive Java*. Pervasive Computing, IEEE. 1(1): p. 82-85.
- HELAL, S. (2002b) *Pervasive Java, Part II*. Pervasive Computing, IEEE. 1(2): p. 85-89.
- INDRANI, M., AMAN, S., AND KENTARO, T. (2006) *Text-Free User Interfaces for Illiterate and Semi-Literate Users*. in *Information and Communication Technologies and Development, 2006. ICTD '06. International Conference on*.
- INSIK, H. AND INGOOK, C. (2001) *The implementation of electronic money for E-Commerce using Java card*. in *IEEE International Symposium on Industrial Electronics, 2001*.
- IVATURY, G. AND PICKENS, M. (2006) *Mobile Phone Banking and Low-Income Customers: Evidence from South Africa*. [cited 22 March 2007]; Available from: Consultative Group to Assist the Poor, UN Foundation & Vodafone Group Foundation.
- LAWTON, G. (2002) *Moving Java into mobile phones*. Computer. 35(6): p. 17-20.

- MADLMAYR, G., DILLINGER, O., LANGER, J., SCHAFFER, C., KANTNER, C., AND SCHARINGER, J. (2007) *The benefit of using SIM application toolkit in the context of near field communication applications*. in *International Conference on the Management of Mobile Business, 2007. ICMB 2007*.
- MCNAMARA, K. (2003) *Information and Communication Technologies, Poverty and Development: Learning from Experience*. in *A background paper for the InfoDev Annual Symposium, December 9-10*. Geneva, Switzerland: The World Bank.
- MOYO, T., THINYANE, M., WRIGHT, M., IRWIN, B., TERZOLI, A., AND CLAYTON, P. (2007) *Bridging the gap for Next Generation Services: Presence Services on Legacy Devices in Southern African Telecommunications Networks and Applications Conference (SATNAC'07)*. Mauritius.
- MWABU, G. AND THORBECKE, E. (2001) *Rural Development, Economica Growth and Poverty Reduction in Sub-Saharan Africa*. *Journal of African Economies*. 13(1): p. i16-i65.
- OESTREICHER, M. (1999) *Transactions in Java Card*. in *15th Annual Computer Security Applications Conference, 1999. (ACSAC '99)*.
- PADE, C. (2007) *An Investigation of ICT Project Management Techniques for Sustainable ICT Projects in Rural Development*. Rhodes University: Grahamstown, South Africa.
- PADE, C., MALLINSON, B., AND SEWRY, D. (2006) *An Investigation of Sustainable ICT Projects in Rural Development*. in *Sangonet 2nd Annual Conference and Exhibition: ICTs for Civil Society*.
- PALMER, R., TIMMERMANS, H., AND FAY, D. (2002) *From Conflict to Negotiation: Nature-based development on South Africa's Wild Coast*.
- POLIKANOV, D. AND ABRAMOVA, I. (2003) *Africa and ICT: A Chance for Breakthrough?* *Information, Communication & Society*. 6(1): p. 42 - 56.
- REUBEN, A. (2006) *Mobile Phones and Economic Development: Evidence from the Fishing Industry in India*. in *Information and Communication Technologies and Development, 2006. ICTD '06. International Conference on*.
- RUSTEN, E. AND RAMIREZ, S. (2003) *Future Directions in Agriculture and Information and Communication Technologies (ICTs) at USAID*. February 2003 [cited 22 March 2007]; Available from: United States Agency for International Development.
- SLAY, H., THINYANE, M., AND TERZOLI, A. (2007) *Mobil-e-Com: A SIM-based Application for Support Second Economy Entrepreneurship*. in *Southern African Telecommunications Networks and Applications Conference (SATNAC'07)*. Mauritius.
- SUN MICROSYSTEMS (2000) *J2ME Building Blocks for Mobile Devices: White Paper on KVM and the Connected, Limited Device Configuration (CLDC)*. Sun Microsystems. <http://java.sun.com/products/cldc/wp/KVMwp.pdf>,
- SUN MICROSYSTEMS (2008a) *Java Card Technology*. [cited 2008 September 29]; Available from: <http://java.sun.com/javacard/>.
- SUN MICROSYSTEMS (2008b) *Java ME at a glance*. [cited 2008 September 29]; Available from: <http://java.sun.com/javame/index.jsp>.
- TERRY, B. (2003) *Measuring Literacy in Developing Countries from an International Perspective*. in *54th Session of International Statistical Institute*. Berlin, Germany.

- THINYANE, M., DALVI, L., TERZOLI, A., AND CLAYTON, P. (2008) *The Internet in rural communities: unrestricted and contextualized*. in *ICT Africa*. Addis Abba, Ethiopia: Nepad Council.
- WIRELESS WORLD FORUM (2006) *World Mobile Statistics*. [Report] [cited 22 March 2007]; Available from: World Forum Group.
- ZATARI, D.I. (2007) *Development of a wireless rural telemedicine network and management system*. *International Journal of Networking and Virtual Organisations*. 4(2): p. 109-117.

2

A Gender Critique of Uganda's Rural ICT Access Policy: Opportunities and Challenges

Aramanzan Madanda, Dorothy Okello, Grace Bantebya – Kyomubendo

Rural access to ICT has been highlighted as key in driving development. It is argued that rural access to ICT boosts production, improves household income, reduces inequalities and widens market options. The Uganda government in 2001 put in place a rural ICT access policy named the Rural Communications Development Policy (RCDP) that provided for a Rural Communications Development Fund (RCDF) through which government subsidises communication investment in areas that are considered unprofitable if left to free market forces to promote universal access. Since 2002 support has been provided to establish telephone points, computer training centres and Internet services.

In 2006 a gender analysis of the Rural Communications Development Policy/Fund (RCDP/F) was conducted in 14 districts of Uganda. The objective was to find out the extent to which the initiatives supported under RCDP/F had provided universal access to rural communications by both women and men and whether the implementation process took into account gender considerations. The findings and subsequent follow-ups in two districts show contradictions and discrepancy between policy conception and implementation, pointing to failure to achieve intended objectives of reaching out especially to women. No women's organisation had ever accessed support. Culture, attitudes and gender blind project selection criteria inhibited females' access to funding. The policy and its implementation did not take into account women's and gender needs. The paper suggests a review and re-conceptualisation of the RCDP/F to remove contradictions so that selected projects benefit women and men equitably.

1. Introduction

Rural access to ICT has been highlighted as key in driving development. It is seen as a way of reaching out to those that would be excluded from the developmental benefits of ICT. A Universal Access Fund (UAF) is one mechanism for motivating and mobilizing private investment into rural areas through subsidies and incentives under a free market telecommunications sector so that services are extended to disadvantaged areas and people. The UAF model encourages private telecommunications operators to compete in extending services to poor rural areas as a means of achieving universal access and country-wide market development from a country's own resources [UCC 2005a].

The UAF which is the Rural Communications Development Fund (RCDF) in Uganda has been acclaimed as a best practice model for funding access for poor rural populations in low income countries [UCC 2005a]¹. The RCDF established

¹UAF is also referred to as the universal service fund, the universal service obligation fund or

under the Rural Communications Development Policy (RCDP) 2001 is meant to act as a means of intervention to ensure that basic communication services of acceptable quality are accessible at affordable prices and at reasonable distances by all people in Uganda. The fund leverages investment and is not a comprehensive solution. Uganda's access fund is contributed to by a one percent levy on revenues of telecommunications, post and courier businesses and has been operational since 2002. The fund extends support to all kinds of ICTs and document delivery services into rural areas. The fund is accessed through a formal application process following set guidelines.

While it was envisaged that by mid 2005, every district would have an RCDF supported internet point of presence (PoP) with high-speed wireless connections available to training institutions, cyber cafes, telecentres and private customers, to date this has not been achieved. Despite the fact that the country's approach has been successful in enabling private operators to exceed targets and to meet some of the objectives in need of special funding, penetration of services is low, inefficient and not in tandem with social and economic transactions for sustainable human development [UCC 2005a &b]. Moreover as this paper argues, there are serious gender omissions in the policy design and implementation that run counter to the realization of Universal access to all women, men, boys and girls which require addressing.

There are compelling arguments for a gender perspective in an ICT access policy to address needs of women and men equally. ICT has a potential for mass transformation, but risks excluding women if their needs are ignored when in developing countries most "are in the deepest part of the digital divide further removed from the information age than the men whose poverty they share" [Olatokun 2008:53]. ICTs are catalysts for women's political and social empowerment and their growth a powerful force in advancing gender equality [Ramilo 2006]. If women are to benefit and contribute more to ICT, efforts at policy and other levels must counter the almost global situation where women are largely excluded. Failure to conduct a gender analysis could unintentionally result in missed opportunities for women and risk ICTs reinforcing women's discrimination and disempowerment [Buskens and Webb 2009]. In Uganda, over 80 percent are rural with women constituting more than half of the total population. If ICT is crucial to development, then women must access the tools so that they are not marginalized. A policy that extends ICT to rural areas is important because it targets the majority most of them poor and less educated than their urban counterparts. ICT has to benefit the majority so as to address access as well as the divide that thrives on gender based exclusion.

2. Methodology

A gender analysis was undertaken in 14 districts of Uganda selected because they were

such related name elsewhere.

districts where RCDF supported projects had been implemented by 2006². Due to difficulties in accessing a list of all beneficiary projects and their spread by district and sub county, projects studied were selected purposively or through snowballing basing on the fact that they ever received RCDF support at least once. Thirteen projects were visited which included multi-purpose community telecentres, school based telecentres, internet cafes, ICT training centres and a University. Most were private businesses while some projects were NGO run or located in public schools.

The assessment objective was to find out the extent to which the initiatives supported under RCDP/F had provided universal access to rural communications by women and men, and whether the implementation process took into account gender considerations. The RCDP/F is used as a case to articulate the risks of ignoring gender in ICT policy and to argue for inclusion of women's specific needs in universal access policy. The focus of the assessment was on utilization and uptake of services offered by RCDP/F supported projects by women and men. Specifically, objectives of the RCDP/F, project selection criteria and benefits were analysed from a gender perspective with a view of suggesting recommendations.

The study design was cross sectional. The approach to collecting and analysing data was qualitative. Data collected was both primary and secondary; the former coming from key informant interviews with proprietors and managers of the RCDF supported projects as well as the end user project beneficiaries. In each project case, at least two end-users one female and the other male were selected. Secondary data was obtained from project proposals, beneficiary records as well as contracts. Literature review was done to locate our example within the broader understanding of Universal access. Policy documents were reviewed including the Uganda Rural Communications Development Policy 2001, The Uganda Communications Act 1997, the National ICT Policy Framework, 2003 and Uganda Telecommunications Sector Policy Review, 2005. Resulting qualitative data was analysed manually. The approach involved deriving themes from interviews and extracting examples and arguments. The data from qualitative interviews was triangulated with secondary sources to draw conclusions. The paper is augmented with findings from a doctoral study in two of the districts namely Iganga and Mayuge [Madanda 2009].

3. Policy Design Contradictions

Uganda's ICT policy framework is better understood by examining a range of policies including the National ICT Policy (NICTP), the RCDP and the Telecommunications Act. The policies emphasise a private sector led competitive development model underpinned by a liberalised free market ideology where forces of demand and supply reign. Theoretically, it is within this framework where the invisible hand of the market is paramount that sporadic growth has been attained in Uganda's ICT sector since mid

² The districts were Kampala, Wakiso, Kayunga, Kasese, Bushenyi, Kabale, Kumi, Soroti, Sembabule, Rukungiri, Mbale, Mayuge, Kibale, and Iganga.

1990s. But this model of ICT development is not free from contradictions. First, liberalisation policy is anchored within the theory of modernisation which unrealistically assumes a free market and perfect knowledge of the market. Furthermore, although a market mechanism associated with increased competition and lowering of prices has brought ICT to poor rural women and men [Hafkin 2002b], liberalisation is often plagued by the chronic problem of market failure as there is nothing that is perfect [Preston 1996]. Moreover, free market ideology that pursues development at all costs has been implicated in worsening income inequalities, jeopardising gender equality and compromising human development [Rist 2008; Peet and Hartwick 1999].

It is not that the UCC that formulated the RCDP was ignorant of the weaknesses of a market model for universalising ICTs. An authoritative publication co-authored by a then UCC commissioner indicated that the likely pitfalls of a liberalised approach to ICT were not an afterthought [Tusubira, Kagawa and Ongora 2005]. The report also indicates that when telecommunications sector reform commenced, there was an explicit recognition, right from the start that the kind of liberalised market envisaged by the government policy framework would lead to the marginalisation of the poorer sections of society and that the UCC identified the challenges of accessibility, affordability and utility and put in place the RCDF. In a characteristic technocratic jargon the report adds: “the holistic approach that has been used especially focuses on the majority of the often – marginalised citizens, consisting of rural communities, women, the physically disadvantaged and the youth” [page 168]. What is amazing is that while women are stressed as marginalised alongside youths and the disabled, hence deserving prioritisation, the RCDP document does not mention women or gender at all. This gender blind formulation of ICT policy is in stark contradiction with the overall national law. For instance the 1995 constitution expressly provides for government provision for marginalised groups including women. The National Gender Policy 1997 and 2007 require all government agencies at all levels to design strategies that address gender. Additionally unlike the RCDP/F, the national ICT policy provides for gender mainstreaming in ICTs including using ICT to uplift disadvantaged groups while taking into account gender balance [Republic of Uganda 2003].

In 2002, Uganda Telecom Limited (UTL) and Mobile Telephone Network (MTN), the then two national telecom operators (NTOs) declared 154 sub counties spread over 35 districts as “uncovered.” The NTOs³ had identified these areas as economically unviable – making them eligible for RCDF⁴. To facilitate the process of universal access, a policy was developed. The RCDP spells out subsidies for

3 Uganda currently (2009) has 5 telecom operators namely CELTEL (taken over by ZAIN), MTN, UTL, WARID and Orange, having ended the duopoly regime in 2005 and licensed new operators.

4 Some of the RCDF supported projects were not necessarily located in sub counties declared as uncovered by UTL and MTN. Indeed some are set up in gazetted urbanities thereby bringing to the fore the contested notion of what is rural or urban.

communication services in the rural areas⁵. The policy covers Internet access provision, information technology (IT) content development, postal services and ICT training. By mid 2005, each district in Uganda was expected to have an Internet point of presence (PoP) with dedicated high speed wireless connections available to schools and other institutions, cyber cafes, telecentres and private customers to make use of within a certain radius of the district centres [UCC 2005a]. As additional incentives the UCC allowed for tariff flexibility where operators can charge up to 50% more than prevailing urban rates in rural areas [UCC 2005a]. This policy stance takes no explicit cognisance of the fact that the rural areas constitute the poor the majority of whom are women. In UA strategies, there is no direct provision for addressing the gender gaps in ICT access. In effect universal access is not universal. Indeed this is explicated by the fact that the areas that have been most difficult to reach are the northern and north - eastern districts which have experienced a two decade plus civil war⁶ and are inhabited by the poorest sections of the people of Uganda most of them women. In the rural areas, however, a challenge still exists due to the cultural life style that keeps females occupied in domestic work, limiting the availability of time for using any ICT other than voice telephony [Tusubira, Kaggwa and Ongora 2005]. From the foregoing, it is clear that a policy like the RCDP that aims to increase access to the majority especially in the rural areas is a smart thing to do. The question of whether or not the current universal access policy framework promotes universal access is the point of contention given that it is prone to problems of excluding the poor especially women.

Further policy contradictions can be unravelled by examining the policy objectives. According to the UCC [2005a], the projects that could justify UA funding need to meet the following criteria: the service targets meet government Universal Access (UA) objectives as “socially and politically desirable”; the overall service supply profile possesses long- term viability (that is, it is sustainable); but in the short-term, the project may be considered financially marginal (that is offering lower rate of return than that needed by private companies), or risky considering the investment required. The specific objectives of the RCDP are to: ensure that all sub counties with at least 5,000 inhabitants have access to basic communication services by 2005; ensure that RCDF resources are effectively utilised to create immediate impact; support establishment of an Internet PoP in every district of Uganda by 2003, where each PoP was to be associated with at least one Internet cyber café; increase the use of ICTs in Uganda, by supporting introduction of ICT use in at least one “vanguard” institution in every district of Uganda, by 2003; promote provision of communication services in rural areas as a profitable business.

5 The RCDP does not explicitly define Rural. It appears rural means any area outside the city Kampala [Parkinson 2005].

6 These areas have since 1986 been under civil war with the Lord's Resistance Army that has negatively impacted any form of development. A peace agreement that created a semblance of peace has not been signed to date

It has to be noted that any effort aimed at increasing access and bringing ICT nearer to the rural people has the potential of reaching the poor and women. However focusing on the sub county headquarters or sticking on a criterion of 5,000 people leaves out many women who are tied up in domestic responsibilities and have no time to travel to the sub county headquarters or the district to access ICTs. Second, to promote provision of ICT services in the rural areas is a good but limited objective in that as is even admitted by the very proponents of the policy, the hard to reach areas are unlikely to be profitable. There will be need to focus on the social benefits rather than profits to reach the majority of the poor especially women. Thus provision of communication services as a profitable business and from a purely economic perspective is limiting.

4. Getting To Know: The Exclusive Information Flow

For potential applicants to know about the UA funding opportunities, they had to be informed. Field data indicated that applicants got information about RCDF in various ways: call for proposals through newspaper adverts; informal contacts with RCDF officials in conferences, seminars and public interactions; through researchers and volunteers [including foreign researchers and students] some of who wrote successful proposals for especially rural districts; insider information from former employees who had prior information about upcoming opportunities; and information from Internet service providers some of who helped to prepare the applications. As expressed by many respondents, there was a feeling that information about the subsidy was insufficiently disseminated and many applicants did not know what exactly constituted the support from UCC. Noteworthy is that while the use of newspapers was commendable for its wider reach, it is clear that such a channel was limited for publicity. It is unlikely that rural based organisations and especially the bulk of community based women's organisations could be informed. Moreover, informal contacts akin to the old boys or old girls' networks came into play in disseminating information. Such mechanisms exclude those who are powerless yet these are supposed to be the target. There is need to specifically target the poor and especially NGOs, CBOs and women's organisations in spreading information. One way is to take advantage of existing women's networks⁷.

Furthermore, following a purely market model in operating the RCDF is inappropriate. First, the existence of the fund as precisely for reaching out to those excluded by the pure market system is in itself testimony of failure of the market mechanism to equitably spread ICT. For instance as noted already the reason for setting up the RCDF was a recognition that telecommunications sector reform with a liberalised market as envisaged by government policy would lead to marginalisation of the poorer sections of society including women, youths and the

⁷ Uganda's civil society organizations have women's network organizations that could be crucial in this endeavour such as the National Association of Women Organizations in Uganda (NAWOU), the Uganda Women's Network (UWONET) and the ICT savvy Uganda Women's Network (WOUNET) that could have been utilized as entry points to reach out to women and girls.

disabled [Tusubira, Kaggwa and Ongora 2005]. The policy implementers needed to live up to this realisation. Therefore, the spread of information needed to follow a model that plugs up the loopholes in the market mechanism. Thus developing partnerships would be more suitable than just asking applicants to “compete” on a criteria that generally favours profit oriented companies. Such a model mostly favours private businesses. It is not surprising these are the biggest beneficiaries with some reporting multiple subsidies while not for profit NGOs and women’s organisations report failure in keeping pace or failure in accessing any subsidy.

Beyond project applicants, we also tracked how the end-users got to know about the RCDF supported projects. From their responses it is clear that the RCDF secretariat’s role in information dissemination was limited. Information moved through friends, relatives and the church, or through a school staff room or general school assembly. If the mode of transmission of information was not informal as through a relative, it was exclusive like through a staffroom for teachers or a church for believers in a certain faith. This point leads to one critical direction. There was need to strengthen publicity through various media channels so that possible beneficiaries could be reached as much as possible.

5. Skewed Selection Criteria And How Women Fell Out

The UCC is mandated by the Communications Act to establish and administer a rural communications development policy and fund [UCC 2001]. UCC’s policy on rural communications development is intended to supplement government efforts by directing development and delivery of communication services in the rural areas aimed at bringing about socio – economic development. Through a competitive tendering process, subsidies are provided to the winning bidders. The kinds of projects financed include: public payphones, ICT training, internet cafes, Internet PoPs, IT content development, telecentres, telephone infrastructure network and ICT research.

The subsidy amounts have an upper ceiling. For internet cafés and ICT training centres, the maximum subsidy is 40% of investment costs up to a maximum of \$6,000. Public access payphones have a maximum subsidy of 50% of the investment costs up to a maximum of \$400 per access site. Upon approval and meeting requirements, which include obtaining a bank or insurance guarantee, the subsidy is disbursed in the ratio of 4:3:3. Forty percent (40%) of the subsidy is disbursed upon signing the agreement; 30% after 50% of the project has been implemented according to an agreed schedule and 30% upon certification by UCC that the project is completed and operational. This implies that there must be co-financing.

The project selection criteria included: presentation of a certified copy of a certificate of registration; presentation of a business plan indicating the viability and sustainability of the project for five years; adequate professional know-how and operational experience; ability to implement the project within a stipulated timeframe; and availability of counterpart funding. Other considerations were relevant experience of the firm, designated professional and technical personnel,

marketing strategy and ability to meet capital requirements

Gender sensitivity was not a requirement. Although two NGOs and one private company had utilised RCDF funds to support gender objectives out of their own volition, all interviewees reported that no requirement for gender was demanded of them during application or implementation. Furthermore, while the criteria for selection were meritorious, the provisions were exclusive. For instance the requirement for experience and counterpart funding excluded the poor and women's organisations that lack experience or counter funds in favour of established male owned ICT firms. Many beneficiaries thought that the financial guarantee conditionality was too high. Some failed to reapply because of the requirement for counterpart funding. This problem was more pronounced for organisations whose objectives contribute to social development rather than business namely NGOs and CBOs. Thus while the mode of operation and guidelines favoured private businesses; they disadvantaged organisations that are not for profit. An inventory of CSOs in Uganda clearly indicates that organisations in this category, many of which have no formal registrations are women's organisations.

During fieldwork we encountered innovative private organisations that used the affordable option of using an insurance company to guarantee their subsidies. But even then, these were private businesses. There seemed to have been no avenue of sharing such information with organisations that had difficulty in finding counterpart funding or financial guarantees. A more proactive action to provide such information to clients and potential applicants would have helped. Some beneficiaries had the perception that although the RCDF had earmarked funds for projects, the fact that it was a subsidy rather than a grant made it very hard to access. There is need to explore the possibility of making the RCDF a grant since it is not expected to be refunded.

Furthermore, the requirement for sound financial plans locked out some applicants. It is common knowledge that although useful, an elaborate financial plan does not always guarantee project success. Neither is beginning with a simplified application procedure a recipe for failure. Some funding agencies have long realised these and find other ways of guaranteeing success such as effective implementation and monitoring systems. Women's organisations in rural areas are particularly constrained by this fact owing to high levels of illiteracy and lack of access to such information that would help them design elaborate implementation plans. To include them therefore requires some more innovative, flexible and inclusive approaches.

6. Through The Filter: The Nature Of Projects Supported

Projects and businesses supported by RCDF often conducted more activities than were directly supported by the fund. Major activities included: Internet and email services; computer typesetting and printing; scanning and type writing (sometimes with manual typewriters); computer training (introduction, Microsoft applications like word, excel, access and PowerPoint), as well as accounting packages (such as sage, tally and quick

books). Others included television viewing (especially for European soccer); Video shows especially educational films and entertainment; and photocopying. Internet and email services were offered by most projects but in some, especially in rural areas, the Internet links were reportedly unstable. Training in typewriting was offered because potential employers of the graduates, mainly schools still used typewriters.

The key objectives of the projects visited varied depending on the nature of their focus. Projects based in rural communities, aimed at social and economic development of the area through use of ICTs. While those in towns aimed at availing ICT services for profit. School based projects aimed at providing students, teachers and in some cases surrounding communities with access to ICT services and application of ICTs in teaching for enhanced learning. Apart from the Mayuge district based NGO, Busoga Rural Open Source and Development Initiative (BROSDI) that had explicit gender objectives in its activities and services, other organisations often had gender blind objectives although SchoolNet Uganda and the Kumi district based Arrow Centre (for profit company) provided specific services aimed at women and girls. It is likely that benefits to females would have been higher if there were specific gender requirements in the funding criteria.

7. Surprise: Women & Girls As Majority Beneficiaries

It has been argued that exclusion of gender in ICT policies is often the norm rather than an exception and that even when policies are developed the real test is the process of implementation and commitment [Jorge 2006]. Even when Uganda's RCDP does not recognise gender and supported projects are not obliged to address gender objectives, in many field examples, women constituted the majority of the final beneficiaries.

In tracking benefits, we inquired about the final project beneficiaries. There were interesting scenarios where females were majority beneficiaries despite the gender blind way in which most projects were designed. Records at many training centres showed that women were the majority trainees. The projects in general targeted a broad range of people that included: civil servants, teachers, community members, cultural leaders, NGOs, health workers, religious leaders and school drop-outs. Others included primary school pupils, women farmers, secondary school and university students, businesspeople, local government employees and tourists.

In Mayuge district, we visited BROSDI an NGO whose aim is to stimulate socio-economic development using open source particularly Free and Open Source Software (FOSS), knowledge sharing and information management. The organisation used computers and the Internet to address poverty through information generation and exchange. ICT was its core project with three other subsidiary projects in education, agriculture and health. BROSDI's education project targeted teachers and school children. The agricultural project, which involved collecting and exchanging local agricultural content targeted women farmers. The health project that had an HIV/AIDS component targeted the youth. The Manager indicated that the project had helped to increase computer awareness. While RCDF funded some ICT inputs,

financial support for the sub projects was obtained from other agencies. BROSDI had got RCDF support thrice: first, to support acquisition of computers; second, to set up an ICT centre and third to acquire an electricity power solution to address the electricity shortage. In this example it was clear that by designing programmes to include women as the target beneficiaries, BROSDI extended its ICT services to women and girls even when it could not be described as a women's organisation.

Another example was SchoolNet which is an education oriented NGO. Some of SchoolNet's RCDF funded projects were implemented in girls' schools. One example was Bweranyangi Girls Secondary School in Western Uganda. As a beneficiary girls' school, female students were primary targets alongside their male and female teachers. The fact that it was a girls' only school constituted one of the criteria of its selection by SchoolNet Uganda. Similarly, SchoolNet chose a boys' only school in the same region namely Kitunga High school. However this was SchoolNet's internal gender sensitive programme implementation policy rather than an RCDF strategy. Based on SchoolNet's initiative, the fund was used to facilitate ICT centres targeting both women and men.

Apart from NGO led initiatives, women also benefited from private businesses. From the Manager Ruhinda ICT centre Bushenyi district, women and girls were part of the beneficiaries just as men. In some cases, women trainees outnumbered men especially senior six leavers and teachers from nearby schools. At the time of the assessment, two females trained at Ruhinda had obtained jobs as secretaries in secondary schools after training in basic computer skills and typewriting. But this was not out of an RCDF special strategy to address women's needs.

In Mayuge district in Eastern Uganda, another private company, Batud ICT training centre, had 42 people trained in introductory computer skills and Microsoft Office applications namely word, excel and PowerPoint of whom 18 were female. Though there was no special treatment to women and girls in particular, females formed a large percentage of the end user beneficiaries. Batud also reported a discounted computer training programme for schools, where both girls and boys were targeted.

There were other examples from Eastern Uganda too. Women and girls were the majority of training clientele at the Tororo branch of the Kampala International Management Institute. Many accessed the centre for skills that could help them get secretarial employment. Men accessed the centre mainly for internet surfing. At Telecom Equator located in Soroti town, women and girls benefited by accessing the Internet and e-mailing. Students used the cafe for research, and the ratio of men to women was estimated at 4 to 6 implying that females were the majority. Students and nurses acquired skills to better their job opportunities. There was an example of a woman who accessed computer services and was trained in Kiswahili language. She was able to read and send her son email in her local language. At the Mbale branch of Nile Computers Limited, which was one of the major beneficiaries of RCDF, it was reported that men used Internet services most though records showed that more women than men had been trained in computer skills though the majority

were unmarried and students.

In Wakiso district of central Uganda, women and girls were reportedly beneficiaries of the Bugema University project though the centre had no gender disaggregated records to show usage by gender. The RCDF supported project at the university targeted all students, teachers and community members. Women had benefited as part of the student, teachers and surrounding community in accessing the ICT services though there were no gender specific implementation guidelines. At Mubende Light Secondary School ICT Centre, in Mubende district, 150 community members had trained of which 80 were women. Trained students were 530, of which 260 were girls. The centre also had a public pay phone, an online electronic library as well as CDs with subject content. Staff had been trained in MS word and access. Mark sheets were computerised and all teachers had e-mail addresses. Ten female and 20 male teachers were trained in basic computer skills and had access to computers. Other contributors were Uganda Industrial Research Institute and the International Institute for Communication and Development (IICD), which gave the school computers, software, initial training and further training in how to manage the centre. Two of the six ICT centre committee members were women. The ICT help desk officer in the school was a woman and girls were encouraged to use ICTs.

To conclude this section two points have to be made. One is that even when the policy is gender blind, it is possible to have a gender inclusive implementation at organisational level largely mediated by gender sensitive actors. This however should not overshadow the overwhelming evidence that gender blind policies often produce gender blind results. Secondly, the predominance of women in basic computer training is the norm in early computer adoption processes. A study in Mayuge and Iganga found that women were the earliest computer technology adopters because of the feminisation of the secretarial profession where they constitute the majority but soon gave way to men as the technology assumed more complex uses [Madanda 2009]. The real challenge therefore is for RCDF or similar policies to ensure sustained universal ICT access to all gender categories.

8. Impact of Rcdf Supported Projects

The impact of the RCDF is difficult to isolate as the changes resulted from multiple initiatives. However, it is possible to pinpoint associated changes including the following:

Access to cheaper ICT services was possible within shorter distances by many including students, teachers and the general public even in remote areas for business, social and educational purposes. Affordable services within shorter distances were especially beneficial to females because many women and girls due to social cultural constraints make restricted travels. Women were reported to be benefiting from internet services by communicating with friends and family members

There was also direct employment creation in RCDF supported initiatives for

ICT trainers and attendants the majority of whom were men. There were also spin-offs reported. For example, businesses had been started by qualifying trainees some for secretarial services, designing posters, document scanning and photocopying. Trained women had got jobs as secretaries, trainers and ICT centre attendants. Men had improved their careers some as teachers, head teachers or administrators.

One area where RCDF ICT initiatives had made a contribution that was potentially very useful was in education. Through supporting the establishment of ICT centres, there had been the exposure of students and teachers to various ICTs in both girls' and boys' schools, which had enhanced learning through Internet and multimedia. Beyond secondary schools, in both Tororo and Soroti districts, students of commercial courses were getting trained from RCDF supported centres. Female secretarial studies students supplemented their skills from the ICT centres as there was no computer training in some of their colleges.

One observation to make was that women had taken lead in responding to computer services. Through the computer skills acquired, many had benefited by getting employment. It has to be noted however, that most of the centres were engaged in secretarial and elementary skills where women are often the majority. Unless there is specific effort, it is likely as is with the management and ownership of these centres that women may be edged out as skills acquisition and ICT use complexity advances.

9. Gender Specific Challenges

Women were constrained by time. According to a key actor of an NGO that targets women "you have to consider household chores. For market days, women have to be in the market. You have to target your activities on non-market days and in the early afternoon like from 2.00pm to 4.00 pm. After four o'clock, women have to go back home, collect firewood and prepare evening meals except if a woman is a widow."

In many instances culture discourages women from taking part in public affairs. For example in Mayuge district we learnt that "even if you invited women only, men come in and want to take over..." Culture underpins control of women by their spouses to the extent that some men are described as "very possessive of their wives" making it hard for them to take part in development programmes generally including ICT initiatives. Conditions in the family reportedly affected women more. For instance in Soroti, an example was given where a woman stopped computer training lessons when her husband was transferred from the district. Generally, women suffer cultural confines which "you have to break to get them into a project" said a project leader. Women wanted to first seek permission from their husbands. There are also financial control problems when resources begin to come in as men want to take charge.

Social roles that girls play in homes demand that they are at home early as they may be required to contribute towards household chores like collecting firewood, fetching water and cooking. Orphaned girls were particularly disadvantaged as they were the ones mostly called upon to do such chores. Girls said they feared staying

out late as this exposed them to rape, sexual assault and the resulting dangers of contracting HIV/AIDS. As such courses scheduled for late evening hours were not recommended for girls because they could not return home late and for married women their husbands could not allow them. Men however were not similarly constrained.

10. Conclusion

Overall, it is important to note that the primary reason for establishing the RCDF was because of the failure of the market based system to reach everyone particularly the marginalized and the very poor sections of society. The point of failure of market oriented systems, their benefits notwithstanding, is well known. Simply put, it is not possible to reach everyone based on the market mechanism alone. In stark contradiction of this well known fact, the criteria for supporting projects was based on the market based assumptions tilted in favour of private experienced male dominated business enterprises against not for profit organizations generally but women's organizations particularly as most of them are focused on social returns or based in the informal sector where elaborate business plans and corporate culture procurement procedures are not the norm. If the fund is to reach the poor, it cannot simply utilize the lenses provided by the market system alone. This alone cannot deliver to the marginalized without a certain mechanism that directs benefits in that direction – that system involves utilizing “gender lenses.”

Neither the RCDP 2001 nor the project selection criteria had any specific gender considerations. Resultantly supported projects did not necessarily address gender concerns. Beneficiary agencies were mainly profit motivated private businesses with a few educational and development oriented NGOs. No agencies that directly accessed funding could be classified as women's organizations. RCDF's lack of a gender sensitive implementation strategy had left out many would be potential women beneficiaries especially as managers and proprietors of funded enterprises particularly for computer training and Internet facilities.

Also the RCDF funding strategy of giving more “commercially viable” projects bigger subsidies as compared to less commercially viable projects in rural areas was a prudent strategy. But simultaneously this left out many rural projects. Henceforth the fund was yet to reach and benefit most of the rural population. If not revised, more will be excluded. As noted by Gurumurthy [2006:615] “the fundamental issue in reaching poor women is not one of profitability of models, but the creation of a set of technology-mediated services and products that allow women to be part of the emerging opportunities.”

It is also important that dissemination of information regarding the RCDP/F be improved. Apart from newspaper adverts and male oriented informal channels, there is need to broaden methods of information dissemination. In particular women and the marginalised populations based in the rural areas that the fund targets cannot all be reached by newspapers. There is need to consider advertising and creating awareness using multiple channels such as radio, the internet, mailing lists, posters, brochures and educational agencies. There is also need to share experiences of

already supported projects regarding the benefits, challenges and lessons learned.

Ultimately the RCDP needs to be reviewed so that specific gender sensitive objectives, strategies and targets are included. The review process should be consultative taking into account effective participation of key stakeholders including the public, private sector, academia, women organisations and civil society. Selection criteria need to be amended to include provisions for gender equality.

References

- BUSKENS I. AND A. WEBB 2009. Introduction. In *African Women and ICTs: Investigating Technology, Gender and Empowerment*, I. BUSKENS AND A. WEBB Eds. Zed Books, London 1-8.
- GURUMURTHY A. 2006. Promoting Gender Equality? Some Development Related Uses of ICTs by Women. *Development Practice* 16, 611-616.
- HAFKIN, N. 2002a. Are ICTs Gender Neutral? A Gender Analysis of Six Case Studies of Multi-Donor ICT Projects; UN/INSTRAW Virtual Seminar Series on Gender and ICTs; Seminar One: Are ICTs Gender Neutral? 1-12 July 2002
- HAFKIN, N. 2002b. Gender Issues in ICT Policy in Developing Countries: An Overview, United Nations Division for Advancement of Women (DAW) Expert Group Meeting, Seoul South Korea 11/-14 Nov. 2002
- JORGE, N. S. 2006. Engendering ICT Policy and Regulation: Prioritising Universal Access for Women's Empowerment. In *Cinderella or Cyberella: Empowering Women in the Knowledge Society*, N. J. HAFKIN AND S. HUYER, Eds. Kumarian Press, Bloomfield, CT, USA, 15 - 47
- KARAMAGI, A.E. 2006. Towards Improving Farmer's Livelihoods Through Exchange of Local Agricultural Content in Rural Uganda. *Knowledge Management for Development Journal* 2, 68-77.
- KASUSSE, M. 2005. Bridging the Digital Divide in Sub Saharan Africa: The Rural Challenge in Uganda. *The International Information and Library Review* 37, 147-158.
- KAWOOYA, D. 2004. Universal Access to ICT and Lifelong Learning: Uganda's Experience. *New Library World* 105, 423-428.
- MADANDA, A. 2006, ICT Liberalisation and Changing Gender Relations in Contemporary Uganda: A Research Agenda. In *Advances in Systems Modelling and ICT Applications*, J. KIZZA, J. AISBETT, A. VINCE AND T. WANYAMA Eds. Fountain Publishers, Kampala 362-374
- MADANDA, A. 2009. *Gender Relations and ICT Adoption in Uganda: A Case of Computers and Mobile Telephones*, Draft PhD Thesis, Makerere University, Kampala.
- MADANDA, A., C. KABONESA AND G. BANTEBYA-KYOMUHENDO. 2007, "Challenges to Women's Empowerment through ICTs: the Case of Makerere University" *Agenda Journal*; Issue No 70
- OLATOKUN, W. M. 2008. Gender and National ICT Policy in Africa: Issues, Strategies and Policy Options, *Information Development*, 24, 53-65 at <http://idv.sagepub.com/cgi/content/refs/24/1/53>, accessed 7/1/2009
- PARKINSON, S. 2005. *Telecentres, Access and Development: Experiences and Lessons from Uganda and South Africa*, Fountain Publishers, Kampala

- PEET, R. AND E. HARTWICK. 1999. *Theories of Development*, Guilford Press, New York and London
- PRESTON, P.W. 1996. *Development Theory: An Introduction*, Blackwell, Oxford and Cambridge.
- RAMILO, C. G. 2006. Beyond Tools: Technology as a Feminist Agenda. *Society for International Development* 49, 68-71
- REPUBLIC OF UGANDA. 1995. *Constitution of the Republic of Uganda*, Entebbe, Government Printer
- REPUBLIC OF UGANDA. 2003. *National Information and Communication Technology Policy, Draft*; Kampala: The Ministry of Works, Housing and Communications, the President's Office and The National Council for Science and Technology.
- RIST, G. 2008. *The History of Development: From Western Origins to Global Faith*; 3rd Edition, Zed Books, London and New York.
- SSEWANYANA J.K. 2007. ICT Access and Poverty in Uganda, *International Journal of Computing and ICT Research* 1, 10-19.
- TUSUBIRA, F.F., KAGGWA I. AND ONGORA J. 2005. "Uganda." In *Towards an e-Index: Household and Individual ICT Access and Usage Across 10 African Countries*, GILLWARD ALISON Ed. The Link Centre, Wits University School of Public and Development Management
- UCC. 2001. *Rural Communications Development Policy for Uganda*, Kampala
- UCC. 2005a. *Funding and Implementing Universal Access: Innovation and Experiences from Uganda*, Fountain Publishers, Kampala.
- UCC. 2005b. *Uganda Telecommunications Sector Policy Review: Recommendations on Proposed Review of the Telecommunications Sector Policy*, Review Report 28/1/05.

3

Pedagogical and Conceptual Design of An E-learning Environment for HIV/AIDS Education

Bada Joseph Kizito, Jarkko Suhonen

In this paper, we present the pedagogical and conceptual design of an e-learning environment, NetAIDS, for HIV/AIDS education in Uganda. We also identify the first concrete design solutions for the NetAIDS environment, which will be later implemented and evaluated. Our focus is to analyse the first steps towards the design of the NetAIDS environment following a development research methodology. Four important design aspects for the NetAIDS environment are identified. First, the design of the environment should be based on sound pedagogical principles. In our case, we suggest a constructivist approach. Second, the design and development of technological solutions is based on the development research method for generation of a prototypical e-learning environment, which uses scaffolds and new web technologies (e.g. web 2.0.) to support Ugandan youth to learn about HIV/AIDS. Thirdly, we present a framework for designing NetAIDS environment in the context of Ugandan schools. Finally, a formative evaluation scheme is presented to provide early feedback from the users so that the designed NetAIDS environment can be modified to suit better suit the needs and requirements of the users. The conceptual and pedagogical design principles for HIV/AIDS e-learning environment can also be applied in other contexts.

1. Introduction

ICT-based HIV/AIDS education and counseling services have been developed in various African countries, such as Uganda and Tanzania [Bloome 2001; Duveskoq 2008]. The existing e-learning approaches for HIV/AIDS education provide limited educational services to children with most solutions presenting static online content, carton-based stories and social networks using email as a means of communication [Bada and Suhonen 2008]. In this paper, we present a conceptual and pedagogical design of NetAIDS e-learning environment, which is aimed to provide a fresh perspective to the HIV/AIDS education and counseling services for Uganda youth. The pedagogical and conceptual design of the environment consists of three main components namely the pedagogical design component, NetAIDS system design and cognition design. Computers are nowadays commonly used in Uganda. There is a need to design novel e-learning solutions for HIV/AIDS education to impart the knowledge of the epidemic to the children so that they can protect themselves against the deadly virus. The contemporary Web2.0 tools and social media offer a lot of opportunities for building motivating and inspiring e-learning environments in which children have a variety of approaches to learn new things using different media such as text, graphics, audio, video and animation. The present paper is a first concrete step towards utilizing the possibilities of novel ICT solutions in Ugandan schools. The paper is constructed

as follows. In Chapter 2, we present the pedagogical perspectives and requirements for the design of the NetAIDS environment. Chapter three includes a conceptual design of NetAIDS environment including three main components human-computer interaction, digital resources as learning objects and technological requirements for the implementation of the environment. In Chapter 4, we discuss about the future work including implementations and formative evaluation plans. Finally, in Chapter 5 we provide a summary and conclusion for the paper.

2. Pedagogical Design Of NetAids Environment

One of the most fundamental aspects of any ICT-based educational environment is the theoretical perspective behind the design solutions to support the learning processes of learners. Different learning theories establish a philosophical and conceptual framework for learning. According to Gredler [1997], learning theories can work as guidelines for planning educational support for learners, like e-learning environments.

We choose a constructivist approach for the NetAIDS environment for two major reasons: First of all, the children in high schools today are growing up in an environment which is affected by AIDS epidemic, they can share ideas on the effects of AIDS on society using online social networks and they can construct stories to educate fellow children. Secondly, children can learn better from their fellow children as compared to passively receiving knowledge from teachers without the active participation of the children.

2.1 Constructivism as a theoretical foundation to the pedagogical design of NetAIDS

Hadjerrouit [2005] affirmed that constructivism has its roots in the constructivist philosophy. Its central figures are Bruner [1999] and Piaget [1969]. The defining characteristic of constructivism is that knowledge can not be transmitted from the teacher to the learner, but it is an active process of construction. A pedagogy that relies on the constructivist philosophy requires a set of pedagogical guidelines and strategies that can be translated into practice:

- **Construction:** Knowledge is actively constructed by the learners through their interactions with the environment, not passively transmitted by teachers [Booth, 2001]. Knowledge is constructed by using the learner's prior knowledge as foundation. Teachers serve primarily as collaborators, guides and facilitators of learning, not as transmitters of knowledge [Crowther, 1997].

Relevance for NetAIDS

Children are able to construct poems and drama for AIDS education. The educational content for AIDS prevention can be accessed by other children via an e-learning environment.

- **Cognitive skills:** In order to be useful for problem solving, knowledge needs to be related to each other. The process of constructing interrelated knowledge structures

requires cognitive skills that learners do not possess, such as analysis and reasoning skills, analogical thinking, reflection, and self-evaluation. Thus in order to scaffold their performance, instructors should identify problem-solving skills that are specific to the subject matter. The development of a personal growth requires learners to think about their knowledge, understanding, ideas, and beliefs about learning [Avraamidou, Zembal-Saul 2002].

Relevance for NetAIDS

It is necessary to introduce children to basic AIDS education with an assessment either in form of questions or puzzle games to test their understanding. The assessment in NetAIDS environment is based on computer games which children can perform to test their understanding of basic AIDS knowledge.

- **Authentic Tasks:** To get students actively involved in knowledge construction, learning activities should focus on realistic, intrinsically motivating problems that are situated in real world tasks [Wilson, 1997]. Rather than applying knowledge to solve abstract problems, knowledge must be constructed in real and useful contexts.

Relevance for NetAIDS

Since the teenagers in Uganda today are growing up in societies devastated by AIDS epidemic, they can be in position to construct AIDS prevention knowledge based on real world experiences.

- **Related cases:** Learners should have access to a set of related experiences and cases from previous learners that a student can draw on to represent their deficient knowledge. Analogical reasoning is the key skill of reusing related cases. It includes a search for similarities and differences between the related cases and the new problem to be solved [Hadjerrouit, 2005].

Relevance for NetAIDS

Debates and drama are common practices in high schools, so the students have skills to organize AIDS preventive education in form of drama and poems from the experiences they have in school drama and debates.

- **Collaboration:** Learning occurs not in isolation but by means of people working together, and as they exercise, verify, and test their knowledge through discussion, dialogue, and playing computer games. Hence learning should take place in an environment that supports collaboration, social negotiation and interaction, because as a learner gains experience in social situation, this experience may verify a learner's knowledge constructions [Bolhuis, 2003; Duffy and Cunningham, 2001].

Relevance for NetAIDS

This practice of collaboration for learning purpose is well facilitated by web 2.0 technologies being proposed for HIV/AIDS education and counseling services.

- **Information Technology:** The design of a constructivist learning environment goes far beyond the computer material itself, but well-designed web-based technology may facilitate constructivist learning when it provides learners the information they

need.

Relevance for NetAIDS

The vision here is to use web 2.0 tools for students to exchange messages in group discussion on issues to do with fight against AIDS epidemic. Students also need to evaluate their own knowledge by playing an online computer game on HIV/AIDS prevention education.

2.2 Other pedagogical requirements

We can also identify practical requirements for the pedagogical design of the NetAIDS environment. First, the children should contribute and share information freely on preventive actions, they should freely communicate among themselves on how to cope up with the trauma AIDS might have caused in affected families. Second, a variety of approaches should be used for communicating preventive information to the children using different media, e.g. instruction as online text and graphics, multimedia lessons recorded from HIV/AIDS instructors and counselors, real world stories constructed by children to educate fellow children and an automated online database for provision of counseling services by fellow students and teachers or AIDS counselors. Thirdly, the NetAIDS environment should be designed to support learning principles suitable for the main audience of the environment, e.g. teenagers and online teacher's tasks. The following pedagogical design aspects are also taken into consideration during the design and implementation of the NetAIDS environment [Vrasidas 2004]:

- Planning and identifying goals and objectives,
- Standards and content for the course,
- Conducting learner and audience analysis,
- Identifying technology requirements,
- Reviewing other similar courses,
- Facilitating the content analysis,
- Reviewing samples of evaluation activities to match objectives and content,
- Examining templates for syllabus design for a variety of levels and selecting the appropriate activities and;
- Choosing the right media attributes to support the objectives of the learning experience.

2.3 Use of scaffolds in the NetAIDS environment

Scaffolding is one of the concrete educational construct based on the constructivist learning theories. The concept of scaffolding is closely linked to the idea of the zone of proximal development [Gredler, 1997; Vygotsky, 1978]. Scaffolding refers to the change of level of support. When a student is learning something for the first time, a skilled person may use direct instruction, but as the student's competence continues to grow the amount of guidance given becomes less. Scaffolding is in most cases used to support students to reach the upper limits of their zone of proximal development.

We have analysed some of the existing web technologies that can be used to scaffold children in HIV/AIDS education. We eventually identified design principles for using scaffolds in the NetAIDS environment.

The evaluation approach we have presented in Table 1 describes how effective the NetAIDS environment achieves the objectives of learning and how learners benefit from it in terms of quality of instruction, usability issues and motivation to use NetAIDS.

Table 1: Learning principles and theories for designing and evaluating content for AIDS Education

Learning Principles	Scaffold design in NetAIDS	Evaluation approach
<ol style="list-style-type: none"> 1. Cognitive development has limitation to a certain range at any given age. 2. Social interaction is the foundation for full cognitive development. 	<p>Provision of online chatting services for children to ask questions to teachers and HIV/AIDS counselors so that they get replies to challenging issues they experience in real life that can lead them in getting HIV/AIDS infection.</p>	<ol style="list-style-type: none"> 1. The type of instructor's response and the time to respond when the student has a question to ask or is experiencing difficulty. 2. Effectiveness of the variety of support in the forms explanation, example, question, summary, suggestion or encouragement
<ol style="list-style-type: none"> 1. Knowledge needs to be presented in an authentic context, i.e., settings and applications that would normally involve that knowledge. 2. Learning requires social interaction and collaboration. 	<p>Online participation by children through demonstrations that educate fellow children to keep away from temptations that leads to HIV/AIDS infection.</p>	<p>Video content analysis. Evaluation of video education in form of interview to children. Questionnaires to assess the learners' knowledge of the interactive stories.</p>
<ol style="list-style-type: none"> 1. The instruction should address the experiences and contexts that motivates the students to learn (readiness), 2. The instruction should be well organized so that the student can learn from it with minimum difficulty 3. The instruction should be designed in such a way that a student goes beyond the given information to come up with new discoveries. 	<p>Children should build stories constructed from their real world experiences. These stories can educate other children. Introduction of online educational resource for access by children (youth) for awareness creation and to impart HIV/AIDS knowledge in them.</p>	<p>Children can be interviewed to find out the points they get from stories in HIV/AIDS preventive education and counseling support services. Provision of computer games for NetAIDS education to evaluate children's basic knowledge in HIV/AIDS education.</p>

3. Conceptual Design Of NetAids

3.1. A conceptual design framework for NetAIDS

The proposed conceptual design framework (see Figure 1) for the NetAIDS environment in the Ugandan school context has three main components namely: (1) the *human computer interface* that describes the physical interaction between the user (high school student/teacher) and the computer, (2) the digital resource which is the actual educational material and the content generated as a result of interactive activities between the learners and the instructors, and (3) the technological requirements for implementation of the NetAIDS environment. The purpose of the conceptual design framework is to provide a comprehensive picture of the socio-technical innovations for supporting HIV/AIDS e-learning environment. The framework can be used to design NetAIDS learning activities, such as creation of locally relevant educational material in HIV/AIDS, sharing of experiences and best practices in HIV/AIDS prevention, collaborating and interacting with peers nationally and internationally and supporting the actions of various actors in changing roles (e.g., student, teacher, facilitator, learning coach, human resource or education manager).

3.1.1 The human computer interface

The human computer interface depicts the interaction between the computer and the user (student/teacher). A user interacts with the computer in the NetAIDS environment when accessing instructional materials for AIDS basic knowledge, playing NetAIDS game, interacting with the social network consisting of peer learners and teachers, exchanging text messages with other users in NetAIDS environment, and accessing multimedia educational content and counseling services.

3.1.2 Digital resources

An important aspect of the conceptual design framework for the NetAIDS environment is the provision of educational content and support services in the form of learning objects and learning tools. A learning object can be defined as “*a resource, usually digital and web-based, that can be used and re-used to support learning. Learning objects offer a new conceptualization of the learning process: rather than the traditional “several hour chunk”, they provide smaller, self-contained, re-usable units of learning*” [Jenny 2006]. A learning tool is a small application designed to support learning, studying and teaching activities. The conceptual design framework proposes that various learning objects and tools should be created to provide a variety of different services for HIV/AIDS education;

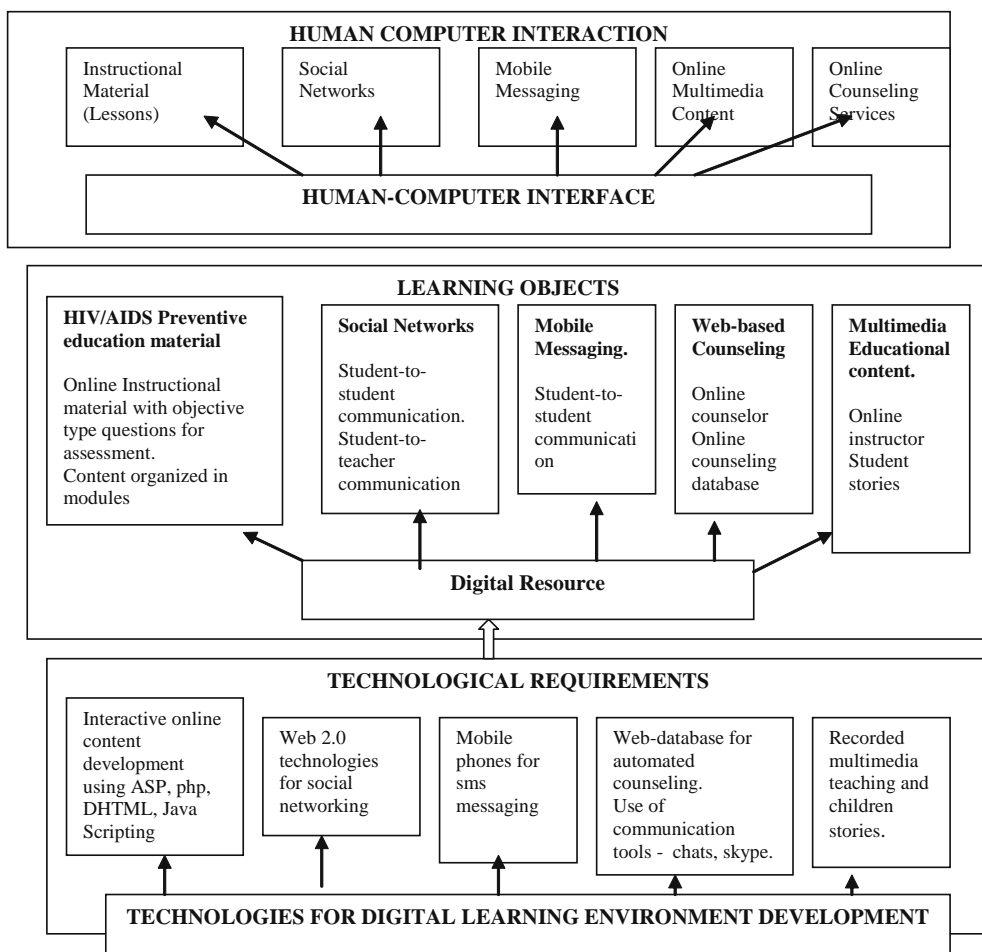
- Learning modules which should be designed for the NetAIDS environment emphasizing the origin of AIDS, ways of spreading, preventive education, positive living, care for AIDS orphans, etc
- Application of Web2.0 tools for creating social networks between children and teachers of high schools. For instance, a wiki tool could be used by children to communicate, discuss, and generate preventive HIV/AIDS knowledge.

- Activating user created educational content – this can be created by building children stories and recorded teachings for online access.
- Counseling services by automating responses to frequently asked questions using web database and online counselor’s services accessible by use of chat tools and Skype.
- Mobile phone messaging by use of mobile phones.
- NetAIDS games for evaluation of children’s knowledge.

3.1.3 Technological requirement

The technological requirements define the implementation solutions and technologies which are used to create the NetAIDS environment. The technological solutions include web 2.0 tools for social networks, Eden windows environment for computer games and database management system. The exact decision concerning the implementation decisions will be made during the later stage of the development process.

Figure 1: The conceptual design framework for NetAIDS environment



3.2. Use scenarios to illustrate the added value of NetAIDS to Ugandan Schools

The following user scenarios provides a concrete picture of how the NetAIDS environment would function in practice:

1. Introductory instructional materials provide an introductory lesson to HIV/AIDS education explaining to children ways of getting AIDS disease and ways of avoiding it, this could also extend to ways of taking care of people infected with AIDS. After the users have taken the introductory session, an online assessment is used to evaluate the knowledge of the children. Another approach to instruction delivery and counseling services can be by providing an online multimedia content in which recorded instructions from teachers and counselors is accessed by school children
2. At any time children can ask questions or the system can prompt the children to ask questions related to HIV/AIDS. The automated frequently asked questions service would compile a response to the questions.
3. Networking among children can be supported by the wiki in which specific discussion topics, for example “HIV/AIDS epidemic: Advice to the children”, can be presented to children to contribute ideas.
4. Use of mobile phones can also promote communication among youth although this is limited as mobile phones are prohibited for children to use in Ugandan schools, but this can still serve children who are having vacation at home.
5. Various NetAIDS games are used to motivate and enhance learning of the AIDS/HIV knowledge. The games can also be used to assess the students’ knowledge of the basic AIDS/HIV concepts.

4. Future Work: Implementation And Evaluation Of NetAids

4.1 Implementation plans of NetAIDS

Educational data has been collected from selected high schools in Uganda. This data collected includes students’ drama and poems. Learning modules have been developed with corresponding computer games for evaluating students’ understanding. The first prototype of NetAIDS is being developed and will be ready by mid July 2009.

4.2 Role of formative evaluation in future development

Formative evaluation is a continuous process throughout the design and use of the digital learning environment. This evaluation seeks to find out the parts of the digital environments that work and the justification that they work [Van Den Akker, 1999; Reeves 2000]. The main purpose of the formative evaluation is to change the functionality and design of the environment based on the feedback received from the expected users of the environment. Hence, the formative evaluation feeds back in to the development process of a e-learning environment. The aim is also to identify other success factors as the popularity of the environment among learners and usability issues. A formative evaluation enables collection of information from the users of the e-learning environment for the purpose of redesigning the system in order to improve

on it. According to [Reeves and Hedberg 2003], formative evaluation should start at the early stages of the development of an e-learning environment. This can be done by gathering learners' opinions, responses and suggestions to practical problems in order to get guidance for future development. The proposed methods for formative evaluation for NetAIDS are user reviews, usability testing, alpha, beta and field tests of prototyped NetAIDS environment with a wide representative of potential users.

5. Summary And Conclusion

An e-learning environment for HIV/AIDS education has a lot of potential of increasing AIDS education outreach to children in Uganda many schools have computer labs with Internet connectivity. When building an e-learning for educating children, the design should not only focus on technology alone but pedagogical and cognitive aspects of the children should be incorporated in the design process. Children can be motivated to learn online when the online content is logically organized. We propose to use constructivist learning principles in the pedagogical design of the NetAIDS e-learning environment for HIV/AIDS education. Additionally, we propose a conceptual design for the NetAIDS environment. The next step in the design and development process is to apply the pedagogical and conceptual design framework in a creative way in order to implement an evolutionary prototype for the NetAIDS environment. Formative evaluation of the early prototype will provide also valuable information related on how to improve and refine the conceptual and pedagogical design of the NetAIDS environment. The pedagogical and conceptual design of the NetAIDS environment presented in this paper can also be applied in other contexts to implement e-learning environments for HIV/AIDS education.

References

- AVRAAMIDOU, L., ZEMBAL-SAUL, C. (2002). Making the case for the use of web-based portfolios in support of learning to teach. *The Journal of Interactive Open Learning*, 1, 2, 1-19.
- BADA, J. K AND SUHONEN, J. (2008). Towards ICT-based Education and Counseling Support System for HIV/AIDS Prevention in Uganda. *In Proceedings of Technology for Innovation and Education in Developing Countries*. 31st July – 2nd August 2008, Kampala Uganda.
- BLOOME, A. (2001). Schools Think about HIV/AIDS. *A World Links online Collaborative Project TechKongja*, July/August 2001. Retrieved March 6, 2008 from <http://www.Techknowlogia.org>
- BOLHUIS, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: a multidimensional perspective. *Learning and Instruction*, 13, 3, 327-347.
- BOOTH, S. (2001). Learning computer science in context. *Computer Science Education*, 11, 3, 169-188.
- BRUNER, J. (1966). *Toward a theory of Instruction*. Cambridge, M.A: Havard University Press.
- CROWTHER, D.T. (1997). The constructivist zone. *Electronic Journal of Science Education*, 2, 2, December 1997, Editorial.
- DUFFY, T.M., CUNNIGHAM, D.J. (2001). Constructivism: Implications for the design and delivery of instruction. In D.H. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology*. The Association for Educational Communications and Technology. 170-198.

- DUVESKOG, M. (2008). *Website of Sura ya UKIMWI Environment*. Retrieved November 25, 2008 from <http://www.Surayaukimwi.com>
- GREDLER, M.E. (1997). *Learning and instruction – theory into practice (3rd Edition)*. Upper Saddle River, New Jersey: Prentice-Hall.
- HADJEERROUTI, S. (2005). Constructivist approach to object-oriented design and programming. *In Proceedings of the Fourth Annual Conference on ITiCSE, 171-174.*
- JENNY, M. (2006). Learning Objects: A new definition, a case study and an argument for change, *Proceedings of the 23rd Annual Ascillite Conference . Who is learning? Whose technology?* Higher Education Development Center, University of Ontago.
- PIAGET, J. (1969). *Judgement and reasoning in the child*. London, Routledge & Kegan Paul.
- REEVES, T.C. (2000). Socially responsible educational technology research. *Educational Technology, 40, 6, 19-28.*
- REEVES, T.C., HEDBERG, J.G. (2003). *Interactive learning systems evaluation*. Engelwood Cliffs, New Jersey, USA: Educational Technology Publications.
- VAN DEN AKKER, J. (1999). Principles and methods of development research. In: J. van den Akker, R.M. Branch, K.L. Gufstafson, N. Nieveen & T. Plomp (Ed). *Design approaches and tools in education and training*. Dordrecht: Kluwer Academic Publishers
- WILSON, B.G. (1997). The postmodern paradigm. In C. R. Dills and A. A. Romiszowski (Eds.), *Instructional Development Paradigms*. Englewood Cliffs NJ: Educational Technology Publications.
- VRASIDAS, C. (2004). Issues of pedagogy and design in e-learning systems. In *Proceedings of ACM Symposium on Applied Computing*, March 14-17, 2004, Nicosia, Cyprus.
- VYGOTSKY, L.S. (1978). *Mind in society*. Cambridge, M.A. Harvard University Press.

4

Assessing Appropriate ICT with ARIS case in Mozambique

Markus Pscheidt, Victor Van Reijswoud and Theo Van Der Weide

ICT projects in developing countries need to integrate requirements of the local context in order to be successful. Appropriate Technology (AT) has been applied in many domains to deliver technological solutions that are suitable to the local contexts for which they are intended. The Appropriate ICT framework has adopted the principles of AT within the domain of ICT. This paper assesses the Appropriate ICT framework in the context of developing and implementing an Academic Registry Information System (ARIS) in Mozambique, and proposes an enhancement to the Appropriate ICT framework in order to facilitate further research on appropriate ICT tools and methods.

1. Introduction

The difficulties reported in Information and Communication Technology (ICT) projects in developing countries [Heeks 2002; Avgerou 2008] suggest a need to improve our implementation of ICT projects. These difficulties include weak sustainability and scalability and missing organizational capacity to pursue change, and lead to underutilized and abandoned projects. In this paper we will examine how the Academic Registry Information System (ARIS) was implemented in Mozambique in order to assess whether or not the Appropriate ICT Framework can resolve some of these difficulties.

Appropriate ICT has been formulated by Van Reijswoud [2009] based on the principles of Appropriate Technology as stated in the AT source book [Darrow & Saxenian]. Appropriate ICT is an attempt to facilitate suitable technologies for ICT projects in developing countries and consists of three components: hardware, software and organizational change management. Many AT artifacts in domains such as energy and water supply are hardware artifacts. The software component has different characteristics. Some of these characteristics fit in well with the AT concept, while other characteristics pose special challenges. Implementing an ICT artifact in an organization is a process during a certain period of time and brings organizational change with it.

All three components - hardware, software and organizational change management - play an important part in the ARIS case study. Guided by the experiences of ARIS up to now, we show that Appropriate ICT is a useful perspective throughout the project lifetime to reflect on challenges and aid in the selection of appropriate tools. Finally, we encourage further research on appropriate tools to provide practitioners with best practices suitable in specific contexts.

The next section will introduce the concept of Appropriate Technology (AT) and the Appropriate ICT Framework. Then the study and the context of the ARIS case are

described. This is followed by observations of the relationship between software and Appropriate Technology principles. After this the study results are presented. Finally recommendations and conclusions are given.

2. Appropriate Technology and Appropriate ICT

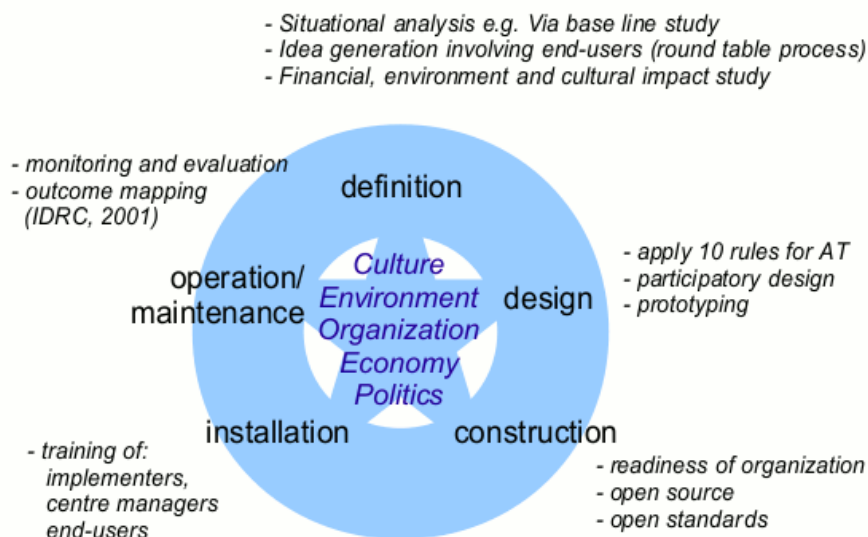
Appropriate Technology (AT) is defined by Wiser Earth as follows [2009]:

“Appropriate Technology (AT) is an applied engineering science suitable to the level of economic development of a particular group of people. Ideally, AT is decentralized, can be used and operated by most of the concerned citizens (i. e. does not require outside operators), uses local or regional fuels and materials in an efficient manner, and involves machines that can be locally repaired. It is sometimes called ‘alternative technology’ and sometimes used for the ‘best choice’ of a technology no matter how complex (‘green technology’)”.

Appropriate Technology has been applied in many domains such as architecture, building, energy and water supply [Darrow & Saxenian, 2009]. There have been a few Appropriate Technology initiatives in the domain of ICT, such as One-laptop-per-child (laptop.org) and the Simputer (www.simputer.org) have tried to provide appropriate hardware, Damn Small Linux (www.damnsmalllinux.org) and UbuntuLite (www.u-lite.org) try to provide software solutions for low-cost hardware. The aspect of organizational change has not received much attention in AT. Appropriate ICT takes a closer look at the complexities of ICT projects and integrates organizational change.

Appropriate ICT is defined by Van Reijswoud [2009] as “the integrated and participatory approach that results in tools and processes for establishing Information and Communication Technology (ICT) that is suitable for the cultural, environmental, organizational, economic and political conditions in which it is intended to be used”. The theoretical considerations on which the Appropriate ICT framework is based consist of the distinction between product - the technical artifact itself - and process of introducing this artifact into the target organization or the respective environment. The process perspective is vital during the implementation phase and is guided by Community Informatics (CI) practices, involving the community itself in the adaptation of the ICT artifact.

Figure 1: Focus of the Appropriate ICT Framework



The proposed Appropriate ICT Framework is based on the traditional Systems Development Life Cycle (SDLC) and extends it with tools and approaches which shall guide the ICT solution to greater appropriateness and thus effectiveness in implementation. A basic set of tools is given and the Appropriate ICT Framework encourages to propose further tools which address cultural, environmental, organizational, economical and political aspects of ICT projects. Figure 1 shows the Appropriate ICT framework extended with possible tools that can support the process. Furthermore, a set of key guiding questions integrate the 10 rules of AT and are structured along the phases of the Systems Development Life Cycle (see Table I).

Table I: Key guiding questions

Hardware	Software	Change mgmt.
Definition phase		
<ul style="list-style-type: none"> - Specific HW requirements in terms of climatological and environmental conditions? - What are the possibilities in terms of enabling factors (Internet connectivity, electricity)? 	<ul style="list-style-type: none"> - What are the needs? - What are the expectations? 	<ul style="list-style-type: none"> - What ICT knowledge levels? - What the financial constraints? - What is the cultural context? - What added value is created? - How is the economic equilibrium affected? - What new ways of working are introduced? - What will the impact be of the system in terms of organizational change? - What is the involvement in the idea generation of key decision makers (political leaders, religious leaders)?

Design phase		
- What is offered on the local market? - What are physical constraints? - What the financial constraints?	- What interoperability needs? - What localization is needed? - What flexibility is expected?	- What are the information needs of the various target groups? - How will these needs evolve? - How do the expectations change?
<i>Construction phase</i>		
- What local skills are available? - Is the equipment protected against physical conditions?	- What local skills are available? - Are features in line with skills? - Are free and open source alternatives considered? - Are the systems well documented?	- Are local skills and knowledge being developed? - Are stakeholders actively involved? - What new ways of working are introduced? - What will the impact be in terms of organizational change?
Installation phase		
- Is all the equipment well protected?	- Has the system been tested with all stakeholders?	- Are all stakeholders involved in training program? - Is the added value made clear?
Operation / maintenance phase		
- Is local capacity sufficient? - Are spare parts easily available?	- Are software maintenance skills available?	- Is a support organization in place? - Is the support organization able to support all stakeholders (e.g. Gender issues)

3. Description of the study

This study builds upon a single-case study of the development and implementation project of the Academic Registry Information System (ARIS) in Mozambique. The functionality of the system is to manage academic information including studies, students, exams and marks, taking into account the specific Mozambican reality and requirements. The project has started in 2005 and this study was conducted at the final stages of the project in 2009. The project has run through all stages of the software development life cycle and at the time of writing the software product is in use. Since the Appropriate ICT framework has been only recently presented by van Reijswoud [2009], it has not been applied during the ARIS project. Through a reconstruction of events of the project, the Appropriate ICT framework is assessed.

The objective of this study is to evaluate the usefulness of Appropriate ICT framework, including its “guiding questions”. The study questions are:

- In which way is the Appropriate ICT framework useful during project lifetime?
- How would the use of the Appropriate ICT framework during the ARIS project have improved the outcome?

- Is the set of guiding questions complete? Can we find suggestions for improvement?
- How can we guide tool development in order to facilitate reuse between ICT projects?

To investigate the study questions, the study made use of the “guiding questions” as formulated by the Appropriate ICT framework. These guiding questions are the means of the Appropriate ICT framework to address relevant issues related to hardware, software and organizational change management. The questions were therefore answered to reflect project challenges. In a second step this picture was contrasted with the actual situation in the ARIS project to evaluate the usefulness and completeness of the picture drawn by the Appropriate ICT framework.

The results of this comparison were strong enough to be used to suggest some improvements to the Appropriate ICT framework.

The sources of evidence for answering the guiding questions for the ARIS project were documents, interviews and participant observation – the latter by assuming a role as a technical expert in the project.

4. The ARIS case

In the following the case of the development and implementation of an Academic Registry Information System (ARIS) at the Catholic University in Mozambique (UCM) will be presented, with the intention to verify the usefulness of the Appropriate ICT framework. This section consists of a description of the context of the case, followed by responses to the guiding questions of the Appropriate ICT framework to verify that a complete picture of project complexity can be developed. Then, the tools which were used during the various cycles of the system development life cycle are described. Finally, some observations are made on the special characteristics of the software development part of the project.

4.1 Context of the case

The case described here focuses on the development and implementation of an Academic Registry Information System (ARIS) at the Catholic University of Mozambique (UCM). The objective of the system to manage academic registry information at Higher Education institutions. The data is stored in a database and can be used to automatically retrieve reports like student lists, certificates, student cards and statistics. The need for such a system has been expressed by several Mozambican universities, since data has been managed mainly with spreadsheets which is labor-intensive, error-prone and not scalable enough for increasing student numbers.

The incremental development model [Davis et al. 1988] was used as the basic software development methodology. After the construction of a partial system, all stakeholders were involved on a regular basis to review progress and provide feedback for further system development.

The information system has been developed by a combined effort of several Mozambican universities, the Ministry of Education and Culture (MEC) and two Dutch

universities. These stakeholders worked together within a development cooperation project funded by the Dutch government and locally coordinated by the MEC. The UCM was one of the beneficiary universities because of its presence in the central and northern provinces of Mozambique. The viewpoint of this case is mainly from the perspective of the UCM as one of the target organizations, but the whole effort was a collaboration between all involved actors, whose input was important for the success in each of the Mozambican target universities. The software development has been done mainly on the Dutch side, in order to provide the necessary competence which was lacking on the Mozambican side. The coordination of most activities like workshops was done by the MEC in Mozambique.

4.2 Hardware, Software and organizational implementation

ARIS has been designed as a client-server architecture and built on Open Source components. Data is stored in a central database. The application logic runs on an Apache Tomcat server providing HTTP content to clients. The clients use a web-browser to access the server. Each user has a separate login and password, and an associated user role to limit the user's privileges in the system.

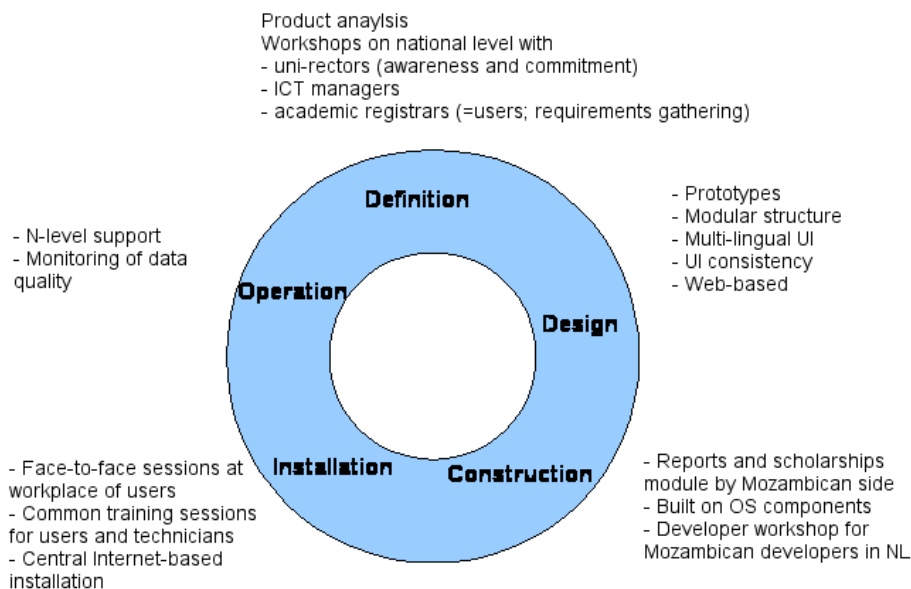
The system is based on Open Source components. PostgreSQL is used as the database and Apache Tomcat as the web server. Furthermore Open Source frameworks like Spring and iBatis are used as building blocks of the software system. Furthermore the system has a modular architecture which comprises a kernel, a reporting module, a scholarship module, a fee module and university specific extension modules.

UCM leased a virtual Linux server from a provider in the United States on which ARIS is deployed. This allows client access from any Internet connected computer via a web browser. Clients use Mozilla Firefox or Internet Explorer to access the ARIS application. The hardware requirements on the client side are a computer with connection to the Internet. All users used Microsoft Windows as an operating system, although this is not a requirement.

UCM has faculties in different cities in central and northern Mozambique. Each faculty has academic registry staff which is trained in common workshops and by on the job. The process is further facilitated with the help of technical experts by importing existing academic data into ARIS, so that the benefits of the new system are more obvious to users. These benefits include automatic creation of certificates and statistics built on data existing in the database.

4.3 Tools used

During the various stages of the SDLC many tools have been applied in order to achieve project results and to deal with the given reality. These are illustrated in Figure 2.

Figure 2: The tools used in the various stages of the ARIS project

5. Observations of SW characteristics

ARIS has a strong software component, so it is pertinent to make some observations about software in relation to Appropriate Technology principles. Many AT artifacts are “hardware” artifacts; for example appropriate cooking stoves. There are several factors unique to software artifacts, as observed in the ARIS case. Some examples are:

- Software production is rather complex [Xia and Lee 2004], which does not easily fit with the AT emphasis of simple technologies. Nauman et al. [2005] give a case study of the complexity of a software development project in a developing country in which they demonstrate considerable complexity in a seemingly simple information system project.

- Because they are complex and need a lot of resources, software systems should not be reinvented by each organization which needs a certain type of software. Uwadia et al. [2006] describe a case study of the collaborative development of information systems with the university environment in Nigeria. The collaborative approach was viewed appropriate for Nigeria because of a high degree of similarity in the core functions and activities needed to support the administration and management of the universities. In such a context, uniform software can be developed at a cost shared by all participating institutions, which can easily be customized to suit each institution’s individual needs. In addition, many universities in developing countries do not have the capacity to develop the required software on their own. For software systems, collaboration should be emphasized and investigated further. But collaborative work without face-to-face contact and without immediately visible benefits for the community involved is not common behavior, so proper ways of organizing collaborative software development need to be investigated and developed.

- There is a design challenge for software systems to allow adapting the system to specific needs of each implementing organization.
- Software needs constant, ongoing development to respond to changing organizational needs [Lehman and Ramil, 2001].
- AT has a "village" orientation [Darrow and Saxenian] but software systems are not typically developed and used in remote villages (which often do not have electrical power). Eventually, the village in this case can be reinterpreted - as organizational communities which, according to AT, shall seek to understand their own needs and how to solve them by their own means.

These observations show the importance of community networking – which is an important aspect of the AT concept. Appropriate ICT orients itself more particularly at the field of Community Informatics [Gurstein 2000], and further tools and methods should be investigated for the community networking aspect, i. e. for the coordination of inter-organizational collaborations for the development and maintenance of ICT systems.

However, software also has characteristics that fit in well with the AT concept: Software production needs small amounts of initial capital; computers and Internet connections are the only investment. It is labor-intensive and has the potential to offer more productive solutions than traditional information management technologies such as pen and paper and plain office applications like word processors and spreadsheets. Furthermore, collaborative software development offers possibilities for local experts and entrepreneurs to get involved if organized properly, e.g. in an Open Source Software approach.

The ARIS case has a strong software part, and one of the conditions for success is that the system is reliable and not error-prone which is a condition for user acceptance of the system. To maintain the code-base constantly in good shape needs experienced software developers. This is a challenge to be achieved in a compatible way to the Appropriate Technology principle that the technology should be maintained by local stakeholders without the need for external expertise.

The high level of skills required to develop and maintain the code-base of a software project bring some serious challenges for sustainability and support. In the ARIS case, no single university is likely to have both the capacity and the acceptance of the other universities to maintain a software system like ARIS on its own. This makes the question of appropriate support and sustainability a crucial one. Donor funding with the development cooperation project was limited to the time needed to develop the system and to do a certain amount of user training. The knowledge transfer and the building of appropriate local structures with accompanying negotiations between the stakeholders will take a longer time. To summarize, the complexity of the system itself is a big challenge for the sustainability of the system and its necessary accompanying support units.

6. Results

In the following the results of the study are presented. First the overall usefulness of the Appropriate ICT framework is reported. Then the two elements of the framework are viewed separately: the “guiding questions” are analyzed for completeness, and guidelines are presented for appropriate tool development.

6.1 Usefulness of Appropriate ICT framework

In our study we have found the following arguments in favor of the usefulness of the Appropriate ICT framework in the ARIS context:

- It gives a good picture of the project reality as it helps to discover ignored aspects.
- The guiding questions can be applied repeatedly throughout the project lifetime to react on changing project realities.
- Analysis through guiding questions aids in proper tool selection.

6.2 Completeness of guiding questions

The guiding questions comprise questionnaires in three areas: hardware, software and change management. The completeness of the questionnaires is considered as follows:

The hardware guiding questions gave a complete picture of project reality by revealing given challenges with Internet connectivity, environmental factors and staff competency.

The software questionnaire covered many important aspects. The following should be considered and integrated:

- In order to analyze the complexity of a project it is important to distinguish if existing software is used or software is developed as a part of the project. In the latter case the complexity of the project is increased and proper staff is needed, even after the construction phase has finished, during operation and maintenance.
- Software requires skills in three different areas: user skills, technical skills for system maintenance and software development skills. This distinction is not present in the questionnaire.

While in the area of organizational change many challenges were revealed, some of which which have not been properly dealt with in the ARIS project, the questionnaire gave a good picture in terms of completeness. The application of the questionnaire showed that the area of organizational change was undervalued in the project as it was set up in relation to hardware and software.

6.3 Tool description

The Appropriate ICT framework in its current form does not provide guidelines for the description of tools, while a good description of tools makes them easier understood and reusable in other projects. We found in the analysis of tools used in the project that there is a common set of characteristics that are helpful to give a concise description of tools, and we suggest to add to the Appropriate ICT framework a set of tool description questions as follows:

- What is the context for which the tool is appropriate?
- What is the problem to be solved by the tool and what is the goal to be accomplished?
- How does the tool relate to local culture, environment, organization, economy and/or politics?
- In which step(s) of the System Development Life Cycle (SDLC) is the tool used?
- What are the necessary preconditions, e.g. availability of resources?
- What potential difficulties can arise as a consequence of the application of the tool (risks)?
- Are there any indicators which can be used to measure the success of the tool?

Table II shows an example how a tool can be described based on the given set of questions.

Table II: Example of a detailed description of a tool

Tool:	Involvement of local software developers during system development: Report and scholarship modules development on Mozambican side.
Context:	Development cooperation project where a software development team in the developed country needs to transfer knowledge to a software development team in the developing country.
Description:	The tool is about giving parts of the system development to the local side; at the beginning small parts and increasingly bigger parts. This is a form of knowledge transfer. The process needs to be initiated by a proper training of the relevant aspects of the system.
Phase:	Construction
Problem addressed:	Missing capacity of local software developers.
Goal of tool use:	Local software developers shall be able to execute basic software development tasks for the application.
Potential difficulties:	<ul style="list-style-type: none"> • Local software developers need some guidance. This is difficult if the competent contact persons are far away in the developed country. This can be facilitated by the presence of a development worker who stays with the community of local software developers. • Conditions for local software developers may not always be appropriate. They may have other duties besides their involvement in the project so they need the commitment by their superiors, sufficient time and proper equipment. • The structure of the software system and its code base needs to allow that people in different locations around the world do not interfere with each other when working on the system. This can be facilitated by a modular structure and a central code repository.

Experience in ARIS:	The practical experience in ARIS is that local guidance and proper working conditions make a big difference. After a three-weeks developer workshop in the Netherlands one team of local software developers got the task to develop a scholarship module. Although the module was specified and designed, the local team did not achieve to continue their work once they team members returned to Mozambique. They had other duties and missing commitment from their superiors. Another example was the development of the reporting module which was successfully developed by another local team supervised by a development worker. After that many other tasks have been done by the same programmer under supervision and increasing communication with the team in the Netherlands via a mailing-list. In this case, the competency is present for the university to further adapt the ARIS for their own needs as well as being a resource for further development of ARIS in general.
---------------------	--

7. Conclusions

The case study confirms that ICT projects in the developing world are confronted with major challenges. Scarcity of equipment, financial means and skills demand careful assessment of the needs and often require alternative approaches. The Appropriate ICT framework promises to be useful to tackle the complexity of ICT projects in developing countries.

In this paper we made observations of software characteristics in relation to Appropriate Technology principles, evaluated the usefulness and completeness of the Appropriate ICT guiding questions and suggested a framework for development of appropriate tools which fit into the Appropriate ICT framework. We recommend further development of appropriate tools in order to aid practitioners in successful ICT project implementations.

References

- AVGEROU, C. (2001). The significance of context in information systems and organisational change. *Information systems journal* 11(1): 43 -63.
- AVGEROU, C. (2008). Information systems in developing countries: a critical research review. *Journal of Information Technology* 23, 133-146.
- BENBYA, H., AND MCKELVEY, B. (2006). Toward a complexity theory of information systems development. *Information, Technology & People*, Vol. 19, Iss. 1; pg. 12, 23 pgs.
- DARROW K., AND SAXENIAN, M. The Complete Appropriate Technology Sourcebook. Available at <http://www.villageearth.org/atnetwork/atsourcebook/index.htm> as of 26. 1. 2009.
- DAVIS, A. M., BERSOFF, E. H., AND COMER, E. R. (1988). A Strategy for Comparing Alternative Software Development Life Cycle Models. *IEEE Transactions on Software Engineering*, Vol. 14, No. 10, October 1988.
- GURSTEIN M. (2000). *Community Informatics: Enabling Communities with Information and Communications Technologies*, Idea Group Publishing.

- HEEKS, R. (2002). Failure, Success and Improvisation of Information Systems Projects in Developing Countries. *The Information Society*, Vol 18(2), pp 101-112.
- HEVNER, A.R., MARCH, S.T., PARK, J., AND RAM, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, Mar; 28, 1.
- JACUCCI, E., SHAW, V., AND BRAA, J. (2006). Standardization of Health Information Systems in South Africa: The Challenge of Local Sustainability. *Information Technology for Development*, Vol. 12 (3) 225-239.
- KIMARO, H. C., AND NHAMPOSSA, J. L. (2005). Analyzing the Problem of Unsustainable Health Information Systems in Less-Developed Economies: Case Studies From Tanzania and Mozambique. *Information Technology for Development*, Vol. 11(3) 273-298.
- LEHMAN, M. M., AND RAMIL, J. F. (2001). Rules and Tools for Software Evolution Planning and Management. *Annals of Software Engineering*, Vol. 11(1) 15-44.
- NAUMAN, A. B., AZIZ, R. AND ISHAQ, A. (2005). Information Systems Development Failure: A Case Study to Highlight the IS Development Complexities in Simple, Low Risk Projects in Developing Countries. *The Second International Conference on Innovations in Information Technology (IIT'05)*.
- UWADIA, C. O., IFINEDO, P. E., NWAMARAH, G. M. ET AL. (2006). Risk Factors in the Collaborative Development of Information Systems for Nigerian Universities. *Information Technology for Development*, Vol.12 (2) 91-111.
- VAN REIJSWOUDE, V. (2009). The role of Appropriate ICT in Bridging the Digital Divide. To be published In: *St-Amant, K. Handbook of Globalization and Digital Divide*.
- WALSHAM, G., AND SAHAY, S. (2006). Research on Information Systems in Developing Countries: Current Landscape and Future Prospects. *Information Technology for Development*, Vol. 12 (1) 7–24.
- Wiser Earth. Available at <http://www.wiserearth.org/aof/144> as of 26. 1. 2009.
- XIA, W., AND LEE, G. (2004). Grasping the complexity of IS development projects. *Communications of the ACM*, May 2004/Vol. 47, No. 5, pp 69-74.

5

The Effect of Cultural Differences on Software Development

D Patel, C Lamson-Johnson & S Patel

This study investigates the cultural issues concerning software development in which the United States (US) and Europe outsource their Information Technology (IT) offshore to emerging countries. Though the benefits of Offshore Software Development (OSD) may result in reduced labour costs for UK and Europeans software development companies; there are concerns about the practice as there are other impacts. The virtual setting of the global environment inhibits team cohesion and interferes with the communication process as the teams collaborate across both regional and functional boundaries. This presents challenges in the cross-cultural relationship of OSD projects as it influences work ethics, task performance and other factors, not just at a management level but at all levels in the organisation. Evidence of the dynamic effect of culture is demonstrated through the experiences of experts associated with OSD; highlighting cultural variables that negatively influence the quality, time and cost of software development.

1. Introduction

Traditionally software is developed within a company's Information Technology (IT) department or a few stages are outsourced to other developers within the same region. In the 1990's, another category of outsourcing, offshoring was adopted as a method of dealing with severe shortages of (IT) professionals in the United States and Europe. With an abundance of skilled IT professionals in emerging countries such as India, software development was transferred to IT offshoring destinations to benefit from the pool of global talent which allowed for quick assembling of fully staffed project teams. This enabled IT projects to be completed quickly, within the schedule time and budget, thus resulting in huge profits for software development companies [Igbaria and Shayo, 2004].

Within a decade, the globalization of many IT firms, such as Microsoft, International Business Machines (IBM) and others, has contributed to software development being transferred to a number of other emerging countries in Asia, including China, Philippines, Singapore, Malaysia and others. The main benefit of OSD for US and European IT firms may be highlighted by the difference in wages for IT professionals. [Igbaria and Shayo, 2004] reported that IT professionals' wages in India were approximately 50% less than a US software development contractor. Though wages in Asia may have increased over a period; evidence still shows that increasing trends in offshoring have caused significant changes to the IT job markets. A global IT offshoring report from Horasis [2005] confirmed that US software development companies have cut full-time jobs by 50 percent removing large capital and valuable jobs from the US job market. In Europe,

the job of basic programming has drastically been reduced as software companies will often choose to subcontract IT projects [Meyer, 2006]. The trend in offshoring has caused many concerns for the IT Industry as it impacts on the process of software development as a whole.

Today, many software development companies are disappointed with the results of offshoring projects proving that OSD is more complex than finding IT skills and resources in low cost locations. According to Meyer [2002], at least 40 percent of software development offshore projects failed in delivering the expected benefits; mainly due to soft issues that cause problems and inhibit project delivery. Currently, there is very little analysis on soft issues such as cultural differences due to the difficulties in measuring the effect of the human factors of the process [Cherry and Robillard, 2008]. As software development heavily relies on human interaction to removed ambiguity from the development stages, it is important to determine the elements that inhibit the success of OSD projects. Studies of global teams in a virtual setting show that cultural differences affect work ethic, work hours, preferred method of communication, revering hierarchy and other factors which may have a negative influence on the experiences of team members [Edwards and Sridhar, 2002]. Therefore an awareness of the cultural factors that may impede the success of Global Software Development (GSD) may assist in effective management of OSD projects.

This study focuses on the cross-cultural context of GSD environment in an effort to determine the effect of culture on software development. In *section 2*, the nature of OSD environment is described and the concept of culture is defined to show how cultural differences may influence the software development process. *Section 3* describes the process used to examine the problems and complexities associated with cultural differences in OSD, while *Section 4* reports the research findings. The findings look at experts' view of how cultural differences interfered with the success of offshoring projects and further demonstrate the dynamics of culture on the global process of software development. Finally, *section 5* concludes by summarizing the main findings of the cultural impact on OSD projects and suggest future work in the area of IT offshoring.

2. Global Software Development Structure and Culture

The traditional approach of face-to-face in-house software development has been shifted to a more "virtual" nature using advanced communication technologies such as teleconferencing, videoconferencing, emailing and NetMeeting. This presents challenges in the management of GSD as team members are from a number of nations with varied social, economic and cultural background and are in diverse locations both onsite and offsite, spread geographically over time zones [Dafoulas and Macauloy, 2001]. Team members also belong to two or more organisations or one organisation and its subsidiaries, in contractual agreements across national boundaries [Huang and Trauth, 2007]. The separation due to the distance between team members interferes with the interaction process including feedback mechanisms, which are necessary for the advancement of the software development stages. For example, the time difference of 10 hours between

the US and India may create difficulties in scheduling of conference calls and meetings. Tasks that are passed backwards and forwards overnight may also encounter delays in clarifying information. Severe communication problems may develop which impacts directly on product delivery [Kobayashi-Hillary, 2005].

The software development stages are communication-intensive especially the requirements stage which is relied on to remove ambiguity from the process. Edwards and Sridhar [2006] confirm that mistakes made in this stage are passed into the latter stages of the development cycle. Even though English is the standard language used to communicate in offshore projects, among the participants there are variations of the language which influences the stages of software development. As a result, some participants are more comfortable with reading than speaking or listening to the language. Thus the team face challenges in pinning down requirements due to diverse interpretation and understanding while relying on sophisticated communication technologies [Huang and Truth, 2007]. The problems encountered in clarifying requirements create delays which impact on the project schedule. Similarly, odd English phrases are reported in the coding stages which require extended time in clarify terminologies and levels of details [Matloff, 2005]. When difficulties are frequently experienced with documentation, it increases the budget which may contribute to the failure of OSD projects.

Culture is a multi-faceted concept which is described by Hofstede [1982] as “the common characteristic which distinguishes the members of one human group from another.” It may be applied in terms of nations or regions or ethnic groups, but also to the collection of human groups within organisations. Global software development benefits from the differences in culture as it promotes creativity, innovation, and enables the adaptability of software to different nations in the global IT markets [Huang and Trauth, 2007; Edwards and Sridar, 2006]. However, cultural differences influence the relationship among GSD teams as it subtly shapes individuals’ perceptions, attitudes and behaviour [Mullins, 2007]. Studies by Olson and Olson [2003] exemplify how the US approaches software development differently compared to the Middle East. Time and deadlines are taken seriously in task oriented cultures, such as the US, than in other relationship orientated cultures, such as in India. Therefore team members from the US will settle down to work more quickly than their colleagues from the Middle East, who will spend more time getting to know each other. This implies that the teams’ impression of each other is unconsciously influenced from the beginning of OSD projects which may affect the day-to-day operation of the team. Subsequently it creates problems for the team to relate to each other which increase complexities in the management of projects; this further impacts the outcome of software development.

3. The Process of Investigating Software Development Offshoring

Data Collection

This study utilizes a semi-structured approach in investigating the present trend of the US and Europe offshoring to emerging companies. In determining the foundation of problems and complexities that are experienced in OSD due to cultural differences, paper

questionnaires were prepared to gather information from known software development companies such as Microsoft Corporation, IBM, Sun Microsystems, Intel and other large software vendors well known for IT offshore outsourcing. There was a very poor response from these large multinational software development companies as telephone conversations and email contacts unfortunately revealed that the time and resources were not available to lend their expertise on the subject. Through consultation with other experts, to maximise the response rates, a decision was further taken to convert the paper questionnaire to an online survey set up on a website that facilitated the collection of responses and provided numerical analysis of the data. The website tools facilitated both open and closed ended questions which were adequate for the transfer of the paper questionnaire. It incorporated Likert scaling which is frequently used in survey research to measure constructs such as attitudes and opinions [Wu, 2007]. This function enabled average scoring of the respondents' level of agreement to questions which was further used to analyse responses after the survey was completed.

Participants

The online survey Part 1 began at the end of June 2008. The online link to the survey was emailed to approximately 200 software development companies explaining its significance in managing OSD projects. The survey consisted of 10 questions of which questions 1 to 5 were generalised questions on offshoring; while questions 6 to 10 were more specific to culture. The slow progress experienced in the participation of software development companies, revealed that the subject of culture may be considered highly sensitive. It is possible that due to the specific nature of the survey questions, organisations may not be eager to express their beliefs for the fear that it may threaten the existence of harmonious relationships with their offshore counterparts. Many organisations may prefer to address a sensitive topic such as culture by using known professionals who provides cross-cultural training for OSD projects [Kobayashi-Hillary, 2004]. Nevertheless, the IT companies that participated revealed valuable information and insights into the developments of IT offshoring. The research findings are revealed in *section 4.1*.

The online research approach initiated a second part to the survey that began in July 2008. This part of the survey targeted knowledge-based experts in IT offshoring, with the aim of further examining in detail the dynamics of culture in the practice of GSD. A similar approach was taken where 10 questions were created that ranged from general questions of outsourcing such as risks and benefits, to more specific ones that included discrepancies between virtual and face-to-face software development collaborations. The survey was sent to approximately 100 experts after a careful selection of the most appropriate experts that could contribute knowledge in the area of culture in IT outsourcing. Informal emails were not as successful and may have been perceived as spam mail. Therefore formal emails were sent to the experts and after a fortnight more experts responded to the survey, than in Part 1. It was evident that the experts trusted the emails which successfully resulted in a professional relationship where they were more willing to provide experiences of cultural differences in OSD projects. Some

experts replied to the emails with openness to provide more information, others also expressed interest in the results of the survey. The research findings are revealed in *section 4.2*

4. The Influence of Culture in Global Software Development

4.1 Findings from Survey Part 1

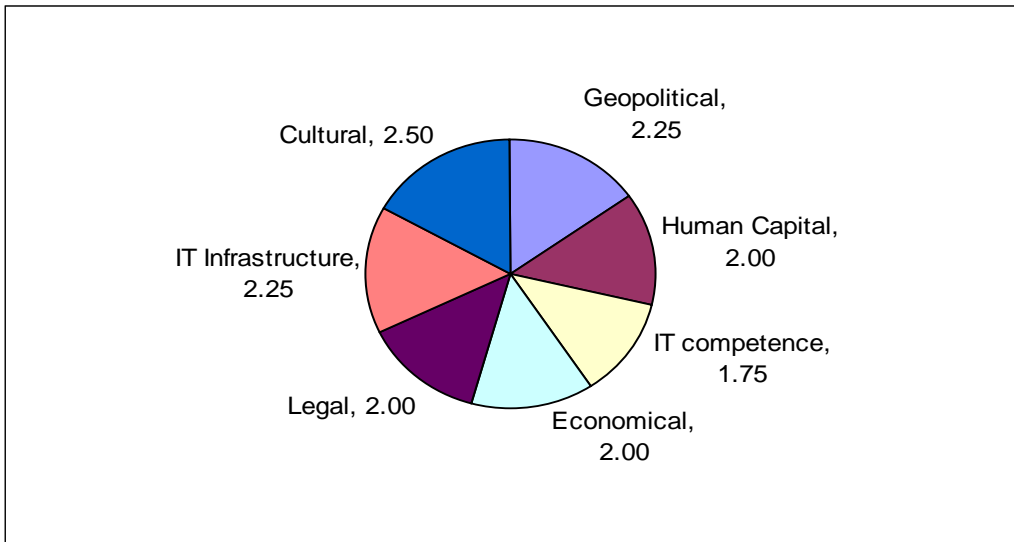
4.1.1 Popular Offshoring Locations

The findings from this part of the survey confirmed the importance of culture in the contractual relationship of GSD. India's language compatibility with the US and Europe has contributed to its popularity of offshore location over China and other Asian nations for a very long period. The responses from the software development companies substantiate that India has been a high quality software development destination. Offshoring to India was described as "cost-effective" and as having "advanced IT" capabilities. The percentage of IT organisations that offshore to China was quite significant although the reason for choosing china was not clearly defined. Although Huang and Trauth, [2007] revealed the potential of china as an offshoring location, India has remained the number one offshoring destination for a long time which is because of its language similarity with the US and Europe. The results confirmed that that similarity in cultures supports organizations' decision making in the choice of offshore locations.

4.1.2 Offshoring Risks

Inadequate risk management may result in IT projects exceeding its estimated time and cost [Sakthivel, 2007]. Moreover in OSD projects, it is crucial for US and European companies to consider the risk factors in deciding an offshore location. Minevich and Richter [2005] identified seven significant risk factors that are associated with offshore locations, geopolitical, human capital, IT competency, economical, legal, IT infrastructure and cultural risks. Based on Minevich and Richter [2005] risk factors, the results from the survey (Figure 1) showed Cultural risk on a scale of 1 to 5 scoring the highest rating average of 2.5. The complex nature of offshore outsourcing requires that the US and Europe analyse the cultural risk for software development destination in relation to language compatibility, cultural similarities, innovation and adaptability. The result signifies that cultural risk in terms of language compatibility is of great importance in determining the possibility of loss to offshore locations. This result is not surprising given that India with its English Language compatibility with the US and Europe, is quiet a popular choice for IT offshoring.

Figure 1: Average rating of Offshoring risks



4.1.3 Benefits of offshoring

Igbaria and Shayo [2004] identified “Low cost labour” as the primary reason for the US and Europe trend in IT offshoring. This theory is supported by Sakhivel, 2007 which pointed out that the IT companies benefited from relatively low cost labour from Asian countries, such as India and China. However the survey findings (Table 1) revealed on a scale of 1 to 5, ‘low cost labour’ scored one of the lowest average ratings of 3. IT Management expert Stan Gibson [2006] has reported a continued wage increase in Asia, specifically in India, and the results may be associated with the steady increase in wages over the period. Although wages in Asian offshore locations are not quite as low as in the 1990’s when the trend began, offshoring is valuable to US and European companies.

Table 1: Average rating of offshoring Benefits

BENEFITS	RATING AVERAGE
Access to Global Talent	3.25
Global Expansion	3.25
Limited Presence	2.5
Quick Completion	3.25
Lower Labour Cost	3.00

The results further substantiate that one of the most valuable benefits of IT offshoring is “access to global IT talent”. This has remained a priority for offshoring as it removes the burden of in-house software development and enables access to offshore resources [Oza et al., 2004]. The underlying theory of “global” software development is confirmed given that the US and Europe IT firms expect “quick completion” of OSD projects within the estimated time and budget. However many companies proved that the benefits of OSD may not be as explicit because the structure of software development is governed by the dynamics of national, regional and functional boundaries [McAllister 2003; Dafoulas and Macaulay, 2001]. The dynamics are displayed in the teams’ relationship which is revealed in the findings of survey part 2.

4.2 Findings from Survey Part 2

The experts that participated in this part of the survey worked or studied in OSD cross-cultural relationships over a period ranging from 5-20 years. Their vast cross-cultural knowledge and experiences enabled informed responses confirming that the geographical features of GSD have an effect on the quality of relationship among OSD team members. The responses provide evidence that software development team cohesion is inhibited due to the differences in perception and interpretation that are associated with national, regional and organisational boundaries. The remainder of this section highlights a number of the experts’ responses for the purpose of demonstrating the effect of cultural differences on the global practice of software development.

4.2.1 Different perspectives

Expert Response: “Power distance and the willingness to take initiatives in a business transformation project.”

The statement implies that the preferences of the Western and the Asian worlds are related to the cultural backgrounds. It is based on Hostede cultural dimensions, which is a conceptual theory that is useful in analysing the cultural characteristics of different nations. For Example, India is regarded as having a large power distance society which means that they have a strong preference for hierarchical team structures, which is the opposite for the societies of the US and the UK that encourage a flatter team structure [Macgregor et. al, 2005]. In India, a subordinate’s perception of the manager is as a “superior” that executes the rights of decision-making and his/her authority will remain unchallenged. Therefore the subordinate’s performance is dependent on the frequency of instruction that is given by his superiors. This attitude of dependency on management is quite normal in the Indian culture but when observed from a Western perspective such as the US and UK where managers and subordinates prefer a more consultative type of relationship, it may be regarded as failure in taking initiatives for the advancement of OSD projects [Narayanaswamy and Henry, 2005].

Expert response: “Typically Indian engineers are less likely to push back on issues or to innovate. American engineers and managers tend to be more out spoken. This leads to things being one sided and where American

outsourcers really don't know the true status of things with the Indian offshore provider."

The findings indicate that the different perspectives of the Western and Asian cultures result in differences in attitudes and behaviours. This factor influences the manager-subordinate perception in cross-cultural relationships in OSD as individuals' unawareness of cultural differences may result in increasingly negative criticisms on both sides. A US manager may criticize the lack of performance of Indian subordinates based on expectations of US teams or Indian subordinates complain for the lack of guidance throughout the software development process based on expectations of Indian Managers. Studies revealed that the different preferences of the Western and Asian societies are related to the characteristics of the educational systems. Huang and Trauth 2007 associate Chinese software developers' unwillingness to speak up in project discussions with Chinese Confucian philosophy and ancient tradition that emphasize authority; where it is not the custom to question a teacher. In contrast to the US where students are encouraged to participate in open class discussions, teachers in China tend not to engage students in open discussions [Matloff, 2005]. The traits have been established in the communication styles, work behaviour and the business cultures of the Western and Asian societies.

4.2.2 Lack of Trust

Expert Response: "When people are physically co-located it is much easier to observe and adapt to local norms (do as the Romans do). Over the phone much of the context is lost."

The expert statement confirms that software development is significantly different when conducted in-house than in a virtual setting. The physical distances between the team members in GSD give rise to misperception and misunderstanding that influences factors such as trust which from the outset creates difficulty in building relationships. From a Western perspective trust is earned based on assessment of competence and the ability to effectively perform. On the other hand, from an Asian perspective trust is based on perceived reliability and in the transfer of shared values and belief [Winkler et al., 2008]. These factors indicate that perceptions are influenced by the way tasks are performed by team members from other cultures. Each subsidiary in multi-site organizations will have its own "site culture" which means that "sites" members will share similar view in regards to the rest of the organisation [Armstrong and Cole, 1995]. Therefore biases that are shared among "site" members may lead to innovative ideas being withheld based on their perception of the rest of the organization. This proves that trust among the team member may be reduced because of the differences in functional, organisational and regional practices. The lack of trust influences the cohesion of the team's which may be difficult to improve given that some team members are onsite while others are in remote locations.

Expert response: "could think of Indian's way of performing only as many of our customers had complaints."

The findings from the expert confirmed that the software development team performance is frequently based on the expectation of the US and UK cultures. The Standard Software Development Methodologies used in offshore projects are from a Western perspective. However, it is important to understand that the application of knowledge and skills in emerging countries such as India is severely different from Western countries which may be misunderstood for incompetence [Edward and Sridhar, 2006; Winkler et al. 2008]. Based on US software development customs the team is expected to openly discuss misunderstandings. In Asian cultures, such as Indian and Chinese, there is a high importance of maintaining mutual harmony within the group and not embarrassing persons in front of the others [Yeo, 2001]. Therefore as direct confrontation is deemed rude and giving frank negative opinion of others is regarded as detrimental to the relationship of the group, there is a tendency for individuals to internalise their opinions to avoid conflicts [Huang and Trauth, 2007]. Another factor that reduces trust in GSD teams is the difference in communication styles. For cultures such as India, 'yes' does not necessarily mean 'I agree' but means 'I heard' which poses questions for US and UK colleagues in terms of how much of an instruction is clearly understood [Matloff, 2005]. Subsequently, India's cultural custom of "nodding the heads" often confused US and European team members. Lack of trust influences GDS team morale reducing the quality of relationship throughout the OSD projects.

4.2.3 The Range in language Competence

Expert Response: "While working with Chinese partners it was clear that their language skills were limited. Conversations were very difficult over the phone. We had to switch to email in order to understand one another."

The cultural context of communication places further pressure on the communication patterns in the software development stages that may be already experiencing difficulty because of OSD dependence on sophisticated communication technology. The communication technologies used in OSD may create "noise", thus causing the context of information exchange to be lost. The concept of "noise" means anything that interferes with the communication process of the team collaborating across geographical boundaries and time zones. Furthermore, though English may be the standard language used among the US, UK and the emerging nations that engage in OSD, team members are from different continents which create communication barriers caused by semantics given that words may be absent, or have different meanings or translation in a particular local language (Kobayashi-Hillary, 2004). Some words and intonation may have different meanings to people from a local region or even a country. Therefore misunderstandings and disagreements may develop which may lead to conflicts that increase complexities in the management of OSD projects.

Expert Response: "Communication and 'business'-communication in particular is impacted by the cultural context. If wrongly understood this will/may lead to misinterpretation."

The research findings based on the experts' perspectives show that the cultural context of the environment impacts on the communication because of the range in language skills that create difficulty in cross-cultural conversations. Thus the team dispersed over multiple time zones using sophisticated communication technology, is a challenge in itself as it influences the knowledge transfers and information exchange. The gradual advancement of the software development stages is impeded as the geographical distance prevents informal communication that creates difficulties in clarifying issues. This leaves room for assumptions in the requirements stage that causes problems in the latter stages of the development [Edwards and Sridhar, 2006]. The range in language competence also affects the documentation of processes including the codes that contains phases or comments which may be inconsistent in meaning in Western and Asian countries [Matloff, 2005]. This results in an offshoring hidden cost, as codes need to be written again which significantly affects the contractual relationships. The factors indicate that the interaction in GSD is affected by cultural context of the environment which in turn will influence the development process as the complexity of the situation will produce disagreements over functionalities, tasks, processes and methodologies which will encourage conflicts.

4.2.4 Conflict and its influence on other variables

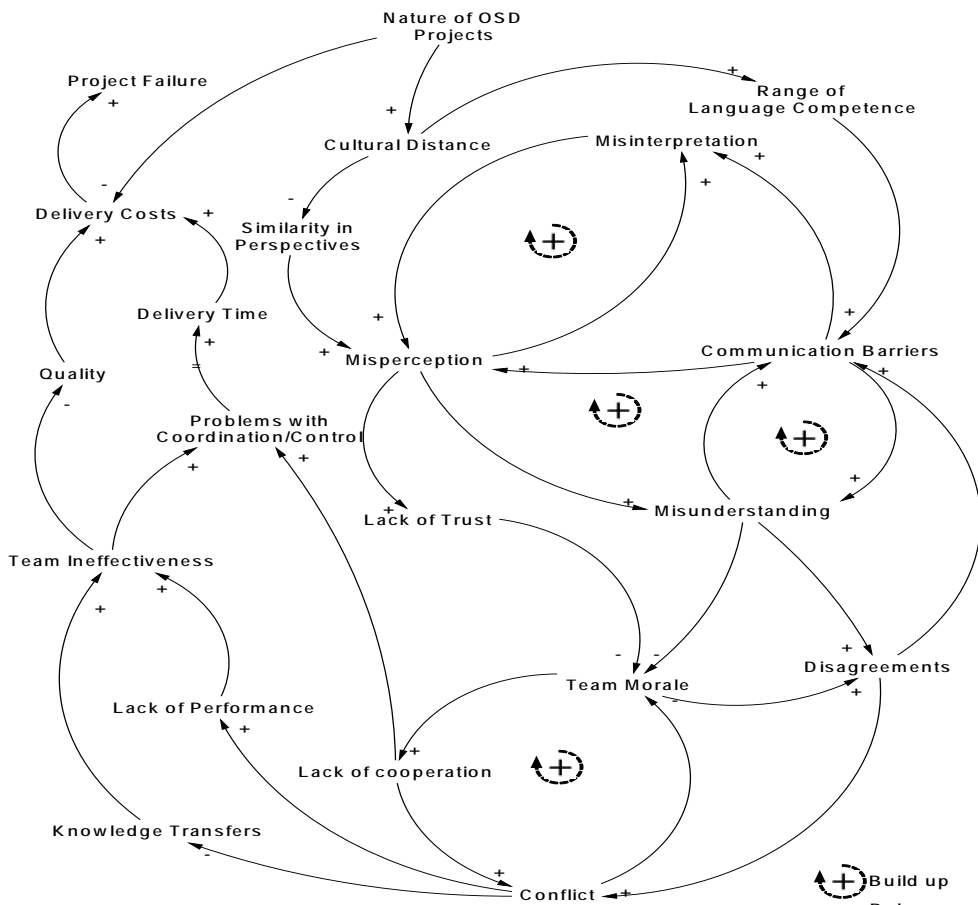
Expert Response: "My research findings show that cultural differences become apparent as behavioural differences in service delivery. It is important to be sensitive to such behavioural differences, as they may negatively affect relationship quality (e.g., if conflicts occur), which in turn has a negative influence on offshoring success (e.g., reduced service quality or project delays)"

The experts' responses confirmed that the success of OSD projects is dependent on the knowledge and understanding of the persons that are employed in offshore contracts. As team members are incorporated from different regions of the world, it implies that different specialist knowledge is to be shared for the success of offshore projects. The difference in preferences, attitudes, behaviour and communication of the Western and Asian societies subtly influences OSD teams' expectations of each other producing conflict that reduces the effectiveness of the team. This decreases knowledge transfers that are critical for the advancement of OSD projects, thus influencing factors such as innovation in the software development process. Persistent conflict adds complexities to the management of the software development projects. This problem further increases misunderstanding and causes teams members to constantly disagree with each other over requirements, methodologies and other issues. Subsequently the continued conflict damages the communication structure influencing other factors that create more difficulty in the coordination and control of OSD projects. In essence, it reduces the level of team performance and affects the quality of product development; hence a time delay occurs that pushes forward the project deadline and adds to the delivery costs. Therefore the spiralling of the cultural factors identified often results in failure of software development offshore projects.

4.3 The influence of culture

The dynamics of cultural differences are experienced at various levels of the organisation where its subtle appearance is demonstrated through the preferences of individuals and groups, work practices and other areas in IT offshoring. The survey findings proved that cultural differences have contributed to a number of problems in OSD projects. The evidence shows that the harmony of the team, the stages of software development and also the management has been affected due to the subtle differences in values and beliefs of multi-national teams. For effective management of OSD projects, it is important to understanding the elements that lie behind the differences. Therefore the variables identified are used in an influence diagram (Figure 2) to conceptualize the complexity of problems associated with cultural differences in the US and Europe offshoring software development.

Figure 2: The Effect of Culture



The influence diagram reveals the basis of cultural distance to include time, distance and cultural differences, which are critical factors to the success of OSD Projects. The dynamics of the variables shows that the perspectives of nations (US, Europe and the emerging countries of Asia) that are involved in IT offshoring are significantly different based on their cultures. The different preferences in hierarchical structures, software development tools, communication, work ethics and other factors pose problems with misperceptions and misinterpretations that affect the level of trust among OSD teams. Additionally, the range in English language competence increases communication barriers in the stages of software development which further escalates into misunderstandings and disagreements. These factors influence the persistence of conflict that continues to deteriorate the team's morale, cooperation, performance and knowledge transfers. Finally, the ineffectiveness of the team obstructs the progress of software development and increase complexities in the coordination and control of offshore projects.

5. Conclusion

The research finding shows that cultural diversity in offshoring may complicate factors that lead to the failure of software development projects. Individuals from the Western (the US and Europe) and emerging Asian nations such as India and China, collaborating across continents will approach situations and tasks from opposite perspectives. The varying perspectives interfere with the manager-subordinate and the peer-to-peer relationships as each nation have different preferences in the practice of software development. This influences the quality of the development process through misperceptions and misinterpretations resulting in poor communication and reduced trust in the relationships. The gradual progress of the software development stages are inhibited due to communication difficulties which includes the range in English language competence of the team members. These human variables interfere with the day-to day team interactions producing misunderstandings and disagreements that lead to persistent conflicts. Persistent conflict influences other variables such as the knowledge transfer, the team effectiveness, performance and, coordination and control increasing the complexities of managing the software development process. A culmination of the challenges experienced decreases the quality of the product, increases the time and overall cost which often result in failure of OSD projects.

The team members of GSD are located in different regions of the world, using sophisticated communications technology, over geographical distances and time zones. This factor restricts the team's cohesion and interferes with the communication process in OSD projects. Additionally, individuals' tendencies and preferences are present from the start of offshoring projects due to national, functional, organisational and soci-cultural backgrounds. The cultural expectations subtly influence how the members relate to each other and based on cultural distance, the nations will not behave in a similar manner. Nevertheless, the main problems experienced in OSD projects lie in the fact that the nations approach communication and work from opposite perspectives. Many organisations that engage in offshoring are unaware of the subtlety of culture that

influences the outcome of software development projects. Therefore this study shows the importance for organisations to develop strategies to mitigate the negative effect of culture as it impedes the success of OSD projects.

Cultural compatibility and similarity is an important factor in the US and Europe offshoring software development as it contributes to the decision making of offshore locations. The study confirmed that India's language compatibility with the US and Europe has assured the country the number one offshore location for quite a long period over China and other Asian nations. Presently, the US and European firms are taking the opportunity to transfer software development to countries in their regions due to the geographical proximity and the cultural compatibility. A report from The Times of India [2005] confirms US "friendly policies" to develop Latin America countries such as Brazil, Argentina, Mexico and Costa Rica into high quality nearshoring locations. The same trend is expected in Europe for Russia, Poland, Romania and Hungary. Though Asian nations, such as India are still quiet competitive as an offshoring destination, nearshoring locations score higher points for communication due to the similarity in languages and cultural understanding [Meyer, 2006]. This result indicates that cultural compatibility is changing the trends in IT offshoring locations. The developments in this area require future research as it will provide further evidence of the dynamic nature of culture. The studies will provide valuable insight into the future of software development outsourcing.

6. References

- ARMSTRONG, D. AND COLE, P. (1995) *Managing Distances and Differences in Geographically Distributed Work Groups, Diversity in work teams: Research paradigms for a changing workplace*, Distributed work. Cambridge: MIT Press pp.187-215.
- CHERRY, S. AND ROBILLARD, P. (2008) The social side of software engineering- A real ad hoc collaboration network, *International Journal of Human-Computer Studies* Volume 66 Issue 7 pp. 495-505.
- DAFAULAS, G AND MACAULAY, L. (2001) Investigating Cultural Differences in Virtual Software Teams, *The Electronic Journal of Information Systems in Developing Countries* 7 (4) pp.1-14.
- EDWARDS, H.K. AND SRIDHAR, V. (2002) Analysis of the effectiveness of global Virtual Teams in Software Engineering Projects, *IEE Computer Society*. In: Proceedings of the 36th Hawaii International Conference on System Science.
- EDWARDS, H.K. AND SRIDHAR, V. (2006) *Collaborative Software Requirements Engineering Exercises in a Distributed Virtual Team Environment*. [Online] Hershey: Idea Group Publishing.
- GIBSON, S. (2006) *Is India losing its wage edge?* [Online] Available from: eWeek.com. <http://www.eweek.com/c/a/IT-Management/Is-India-Losing-Its-Wage-Edge/>.
- HOFSTEDE, G. (1982) *Culture's Consequences*, international differences in work related values. Abridge ed. Newbury Park: Sage Publications.
- HORASIS (2005) *Release of the global outsourcing report*, The Global vision Community. [Online] Available from: Horasis.org http://www.horasis.org/press_releases_all_4.php.
- HUANG, H. AND TRAUTH, E. (2007) *Cultural Influences and Globally Distributed Information Systems Development: Experiences from Chinese IT Professionals*. In: Proceedings of ACM SIGMIS-CPR

- Conference, St Louis, Missouri, USA, April 2007 pp.36-45.
- IGBARIA, M. AND SHAYO, C. (2004). *Strategies for managing IS/IT personnel*. Hershey, PA, Idea Group Pub.
- KOBAYASHI-HILLARY, M. (2004) *Outsourcing to India: The Offshore Advantage*. Berlin: Springer.
- KOBAYASHI-HILLARY, M. (2005) A Passage to India, Pitfalls that outsourcing vendors forgot to mention, *Queue* pp. 56-60.
- MACGREGOR, E., HAIEH, Y. AND KRUCHTEN, P. (2005) *Cultural Patterns in Software Process Mishaps Incidents in Global Projects*. Paper presented on Human and Social Factors of Software Engineering from ACM, St. Louis, Missouri, USA.
- MATLOFF, N. (2005) Offshoring What Can Go Wrong?, *Perspectives*, pp.39-45.
- MCALLISTER, N. (2003) *Offshore Outsourcing: Navigating the Globe*. Offshore Development
- MEYER, B. (2002) The unspoken Revolution in Software Engineering, *The Profession* pp.124, 121-123.
- MEYER, T. (2006) Offshoring to new shores – Nearshoring to central and Eastern Europe, *Deutsche Bank Research*, pp. 1-12.
- MINEVICH, M AND RICHTER, F. (2005) *The Global outsourcing Report-Opportunities, costs and risks*. The CIO Insight Whiteboard.
- MULLINS, L. (2007) *Management and Organisational Behaviour*, 8th ed., FT/ Prentice Hall.
- NARAYANASWAMY, R. AND HENRY, R. (2005) *Effects of culture on Control Mechanisms in Offshore Outsourced IT Projects*. In: Proceedings of ACM SIGMIS-CPR Conference, Atlanta, Georgia, USA, April 2005 pp.139-145.
- OLSON, J. AND OLSON, G. (2003) Culture Surprises in Remote Software Development Teams, *Focus Distributed Development*. pp. 52-59.
- OZA, N., HALL, T., RAINER, A. AND GREY, S. (2004) *Critical Factors in Software Outsourcing – A Pilot Study* In: Proceedings of Workshop on Interdisciplinary Software Engineering Research, Beach, California, USA, April 2004 pp.67-71.
- SAKTHIVEL, S. (2007) Managing Risk In Offshore Systems Development, *Communications of the ACM*, 50 (4) pp.69-75.
- THE TIMES OF INDIA (2005) *India is outsourcing to compete with nearshoring*. [Online] <http://timesofindia.indiatimes.com/articleshow/1338040.cms>
- WINKLER, J.K., DIBBERN, J. AND HEINZL, A. (2008) The impact of cultural differences in offshore outsourcing: Case study results from German-Indian application development projects, *Information Systems Front*, 10, pp. 243-258.
- WU, C.H. (2007) An Empirical study on the transformation of Likert Scale data to Numerical Scores, *Applied Mathematical Sciences*, 1(58) pp.2851-2862.
- YEO, A. (2001) Global-Software Development Lifecycle: An Exploratory Study, *CHI 2000*, 3, (1) pp. 104-111.

Endnote

The diagram in **figure 2** is based on the Systems Thinking Methodology. It conceptualizes the complexity of problems associated with cultural differences in OSD projects. The minus (-) signs indicates that there is a negative effect on the defined variable while the plus (+) sign indicates the opposite. The result is highly subjective to the variables identified in analysis of the effect of culture on the global practice of software development.

6

Service-oriented Architectures as a Vehicle for ICT in developing Countries: An Awareness Campaign

Agnes F. N. Lumala, Benjamin Kanagwa, José Ghislain Quenum and Jude T. Lubega

Service-oriented architecture(SOA) is one of the ways to build applications today. Indeed current applications of SOA type are an aggregation of several smaller applications. Such applications are synonymous with Web 2.0. The cost of consuming and using existing SOA services is much lower than establishing a new infrastructure to provide the same services and there exists free and open source services. Such a technology is appropriate for developing countries where technical and financial resources to setup and support ICTs are severely constrained. Therefore we organized a contest to create awareness about SOA as alternative application development strategy and to expose university students in Uganda to SOA technologies. In this paper we discuss the SOA contest and we give lessons learned from the SOA awareness as a whole.

1 Introduction

The last decade has witnessed a significant shift from traditional software architectures to service-oriented architecture (SOA). Thanks to the Internet revolution, services developed on one site could be assembled on a different site and could be accessed on a third one. Indeed current applications of SOA type are an aggregation of several smaller applications. Such applications are synonymous with Web 2.0.

Due to the tremendous progress in software development brought in by SOA, one can flexibly and transparently combine a book order and its shipment, a flight ticket to a hotel reservation and even a car reservation at the destination. The list goes on and on.

Moreover, the cost of consuming and using existing SOA services is much lower than establishing a new infrastructure to provide the same services and there exists free and open source services. By using SOA technologies the cost of the service, security, maintenance is greatly reduced. However you need to be aware of the cost and take care of basic technological requirements before you enjoy benefits of the technology. For example the Amazon storage service costs \$0.180 per GB for the first 50 TB / month of storage used. We therefore see SOA technologies as an important approach to support ICT use in the developing world.

Given the popularity gained by SOA in recent years, Makerere University Faculty of Computing and Information Technolofy (FCIT) organized a contest about SOA implementation. This was to create awareness and exposure to SOA among our

students. We offered three applications (see section 2) and requested students to find data from/for any of these applications using Yahoo! Pipes and Amazon S3. The target participants were masters and bachelors students (with a background in information technology) from different universities in Uganda. However, only students from Makerere University and Mukono University expressed interest and participated in the contest. The winner of the contest and the first runner up were to receive a prize of a simple mobile handset. All the contestants would have a chance to attend lectures for two weeks by a SOA specialist who was coming in after the contest. Besides, the contest organizers offered them technical support as they developed the different applications. The contest organizers developed the backend of the applications involved, and the students were expected to develop the front end of the applications involved. Before discussing more about the soa awareness, it is important to introduce some basic technologies that are necessary for building a soa-based application.

1.1 SOA Technologies

There are several tools and technologies available to support the implementation of SOA. They focus on the core tenets of SOA, that is to describe, identify and consume the services. In web services, *Web service description language* (WSDL) is designed to describe services in way that they can be used by others. WSDL is a web service standard by W3C for describing the syntactical structure of a services. More specifically, the WSDL specifies the operation names and parameters. Details about WSDL can be found in [Chinnici, *et. al.*, 2007, Curbera, *et. al.*, 2002].

To implement SOA there is a need to transfer data across systems. *SOAP* (Simple Object Oriented Access Protocol) is one of the widely used technologies to package data into envelopes. In web services, an envelope contains information about the service to be invoked and the information required. The SOAP specification can be found in [Box, *et. al.*, 2003] and usage in relation to other technologies is illustrated in [Curbera, *et. al.*, 2002].

Another technology is *REST* (REpresentational State Transfer). Using *REST* an application can send requests and responses using HTTP(Hyper Text Transfer Protocol) verbs such as GET, POST, PUT and DELETE. In RESTful terminology a web service is called a resource. These resources can be located using a Uniform Resource Identifier (URI). Refer to [Fielding, 2000] for more details on REST.

The *Web feeds* provide simple means for dynamic update of information on the web. A web feed presents data in format (usually XML) that can be accessed directly by interested users. Through a web feed, you provide a mini database containing headlines and descriptions of what's new on your site. RSS (Really Simple Syndication or Rich Site Summary) is one of the many formats that belongs to the family of web feeds. Interested users can then link to the URI for the feeds. With this simple Web feed technique, on top of displaying your news on other sites and headline viewers, RSS data can flow into other products and services like PDA's, cell phones, and many others. Some tools such as Yahoo! pipes provides means of aggregating feeds. Together with

Atom, they are the leading technologies about feeds publishing on the Internet. Despite their widely admitted limitations we adopted RSS here for the sake of simplicity.

Amazon Web Services (AWS): Amazon is well known for a vast experience in offering services especially for their online store. They decided to extend this experience to web application developers and in July 2002 they launched *AWS*. The aim was to provide a set of infrastructure services that together form a reliable, scalable, and inexpensive computing platform “in the cloud”. *AWS* is a collection of web services offered by Amazon.com over the internet. These services include:

- Amazon Elastic Compute Cloud (*Amazon EC2*), a web service that provides resizable compute capacity in the cloud.
- *Amazon SimpleDB*, a web service providing the core database functions of data indexing and querying.
- Amazon Simple Storage (*Amazon S3*) is storage for the Internet.
- *Amazon CloudFront*, a web service for content delivery.
- Amazon Simple Queue Service (*Amazon SQS*) which offers a reliable, highly scalable, hosted queue for storing messages as they travel between computers.
- *AWS Premium Support* which is a one-on-one, fast-response support channel to help you build and run applications on *AWS* Infrastructure Services.

They are all designed to make web-scale computing easier for developers. These services can be accessed over *HTTP* using *REST* and *SOAP* interfaces. *AWS* offered on a per usage billing basis. For example it costs \$0.180 per GB for the first 50 TB / month of storage used. In order to utilize these services, you need to sign up with Amazon giving credit card information before you can access the services. For purposes of this contest we demonstrate the use of *Amazon S3* to avail a digital learning repository *DLR* which can be used to build e-learning applications.

Amazon Simple Storage Service (Amazon S3)

This service provides unlimited storage on the internet through a simple web service interface. Storage is organized in terms of buckets which contain objects. A user may own up to 100. In any bucket, you may write, read, and delete objects from 1 byte to 5 gigabytes in size with accompanying metadata. There is no fixed limit on the number of objects you can store. Buckets and objects can be created, listed, and retrieved using either a REST-style HTTP interface or a SOAP interface. Currently most ftp clients support manipulation of Amazon S3 buckets. For the *DLR* we used the GUI-based *bucketexplorer* and *S3 Fox* extension of Firefox. The three concepts are buckets, objects and keys.

Buckets A bucket is simply a container for objects stored in Amazon S3. Every object is contained within a bucket. For example, if the object named `homeindex.html` is stored in the `courseA` bucket, then it is addressable using the URL `http://courseA.s3.amazonaws.com/homeindex.xhtml`. Buckets serve several purposes: they organize the Amazon S3 namespace at the highest level, they identify the

account responsible for storage and data transfer charges, they play a role in access control, and they serve as the unit of aggregation for usage reporting.

Objects Objects consist of object data and metadata. The data portion is opaque to Amazon S3. The metadata is a set of name-value pairs that describe the object. These include some default metadata such as the date last modified, and standard HTTP metadata such as Content-Type. The developer can also specify custom metadata at the time the Object is stored.

Keys A key is the unique identifier for an object within a bucket. Every object in a bucket has exactly one key. Since a bucket and key together uniquely identify each object, Amazon S3 can be thought of as a basic data map between «bucket + key» and the object itself. Every object in Amazon S3 can be uniquely addressed through the combination of the Service endpoint, bucket name, and key, as in <http://courseA.s3.amazonaws.com/homeindex.xhtml>, where «courseA» is the name of the bucket, and «homeindex.xhtml» is the key.

Amazon provides technical documentation about the storage service.

Yahoo! Pipes

In mid 2006, in order to cope with the growing interest in social networking and Web services, Yahoo! released Yahoo! Pipes. Yahoo! Pipes is a web application that makes it easier to find information in an RSS feed. We use one of the applications for the contest (see section 2) for *Cinnamon*, a virtual reading group web application to demonstrate *Yahoo! Pipes*.

The remainder of this paper is organized as follows. In section 2 we describe the applications and tools used stating the role of the students for each of the applications. Section 3 discusses how the contest was organized. Section 4 gives the challenges we faced during the SOA awareness and section 5 concludes the paper.

2 Applications and tools used

For the purpose of the contest, we introduced two widely adopted tools: *Yahoo! Pipes* and *Amazon S3* (part of the AWS family). The reader should recall an overview of these tools in section 1.1. We developed three applications for purposes of the contest. Other than utilization of SOA technologies for their own sake, we targeted applications that could be developed in a relatively short time, interesting and easy to work with by the students, and extendable into fully fledged applications. In this section we describe each of the applications and we state the expected role of the students for each application.

2.1 Virtual Reading Group

In most research centers, a reading group is one where researchers share their recent readings and views about a topic of interest. In its usual setting, it is an active group where members meet periodically. In this application (*Cinnamon*), we intended to create a virtual one where users can post comments and questions about their latest readings. Other users can contribute as well. The challenge here consisted of finding discussion tracks following different criteria.

2.1.1 Basic functionality of Cinnamon

Cinnamon is a WebApp to support discussions in a research. Generally, members of a research group have to share their ideas and insight about research topics. *Cinnamon* creates a virtual community to support this kind of activity. *Cinnamon* has been developed using two open source Web development frameworks, *Sproutcore* and *RubyonRails* also known as *Rails*. *Sproutcore* (*Cocoa*) is the *Objective-C* based programming environment for Mac OS.-inspired javascript framework for creating Web applications that look and feel like Desktop applications (see *AppleInsider* blog). On the other hand *Rails* is a fully fledged framework to develop webapp using the Ruby programming language. Both frameworks support MVC (model-view-controller), AJAX, etc.

As depicted in figure 1, a user willing to use application should first log into the system. New users can join by signing up, using the same page. They just need to check the signup button as shown in figure 2. Note that the email address is validated against the usual email address pattern.

Figure 1: Overview of Cinnamon: Login Phase

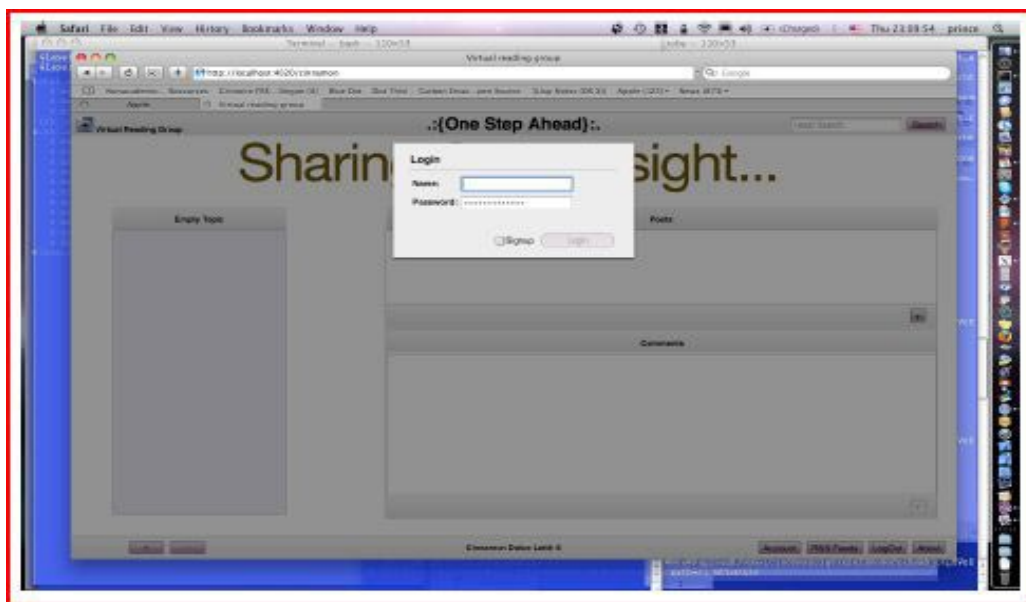


Figure 2: Overview of Cinnamon: Signup

Login

Name:

Password:

Confirm Password:

First Name:

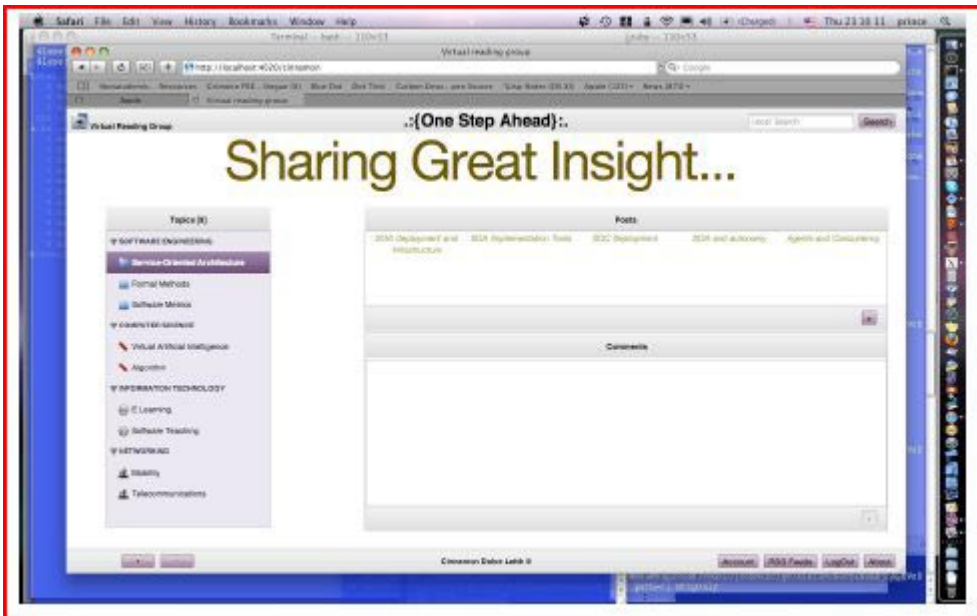
Last Name:

Email Address:

Signup

After logging or signing up successfully an interface as depicted in figure 3 appears.

Figure 3: Cinnamon after login



This interface allows to do one of the following:

- Create new topics or delete existing ones. Note that a topic is a concept discussions in a research group revolve around. Topics can be searched for using the search text field and button in the upper right side of the page. When the topic is found it is automatically selected and only post related to that topic will appear.

- For a selected topic, create, modify or delete entries. An entry is a post. It is considered as an entry point in a discussion track. An entry has a title and a content. It can be backed by references (authors, title of the resource, event where it appeared or was delivered, etc.).
- For each entry, participants of a research group can add comments.
- Besides, a user can also update their account information by clicking on the Account button at the bottom of the page. They find out about new version (and any other news about Cinnamon) by clicking the About button. Finally, they can generate RSS feeds, a summary of the activity going on in the research group.

2.1.2 Feeds Generation

To generate an RSS feed, a user simply clicks on the RSS Feeds button. Sproutcore then sends a (RESTful) AJAX request to the Rails server. This triggers the generation of the feeds in “public/feeds/RSS_feed.xml” on the Rails side. Originally, the structure we would like for our feeds is depicted as follows.

```
<!-- we do not display feeds header here -->
<posts>
  <post title="" author="" topic="">
    <article>entry content</article>
    <comments>
      <comment author="">comment</comment>
      <comment author="">comment</comment>...
    </comments>
  </post>...
</posts>
```

However, because we had to comply with RSS specifications, we flattened the feeds to the following model.

```
<?xml version="1.0"?>
<rss version="2.0">
  <channel>
    <item>
      <title>News for September the Second</title>
      <link>http://example.com/2002/09/01</link>
      <description>other things happened today</description>
    </item>
    <item>
      <title>News for September the First</title>
      <link>http://example.com/2002/09/02</link>
    </item>
```

```
</channel>
```

</rss>In the following section, we introduce *Yahoo! Pipes* using the simplified version of our feeds.

2.1.3 Yahoo! Pipes and Cinnamon

Yahoo! Pipes offers the possibility to fetch information from different sources, compose them in a way suitable to ones own needs and publish the output. The WebApp comes with an editor. You need a Yahoo! account to use the editor.

To create a pipe, you will use the pipe editor offered by Yahoo! . The editor consists of three elements: *Library*, *Canvas* and *Debugger*. The library contains a set of modules. A module performs a specific task, e.g., treating data sources, translating items from one language to another, etc. The canvas is the working area for assembling and testing the pipes. Finally, the debugger offers you the possibility to inspect the output of a pipe.

Figure 4 depicts an excerpt of the feeds generated in Cinnamon.

Figure 4: Cinnamon: RSS Feeds

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0">
  <channel>
    <title>Cinnamon:Virtual Reading Group</title>
    <link>http://www.cit.ac.uk</link>
    <language>en-us</language>
    <item>
      <title>SOA Deployment and Infrastructure</title>
      <description>Just want to edit this entry as a test</description>
      <author>Michael Scott</author>
    </item>
    <item>
      <title>SOA Implementation Tools</title>
      <description></description>
      <author>Ryan Howard</author>
    </item>
    <item>
      <title>Static Analysis vs Model Checking</title>
      <description></description>
      <author>Dwight Schrute</author>
    </item>
    <item>
      <title>SOC Deployment</title>
      <description>How do we make sure about the safety of the various we use to deploy an SOA-based system</description>
      <author>Michael Scott</author>
    </item>
    <item>
      <title>SOA and autonomy</title>
      <description>Have you ever thought about web-services auto-organizing themselves</description>
      <author>Michael Scott</author>
    </item>
    <item>
      <title>Agents and Concurrency</title>
      <description>Hi folks. Is it possible to borrow ideas from the concurrency world and apply them to the notion of
agent</description>
      <author>Michael Scott</author>
    </item>
    <item>
      <title>Model Checking</title>
      <description>How to reconcile model checking techniques with new program analysis approaches</description>
      <author>Michael Scott</author>
    </item>
    <item>
      <title>Static Program Analysis</title>
      <description>I know static program analysis sucks because of the state space. But what if we bring in new ideas from
simulation</description>
      <author>Michael Scott</author>
    </item>
  </channel>
</rss>
```

To create a pipe displaying information about the discussion tracks in Cinnamon, you can follow the steps as discussed below.

- select the *Fetch* module and enter the URL of Cinnamon RSS feeds. As you may have guessed, this module reads RSS feeds from the URL provided.
- to filter the information gathered from Cinnamon, you can add a filter. The latter defines rules to block or permit feeds related to keywords entered in the appropriate text field.
- finally, you can connect the output of the filter to the input of the output pipe. The connections are similar to composing services: input, output and constraints on the possible composition.
- you can now save the pipe and add it to your Yahoo! page. Note that search engines can search your pipes. Therefore, we recommend you add properties to your pipe that facilitate the search. These are name, description and tags.

2.2 E-Notice Board

In this application, there is an electronic board, where notices are displayed for some time. A notice can be a textual information (pdf, doc, etc), a sound or a video. A user can register for some type of notice and be notified once they are available on the board. The challenge in this contest is to find different notices, which are interrelated given a topic-name or keyword and which are still temporarily valid. The task of the participants was limited to extracting the RSS feeds to combine them into a single *notice board* like application. The fundamental principle of service-orientation is seen in terms of the autonomous nature of the feeds and ability to compose the feeds. The composition of simple RSS feeds provides a good basis to explain two concepts. First, it explains the concept of composing services into more complex services. Second, it explains the concept of exposing a service as a resource for use.

Further, the fact that a single feed can be used by different notice board applications puts emphasis on the concept and advantages of reuse. Reusability is one of the fundamental concepts in service-orientation. Moreover, the subscribe-push concept of feeds is good basis to introduce the need for advanced forms of registry and discovery techniques including semantic capabilities to enable full automation of services. More details on the need for semantic capabilities can be found in [McIlraith and Son, 2002, Fensel *et. al.*, 2002]. In figure 5, we show the e-notice board resources to be extracted and composed into a notice board application with form factor as in the bottom right part of the figure.

Figure 5: e-noticeboard Resources



Information Resources

MAKERERE UNIVERSITY, FACULTY OF COMPUTING & IT

FEEDS

	General notices and information
	Information and news for undergraduate students
	Information for Post graduate students

my notices

2.3 Digital Learning Repository(DLR)

DLR is a digital repository which provides services to users seeking specific learning contents of their choice. It accomplishes the latter by acquiring contents deposited into a repository built on an ontological understanding of the learning object (LO) concept [Hamel and Ryan-Jones, 2002, Weitz *et. al.*, 2004]. Content developers therefore design content to suit the LO with varying sizes, formats, quality and educational values. The learning objects used for the contest are XHTML files. We used Amazon S3 to store the different learning objects. A bucket contains all the learning objects related to a single subject. These learning objects can be accessed using URLs. The challenge was to find the exact and appropriate content for e-learning for both a tutor and a student.

3 SOA Contest Organization

We started by developing the backend of all the applications. The students role would then be to come up with the front end of a chosen application. The students were allowed to work in groups.

The reader should remember that Cinnamon was developed using *Rails* and *Sproutcore* web application development frameworks. For the e-Notice Board, the resources were provided statically using standard RSS format. For DLR, we created the learning objects as XHTML files and we employed the Amazon S3 service for storage. The Amazon S3 service provides a way of accessing services using technologies and techniques that are commonly used in web services.

We created a website www.cit.ac.ug/soato to provide detailed information about the contest. It is through this website that contestants registered for the contest, got

documentation about the different tools, technologies and applications related to the contest. Here they also got technical support from the organizers.

The contestants were given a duration of 1 month to work with one of the applications. The contestants were allowed to upload the most current versions of their applications. This allowed us to monitor the progress of the contestants, thereby providing necessary technical support.

4 Challenges

In this section we highlight challenges faced during the SOA awareness contest. We divide the challenges into technical challenges, social and economic challenges.

4.1 Technical Challenges

The following challenges that we met are common in developing countries. However, in order to benefit fully from the use of SOA technologies for application development, it is important that they are resolved.

- i. Internet Services:* It was difficult to work with technologies that employ the internet with very low bandwidth available in most universities. For example the amazon infrastructure is a purely web-based infrastructure which requires good internet. The internet services were intermittent. Therefore completion of long internet transactions was almost impossible. In order to carry out the contest we had to pay up in internet cafes to complete some of the online transactions.
- ii. Connectivity:* Remote connection to and through the local network was difficult. This restricted participants from working off the university site. For example it was not possible to connect using *Secure Shell (SSH)*. We only managed to resolve this with Makerere university which allowed temporary remote connection for the sake of the contest.
- iii. Physical infrastructure:* Some of the applications required the use of a server and it was challenging accessing one. A server allocated for use during the contest but it was done late and this lengthened the time set for the contest. *Note:* The use of some SOA technologies requires skilled manpower. Therefore it is important that skilled manpower is available before planning for use of SOA technologies for application development.

4.2 Social and Economic Challenges

- i. Choice of applications:* The first challenge was to craft suitable applications for students with less exposure to Information Technology (IT). For instance students from the School of Industrial Fine Art expressed interest, but it was not easy to explain an advanced IT concept (SOA) due to low IT essential skills. This led us to restrict the awareness to students with a background in IT.
- ii. Standardization:* For some technologies the way to consume and use services differed from popular techniques in web services. Therefore more time was used to understand the same concepts.

- iii. *Collaborative communication*: The organization was mainly virtual because of the holiday season. However in a developing environment, working virtually is met with unpredictable situations for example power outages, low bandwidth, lack of remote access and so forth. Therefore it was a challenge to utilize virtual communication to coordinate project activities.
- iv. *Publicity*: SOA being a new concept in the contest environment, it was challenging to encourage students to participate in the contest. We used prizes to attract participants to the contest.

5 Conclusion

There are several services that are offered via the web (SOA) that can be appropriately utilized for ICT in a developing world. The services are both available as proprietary and open source for application development. In this paper we discussed a contest conducted to serve as an awareness campaign for SOA technologies as an alternative to application development. The contest targeted students from all universities in Uganda. However, only students from Makerere university and Mukono University showed interest and participated. In future we would like to find out why the other students did not respond at all to such a campaign. At the end of the contest, the participants were more knowledgeable about SOA in general and SOA technologies in particular. This they demonstrated when they presented their work for evaluation as part of the contest. From the contest, it was clear that despite the various challenges (intermittent internet especially) that are inherently faced by Uganda as a developing country, training in SOA technologies can be done to build enough skill for future use for application development.

The reader should note that a development environment for SOA-based applications requires various components for service request, service discovery and selection, service composition and deployment. However, the discussion of these components is beyond the scope of this paper.

Acknowledgements

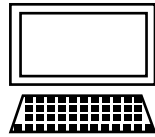
We wish to acknowledge Amazon.com which allowed us to use AmazonS3 for the contest without paying. We also acknowledge Mr. Maxwell Omwenga who developed and maintained the SOA awareness website.

References

- BOX D. , EHNEBUSKE, D., KAKIVAYA, G., LAYMAN, A., MENDELSON, N. , NIELSEN, H. F., THATTE, S. AND WINER. D. (2003). Simple object access protocol (soap). <http://w3c.org/TR/SOAP>, Version 1.2.
- R. CHINNICI, J.-J. MOREAU, A. RYMAN, AND S. WEERAWARANA. Web services description language (wsdl) version 2.0 part 1: Core language. <http://www.w3.org/TR/wsdl20/>, 2007. W3C Recommendation 26 June 2007.
- F. CURBERA, M. J. DUFTLER, R. KHALAF, W. NAGY, N. MUKHI, AND S. WEERAWARANA. Unraveling the web services web: An introduction to soap, wsdl, and uddi. *IEEE Distributed Systems Online*, 3(4), 2002.

- D. FENSEL, C. BUSSLER, AND A. MAEDCHE. Semantic web enabled web services. In *International Semantic Web Conference*, pages 1–2, 2002.
- R. T. FIELDING. *Architectural Styles and the Design of Network Based Software Architectures*. PhD thesis, University of California at Irvine, 2000.
- C. J. HAMEL AND D. RYAN-JONES. Designing instruction with learning objects. *International Journal of Educational Technology*, 3(1), 2002.
- S. MCILRAITH AND T. C. SON. Adapting golog for composition of semantic web services. In *In Proceedings of the Eighth International Conference on Knowledge Representation and Reasoning(KR2002), Toulouse, France*, pages 482–493, 2002.
- F. WEITL, R. KAMMERL, AND M. GÖSTL. Context aware reuse of learning resources. In *Proceedings of ED-MEDIA 2004, World Conference on Educational Multimedia, Hypermedia & Telecommunications*, Lugano, Switzerland, June 2004. angenommen als Konferenzbeitrag.

PART FOUR



Data Communication and Computer Networks

1

(Inter)net Neutrality: Your Voice Matters

Joseph Kizza Migga

The growing use of the Internet coupled with the rapid growth in applications' bandwidth requirements are pushing network owners (ISP) to a point where they can no longer treat all Internet content, sites, and platforms equally. This in turn is leading providers to start employing proprietary protocols or to enter into exclusivity agreements with content providers that may reduce the transparency and hence the neutrality of the Internet. Current network neutrality rules forbid network operators to discriminate against third-party applications, content or portals or to exclude them from their network. However, there is mounting evidence that providers are circumventing these rules. The debate has erupted for advocates of more network neutrality regulation to stop this from escalating and there are those opposed to new rules for net neutrality. Those supporting a strong net neutrality argue that in the absence of network neutrality regulation, there is a real threat that network providers will discriminate against independent producers of applications, content or portals or exclude them from their network. This threat then will reduce the amount of innovation in the markets for applications, content and portals at significant costs to society.

Those opposed to net neutrality argue that the regulatory tools needed to implement network neutrality are likely to prove ineffective in a world in which communications technologies are increasingly changing. Their most important argument though is that network neutrality threatens to make things worse by reinforcing the sources of market failure in the last mile and dampening incentives to invest in alternative network capacity.

The paper explores and advances the debate over network neutrality, highlighting important limitations of both arguments and showing other alternatives.

1. Introduction

Net neutrality as a principle or as a goal is a very controversial but interesting issue of discussion today. As a society, we are getting more and more included in the information revolution. Society, in varying degrees, is getting more and more dependant on computing systems and automated devices for day to day life tasks. Most of these devises and the information we get have all come to depend on the internet. The future, for those of you who care to think ahead, is going to depend more on the evolving internet of today. We are likely to success or fail based on the success or the failure of the internet. Yes we have a lot to think about the future of this internet. We should and indeed must care about who controls it, will be denied equal access, what privileges and rights will we have and what content of the internet will we be allowed to own and use? All these concerns and more make up what is net neutrality. Net neutrality has varying definitions depending on the person you ask based on their philosophy and expectations of the future internet. Many people involved enough today and understand the risks and involvement of society with the internet, tend to be advocates of net neutrality

and the definition they will give you will reflect their shade. There are two sides to the debate. On one side of the net neutrality fence are the telcos and cable companies that are investing big bucks to offer consumers access to their broadband networks. On the other side are the application providers who are responsible for creating Internet content – “the Google’s, Yahoo’s and eBay’s of the world. Based on these camps, three definitions of net neutrality emerge according to Wikipedia [2009].

The first definition of network neutrality deals with the network design principle based on the idea that the internet is a useful public information network, just like ports and railways lines, whose content, sites, and platforms should be treated equally and every user should have unrestricted access. It emphasizes that internet users should be in control of what content they view and what applications they use while on the Internet. The internet has operated according to this neutrality principle since its earliest days. Fundamentally, net neutrality is about equal access to the Internet. No one, including broadband carriers should have sole, even partial control of the internet and be in position to detect terms of use of the internet.

The second definition focuses on the quality of service (QoS) of the internet. Internet users and applications come in different categories. There are those who use a lot of bandwidth and there are those with applications that are real time like streaming data and video that require pre-reservation of the bandwidth. For all these to work peacefully and successfully, when there is congestion in the network and network capacity is insufficient, there is a need for limited discrimination of users and applications based on their need. Quality of service in computer networking deals with resource reservation control mechanisms under network conditions when there is congestion. It is the ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow. For example, quality of service guarantees are important if the network capacity is insufficient, especially for real-time streaming multimedia applications such as voice over IP and online games since these often require fixed bit rate and are delay sensitive, and in networks where the capacity is a limited resource, for example in cellular data communication.

The third definition also involves quality of service but this time allowing for a higher charge on those requiring a higher QoS as long as there is no exclusivity in service contracts.

The rest of the paper is as follows. Section 2 we give the background to the net neutrality debate, section 3 we define and explain net neutrality, section 4 we give a detailed discussion and outline and analyze the issues in the net neutrality debate including suggestions and useful alternative solutions to those advocated by the opposing sides. Finally section 5 is the conclusion.

2. Background

The history of net neutrality is not long. In fact, the phrase “net neutrality” (or “network neutrality”) did not come into use until somewhere in the early 2000s. There are those who argue that although the phrase may not have been there but the idea and principle of net neutrality were there. For example, the principle of public carrier neutrality

was bestowed by the U.S. Congress on all public carriers and infrastructure like roads, canals, electric grids, trains, and telecommunications. Since the 17th century, there has been regulations in U.S. that forbid discrimination on transport networks that serve the public interests because they catalyze entire industries. It was deemed that any type of discrimination on such networks would have ripple effects across the nation. The same logic now applies to the internet because it is a public carrier. For generations, these regulations have been ensuring fairness in the market place and protecting the consumer.

The internet was started by the United States Department of Defense through funded research and it was initially governed by an Acceptable Use Policy (AUP) prohibiting commercial activity. It was first privatized and the AUP lifted for commercial users in the early 1990s. The privatization of the internet and the rifting of AUP, immediately created an internet driven by market forces with limited to no central control. In its early years, the issue of QoS did not affect the Internet. The issues of QoS first appeared in the Internet2 project but did not receive that much attention until 2003 when Prof. Tim Wu of Harvard Law School, published a proposal for a net neutrality rule in which he considered network neutrality in terms of neutrality between applications, as well as neutrality between data and QoS-sensitive traffic, and proposed some legislation to potentially deal with these issues [Wu, 2003]. Based on the discussion generated by Wu's paper, in 2005 the U.S. Federal Communications Commission (FCC) introduced and enforced the network neutrality principles in a documented case of abuse involving Madison River Communications, a small DSL (Digital Subscriber Line) provider that briefly blocked VoIP (Voice over Internet Protocol telephone) service Wikipedia [2009]. In the same year, the FCC further adopted a policy statement in which it adhered to the four principles of network neutrality. The four principles were [Federal Communication Commission]:

- *To encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet, consumers are entitled to access the lawful Internet content of their choice.*
- *To encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet, consumers are entitled to run applications and use services of their choice, subject to the needs of law enforcement.*
- *To encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet, consumers are entitled to connect their choice of legal devices that do not harm the network.*
- *To encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet, consumers are entitled to competition among network providers, application and service providers, and content providers.*

On the legislative side in U.S., there has been several bills in Congress with varying success. Many failed. Some of the attempted bills included:

- *Internet Freedom Preservation Act of 2008*, contains no requirements for regulations on the Internet whatsoever. It does, however, suggest that the principles which have guided the Internet's development and expansion are highly worthy of retention, and it seeks to enshrine such principles in the law as guide stars for U.S. broadband policy. This is pending.
- The *Internet Freedom and Nondiscrimination Act of 2006* that would have made it a violation of the Clayton Antitrust Act for broadband providers to discriminate against any web traffic, refuse to connect to other providers, or block or impair specific (legal) content. It would also have prohibited the use of admission control to determine network traffic priority [Wikipedia, 2009]. This Act, however, failed to become law.
- The Communications *Opportunity, Promotion and Enhancement Act of 2006*, referencing the principles enunciated by the FCC and authorizing fines up to U.S. \$750,000 for infractions, was passed by the full House of Representatives but failed to pass in the U.S. Senate [Wikipedia, 2009].

After many legislative failures in the U.S. Congress to pass legislation to protect net neutrality, its advocates took to courts. This led to several landmark court rulings on net neutrality including:

- The Comcast, a U.S. Cable Television and a major internet provider was sued and was found to be blocking or severely delaying *BitTorrent* (bitTorrent is a peer-to-peer (P2P) file sharing communications protocol) uploads on their network using a technique which involved the network creating 'reset' packets (TCP RST) that appeared to come from the other party [Cheng, 2007]. Comcast reached an out court settlement and agreed to adopt a protocol-neutral stance and explore ways to "more effectively manage traffic on its network at peak times." [Wikipedia, 2009]. This case, and others, led the FCC to issue a ruling that Internet service providers (ISPs) cannot ration service to heavy users of the Internet, indicating that Comcast broke the law when it slowed the transfer of video files among a group of its customers to ensure that other customers had adequate bandwidth. This ruling and others following it has given net neutrality advocates hope that the net will remain free for some time, however long that is.

3. The Notion of Net Neutrality

The notion of net neutrality is comprised of three distinct issues, as defined by Information Technology and Innovation Foundation (ITIF): *Transparency*, *Blocking* and *Tiering*. *Transparency* deals with clarity of broadband provider usage policies, an issue that has yet to gain full public attention. *Blocking* relates to the practice of degrading or blocking consumer access to content and applications. Finally *tiering* addresses the question of whether broadband providers can charge content providers for access to their networks [Barry, 2006]. These issues are at the core of the net neutrality debate.

4. The Debate About Net Neutrality

There is no question that the explosive use of the internet has resulted in headaches for internet service providers (ISPs), usually big telecommunication companies. Together with the explosive use of the Internet, there is a growing, stochastically changing and unpredictable user needs in content and bandwidth use. This is creating serious and unpredictable congestion problems in the network. Thus the call for the way to manage this problem. This had led to the QoS issue in the debate for net neutrality. The question is how to create fairness in bandwidth use when there are varying interests and motives of users?

4.1 The Role of Net Quality of Service (QoS)

Recall from the introduction that quality of service within the computer networking communities as a set of internet resource reservation control mechanisms under network conditions when there is congestion. It can also be taken as the ability to provide different priorities to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow.

In an ideal internet working environment, all users and applications would be able to request for services requiring any amount of bandwidth and they would get them with no questions asked. However, in the real world under which the current internet is operating, nothing like that exists. First bandwidth is limited. Second, the needs of users and applications are growing and the amount of bandwidth required for those applications is growing exponentially. In fact providers are currently faced with a growing list of applications that are bandwidth guzzlers including [Wikipedia, 2009]:

- streaming multimedia
- IPTV (Internet Protocol Television) - a system where a digital television service is delivered using Internet Protocol over a network infrastructure, which may include delivery by a broadband connection.
- IP telephony or Voice over IP (VOIP) which may require strict limits on jitter and delay
- Video Conferencing (VTC) which normally requires low jitter and latency
- Alarm signaling such as Burglar alarms
- Dedicated link emulation which requires both guaranteed throughput and imposes limits on maximum delay and jitter
- A safety-critical application, such as remote surgery which may require a guaranteed level of availability.
- Online gaming, such as fast paced real time simulations with multiple players.

All these types of service, called *inelastic* because they cannot scale to whatever bandwidth is available and run, require a certain minimum level of bandwidth and a certain maximum latency to function. In cases like these, the providers are faced with serious bandwidth management problems and a need for good QoS.

Delivering and maintaining good QoS to meet user needs outlined above, may require several approaches including [Wikipedia, 2009]:

- An advance contractual agreement called a service level agreement (SLA) between network users and providers which specifies guarantees for the ability of a network/protocol to give guaranteed performance/throughput/latency bounds based on mutually agreed measures, usually by prioritizing traffic.
- A process resource reserving system similar to the Resource ReSerVation Protocol (RSVP), described in RFC 2205, in TCP to reserve resources across a network for an integrated services internet.
- Traffic/packet shaping to control network traffic in order to optimize or guarantee performance, lower latency, and/or increase usable bandwidth by delaying packets that meet certain criteria.
- Scheduling algorithms like weighted fair queuing (WFQ), class based weighted fair queuing, weighted round robin (WRR), deficit weighted round robin (DWRR), and Hierarchical Fair Service Curve (HFSC) to provide access to threads, processes or data flow to system resources like processor time, communications bandwidth, and others.
- Congestion avoidance techniques like policing (marking/dropping the packet in excess of the committed traffic rate and burst size), explicit congestion notification (ECN) and buffer tuning.

4.2 Arguments for network neutrality

After understanding what is at stake in the net neutrality debate, let us now look at the current arguments in the debate. Those who advocate for network neutrality have several issues that they fear may mean the end of net neutrality according to William [2006]. These include:

4.2.1 Content Blocking

Proponents of net neutrality content that with it, ISPs would become Internet gatekeepers, deciding at will which application or user to go first and fast which may lead to unfairly discriminate and control which data they prioritize, such as data from their own sponsors or media interests. This is not an isolated or hyped fear; there have been cases in the U.S. where this has happened in including [Wikipedia, 2009]

- *The Madison River Experience* - In the Madison River experience, the Madison River Communications' customer Doug Herring was traveling through Tennessee in November of 2004, Mr. Herring attempted to call his wife at their Alabama home but could not get through. Mr. Herring's frustration turned to outrage when he discovered the reason behind his inability to call home Madison River Communications, his DSL provider, had instituted a policy of blocking rival Voice-Over-Internet Protocol (VoIP) Internet phone services [William, 2006].

- In April 2006, Time Warner's AOL blocked all emails that mentioned www.dearaol.com, an advocacy campaign opposing the company's pay-to-send e-mail scheme. An AOL spokesman called the issue an unintentional glitch [ZDNet News, 2006].
- In February 2006, some of Cox Cable's customers were unable to access Craig's List because of a confluence of a software bug in the Authentium personal firewall distributed by Cox Cable to improve customers' security and the way that Craig's List had their servers misconfigured. Save the Internet said this was an intentional act on the part of Cox Cable to protect classified ad services offered by its partners. The issue was resolved by correction of the software as well as a change in the network configuration used by Craig's List [Matt, 2006].
- In September 2007, Verizon Wireless prevented a pro-choice organization from sending text messages to its members coordinating a public demonstration, despite the fact that the intended recipients had explicitly signed up to receive such messages [New York Times, 2007].
- In October 2007, Comcast was found to be preventing or at least severely delaying uploads on BitTorrent.

4.2.2 *The Move toward a Two-Tiered Internet*

There is also fear of the emergence of a “two-tiered” Internet or an Internet in which select content and applications would be offered at higher speeds. Although this concept is not new, however, it has been, of late, moved to the fore front of the debate. This is a result of DSL providers started to aggressively press for U.S. congressional approval of a two-tier scheme that will use one tier with a fee-based reserved pipe to provide faster platform for content and application companies. The other tier (slower) will be for the public [William, 2006].

4.3 Arguments against network neutrality regulations

As would be expected, net neutrality is mostly supported by intended user groups and in various locations around the globe is arguing for regulations that will continue to protect net neutrality. However, this is not the case with many providers. Many of them, especially telecommunication companies, are opposed to net neutrality regulations claiming that these regulations may indirectly prevent the expansion and improvement of Internet access for their customers, who are demanding an increasing amount of bandwidth. In fact there many valid reasons for a non-neutral network including [William, 2006]:

4.3.1 *Blocking Content and A Two-Tiered Internet May not Solve All Problems*

Both of these methods would definitely drop competition in the broadband Internet access market, posing the danger of limited market power resulting in loosing customers. Losing customers results in long term profit losses and a low return on provider investments.

4.3.2 *The Benefits of a Non-Neutral Internet*

There are several customer benefits to a non-neutral network including [Matt, 2006]:

Broadband Expansion

Many countries around the world are trying or planning to invest an unprecedented amount of money to bring about universal broadband to their citizens. There are huge national benefits for this. In U.S. it is estimated that this would bring in an estimated \$300 billion a year [William, 2006]. The amount of investment into this venture may require billions of U.S. dollars. This money either must come from government or through private investments. Assuming private investment is the preferable, there must be enough investor incentives to be able to raise this amount of money to build the networks. Blocking and regulations are not these incentives.

Other Benefits from a Non-Neutral Net

With a neutral network, there is likely to be additional competition in the application/content market. Such competition may lead to lower prices, increased variety, and a larger incentive for companies to innovate. This lack of incentives threatens to make things worse by reinforcing the sources of market failure in the last mile and dampening drive to invest in alternative network capacity.

4.3.3 *Inadequate Technology to Maintain a Neutral Network*

There is a strong belief that the regulatory tools needed to implement network neutrality are likely to prove ineffective in a world in which communications technologies are increasingly changing.

4.4 Alternatives to the net neutrality debate

Not everyone is in one of the last two opposing camps. For example many U.S. technology trade associations have remained noncommittal on the issue and a large chunk of U.S. internet users; either do not want to pay attention or are not aware of what is going on. Outside the U.S., there is little to no debate about net neutrality for a variety of reasons.

Among those who are following what is going on but are not interested in taking a side are proposing a variety of mid-of-the road alternatives to those proposed by the two opposing camps. For example, according to Wu [2003], total prohibition is bad. There is a need for some degree of discrimination but we need to have a grip on good and bad discrimination. Some of their proposals include:

4.4.1 *Eminent Domain Against ISPs*

The thread of eminent domain against ISPs, according to Andy Kessler, is very effect and much better than new legislation [Kessler, 2006]. In eminent domain, the state possesses the power over all property within the state and it can appropriate property for a public use. In some jurisdictions, the state can delegates eminent domain power to certain public and private companies like utilities, such that they can bring eminent domain actions to run telephone, power, water, or gas lines. In most countries, however, the owner of any appropriated property is entitled to reasonable compensation based on the fair market value of the property.

4.4.2 *Creative Commons*

The concept of Creative Commons is to make web content free for all to use. This way, information can be shared anywhere in the world, at any time, by any one, with no worry of limitation. With Creative Commons we keep the Internet as an open democracy without placing regulations like both sides of the Net Neutrality debate want to do.

4.4.3 Government as “Net Nanny”

Another approach is to make the government be a “Net Nanny” and let it baby sit the internet arbitrating between the warring sides. This may require governments to pass some laws and let bodies like the U.S. FCC monitor and supervise the internet.

4.4.4 Managed Services

Under the managed services model, the ISPs must agree to offer a basic amount of open and non-discriminate data pipe access, then the provide would be allowed to charge for content that goes across their networks just like rail lines and ports do. The approach consists of a three-prong strategy that includes: effective consumer protection measures, sound competition policy oversight, and conditioned tax incentives [Barry, 2006].

- 1 For consumer protection, the government requires ISPs to state their broadband access usage policies in clear terms that specify level of bandwidth, amount of latency and limitations consumers may face in accessing content or services. The government regulatory body like the FCC would monitor these policies and take action against companies that fail to comply.
- 2 The competition model policy calls on governments or government agencies like the U.S. FCC to address competition policy issues after-the-fact and require the Commission to manage relevant proceedings on an expedited basis.

Offer financial incentives to companies that invest in broadband networks, as long as those companies offer consumers an open Internet pipe with speeds at least as fast as the evolving regulatory agency definition.

4.4.5 Structural Separation.

Structural Separation is a law that requires that no ISP may have any financial interest in any of the content carried by that network and that any network operator’s network management is for the sole purpose of running the network. This may keep regulating bodies out.

4.4.6 Use of Antitrust Laws

Where competition breaks down or does not exist at all like in a neutral network, antitrust laws can be used to provide consumers the necessary protection. For example in U.S., in markets where a broadband provider may enjoy monopoly power, the Sherman Act will provide a remedy for anti-competitive behavior and an incentive for the monopolist not to engage in harmful business practices. Also in markets where competition breaks down due to collusion among broadband providers, the Sherman Act again provides a remedy and steep punishment for the guilty parties. The presence of this legal mechanism to address potential harms in the absence of competition makes additional regulation

unnecessary [William, 2006].

4. Conclusion

We have introduced the debate of net neutrality. We surveyed the many valid reasons given by both the advocating and opposing sides of net neutrality. So far the debate seems to be pitting first the webes, large bandwidth intensive web companies and a few individual consumers against correspondingly wealthy telcos, large telecommunication companies, who are watering for a hand in the webes money pot. We propose that governments should come in and try to bring fairness to the playing field by introducing some regulations that wont make the telcos, the omniscient gatekeepers of the global internet upon which all of us have come to depend and will not make the webes haggle the limited bandwidth for their high paying customers,

5. References

- “AOL charged with blocking opponents’ e-mail”, ZDNet News (2006-04-13). Retrieved on 10 July 2006.
- BARRY, A. J. (2006). Net Neutrality Alternative Proposed, <http://www.pcmag.com/article2/0,2817,1970356,00.asp>
- CHENG, J. (2007). Evidence mounts that Comcast is targeting Bittorrent Traffic, ARS Technica. <http://arstechnica.com/news.ars/post/20071019-evidence-mounts-that-comcast-is-targeting-bittorrent-traffic.html>
- FEDERAL COMMUNICATIONS COMMISSION, http://fjallfoss.fcc.gov/edocs_public/attachmatch/FCC-05-151A1.pdf
- KESSLER, A. (2006). “Give Me Bandwidth...”, The Weekly Standard. Retrieved on 9 July 2006.
- MATT, S. (2006). Craigslist Blocked by Cox for Three Month. MyDD Direct Democracy. <http://www.mydd.com/story/2006/6/8/144357/7525>
- “VERIZON REJECTS TEXT MESSAGES FOR ABORTION RIGHTS GROUP”, New York Times (2007-09-27).
- WIKIPEDIA. http://en.wikipedia.org/wiki/Network_neutrality
- WILLIAM, G. L. (2006). The End of Net Neutrality . Duke Law and Technology Review, Duke L. & Tech Rev. 0015., <http://www.law.duke.edu/journals/dltr/articles/2006dltr0015.html>.
- WU, T. (2003). “Network Neutrality, Broadband Discrimination”. Journal of Telecommunications and High Technology Law 2: p.141. doi:10.2139/ssrn.388863. SSRN 388863.

2

Security Issues in Ambulatory Wireless Sensor Networks (AWSN): Security Vs Mobility

Drake Patrick Mirembe and Maybin Muyeba

This paper presents a high level design of a typical Ambulatory Wireless Sensor Network (AWSN) and its security requirements. It presents a theoretical analysis of the effects of mobility and dynamism of nodes in relation to the delivery of (authentication, reliability, data freshness, and integrity) services by routing protocols. The analysis focuses on the Ariadne protocol and uses the Random-Way Point as the mobility model. The delivery of security services is modeled in terms of packet delivery ratios, link re-connections, and end-to-end delays. The logical analysis reveals that, the rate of node mobility and dynamism affects the delivery of security services. The paper concludes with the emerging research directions, which form the basis of the future work. Note that this is work in progress.

1. Introduction

Globally there is a growing need to improve access to quality healthcare services. As a result, a number of research groups are exploring ways of integrating wireless sensor technologies in the delivery of health care services. Notable projects include CodeBlue [Shnayder et al. 2005], MobiHeath [MobiHealth 2008], and MyHeart Project [Phillips 2008]. The desire to develop ambulatory sensor technologies is fueled by growing populations and the increase in natural disasters in recent years [Cred 2008]. In disaster emergency response, the need for reliable information acquisition, processing and dissemination technologies is an important component in the response strategy and overall case management. The availability or unavailability of patient physiological data can mean either saving a life or losing it. These have been the driving force behind the development of wireless medical sensor networks in general.

In the Wireless Sensor Network (WSN) research community, the phrase “Medical Sensor Network” has been widely used to describe wearable wireless sensors, but we find the concept restrictive as it does not embrace other mobile devices like PDAs that interface with Wireless Sensors (WS). Thus, in this research a term “Ambulatory WSN” is coined to refer to an integral system of wireless (mobile or/and fixed) sensors and interface terminals (like PDAs) that enable care providers to monitor the physiological status of a patient.

Until recently, research on MSN or simply AWSN has focused on making the technologies work at the expense of security and reliability. AWSN areas that have received a lot of attention include; stimuli acquisition and analysis [MobiHealth 2008; Phillips 2008 and Shnayder et al. 2005], reliable power supply [Muneeb et al. 2006 and

Shnayder et al. 2005], and optimization of communication protocols [Jamal and Kamal 2004; Intanagonwiwat et al. 2003; Walters et al. 2006 and Yu et al. 2001]. Nevertheless, security of sensor networks in general has attracted a lot of attention [Karlof et al. 2004; Krishna et al. 2006; Lorincz et al. 2004; Sastry and Wagner 2004 and Warren et al. 2005], and a number of interesting security oriented protocols have been proposed including; SPINS [Perrig et al. 2002], INSENS [Deng et al. 2006], SAR [Seung et al. 2001], and Tiny Sec [Karlof et al. 2004].

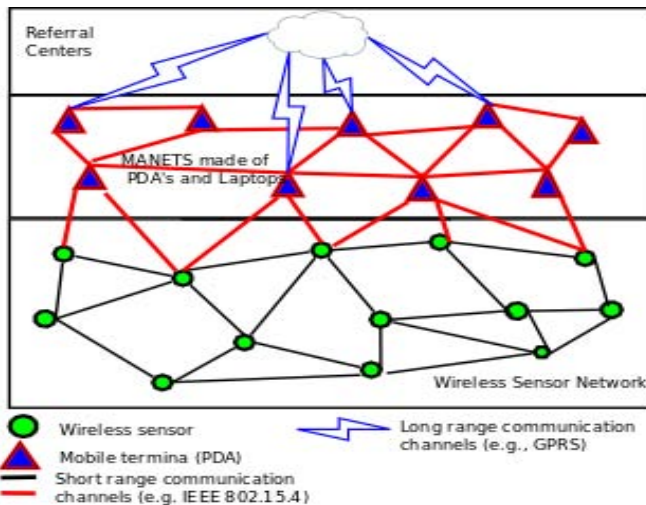
Note that, most of the current WSN security oriented protocols have focused on securing data generated and disseminated by static WSN. In theory, static WSN face different threats compared to mobile sensor networks like AWSN. The chances of compromising AWSN are increased by the mobility and dynamism of nodes. Therefore, there is need to develop new approaches such as those proposed in [Kambourakis et al. 2008 and Ng et al. 2006].

The rest of this paper is organized as follows; Section 2 presents an abstract scenario of the application and Section 3 discusses the effects of mobility on the delivery of security services by Ariadne protocol [Yih-Chun et al. 2002]. Section 4 presents the emerging issues and concludes the paper with future research directions.

2. Abstract Scenario

To help the reader visualize the application domain, an abstract scenario of various components of the service and key use case transactions are presented in Figure 2. The set-up described in figure 2 is based on a number of design decisions about AWSN service. Although most of the design decisions are based on user requirements, some are purely based on the perceived future deployment environment.

Fig 2: Abstract Scenario



Imagine a set-up where small light weight wireless sensors are attached to mobile objects under observation (e.g. patients in a disaster area or animals in a zoo). In such a set-up, care providers can use mobile terminals like PDA's, Smart phones or Laptops to read data from wireless sensors attached to the object over a wireless interface like Zigbee [Zigbee-Alliance 2008]. For the deployment model described above, the following design decision must hold;

1. Assume that objects under observation occupy a vast area, perhaps hundreds if not thousand of square meters.
2. That the effective communication range of wireless sensors is short (say <10m).
3. That network nodes are heterogeneous (consists of high performance and low performance nodes)
4. Nodes in the network obey the random mobility model.
5. Wireless sensor nodes far from mobile terminals (sink node) must relay their data over a mesh like network in order for their data to reach the intended recipients.
6. It is assumed that both Sender and Receiver nodes are mobile or one of them mobile and the other fixed.
7. Care Providers are envisioned to carry mobile terminals that can share data on peer-to-peer scheme, hence forming a MANET.
8. Wireless sensors can be lost.
9. The broadcast domain is assumed to be a natural environment with abstractions.
10. Sensors and Mobile terminals can not be trusted
11. Sensors deployed in the abstract scenario are assumed not to have tamper resistant hardware.
12. The communication model between nodes is multi cast;
 - i. Sensor to Sensor
 - ii. Sensor to Base Station (*high performance device e.g. PDA*)
 - iii. Base Station to Sensor
 - iv. Base Station to Base Station

In a typical disaster emergency response [Tia Gao et al. 2008], multiple care providers could be roaming over the field in which patients under observation are resting. As they move about in the field, their mobile terminals scan the environment at a certain frequency, locate wireless sensors, establish connections, download data, correlate and analyze it. Sensors relay their data to the mobile terminals either;

1. When an interesting event occurs
2. Periodically
3. Or when queried by users (mobile terminal requests).

In the emergency response deployment, patients are constantly being transferred from locations where they are resting to points of care, hence inducing mobility of sensor nodes as well. The fluidity of the network and the random mobility of nodes make securing data collected and relayed by nodes in the field difficult. Moreover the scenario could have multiple response teams, with no prior communication knowledge; hence the need for an ad hoc trust management scheme based on sound security protocols.

When long range wireless communication technologies like GPRS become available in the operation field, mobile terminals can then relay the correlated data to the response center server. While at the center, the data can be integrated in the general care management information system for future references.

2.1. Security Requirements

To understand the security requirements of a service, it is imperative to first identify the key stakeholders and assets that may need protection. The key stakeholders of AWSN service include; patients, care providers, technology suppliers and the authority. While key assets include; application data (sensor readings), control packets, wireless sensors, sensor application programs, mobile terminals, and cryptographic keys. Based on the above background, the following are the security requirements for the AWSN service;

1. Guarantee of reliability and availability of the service
2. Maintenance of authenticity of nodes and data
3. Preservation of integrity of data
4. Preservation of privacy of users and their data and
5. Authenticity of patients

In general all stakeholders are interested in the reliability of service. Patients would like to be assured that correct data about their physiological status is being reliably received by authentic care providers. While care providers and technology suppliers would not like to be litigated for failure of the technology and the authority is interested in quality of service. During the delivery of service, the care providers are interested in authenticity of data, an assurance that the readings on their screens are from patients they are interested in, NOT from fraudulent objects. Besides authenticity of data, patients and the authority are also interested in the preservation of patient privacy as recommended in the Health Insurance Portability and Accountability Act of 1996 (HIPAA) [US Gov. 2008]. All these requirements make the design and development of security protocols for AWSN a must have for successful deployment of the service.

2.2. Where to Deploy Security Services

Ambulatory Wireless Sensor Networks are unique from ordinary Mobile Ad Hoc Networks (MANET) in terms of device configuration, user requirements, deployment architectures, variety of technologies integrated, and nature of service provisioning among others. Therefore, the application of conventional MANET security mechanisms to secure AWSN services may not be applied unswervingly without modifications.

Hence two approaches have emerged in recent years;

1. Adaption of existing MANET security protocols to AWSN
2. Development new protocols to cater for the emerging needs and constraints of AWSN.

Whichever approach one chooses to follow, the fundamental question is at what level in the communication stack should security be enforced? Researchers for various reasons have made different preferences and each preference has a benefit and a cost.

In general, network and medium access implementation of security mechanisms are preferred because of their support of in-network. In-network processing is common in AWSN as it forms a basis of data aggregation. Deploying security mechanisms at lower layers help mitigate attacks before they can consume significant network resources. While researchers advocating upper layer security deployment (application layer) argue that, it is less expensive to implement, since it does not constrain mandatory processing of security services by lower layers. From Section 2 it is clear that AWSN integrates high performance devices (e.g, PDAs) and resource constrained devices like the MicaZ motes; thus security protocols defined for ordinary MANETS [MANETS WG 2008] e.g, Ariadne [Yih-Chun et al. 2002] could be scaled down to deliver security services in resource constrained motes as well, hence the motivation to focus on Ariadne. Besides, the choice of Ariadne was influenced by the fact that the protocol is built to work in dynamic ad hoc networks like AWSN. To our knowledge, until now Ariadne is the most comprehensive proposal for secure routing in mobile ad hoc networks like AWSN. Since Ariadne is based on Dynamic Source Routing protocol (DSR) [Johnson 1994] and TESLA [Perrig et al. 2002], its good to assume its effectiveness largely depends on efficient integration of the two modules. Note that the effectiveness of a routing protocol also depends on a Medium Access Control (MAC) layer protocols like MMAC [Ali et al. 2005], key management scheme, choice of cipher, and reliable time synchronization scheme.

Table 1: Where to implement security services

Communication Layer	Nature of protocol	Merits	Demerit
Application	Authentication and Authorization	Only applied on Demand.	Delay in processing, & increased power consumption
Network	Encryption, Reputation Schemes, Identity management and Authorization	Supports End-to End security Enforces baseline security	Increases power consumption, does not support in-network processing like continuation programming
Media Access	Time Synchronization, Dynamic Media Access, Error corrections	Supports in-network processing(A link layer security mechanism can detect unauthorized packets when they are first injected onto the network)	Provides generic security services not specific to individual application needs
Physical	Spread-Spectrum and Priority Messaging	Resistant to jamming, fading and eavesdropping	High energy costs, constrained, limited frequency domain

3. Effects of Mobility on Performance of Ariadne Protocol

Ariadne is a demand driven secure ad hoc routing protocol that is built on two protocols; (1) Dynamic Source Routing (DSR) [Johnson and Maltz 1996] for routing and (2) Timed Efficient Stream Loss-tolerant Authentication (TESLA) [Perrig et al. 2002] for authentication. Although Ariadne protocol can use other forms of authentication like Message Authentication Code (MAC) [Pfleeger and Pfleeger 2003] and Digital Signatures [Pfleeger and Pfleeger 2003; Tilborg 2003], the default scheme is TESLA. The TESLA authentication scheme uses delayed disclosure of the shared key to achieve asymmetric cryptography over symmetric cryptography. The design goals of Ariadne protocol can be summarized as follows;

1. Perform basic routing tasks, i.e.;
 - i. Route establishment
 - ii. Route maintenance
2. Perform node identity management
3. Preserve the integrity of packages and
4. Be efficient

Since secure routing in theory refers to secure establishment and maintenance of communication links, the following subsections highlight the operations of Ariadne protocol in that regard.

3.1. Basic Ariadne route discovery

The Ariadne route discovery scheme is built on DSR route discovery protocol. During the route discovery phase, a Sender node S wishing to establish a connection with a Destination node D, prepares a route request packet. The packet contains eight fields: $\langle \text{Route Request, initiator, target, id, time interval, hash chain, node list, and MAC list} \rangle$. The initiator refers to the address of the Sender while the target refers to the address of the Destination. Every route request initiated by the Sender must be tagged with an "id", this parameter is used for accountability purposes. Because the Ariadne protocol relays on TESLA to provide authentication, which itself uses delayed disclosure of symmetric keys to achieve asymmetric cryptography, the sender of the route request defines the upper bound delay interval T, in which it expects keys to be revealed. The node list which is initially empty is updated by intermediate nodes along the path of the request by appending their names, same holds for MAC list. When an intermediate node receives a route request packet, first it checks if it is the intended destination, if it is not, it checks from its local table the matching pair of initiator and request id. If it sees the value in the local table, the node discards the packet. Otherwise, it checks if the time interval is valid, if not valid, the node discards the packet as well. If the packet passes the above two tests, the node appends its own address on the node list in the packet, computes a new hash chain and MAC, appends the initial and id values of the request into local table, and rebroadcasts the packet.

On receiving the packet, the destination node first checks the validity of the request, by determining whether the time interval of the keys specified is valid, and that the hash chain in the packet must be equal to

$$H \left[n_n, H \left[\dots, H \left[n_1, MAC_{K_{sd}} \text{ Sender, Dest, id, time-interval} \dots \right] \right] \right]$$

Where n_i is the node address in the node list at position i in the request packet? The route reply packet has the structure: $\langle \text{Route Reply, initiator, target, id, time-interval, node list, MAC list, target MAC, key list} \rangle$. The *target*, *initiator*, *time-interval*, *node list*, *MAC list* fields are set to the corresponding values in the route request packet.

3.2. Basic Ariadne route maintenance

When an Ariadne intermediate node forwards a packet to the next hop along the source route, returns an error to the original sender if it is unable to deliver the packet after a limited number of retransmission attempts. A route error packet in Ariadne contains six fields $\langle \text{Route Error, initiator, receiving address, time-interval, error MAC, and recent TESLA key} \rangle$. The initiator is the node which has encountered the error; the receiving address is the node which the intermediate node failed to reach when attempting to forward the packet. The time-interval is the TESLA link latency time and the error MAC is the hash of all fields in the error packet computed using the most recent TESLA key of the intermediate node encountering the error. Finally the node must define the destination address which is the address of the original sender of the packet. The error packet is

forwarded towards the destination as the normal packet which each intermediate node performing node packet and key validation before forwarding the packet towards the destination. The destination processes the error packet the same way it processes the route request, except the processing node it does not have to generate reply. The only action it takes is to remove the route from its cache memory.

3.3. Mobility Vs Security Services

This section presents a logical analysis of Ariadne the security mechanism. The analysis focuses on the impact of mobility of nodes and the fluidity of the network on the delivery of security services of Authentication, Integrity, Reliability, Confidentiality, Message freshness and Non-repudiation.

3.3.1. Mobility Vs Authentication

Hypotheses:

Authentication can only be maintained if the average end-to-end link delay does not exceed a certain threshold value.

Explanation:

Let a Packet P take n seconds to travel from Node A to B. If the global upper bound packet delay between Node A and B is q seconds, called threshold value, P will be dropped if $n > q$. If P was an authentication packet, the authentication protocol will have failed, since credentials will not have been shared among protocol principles. Let Nodes move at a constant speed S , and let the relative distance between node A and

B be D , then from $Speed(S) = \frac{Distance(D)}{Time(T)}$, we have $T = \frac{D}{S}$, if $T > q$ then authentication will fail because of lack of freshness of the authentication packet. Thus, authentication can only be maintained if the delay does not exceed the threshold value of q . Since

nodes in AWSN application move at a random speed, it is had to maintain q for all connections. Thus, a good number of authentication requests for AWSN service running Ariadne will fail. Note that Ariadne relies heavily on time synchronization to authenticate both packets and nodes, because of mobility in the network and dynamism of nodes i.e., the rate of node joining and leaving, it is our conclusion that a large number of authentication requests will fail. In Ariadne, there is another key requirement, the loose time synchronization scheme.

3.3.2. Mobility Vs Reliability

Hypothesis:

Since Ariadne depends on DSR and TESLA as the base protocols, it inherits the limitations of the two

Explanation:

One of the drawbacks of DSR is that it lacks a mechanism to repair broken links locally, thus route cache information may sometimes result into inconsistencies

during the route reconfiguration as new nodes join the network while others leave, besides changing their physical locations. Note that Ariadne requires every packet be authenticated before being processed, this requirement is not ideal for AWSN where there is high mobility and node dynamism because the link connection setup delay is higher than in ordinary lookup table protocols. Also due to source routing scheme and TESLA, the sender must know the whole route information and the cost of knowing this information is propositional to the route path length and it often times takes time to build, which by nature of mobility it may be absolute when it is received by the source as new information. Therefore, while DSR may perform well in static and low-mobility environments, its performance degrades rapidly with increasing mobility and node dynamism. We anticipate high package delays and low package delivery ratio for the following reasons

1. Nodes changes location information at a relatively high frequency
2. Node join and leave the network at random
3. Authentication of all packets in the network by TESLA increases delays and for long routes results into low rate of packet deliver as packet are dropped along the path because of delayed arrival
4. Ariadne has no mechanisms to mitigate replay attacks by compromised nodes
5. In a highly dynamic and mobility network, it is difficult to maintain time synchronization signature in the bounds defined for TESLA to remain effective.

3.3.1. *Mobility and dynamics Vs Non-repudiation*

Hypothesis:

Since the authentication protocol for Ariadne relies on a receiver's ability to determine the senders keys, which are already published based on the loose time synchronization between the two nodes. Then if the two nodes are not well synchronized, it will be had for the receiver to verify senders keys and hence to enforce non-repudiation of entities.

Explanation:

In digital security non-repudiation refers to a service that provides proof of the origin and sometime integrity of data or packet. Mobility and dynamism issues in AWSN make the verification of message source and sometimes integrity of data difficult.

3.3.2. *Mobility Vs Confidentiality, freshness and data integrity*

In AWSN it is difficult to maintain the freshness of all messages in the network particularly for the application we envision in this research. According to the operations of Ariadne protocol, two types of packets are transmitted on the network, i.e., application datagram's and control packets. The challenge of keeping these packets confidential lies in the key management framework. Since node join and leave the network at random and these nodes might have prior knowledge of each other, then we assume it difficult for Ariadne protocol to guarantee confidentiality of data. About integrity, since intermediate nodes

perform data aggregation from multiple nodes, we will investigate if it is possible to maintain sound message integrity in such an environment.

4. Emerging Security Issues and Conclusions

Our preliminary work has revealed a couple of emerging research directions. According to [Kambourakis et al. 2008 and Muneeb et al. 2006] the sensor network research community has ignored at least until now the effects of mobility on the delivery of security services. As highlighted in Section 1, most efforts to secure WSN have focused on designing solutions for static networks not the highly mobile heterogeneous network we envision in this research. In theory, mobility of nodes complicates bandwidth utilization because of too many link re-connections. The excessive re-connections do not only raise routing overheads, but they also raise concerns about fairness, packet delays, authentication, authorization and accountability and the overall reliability of service.

Consequently, whoever is designing a secure protocol for AWSN must investigate the effects of mobility and network dynamism on authentication, authorization, accountability, identity management, reliability, key management and time synchronization schemes? We are also fully aware that robust security does not only depend on sound security protocols, but it also on sound user ethic. We will discuss other factors that affect optimal deployment of security controls in our future works. The following are some of the emerging research directions from our preliminary investigation which form the core of our immediate future research direction;

1. Does mobility and dynamism of nodes affect how ARIADNE protocol delivers its security services?
2. How reliable is MMAC protocol with increasing mobility and density of nodes?
3. Considering mobility and network dynamism, what would be the optimal layer in the communication stack to implement security services?
4. How can we best address concerns about user privacy in AWSN?
5. How can we efficiently manage identities of nodes and objects being monitored?
6. How can we efficiently manage cryptographic keys in such a dynamic and resource constrained environment?

This is so much work in progress, in the next stage we are going to simulate hypotheses in Section 3, with NS 2 network simulator to generate empirical results and validate theories concerning the effects of mobility on delivery of security services. We also plan to analyze the authentication properties of Ariadne using *AVISPA*, a tool for analyzing security protocols. Also a new secure routing protocol for AWSN will be proposed, and have its performance analyzed. Besides, technical consideration of security analyze user behaviors which might affect optimal deployment of security protocols in AWSN but concerns about optimal data acquisition by sensors are out of the scope of our

immediate future work. We believe our work will contribute to the understanding and development of optimal security and privacy enhancing techniques for AWSN which are vital for the adaption and deployment of AWSN services.

5. Acknowledgments

We would like to acknowledge the support Faculty of Computing and IT, Makerere University and NUFFIC. Special thanks go to members of the digital security research group, Radboud University especial Dr. Richard Brinkman, who guided us during the initial work. We also commend Prof. Fisseha Mekuria for his insights in mobile computing and the security requirements of mobile services.

References

- ALI, M., SULEMAN, T., AND UZMI, Z.A. MMAC: a mobility-adaptive, collision-free MAC protocol for wireless sensor networks. 2005. In Proc. Of Performance, Computing, and Communications Conference, 2005. IPCCC 2005. 24th IEEE International. 401- 407.
- CRED, Naural Disaster Trends, <http://www.cred.be/>, accessed in Oct, 2008
- DENG, J., HAN, R. AND SHIVAKANT, M. 2006. NSENS: Intrusion-tolerant routing for wireless sensor networks. In Computer Communications : Dependable Wireless Sensor Networks, Elsevier B.V. 29:2, 216-230
- INTANAGONWIWAT, C., GOVINDAN, R, ESTRIN, D, HEIDEMAN, J, AND SILVA, F. 2003. Directed Diffusion for Wireless Sensor Networking. IEEE/ACM Transactions on Networking TON. 11:1
- JAMAL, N.A. AND KAMAL, A.E. 2004. Rounting Techniques in Wireless Sensor Networks: A Survey. In IEEE wireless communication.
- JOHNSON, D.B. 1994. Routing in Ad Hoc Networks of Mobile Hosts. In Proc. Of the Workshop on Mobile Computing Systems and Applications. IEEE Computer Society. 158-163.
- JOHNSON, B.D. AND MALTZ, D.A. 1996. Dynamic source routing in ad hoc wireless networks. In Proc. of Mobile Computing, 1996, Kluwer Academic Publishers, 153—181
- JOVANOVIĆ, E., MILENKOVIĆ, A., OTTO, C. AND GROEN, P.C. 2005. A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. Journal of Neuro Engineering and Rehabilitation, 2(6):10.1186/1743-0003-2-6.
- KAMBOURAKIS, G., KLAOUDATOY, E. AND GRITZALIS, S. 2008. Securing Medical Sensor Environments: The CodeBlue Framework Case. *Security and Communication Networks*. 1:5, 375-388
- KARLOF, C., SASTRY, N. AND WAGNER, D. 2004. Tinysec: A link layer security architecture for wireless sensor networks. In Proc. Second ACM Conference on Embedded Networked Sensor Systems. 162 – 175
- KRISHNA, V.K. AND SANDEEP, G.K.S. 2006. Security for pervasive health monitoring sensor applications. Fourth International Conference on Intelligent Sensing and Information Processing, (ICISIP 2006). 197–202.
- LORINCZ, K., MALAN, D.J., FULFORD, T.R.F., JONES, A., NAWOJ, A., CLAVEL, V., SHNAYDER, M. MAINLAND, G ., WELSH, M AND MOULTON, S. 2004. Sensor network for emergency response: Challenges and opportunities. Pervasive computing, IEEE CS, 3:4, 16–23.

- MANETS WG, IETF <http://www.ietf.org/html.charters/manet-charter.html>, accessed in Oct, 2008
- MALHOTRA, K, GARDNER, S. AND REES, D. 2005. Evaluation of GPRS enabled secure remote patient monitoring system. In Proceeding of the 12th Conference ASMTA
- MOBIHEALTH, Innovative Mobile Service Application in Health care. 2008. DOI= <http://www.mobihealth.org/>, (Accessed on 23th September 2008)
- MUNEEB, A., UMAR, S., DUNKELS, A., VOIGT, T., ROMER, K., LANENDOEN, K., POLASTRE, J. AND UZMI, Z.A. 2006. Medium Access Control Issues in Sensor Networks. ACM SIGCOMM Computer Communication Review. 36,2 (April 2006). 33-36
- NG, H.S., SIM, M. L. AND TAN, C. M. 2006. Security issues of wireless sensor networks in health care applications. BT Technology Journal 24:2,138-144
- PERRIG, A, SZEWCZYK, V., WEN, R., CULLER, J.D., AND TYGAR, D. 2002. Spins: Security protocols for sensor networks. Wireless Networks Journal (WINET). 8:5,521–534
- PERRIG, A., RAN CANETTI, J. D. , SONG, T.D. 2002. The TESLA Broadcast Authentication Protocol. In CryptoBytes, 5:2, 2-13
- PFLEEGER, C.P. AND PFLEEGER, S.L. 2003. Security in Computing. 0-13-035548-8. Prentice Hall, third edition
- PHILIP'S RESEARCH LABS, MyHeart Prohct. 2008. Achaan Germany, DOI= <http://www.hitech-projects.com/euprojects/myheart/home.html>, (Accessed on 2nd October, 2008)
- SASTRY, N AND WAGNER, D. 2004. Security considerations for IEEE 802.15.4 networks. ACM WiSE'04.
- SEUNG, Y., NALDURG, P. AND KRAVETS, R. 2001. A Security-Aware Routing Protocol for Wireless Ad Hoc Networks. In Proc. of ACM Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc). 286—292
- SHNAYDER, V., BOR-RONG, C., KONRAD, L, THADDEUS, R. F., FULFORD, J., AND WELSH, M. 2005. Sensor networks for medical care. Technical Report TR-08-05, Division of Engineering and Applied Science, Harvard University. (April 2005).
- TIA GAO, P., SELAVO, C, YIN CHEN, L., GIL KO, J, JONG HYUN, L., TERZIS, A., WATT, A., JENG, J., BOR-RONG, C., LORINCZ, K. AND WELSH, M. 2008. Wireless Medical Sensor Networks in Emergency Response: Implementation and Pilot Results. IEEE Conference on Technologies for Homeland Security, WALTHAM, M.A. 187-192
- TILBORG, H.C.A. 2003. Fundamentals of cryptology, A Professional reference and interactive tutorial. 0-7923-8675-2. Kluwer academic publishers, third edition
- US, GOVERNMENT, “The Health Insurance Portability and Accountability Act of 1996 (HIPAA)”, <http://www.hhs.gov/ocr/hipaa/> accessed in Oct, 2008
- WALTERS, J.P., LIANG, W., SHI, Z. AND CHAUDHARY, V. 2006. Wireless sensor network security: A survey. Security in distributed, Grid and Pervasive Computing, YANG XIAO, (Eds.)
- WARREN, S. LEBAK, J. YAO, J., CREEKMORE, J., MILENKOVIC, A. AND JOVANOV, E. 2005. Interoperability and security in wireless body area network infrastructures. In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China.

- YIH-CHUN, H., PERRIG, A. AND JOHNSON, D. B. 2002. Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks. In Proc. Of the 8th ACM International Conference on Mobile Computing and Networking
- YU, Y., GOVINDAN, R. AND ESTRIN, D. 2001. Geographical and energy aware routing: A recursive data dissemination protocol for wireless sensor networks
- ZIGBEE-ALLIANCE. <http://www.zigbee.org>, September 2008.

3

Measurement and Analysis of Copper Access Network for ISDN Basic Rate Interface

Grace Bakaki and Idris A. Rai

Improper combination of copper pair properties like gauge, longitudinal balance, loop resistance, insulation resistance and distance for Integrated Services Digital Network Basic Rate Interface (ISDN BRI) leads to users experiencing connectivity problems and fluctuating link speeds. Installations, configurations, troubleshooting and maintenance of ISDN BRI services can be very complex to the technical team of the service provider. There may be many reasons causing this in developing countries. Troubleshooting at Uganda telecom, for instance, has been cumbersome due to the absence of manuals that guide technicians on proper installation of copper lines.

In this paper, we report measurements study to analyze and assess the impact of the copper pair parameters/properties combination to the copper line speeds. The study aimed at simplifying the installations, configurations, troubleshooting and maintenance of the ISDN BRI services on the side of the technical people. This was achieved by taking measurements at different parameters (distance, gauge, longitudinal balance, loop resistance, and insulation resistance) in order to come up with optimal values for the design of the computer based program for optimization of data connectivity in ISDN BRI. We use the measurement results to design and develop a computer based program that can be used to verify the measurements and identify the expected optimal parameters in troubleshooting.

1. Introduction

Integrated Services Digital Network Basic Rate Interface (ISDN BRI) is commonly used as a solution to provide low bandwidth Internet access to small offices or dial-in users with traditional analogue dial-in services. In general, it is a network that provides end-to-end digital connectivity to support a wide range of services including voice and data services. Some readers, mainly from developed world, may wonder why we are still talking about ISDN technology, while may appear obsolete in developed world; ISDN is still used as Internet access technology in developing countries albeit at a very small scale. In this paper we are not proposing using copper as a communication medium, but we are analyzing the performance of the currently used copper system / technologies (ISDN BRI) and what we show are the typical mistakes that can be done in developing countries due to lack of expertise and negligence.

ISDN allows multiple digital channels that operate simultaneously through the same regular phone wiring used for analog lines. Latency is much lower on an ISDN line than on an analog line. The technology is implemented over copper access network for voice and data communication; however it should be implemented with proper combinations

of copper pairs' properties/parameters like *gauge*, *longitudinal balance*, *loop resistance*, *insulation resistance*, and the *distance*. Distance refers to the length of a pair of copper conductors from the central telephone office to the customers' premises. Longitudinal balance is the electric symmetry, with respect to ground, of the two wires of a pair or an expression of the difference in impedance of the two sides of a circuit [Technical, 2007] Loop resistance is the total resistance of a thermocouple circuit caused by the resistance of the thermocouple wire (total resistance of a closed circuit of two wires measured at one end). Insulation resistance is the DC resistance expressed in ohms measured between any electrical connector pins or lead wires and the transducer body or case [Grace, 2009].

In the event that ISDN BRI connectivity is not working well the developers of the technology developed error codes and their likely causes [ITU, 2002] which are supposed to assist the technical team in coming up with the proper installations, repairs/troubleshooting and maintenance. Despite using error codes, installation manuals of ISDN BRI and several trainings to the technical team of the service providers, users experience connectivity problems and new installations at times take long to be completed. This was got through a verbal interview we carried out on both the technicians and clients of ISDN BRI technology at *Uganda Telecom*, a telecom provider in Uganda. However, these manuals and guidelines for the existing copper are no where to be found, and it is not clear if the guidelines were properly used when installing them.

ISDN users have been constantly reporting connectivity problems ranging from low speed to poor connectivity. This is the main motivation of this work: Firstly, we carried out a measurement study for 150 copper cables using a dynatel 965 test meter to determine the dependency of speed to the copper parameter values. Dynatel 965 test meter is used to determine the speed of copper cable in addition to inherent parameters of the copper at any distance from the central office of the ISP. Analysis of the measurement results enabled us to know the dependency of link speeds to copper parameters. We used the measurement results to implement a computer based tool that can be used to automatically check if the values from the Dynatel are optimal.

This paper is structured as follows. Section 2 describes how the measurements were done, and Section 3 outlines the results of the measurements. In Section 4, analysis of the results is explained and we briefly describe the computer based tool that was developed as a result of the measurement analysis. Finally, we conclude the findings in Section 6.

2. Measurements

We measured parameters for 150 copper cables of Uganda Telecom that support ISDN connectivity to users. As mentioned earlier, a Dynatel 965 test meter was used to give the values of all parameters of a given copper cable, namely gauge, longitudinal balance, loop resistance, insulation resistance. We measured the distance from the ISDN central operational office of the telecom provider to the measurement point (at the end user) in order to assess the impact of the combination of the copper pairs' parameters to the

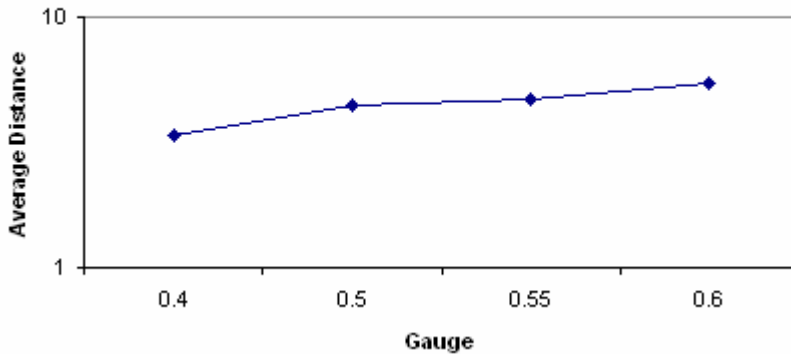
distance. The 150 lines were chosen because users of those lines had reported faulty behaviors more than four times in the past two months, information (Location and technical details) about the lines was collected so that measurements could be made. Each ISDN BRI line in question was disconnected from the switch and the network terminal was also disconnected at the customer's premises so that the copper pair was free from any electrical signal, then the measurements were taken by a digital meter.

A 1 MB test file that we named `testfile.tar.gz` was put on one of the servers at the service provider's end, the ISDN BRI line was reconnected and the test file was downloaded at the customer's end. This allowed us to record download rates (data transfer rates) at three different intervals during the download (i.e., at the beginning, middle and the end). The average download rate was taken as the data transfer rate at the measured parameters / properties and the results were tabulated. Latter all the data was analyzed to establish the dependency of link speed to various parameters of copper cables and the distance between end user (measuring point) to the server.

The results of the measurements exhibited two categories; those with data transfer rates above 50 kbps and those that were below. We consider 50 Kbps as the best line speed because bearer channel of ISDN is 64 Kbps, with assumed overhead of 14 Kbps, then the optimal link speed is 50 Kbps for one bearer channel of ISDN BRI to be fully optimized. The results that exhibited data transfer rate above 50 kbps were used for the implementation of computer based program for the optimization of ISDN BRI under fixed gauge with varying distance, loop resistance, insulation resistance and longitudinal balance.

3. Analysis of Measurement Results

In this section, we present analysis of the measured data in order to determine the variation of line speed as a function of increasing values of the parameters of the copper line and the distance. We classify the analysis into different categories, the basis was on copper pairs' gauge / property under study. Data was grouped and analyzed into two groups, namely, 1) data that resulted in link speed above 50Kbps, and 2) data that led to link speed of less than 50 Kbps. Our analysis in this paper will emphasize the first case. For analysis of complete data set the reader is referred to [2]. Four gauge values (0.4mm, 0.5mm, 0.55mm (Mixed gauge), and 0.6mm) were put into consideration because there are only four copper pair gauges being used by the service provider.

Fig. 1 The average distance as a function of gauge

The average distance for different gauges for all ISDN BRI lines (below and above 50 kbps data transfer rate) were collected and a graph of average distance as a function of gauge is seen in Figure 1. The figure shows general dependency between distance and gauge. It can be observed from the Figure 1 that the gauge increases as the distance increases. These results suggest that the installation was deliberately done to favor higher gauges for longer cables in average sense.

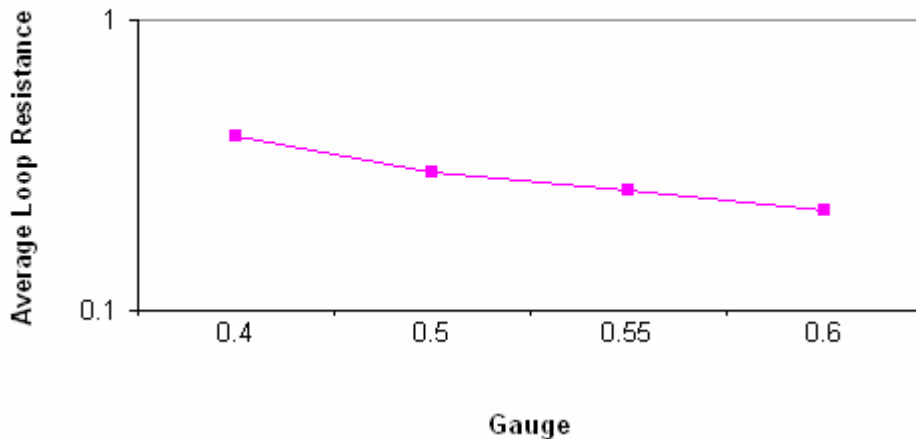
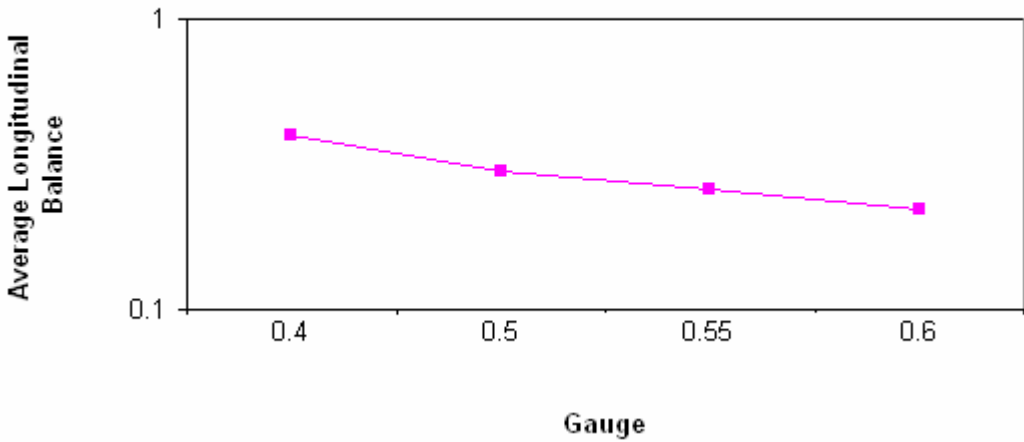
Fig. 2 The average loop resistance function of gauge

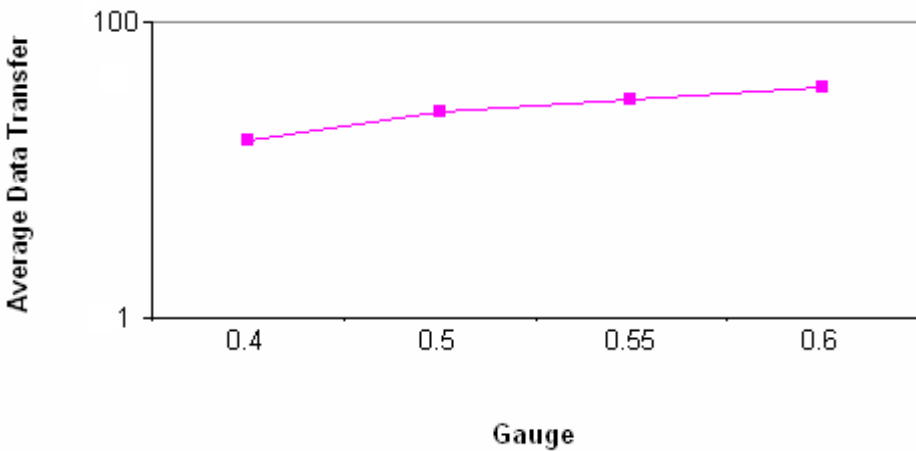
Figure 2 shows the average loop resistance as a function of gauge. It can be observed from the figure that as the gauge increases the average loop resistance decreases. Similar trend can also be observed for the longitudinal balance from Figure 3; that as gauge increases the longitudinal balance decreases.

Fig. 3 The average longitudinal balance as a function of gauge



The average data transfer rate as a function of distance for copper conductors was also plotted and it is observed from Figure 4 that for all data transfer rates, as the distance increases the data transfer rate decreases. This show that distance has an impact on the data transfer rate, the further you go the lower the data transfer rate or speed. This implies that a copper cable with higher gauge value is required to extend the ISDN BRI services to subscribers located a longer distance from the service provider and the bigger the gauge the better the service.

Fig. 4 The average data transfer as a function of gauge



4. Computer Based Tool

The computer based tool for optimization of ISDN BRI was developed after the analysis of the results of measurements carried out using the *confidence interval* (μ). Confidence Interval refers to a condition given that if a and b are limits of a parameter, then $a \leq \mu \leq b$, where a and b are called confidence limits, a being the lower confidence limit while b is the upper confidence limit. The probability that the true population parameter will lie within the stated interval is denoted by $(1-\alpha)$ and this is referred to as confidence coefficient or degree of coefficient. Standard deviation on the other hand is a measure of the dispersion of outcomes around the mean of a given sample size denoted by sigma well as mean refers to the mathematical average of a given sample size. In our case, optimal values are measurements which exhibited data transfer rates or link speeds above 50 kbps.

The confidence interval values for each parameter were used the confidence limits under which we are confident that ISDN BRI can optimally work. It was also used in the analysis of the measurements in order to come up with optimal ranges under which ISDN BRI should operate. Table 1 shows the optimal values or degree of confidence after measurements and analysis.

The results of the analysis were further filtered in order to come up with measurements which were used as a yard stick for the development of the computer based program. Then Windows XP was used as an operating system, Visual Basic 6 as the software and a Laptop as the hardware for the computer based program.

Table I: Degree Of Confidence

Gauge	Distance Km	LT/BAL	L/Res	INS/RES
0.4	$(2.7 \geq \mu \leq 4.1)$	$(-53.3 \geq \mu \leq -67.1)$	$(1121.7 \geq \mu \leq 1311.5)$	$(375.4 \leq \mu \leq 644.4)$
0.5	$(3.5 \geq \mu \leq 5.4)$	$(-61.4 \geq \mu \leq -69.4)$	$(1176.3 \geq \mu \leq 1350.7)$	$(530.7 \geq \mu \leq 766.1)$
0.55	$(4.66 \geq \mu \leq 6.74)$	$(48.8 \geq \mu \leq -65.2)$	$(1176.3 \geq \mu \leq 1342.9)$	$(497.4 \geq \mu \leq 791.8)$
0.6	$(4.5 \geq \mu \leq 6.5)$	$(56.8 \geq \mu \leq -73.2)$	$(1204 \geq \mu \leq 1374)$	$(550 \geq \mu \leq 766)$

Table I shows the degree of confidence for measured distances, longitudinal balance (LT/BAL), loop resistance (L/Res), and insulation resistance (Ins/Res) as a function of gauge. These are the results which were used in the development of the computer based program for optimization of ISDN BRI. For distance, the upper limits to the nearest whole number were used as the maximum distance for optimal data transfer except for a 0.55 mm (mixed) gauge the mean of 0.55 was considered since it was a mixed gauge. The table shows that the maximum longitudinal balance for gauge 0.6 mm is -56.8 dBm (approx -57 dBm) and was used as the benchmark for all gauges since longitudinal balance increased with decreased in gauge. In the loop resistance, the upper limits to the nearest whole number were used as the benchmark. For insulation resistance, the 375.4 mega ohms (approx 375 mega ohms) was used as the benchmark for all gauges as the minimum insulation for optimal data transfer.

The tool was tested in practice in conjunction with the dynatel and it was shown to comply with the expected results from the measurements. In general, the tool can help technicians when troubleshooting different lines. It provides a basis for the expected copper parameters as well as the expected line speed.

5. Conclusion

The analysis of copper access networks was performed in order to obtain optimum parameters under which ISDN BRI provides good performance. And after the analysis of measurements, results showed that as the gauge values increased, the distance increased hence a bigger gauge was required to extend the services to a longer distance.

The measurement results showed that as the gauge increased the loop resistance decreased, hence more and more links offered high speeds in low and low values of loop resistance, in order to achieve optimal data connectivity, a maximum of 1312 Ohms for a 0.4mm gauge, 1351 Ohms for a 0.5mm gauge, 1342 Ohms for a 0.55mm gauge, and 1374 Ohms for a 0.6mm gauge was required maximum loop resistance and anything above those values exhibited a drop in data connectivity.

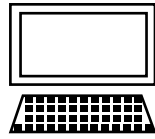
For the increased gauge, the longitudinal balance decreased hence more and more links offered high speeds in lower and lower magnitudes of longitudinal balance values, so to achieve optimal data connectivity, -57 dBm was the maximum longitudinal balance required for all gauges. As the gauge increased the insulation resistance increased hence high speeds were attained in high magnitudes of insulation resistance, so to achieve optimal data connectivity, 375 mega ohms was the minimum insulation resistance required for all gauges. We also found out that distance had an impact on the data transfer rate, as distance increased, less and less links offered high speeds, in order to achieve optimal data connectivity, a maximum of 4 Km for a 0.4mm gauge, 5 Km for a 0.5mm gauge, 5.5 for a 0.55mm gauge, and 6 Km for a 0.6mm gauge was the required maximum distance and anything above those values exhibited a drop in data connectivity. Measurements results were used to develop a computer tool that provides a base line for expected measurements values of copper parameters and line speed that help technicians to troubleshoot copper lines.

Measurement results suggest that the reported connectivity problems by end-users are probably due to improper installation of copper lines. The scope of this paper was limited to copper pairs parameters/properties combination for ISDN BRI connectivity, however further investigations can be carried out to cover other technologies like Asymmetrical Digital Subscriber Line (ADSL) which use copper access networks.

References

- ATIS, T. 2007. *Performance and Signal Processing: Longitudinal Balance, Glossary and definition*, ATIS TELECOM, Washington DC, USA
- Grace B. 2009. *Analysis of Copper Access Networks for Data Connectivity Optimization in ISDN, Basic Data Rate Interface*, MSc. Project, Makerere university, Kampala, Uganda.
- ITU, 2002. *Standard Q.931 disconnect cause codes, Q.931 ISDN*, Quintum Technologies, USA
- TECHNICAL, F. 2007. *Products: Loop resistance, Telecom notes*, Flw Technical glossary, E-Costa Mesa, CA.

PART FIVE



Computer Science

1

Apprenticeship Environment and Co-operation Based on the Grid and Web 2.0 Designed for Training Communities with Common Interest Centres

Bernabe Batchakui , Claude Tangha , Roger Nkambou And Thomas Ndie Djotio

Common interest centers specialized in training communities, such as secondary schools, university training institutes, have heterogeneous, rich and varied resources in contents. Yet in their assessment, these communities present very disproportionate results. The present article comes out with an apprenticeship and co-operation setting which is based on Grid technologies and uses Web 2.0. The new environment is designed to ease the communication between the communities and thereby reduce the gap between them. To achieve this objective, a set of technologies and tools is used. It includes domain ontology for the validation of course contents, intelligent agents for communication and guidance, extended Moodle platform for contents surfing, multimedia message services (MMS) for transfer and retrieving grid courses through the mobile phone.

1. Introduction

Today, emerging technologies like the grid and web 2.0, significantly improve access to resources. For instance with grid, it is possible for a user to run in a transparent manner a software in a remote computer without having to worry about material or software constraints (capacity of the processing unit, memory, computing power, etc.). That responsibility is incumbent on the grid load. Web 2.0 on its part enables the internet user to impact on the contents that it receives, thereby making him not only a mere consumer, but also a producer of resources. These characteristics of the grid and web 2.0 meet the present requirements of e-learning consisting in the collaboration and co-operation between all the participants and the learner as the centre of his apprenticeship.

Furthermore, many training communities that share common interests like secondary schools and universities have numerous, heterogeneous resources rich in contents. Nevertheless, the problem encountered in reviewing these communities is that the latter present, for the same course contents, disproportionate results. For instance, in Cameroon, for the same national result, the success rate is some times far above average for certain schools and below average for others; this particular case can be observed through the statistics of the 2007 official exams published by the “Office du Baccalauréat du Cameroun” [BEBEC 2008]. In order to reduce the gap between these communities, our hypothesis is that, resources of each of these communities reside in the computer connected on the Internet. This article thus puts forth an apprenticeship

and collaboration environment named « Grid Environment for Learning and Sharing Objects between Training Communities » GELSOTC, designed to facilitate inter-community communication.

GELSOTC is based on the grid infrastructure. It federates the following tools and technologies: domain ontology for the research and the validation of contents, intelligent agents for communication and tutorship, Moodle platform extended by a module of content structuring and Web 2.0 concepts for contents surfing, and Multimedia Message Services (MMS) for depositing and withdrawing grid lectures via the mobile telephone. Its main objectives are:

- To enable training communities with common interest centres to gather in networks in order to exchange training resources, create and put at the disposal of the network consensual contents, pedagogically rich and varied, to validate contents created by domain ontology, etc.
- To enable constituted members of networks: to learn, enrich various training contents of their networks, organize themselves into self aid groups, and seek available resources of their networks.

We have structured this article as follows: the first part presents the state of the art on the field of grid services for learning, a glance at the Web 2.0 and its advantages for e-learning, a glance at the Moodle platform, a brief definition of the ontology and the intelligent agent. The GELSOTC environment is specified in part two while the third part concerns the implementation of GELSOTC middleware. The article ends with a discussion, a conclusion and some perspectives.

2. State of the Art

2.1 Computer grid and its apprenticeship capacities

2.1.1 Grid computing.

From the numerous definitions of grid computing, we have retained Ian Foster and Keselman's who define computer grid as an environment operational and heterogeneous architecture systems whose access is provided to the user under a unified form by a software called Middleware [Foster 2001, Batchakui and al. 2007]. The grid is often used to share free computing power of connected computers. It is based on a standard architecture called OGSA (Open Grid Service Architecture). It deals with architecture in layers whose services form the basis of communications [Dugenie 2005, Kesselman 1998]. OGSA has a series of technical specifications which help to define an infrastructure in order to integrate and manage services within an organisation which is virtual, spread out and dynamic [Foster et al. 2002].

Today, grid usage has spread out to other domains. However, our interest is on e-learning. Apprenticeship grids as they are often referred to help to share out learning resources, give room for the inclusion of constraints such as : 'information access only when the need is expressed, the management of the access to the distribution network and safety. The above-mentioned constraints constitute the requirements of a learning

model on information transfer. Apprenticeship grids are, on an architectural point of view, based on the OGSA extension called GLS (Grid Learning Services) [Nkambou and al. 2005]. This is an OGSA to which a layer has been added comprising basic services that enable sharing as perceived by e-learning (collaboration, e-qualification, co-ordination, etc). The objective of GLS is to facilitate the contents broadcast and the sharing of apprenticeship objects. Grid presents so many advantages for apprenticeship.

2.1.2 *Advantages of the grid for teaching.*

The grid is supported by a distributed architecture, as solution to the problems of the interoperability and heterogeneity. Its advantages are numerous. It permits: the access of a larger number of learning objects geographically dispersed, the coexistence of heterogeneous resources, the dynamic management of learning resources, the reduction of costs related to maintenance, the sharing of resources to enhance the collaboration between learners and the authors of the learning objects [Pankratius 2003], the sharing of the knowledge of the learner personalized and contextual. Also, it creates a learning environment centered on communication and collaboration and consequently, reduces considerably the distance between the entities or actors of the training.

2.1.3 *The grid environments for Learning.*

Grid systems have been conceived in the domain of e-learning and their main objective is to enable their users to collaborate through exchange of learning objects. The stress this time is put on the learner: the systems designed save the latter from the overload of information in registering into the virtual organization and the contents offered take into account the profile of the learner. The environments that we have studied, those that have mostly struck us with respect to our objectives are the following:

- DIOGENE [Cri 2002, Woolf 2005]: it is a platform resulting from a project financed by the European Commission which has as objective to design and realize an innovative environment for the brokerage of the online training services, then evaluate it in real situation. It serves as a support to the learner all along his learning process, from the definition of the objectives up to the validation of the results, passing through the construction of the personalized training channel. It offers innovative functionalities, for example, the probabilistic modelization of the learners, the personalized adaptation of the training channel, cooperative learning and online guidance, the intelligent and assisted definition of the training objects, dynamic learning strategies, the management of learning objects, etc. The technologies used are metadata and ontologies.
- Customized Grid Learning Services [Woolf 2005]: it uses the techniques of web semantics, ontology, and autonomous agents to supply a personalized approach of the services of the learning grid. The learners and the teachers access the course according to their profile.
- GRID-e-CARD [Gouarderes 2005], which is a system based on the grid infrastructure and a set of P2P agents with the objective to bring together,

in Virtual Learning Communities (VLC) the actors or entities of a learning environment on the basis of their signatures (knowledge acquired, future objectives, desired learning services, ...) for collaboration purposes.

- DyColo (Dynamic Collaboration Learning Object), is a collaboration platform based on the Grid Learning Service which permits the learning entities and teachers to collaborate with the particularity that the entities which share resources are grouped together thanks to the services of e-qualification integrated into the GLS [Mireille 2006].

The contributions presented above are indeed very interesting but for most part not yet effective. The groupings proposed for these environments are virtual organizations with the inconvenience that they can have a very high number of members which does not ease proximity learning. Practically there exist no environment that offers the possibility to the communities having the same centre of interest to be grouped in order to join their efforts to converge towards the best results when they are evaluated.

2.2 The Web 2.0 and its advantages for e-learning

The Web 2.0 is an information-pooling concept which modifies our way of working and interacting with information. With the Web 2.0 the user is at the centre of the exploitation of the information given that he participates actively in the enrichment of the content that he receives. Contrary to the Web 1.0 which proposes only internet pages to be consulted, the Web 2.0 allows the user to react to information found on the internet. Access to the tools is at the centre of the utilization of the Web 2.0 for education: for example we have, wikis, blogs, RSS lines, aggregators or personalized welcome pages [Dremeau 2007].

Learning is said to be effective when the activity of the learner is highly solicited, hence the necessity to reinforce the teaching model in use by pedagogic strategies that permit the student to be at the centre of the learning process. The Web 2.0 offers the possibility to the actors of the training to impact the content, through their remarks and annotations. It provides the learner with a flexibility and a certain learning autonomy. For example the learner can subscribe as a student to the RSS streaming to benefit from updates of the learning domain.

There exist environments of e-learning based on the Web 2.0, including among others [Dremeau 2007, Perruzza 2008, Thomson 2005] :

- “Wikipedia”, which is a freely distributable online encyclopedia project that any user can ameliorate. The content proposed is published under GFL(GNU Free Document Licence),
- “edtags” where teachers and students list the links and articles that they put at the disposal of the community,
- “google for educators” from Google which provides educators with a space where they can find all the google applications, tools and services used for pedagogical aims,

- “facebook” which permits the learners to exchange their notes and courses,
- “Podcast” of American Universities which distributes in audio or video forms their courses on the Net.

The content is the central resource of a learning environment. Putting the course files online is not sufficient, we need to think of an actual intelligent space to use the content. We saw that there exists a solution with the platforms based on Web 2.0 but these platforms do not refit on a grid infrastructure, consequently their usage depends on the system that shelters them.

2.3 Brief general idea of Moodle

Moodle is an Open Source environment of e-learning, developed by Martin Dougiamas and the Moodle community. It manages the contents and creates interactions between teachers, learners and pedagogic resources through the network. It is within the pedagogical logic of the social constructivism that postulates that knowledge is constructed in the mind of the learner and not retransmitted in a static manner through textbooks or trainers [Wikipedia 2008]. The management tools are centered on the following principal modules: user, course management, test, assignment, wiki, opinion, and forum.

Moodle is more generic than the web 2.0 based environments cited above. It has indeed interesting functionalities for an e-learning environment, yet the learner-content interaction remains a problem. Indeed, the lessons available on this platform like in many learning environments are in the doc, pdf or html formats. In fact, a course is a set of knowledge units (or notions). From this view point, the follow up of the learner is reinforced if the teacher has the possibility of presenting to the learner the content that brings him gradually to the end of the fixed objectives. This is only possible with a structured content. Our contribution is therefore to reinforce the learner-content interactions by integrating to Moodle, a transparent module for the user capable of transforming all non structured content to structured content accompanied by a navigation interface in the content based on web 2.0. xMoodle obtained has just by its structuring reduced the cognitive load of the learner and reinforces the learner-content and teacher-content interaction of Moodle. Consequently, it reduces to a greater extend the transactional distance [Moore 1996].

2.4 Ontology

The ontology tackles the problem of reuse and sharing of knowledge while privileging the explicit representation of the sense. It is defined as an explicit and formal specification of a conceptualization of a domain of knowledge making the object of a consensus [Gruber 1993, Pierra 2003]. It permits among others: the communication between humans for it favors the sharing of terms, the communication between humans and the enterprise for it standardizes the vocabularies, the indexing and the research of information.

To facilitate the understanding and the sharing of the knowledge between training communities, we need a tool that allows the actors to be okay on the terms they use. A learning environment has to permit the learner to access easily the concepts which he needs. The courses put at the disposal have to respect the standard defined by the communities that collaborate. In the framework of the present collaboration environment, we have chosen the ontologies to respond to this problem.

2.5 Intelligent agent

There are several definitions of an agent. Within the framework of this present work, we will choose that of Wooldrige and Jennings who define an agent as a computer system, situated in an environment, and that reacts in an autonomous manner to attain the objectives (goals) for which it was designed [Wooldridge 1995]. An agent is said to be intelligent when it is conceived to execute tasks based on its proper knowledge and message that it has received.

The agent service of GELSOTC comprises two types of intelligent agents:

- A recommendation agent: It plays the role of tutor; he helps the learner by conducting him to the end of his initiation. For example advise a learner a particular work group that can help the latter.
- A notification agent: It sends notification messages to the members that collaborate. For example, he can remind a group of the work which they took an appointment to work, specifying the planning which the members of the group established and even the theme on which they promised to debate. It sends messages to mobile phones of members and by e-mail.

Our contribution is a solution to the collaboration between heterogeneous environments through our infrastructure of the grid. Consequently, it favours the collaboration between the intelligent agents of the systems connected to the grid. A recommendation agent of a given system can for example solicit the support of a recommendation agent of another system. We will detail these concepts of “grid-agent” in the framework of another article.

3. Specification of the Gelsotc Environment

GELSOTC is a software collection that regroups a platform of content management called “xMoodle 2.0” and the GLS services :

- « xMoodle 2.0 » is the extended Moodle plate form based on the Web 2.0. Extended for it is being integrated complementary functionalities that comprises: a navigation interface in the content of the training, a service of a particular group called the study group, and a forum adapting the present context of the collaboration.
- The services of GLS are among others:
 - o The collaboration: It is a conversation process that goes beyond a simple exchange of information for it takes into account the social context of learning. Collaboration implies the adhesion of the community, the

distribution of tasks, and the sharing of knowledge in order to attain a common goal. As an example of collaboration in the framework of this work, we have the service “ontology of the domain”. The other collaboration services as instant messaging, electronic conferences, the forum, and the study group are integrated into the xMoodle 2.0;

- o Communication: when it has to do with direct communication between the actor entities (humans, hardware, or software) of GELSOTC, we talk of a peer to peer communication; in this case software agents are used for communication. For example the agent of recommendation plays the role of counseling and the agent of notification informs the learner of the rendezvous in a study group. More so, Multimedia Message Services also enable the communication with the server towards deposit and withdrawal of content.
- o E-qualification : it enables the individual evaluation of the actors during their evolution within their community. To this effect, it provides the learner with the best virtual organization, i.e. that corresponds to the learning profile.

3.1 Objectives of GELSOTC

The objectives of GELSOTC are situated at two scales: a macroscopic scale where it has to do with laying emphasis on the activities carried out in each training community; and a microscopic scale where there is the need for the exploitation of the existing services in order to respond to the needs of the distribution of the learning objects of the different communities.

3.1.1 Macroscopic Scale.

On the macroscopic plan, GELSOTC puts at the disposal of the communities a common centralized unit called xMoodle 2.0. This is an extended Moodle[Wikipedia 2008], comprised of a structuring module of the contents and a web interface reusable to navigate in any structured and shared content of our learning environment.

GELSOTC establishes relations between the following five types of actors: the learner, the teacher, the author, the manager of the community, and the administrator. Seen from the community angle, it enables the constituted collaborations network to :

- publish the pedagogic contents (courses, evaluation, etc) of the different communities,
- share their seminars and pedagogic meetings through the video conference service,
- organize evaluation sessions between the training levels,
- inform a given community on its evaluations through statistical data,
- Propose their laboratory or their library through the virtual reality service,
- facilitate the availability of competences through the services of recommendation agents.

- Inform the learners, teachers, or community managers of the availability of important resources: ensuring a permanent technological watch with respect to a particular need of the moment. The actors receive, when necessary, information through their mobile phones and they can equally access contents through the same channel.

3.1.2 *Microscopic scale.*

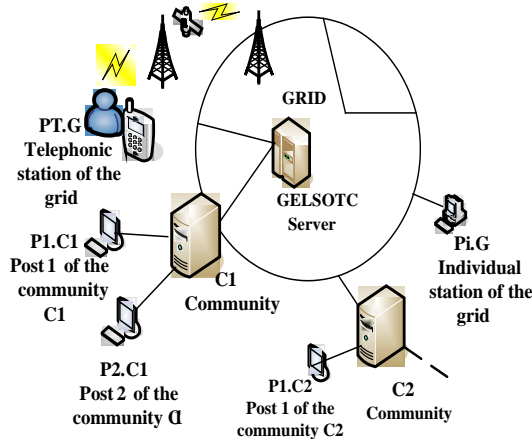
On the microscopic plan, with a regard to the actors of the constituted network, GELSOTC enables the latter to :

- constitute themselves in study groups comprising learners from different communities. These learners can plan their study sessions and be notified of the program and the study time through their mobile telephones by the notification agents of the system.
- to access resources of their study level in the constituted network. For a given study level of the network, we have as much content as the communities of the network,
- to access tutor-type resources or assistance through the recommendation agents,
- to navigate in varying course contents validated by the ontology of the domain,
- research learning resources through a research service based on the ontology,
- visualize the learning objects belonging to their virtual community,
- etc.

3.2 **General functioning**

3.2.1 *Logical architecture of the GELSOTC interconnection.*

The communities connect to the GELSOTC server and form with it the Grid. The network thus constituted consists of a server, the set of computers of the communities, individual computers, and mobile telephones. The Figure 1 presents the logical architecture of the constituted network.

Fig. 1. Logical Architecture of the GELSOTC interconnection

At the middle of the above architecture (Fig. 1), we have a principal server, which will contain common services, for example the management of the learning content xMoodle 2.0, the service of “recommendation agent”, etc. At the peripherals we have the secondary servers that represent the training communities that constitute a collaboration network ($C_1, C_2, C_3, \dots, C_n$).

There exist three forms of connections to the grid:

1. A user can be connected to the grid from a post of the local network of his training community. For example the post P1 of the community C1 (P1.C1).
2. A user can be connected to the grid through his personal post. For example the post P_i of the grid (P_i, G).
3. A user can be connected to the grid from his mobile telephone. For example the post PT of grid (PT.G). The connection through the mobile telephone passes through a GSM modem that is not represented in the diagram. This form of connection has a very important role for a country like Cameroon in which the use of mobile telephones is expanding day after day. The communities that don't have internet can connect to the GELSOTC server through their mobile telephones [Djotio 2009]. A stock zone is reserved to them to allow them stocking their resources. And so we can constitute a virtual grid. More so, the users can access the content, receive notifications or latest information using their mobile telephones.

Such an organization in terms of interconnected equipments is mapped a logical organization of the human entities that constitute the collaboration network. This second form of organization that present the functional view of GELSOTC is represented by Figure 2 below.

3.2.2 General functioning diagram of GELSOTC

Fig. 2. General functioning diagram

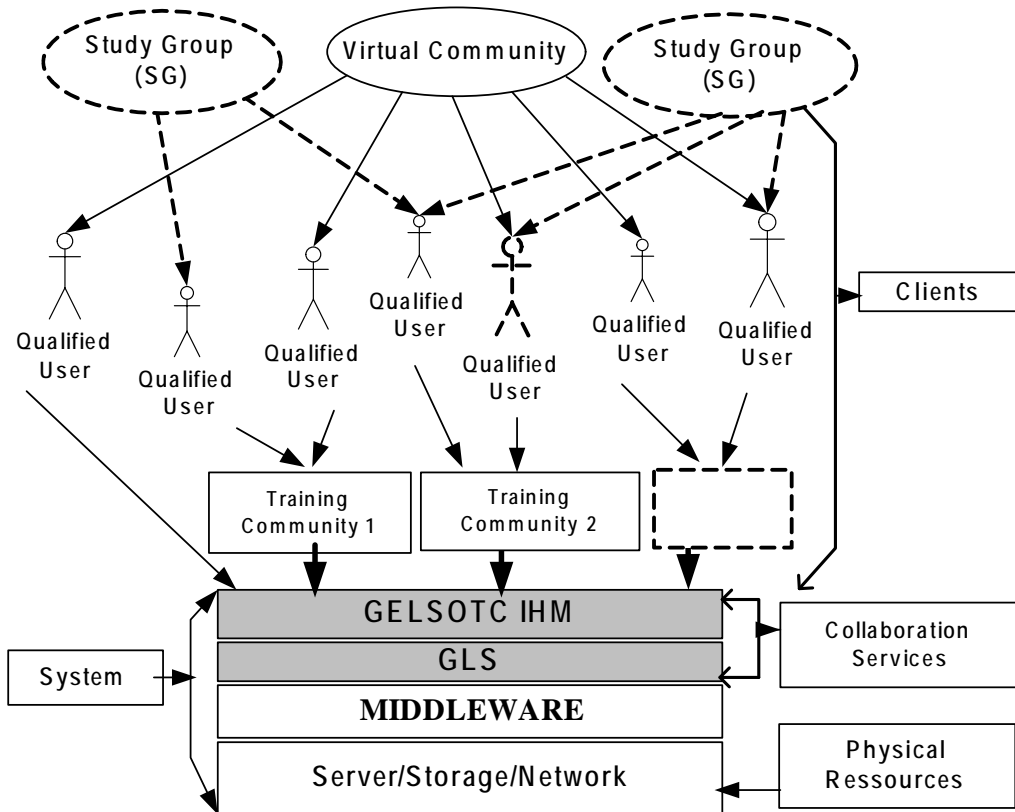


Figure 2 is a set of two big blocks:

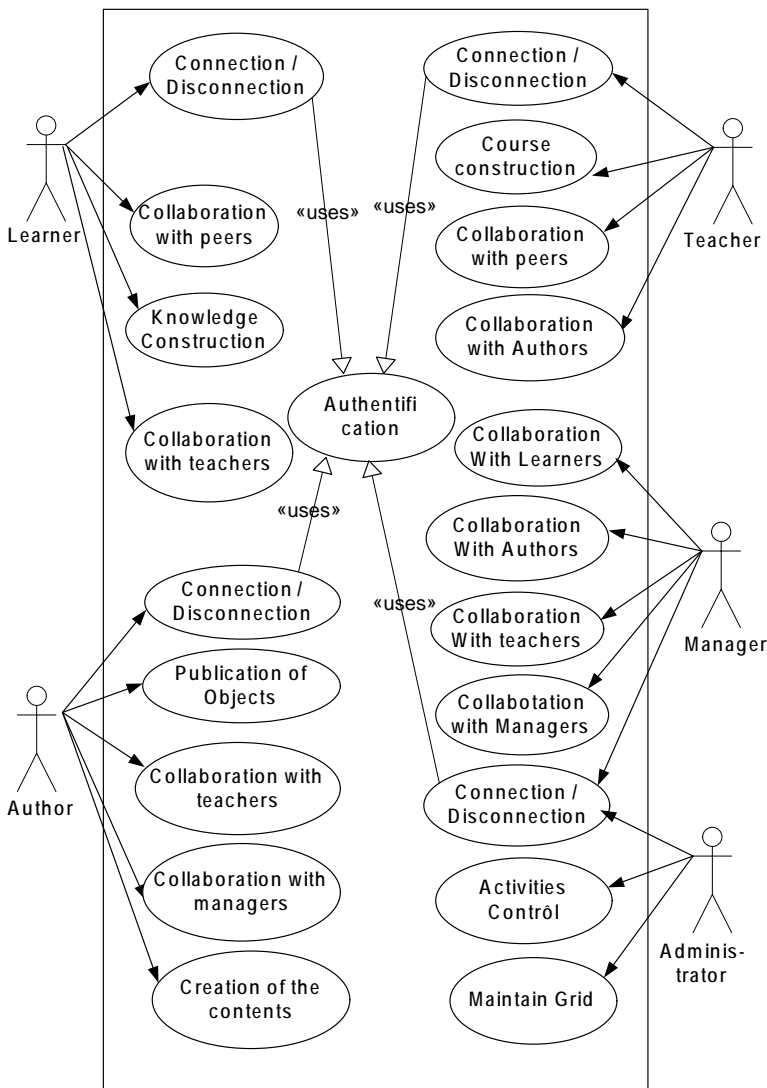
- a block “clients” composed of four entities: the user, the study group, the virtual community, and the training community:
 - o The user is a qualified user (the system masters its profile); he is either a member of one of the communities or a guest to the grid.
 - o The study group is a regrouping restricted to at most 5 learners of the constituted network who have decided to work program in common.
 - o The virtual community is similar to the virtual organization, it has to do with a regrouping of persons having the same profiles around the same resources. The number of members is unlimited.
 - o “Training community P” represents a community in the real sense that has decided to form with some of its peers a collaboration network.

- a block “System” is composed of layers of the grid infrastructure : a lower layer composed of physical and logical resources (data and resource servers) ; a middle layer (Middleware) based on the standard OGSA; a GLS layer comprising the collaboration services, communication and e-qualification, etc; and the GELSOTC layer that is the shop-window or the interface with the user.

3.2.3 Use case diagram.

Figure 3 presents the use case diagram. The user interacts with the system through the services. He can be a member of a community of the network or not, however he integrates one of the following five roles: learning, teaching, author, community manager, and administrator.

Fig. 3. Use case diagram



3.3 Technical Architecture of GELSOTC

GELSOTC has a layered technical architecture as represented by Fig. 4.

Fig. 4. GELSOTC technical Architecture

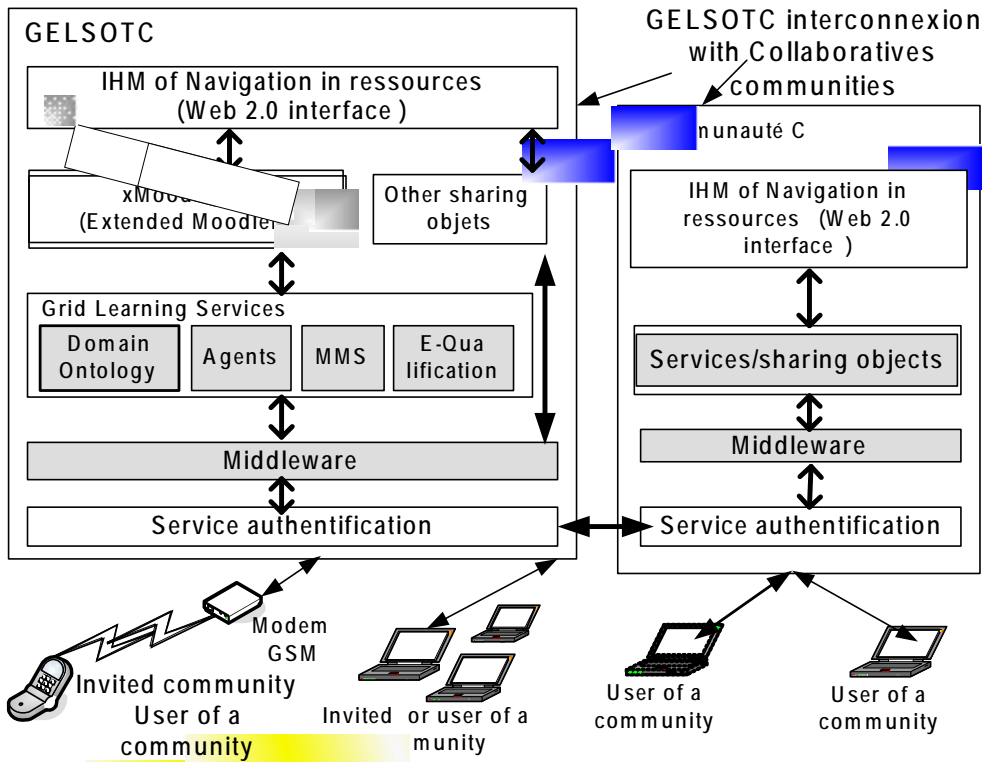


Figure 4 is essentially made up of three functional blocks:

- a central bloc composed of:
 - An authentication layer that permits the interconnection of guests or communities to the Grid.
 - A MiddleWare Grid layer which is the central element of the grid, in charge of communication and data transfer between the computers, components or programs. It is composed of the layers of the architecture.
 - A GLS layer that comprises the services of : ontology of the domain that validates the content, recommendation agents or tutor counselor and the notification agent, MMS (Multimedia Message Services), etc. This layer is the centre of the collaboration service, of e-qualification, and communication.
 - An application layer, which is composed of the xMoodle 2.0 services. xMoodle 2.0 is an LMS (Learning Management System) that is situated at the heart of GELSOTC because it provide contents. It integrates a study group and an interface based web 2.0 used to navigate in contents.

- o And a navigation IHM on the resources of the grid, that constitutes the desktop of a user connected to the grid.
- A peripheral block in the direction of the compound communities:
 - o An authentication layer that permits a guarantee of security of access to the resources of the customer.
 - o A Middleware layer, in charge of the communication and transfer of data between the computers, components or programs.
 - o An agent layer composed of a management agent of the connected community who is the virtual representative of the community, and an actor agent that represents the user of the community connected to the grid.
 - o A layer of services or shared objects that the community makes available in the framework of the collaboration.
- A peripheral block in the direction of the user: these users have at their disposal internet connection on their working places and are connected directly to the grid; the users that don't have internet can be connected to the grid from their mobile phones.

4. Implementation and Results

We have presented in the section 3.3 the layered technical architecture of GELSOTC. The heart of the present environment is composed of the GLS and the middleware layers. Now our interest is centered on the implementation of the middleware layer. We have opted in the framework of this work on the association of technologies JINI and RMI. So, our middleware is essentially composed of the sub layers RMI, JINI, and communication.

4.1 The RMI sub layer

It is the basic middleware sub layer. RMI (Remote Methode Invocation) is a mechanism for the invocation of a remote object method. It is an evolution of RPC (Remote Procedure Call) [Sun 1999]. It is application of the JAVA object methods. With RMI, each remote object needs the use of stub by the client and a skeleton by the server for the good functioning of communications. For our middleware, RMI is used transparently by JINI. The generations of stub, skeleton and the registering in the RMIRegistry are under the care of JINI mechanisms.

4.2 The JINI sub layer

JINI is an API developed by SUN which adds to the JAVA programming model the notion of service [OAKS 2000, Sun 2000]. A JINI network is generally called a community in the sense that it regroups the entities that have access to each another. We distinguish principally two types of entities: the service and the client. The service is an entity capable of realizing a certain task for the set of entities that are members of the community. The client on its part uses the said services. Each community regroups its members around a centralized service called lookup. Its principal role is

the management of the discovery of services that are found around it. Its participation as an intermediate communication entity between the client and the service, is limited to the discovery process. Once established, the communication is carried out directly between the client and the service. From a technical point of view, JINI offers a software infrastructure, permitting the JAVA objects, called services, to discover and use themselves spontaneously.

4.3 The communication sub layer

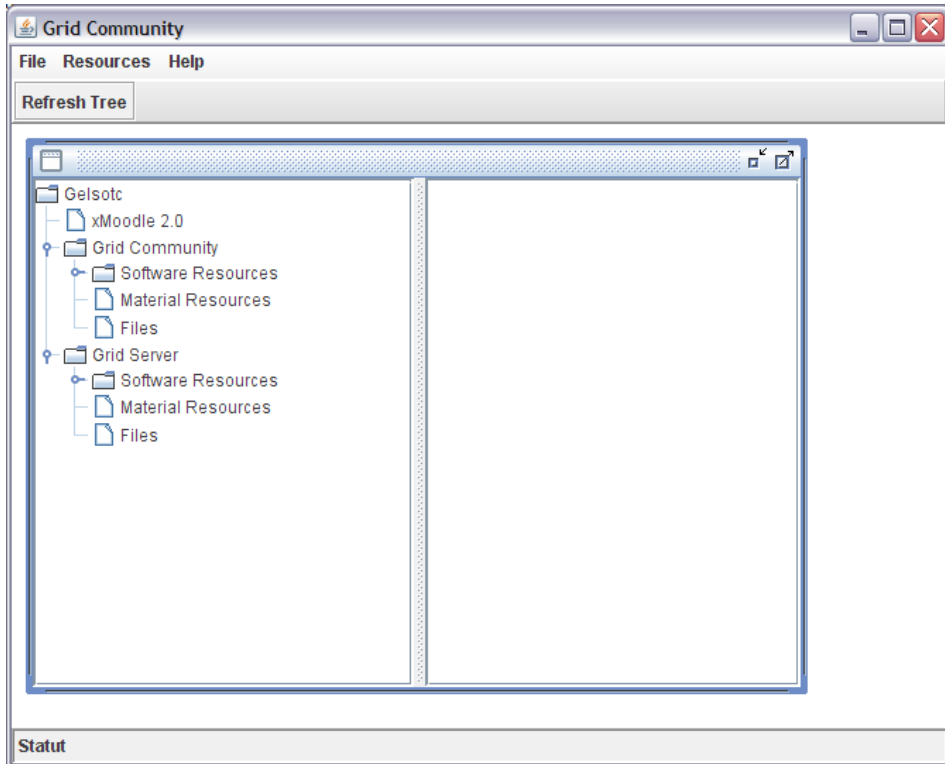
The sub layer of the communication service allows principally the management of resources. It establishes the link between JINI and the resources of the communities. In fact, JINI services invoked in the previous section are considered here as resources. The sub layer of communication service allows the display under a tree form all the resources of the machine on which it is installed and equally the communities connected to the grid. It also permits the management of resources by loading them into a sort of warehouse (for example a folder of the machine on which is launched the intergiciel and which is visible in the JVM classpath).

4.4 Results

The middleware that we have developed are installed on the servers of the community and the principal server. Putting in place the middleware needs the installation of the pile of layers presented (RMI, JINI, and communication) on each server. In a practical manner, in each server runs a Lookup service for the registration of resources that the members of the community will publish. The grid is realized in such a way that when a Lookup is added or deleted (which means that one server is off, removed from the grid or started), then all the components of the middleware situated on all the other posts (principal or community server) are notified and the bank of resources available on the grid is updated. In this way, the user can therefore notice that new resources were added or have disappeared. At the level of each community or of the principal server (the machine on which is installed xMoodle) the authorized user can add or remove resources. The description of these resources will be registered in the corresponding Lookup.

Figure 5 presents the access interface to the resources of the communities of the network. It has to do with the view of a connected community: we note the presence of xMoodle, the resources of the community, and the resources of the central server.

Fig. 5. Interface of access to resources : community view



In this interface, the “Resources” menu allows the management of resources (updating of resources). The “Refresh Tree” allows refreshing and shows the communities which are connected to the grid with time.

5. Discussions

GELSOTC is a platform designed to allow the constituted training communities to form networks of help: the communities can put at the disposal of their members, pedagogically rich and varied training contents. It is used to increase the performance of the communities of the constituted networks.

The main idea is to transplant an application of content management on a layer of the basic services used by the learners and teachers in training situation :

- The learners through their computer reach the different communities where they find the corresponding resources to their profiles. They sail in a transparent manner in the contents formation without worrying about the system that shelters these contents.
- The teachers using their computer to put their teachings at the disposal of the communities. They can converse with the learners who are connected.
- The communities through their manager put at the disposal of the other communities a software resource, for example a virtual laboratory and virtual library.

The specifications above show that GELSOTC would like to transform the office of:

- The learner in a virtual class. The learner can: have several teachers in his class, converse with the learners of his level, and especially form a study group with the learners of his choice.
- The teacher in a virtual school with virtual classrooms where he offers some teachings. The teachers can put their lessons at the disposal of the students of a given level, receive the feedback on the teachings published, teach several learners at a time, communicate with communities managers on pedagogical topics, etc.
- The community manager in a virtual school where he manages learners, teachers, and resources at his disposal for a good running of the school. He can put at the disposal of the other communities the educational resources, inquire about the evolution of a virtual given class of his community, etc.

When we refer ourselves to the computer grid, it follows that the user can access in a transparent manner the resources which he needs. The above description shows that the actors don't have to give an account of the technical disposals that permit them enjoying the resources put at their disposal. Consequently, GELSOTC wants to be a model of Grid for Learning Management Systems (GLMS).

6. Conclusion And Perspectives

The present work is the integration within a single environment that we have baptized GELSOTC, a technology of grid and Web 2.0 concept. The proposed environment allows regrouping the communities into a common centre of interest in the mutual-aid network and to widen to the maximum the exchanges between these communities. Therefore this provides the needs in educational resources of the communities of the networks formed. The communication and the interaction between the members of the different communities of formation are improved thanks to a panel opened to collaboration services:

- The learners of the different communities of the network have the possibility to collaborate following their knowledge level: they can share the points of view on a common topic from the forum, they can constitute a study group. At the time of the sessions of collaboration within such group they can contribute to the enrichment of the content.
- The trainers of the communities have the possibility to harmonize their teaching at the end of the assessments made in common in the network of exchange or proposed in the different communities, to bring their suggestions on the contents proposed by their colleagues.
- The managers of the communities have the possibility to share their experience with their equals, to value their own community of formation from the results of assessment common and statistics of evolution of the learners in the groups of survey formed, etc.

This work is a contribution to the value added conception of a Grid Learning Service (GLS) (with the presence of services as MMS, the study group, etc.). It reduces the transactional distances between the members of the grouped communities and consequently reduce the gap between them. We are very much interested in the specification of GELSOTC. The continuation of our works will be on the implementation: first the implementation of the middleware that coordinates the activities of the grid, and next the implementation of the interface layer (IHM of navigation in resources) and basic services such as the agent and ontology services. The project will end with the gradual integration of the usual collaboration services as messaging, chat, study group, etc.

Acknowledgments

We are grateful to numerous colleagues and friends for discussions on the topics covered in this pages, in particular (in alphabetical order, with apologies to anybody we've missed) Fomban Collins Ken, Khan Tega Martin, Nashipu Julius, Taku Alemandji John, Njanwouo Boris, Yekel Augustin,.

References

- BATCHAKUI, B., AND AL. 2007. Environnement de collaboration basé sur le Grid Learning Services (GLS) pour les communautés de formation à centre d'intérêt commun. *In Proceedings of the 4th LORNET Annual Scientific Conference (I2LOR)*, Montreal, 7 november 2007. Available at <http://www.lornet.org/Nouvelles/I2LOR2007ProceedingsDay1/I2LOR2007ProceedingsDays3/tabid/574/language/en-US/Default.aspx>.
- BEBC -Baccalaureate Examination Board of Cameroon . 2008. Charts of the exams 2007. *Cameroun-tribune of Tuesday January 15, 2008*. Available at <http://www.camexamen.com/download/palmares-examens-2007-etablissements-cameroun.pdf>,
- DJOTTO, T. N., AND AL. 2009. VeSMNetMon(VeSMp Network Monitor): a VeSMp-based Tool for Network/System remote administration. *In proceedings of IWSS-2009 -International Workshop ARES/CISIS 2009*, Paper ID 288, for March, 16th - 19th 2009; ppxxx-xxx. Available at <http://www.nfc-research.at/clessec/>, http://sebastienaumonier.blogspot.com/2009_01_01_archive.html; 03 April 2009
- CRI. 2002. DIOGENE, from information to knowledge. 1- april. Available from <http://conseil-recherche-innovation.net/environnements-distribues-pour-le-learning-et-grid-technologies>.
- DREMEAU, I. 2007. Panorama du Web 2.0, december . Available at <http://www.scribd.com/2153936/doc/panoramama-Wb-20/>
- DUGENIE, P., 2005. Orientation et usage de l'Architecture de Services Grille OGSA. *In proceedings of the MajecSTIC 2005*, pages 283-290, Rennes, France, November 2005. IRISA IETR – LTSI
- FOSTER, I. AND AL. 2001. The Anatomy of the Grid, Enabling Scalable Virtual Organisations. *International Journal of Supercomputer Applications*. Available at <http://www.globus.org/alliance/publications/papers/anatomy.pdf>
- FOSTER, I. , AND AL. 2002. The physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. [pdf] draft document.
- GOUARDERES, G., AND AL., 2005. The grid e-card Architecture to share collective intelligence on the grid. *Applied Artificial intelligence*, October-december , vol 19 No 9, p. 1043-1043.

- GRUBER, T. 1993. A translation approach to portable ontology specification. *Appeared in Knowledge Acquisition*, 5(2):199-220. Available at <http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>
- KESSELMAN, C., AND FOSTER, I. 1998. The Grid, Blueprint for a new Computing Infrastructure. Morgan Kaufmann Publishers, first edition, 1998.
- MOORE, M.G, AND KEARSLEY, G. 1996. Distance education : a system view. Boston, MA: Wadsworth Publishing. 5. *Gunawardena, C. N. (2003)*
- NKAMBOU, R., AND AL. 2005. Toward Learning Grid Infrastructure: An Overview of Research on Grid Learning Services. *Applied Artificial Intelligence*, vol.19 No.9, p.811-824.
- MIREILLE, A. 2006. Les services de collaboration à travers la grille d'apprentissage. memory of master in data processing option engineering of knowledge. University of Quebec Montreal, department of data processing (LGDAC), december .
- OAKS, S., AND WONG, H. 2000. JINI in a Nutshell. *O'Reilly, first edition, March 2000*. Available at <http://www.livres-chapitre.com/-10GUS3/-OAKS-SCOTT/-JINI-IN-A-NUTSHELL.html>
- PANKRATIUS, V., AND VOSSSEN , G. 2003. Towards ELearning Grids: Using Grid Computing in Electronic Learning. In proceedings of the IEEE Workshop on Knowledge Grid and Grid Intelligence (in conjunction with 2003 IEEE/WIC International Conference on Web Intelligence/Intelligent Agent Technology, p.4-15.
- PERRUZZA, L. 2008. Le web 2.0 dans le e-learning et le partage des connaissances. Available at <http://www.scribd.com/doc/2422208/web-2-0-dans-le-elearning-et-le-partage-des-connaissances> - 144k
- PIERRA, G. 2003. Context-Explication in Conceptual Ontologies: The PLIB Approach. *In Proceedings of CE'2003 : Special track on Data Integration in Engineering*.
- WOOLF, B., AND ELIOT, C. 2005. Customized the intructionnal GRID. *Applied Artificial Intelligence, October-December , 19:825-844*. Available at <http://www.informaworld.com/smpp/content~content=a725750364~db=all>
- SUN, M. 1999. Java™Remote Method Invocation Specification. Available at <http://java.sun.com/j2se/1.3/docs/guide/rmi/> or <http://www.dhpc.adelaide.edu.au/reports/072/dhpc-072.pdf>
- SUN, M. 2000. A Collection of Jini™Technology Helper Utilities and Services Specifications. October 2000. Available at <http://www.sun.com/jini/specs/jini1.1html/collection-title.html>.
- THOMSON, B. 2005. What is it with Wikipedia ? *BBC, December 16, 2005, 12:24 GMT*. Available at <http://news.bbc.co.uk/1/hi/technology/4534712.stm>
- WIKIPEDIA. 2008. Moodle. Available at [http://docs.moodle.org/fr/A_propos_de Moodle](http://docs.moodle.org/fr/A_propos_de_Moodle).
- WOOLDRIDGE, M., AND JENNINGS N. R.. 1995. Intelligent Agents: Theory and Practice. *Knowledge Engineering Review 10(2)*. Available at <http://www.csc.liv.ac.uk/~mjw/pubs/ker95/ker95-html.html>

2

Developing a Set of Requirements for Algorithm Animation Systems

Jean Greyling

The learning and analysis of algorithms and algorithm concepts are challenging to students due to the abstract and conceptual nature of algorithms. Research conducted at the Nelson Mandela Metropolitan University in Port Elizabeth, South Africa focused on the design of an extensible algorithm animation framework to support the generation of interactive algorithm animations. A comprehensive literature study showed that no unified and common instrument for evaluating the pedagogic effectiveness of algorithm animation systems exists. An output of this research was the compilation of a list of requirements that could serve as such an instrument. The value of this list is proven through the evaluation of extant systems as well as in developing a prototype in order to prove the effectiveness of the animation framework.

1. Introduction

Algorithm animation is a form of technological support tool which encourages algorithm comprehension by visualizing algorithms in execution. Algorithm animation can potentially be utilised to support students while learning algorithms.

Despite widespread acknowledgement for the usefulness of algorithm animation in algorithm animation courses at tertiary institutions, no recognized framework exist upon which algorithm animation systems can be effectively modelled. Research conducted at the Nelson Mandela Metropolitan University in Port Elizabeth, South Africa focused on the design of an extensible algorithm animation framework to support the generation of interactive algorithm animations [Yeh, 2006]. A comprehensive literature study showed that no unified and common instrument for evaluating the pedagogic effectiveness of algorithm animation systems exists. Before designing a framework, it was thus important to compile a set of requirements for an algorithm animation system. This paper reports on the process that was followed in compiling and applying this set.

A literature study identifies features that increase the instructional value of algorithm animation systems. The identified features are organised into a set of requirements for analysing algorithm animation systems (Section 2). Selected extant algorithm animation systems are then evaluated making use of these requirements (Section 3). The paper concludes with a discussion on the prototype system that was developed and how it supported the requirements (Section 4).

2. Desirable requirements for an Algorithm Animation System

The focus of algorithm animations has moved beyond merely showing students an algorithm in the hope that they will understand and retain some of the algorithm concepts illustrated. Current emphasis is placed on identifying factors which will increase the instructional value of algorithm animations [Stasko, Badre and Lewis 1993; Saraiya 2002].

This section reports on a literature study on algorithm animation features that would increase the pedagogic effectiveness of an animation system or broaden its usefulness within a teaching environment. The findings are summarized and presented as a list of requirements which can be used to measure the effectiveness of a system.

2.1 Requirements based on Levels of Engagement

Studies have suggested that rather than just letting students view an algorithm animation passively, better learning results may be obtained by allowing students to engage interactively with the animation [Hundhausen 2002; Naps, Fleischer, McNally et al. 2003]. Naps et al [2003] defines taxonomy of students' interaction with algorithm visualisations based on six levels of engagement:

No viewing No algorithm visualisation is utilised.

Viewing Algorithm animations can be viewed as a surveillance video that records and displays the execution of an algorithm. When users investigate an algorithm, they may slow the video down to better examine a particular event, speed through events which offer no further contribution to the investigation, or step through key events one at a time.

Responding The system should support the activity of letting students make predictive answers by running animations in discrete steps, thus allowing the students to pause before each interesting event in the animation to predict the next algorithm action [Anderson and Naps 2000].

Changing Students and instructors should be allowed to input custom data into the algorithm. Instructors will thus be able to demonstrate algorithm specific characteristics to students, such as best-case and worst-case performance scenarios [Naps, Eagan and Norton 2000; Saraiya 2002].

Constructing Creating an animation of the algorithm under study would induce students to have a deeper understanding of the algorithm's operations, since students must learn the algorithm with the intent of sharing their understanding of the algorithm concepts to an audience [Hübscher-Younger and Narayanan 2003].

Presenting Presenting visualisations to other students to stimulate discussions on the given topic.

The features are identified and organised based on four levels of engagement, namely viewing, changing, responding and constructing. The first and sixth level of engagement - No viewing and Presenting - are not seen as applicable in this discussion. No viewing is essentially the absence of algorithm animations. Presenting involves the

learner demonstrating an algorithm animation. It is thus an activity generally performed by the instructor to aid learners.

2.2 Complementary Requirements

A number of additional algorithm animation system requirements are identified which are believed to enhance the pedagogic effectiveness and usefulness of the system. Each of these requirements is discussed below.

Smooth animation aids the student intracking changes between discrete steps of an algorithm [Stasko 1998b]. This feature forms a fundamental part of algorithm animation. In certain cases, such as when large datasets are being viewed, students should be able to disable animations and view discrete steps of the algorithm [Rößling and Naps 2002].

Analysis features can aid students in better understanding the efficiency of an algorithm and the relative performance differences among various algorithms [Gloor 1998]. Relative performance can be illustrated by running several algorithms simultaneously, thus letting the students compare the differences visually [Naps, Fleischer, McNally et al. 2003].

Multiple views of an algorithm may be used in different approaches to aid students. Students may find the use of certain metaphors easier to understand, and thus prefer a certain approach of animation [Gurka and Citrin 1996]. Different views may also be used to illustrate algorithm executions at different levels of abstraction, or demonstrate different characteristics, such as operational or performance trends [Wilson, Katz, Ingargiola et al. 1995; Naps, Fleischer, McNally et al.2003].

Additional materials accompanying algorithm animations may increase the instructional effectiveness of the animation. The materials may include simple textual explanations, pseudo-code or source code views [Rößling, Schüler and Freisleben 2000]. Alternatives include using multimedia elements, such as audio and video of instructors explaining the algorithm [Stasko, Badre and Lewis 1993].

2.3 List of Requirements

From the preceding discussion, a number of features are identified as the requirements for a pedagogically effective system. The identified requirements appear in Table 1. The effectiveness of an algorithm animation system to complement the students' study of algorithms is determined by the system's ability to engage the students in an active learning process (categorised as requirements R1 through R6), and system features which either provide additional information to enhance comprehensibility of the animation, or increase its usefulness in an educational environment (requirements R7 through R10).

Table 1: List of identified requirements

Requirements for Algorithm Animations R1: Allow speed control of algorithm animation
R2: Allow rewinding of the animation
R3: Accept user input data for the algorithm
R4: Provide questions to predict algorithm behaviour
R5: Allow stepping control of algorithm animation
R6: Support construction of animation by students
R7: Support for smooth motion
R8: Include capabilities for comparative algorithm analysis
R9: Provide multiple views of an algorithm
R10: Provide additional instructional material

3. Evaluation of Extant Systems

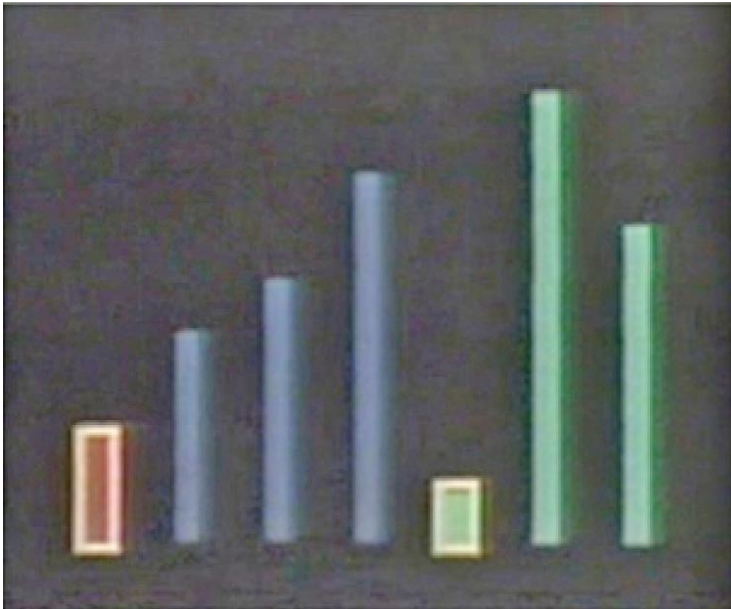
The compilation of a set of requirements for Algorithm Animation Systems addresses the concern that there is no unified and common instrument for evaluating the pedagogic effectiveness of algorithm animation systems. In order to highlight the value of the list, the reported research made use of the list to evaluate the following systems:

- Sorting Out Sorting
- Brown University Algorithm Simulator and Animator – BALS
- Generalised Algorithm Illustration through Graphical Software – GAIGS
- Java Collaborative Active Textbook – JCAT
- SAMBA/JSAMBA
- Java and Web-based Algorithm Animation – JAWAA
- A New Interactive Modeller for Animations in Lectures – ANIMAL

This paper will report on three of these evaluations.

3.1 Sorting Out Sorting

The “Sorting Out Sorting” video [Baecker 1998], although not strictly defined as a system, is nevertheless, worthy of mention. The video employed a number of features which were unprecedented at the time in demonstrating sorting algorithms. These features include the use of various visual metaphors, including animation, colour, audio, and voice-over commentary. The operations of nine sorting algorithms were illustrated, followed by time versus data size performance graphs typically found in textbooks. The nine algorithms were then run simultaneous in a race to compare and contrast their performance characteristics.

Figure 2: “Sorting Out Sorting”

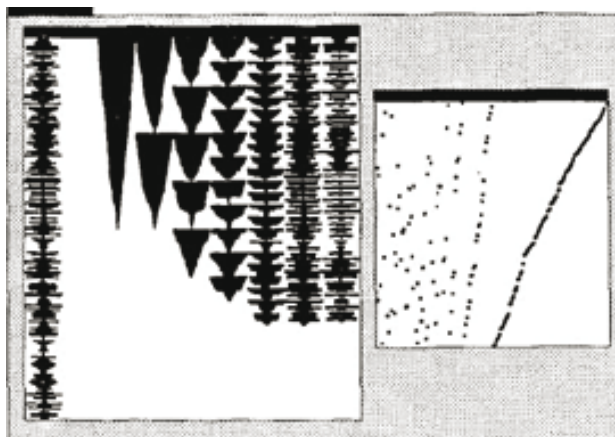
The video is identified as supporting smooth motion (R7). Comparative algorithm analysis and multiple views of an algorithm are included (R8, R9). Additional material is provided through audio narratives (R10). Because the animations are produced as a video, all requirements are essentially of historical nature and thus regarded as partially supported.

3.2 Brown University Algorithm Simulator and Animator II (BALSA)

The BALSA animation system [Bazik, Tamassia, Reiss and van Dam 1998] can be regarded as a concept prototype for all current systems due to the novel design concepts utilised by the system. The system was designed and implemented to integrate into Brown University's electronic classroom concept.

The system supports dynamic input to generate animation (R3), however, literature did not specify if the feature is directly accessible by students. No details were available on the level of animated motion support by BALSA. The system allows for speed control of algorithms (R1) and stepping through animations (R5). As shown in Figure 1, the system also supports capabilities to compare algorithms (R8), and show alternative animation views (R9).

Figure 1: BALSII - Illustrating a mergesort using a clouds view, and a bar chart to show consecutive states of the data.



3.3 Java And Web-based Algorithm Animation (JAWAA)

JAWAA [Pierson and Rodger 1998; Akingbade, Finley, Jackson et al. 2003] is an algorithm animation system which employs a scripting language. The visual objects and associated commands available in JAWAA are designed for the animation of algorithm operations, with specific support for data structure objects like arrays, stacks, queues, pointers and linked lists.

Students have speed control of animations (R1). The scripting language graphic system gives more accessibility for constructing various animations (R6). The system is implemented in a Java™ applet, the embedding of static material into the client webpage is thus possible (R10). Smooth motion is supported (R7).

Evaluation of the seven systems listed showed that all the systems only satisfied a subset of the requirements reported on in Section 2. This emphasised the need for a framework that would support all these requirements.

4. Functional Prototype System

In designing a framework for Algorithm Animation, the reported research aimed at addressing the limitation that no extant system satisfied all the requirements identified from literature. In order to demonstrate the effectiveness of the framework, a functional prototype system was developed. The prototype system was developed to animate the quadratic and $O(N \log N)$ sorting algorithms commonly taught in introductory algorithm curricula.

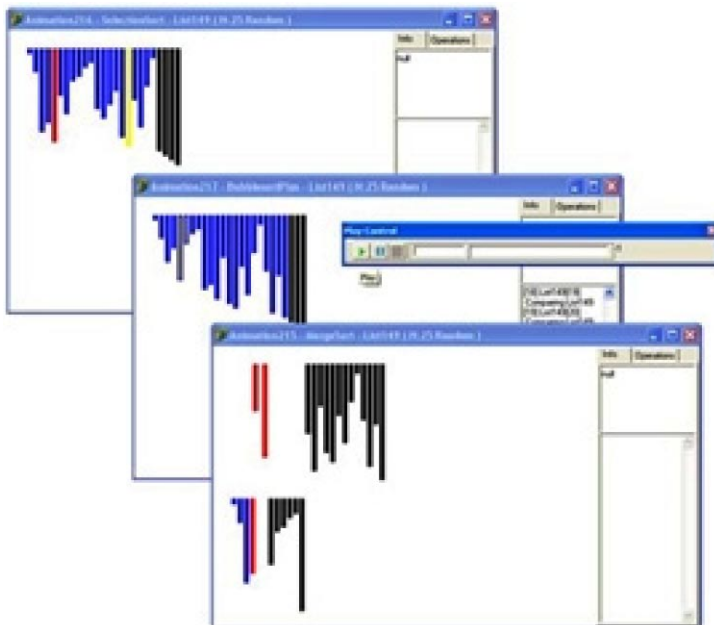
User interaction with the algorithms and data are provided through a data layer interface which supports the requirement of allowing the user to input data for the algorithm (R3).

The design of the prototype centres on the concept of using a unified algorithm animation desktop, which allows the user to control the generation and display of

animations. The desktop acts as a centralised placement area for animation panels, each of which presents a particular scenario. This feature of the prototype supports the requirements of speed (R1) and stepping (R5) control of algorithm animation, smooth motion (R7) as well as the construction of animation by students (R6).

The algorithm animation desktop further provides an integrated platform for comparative analyses, allowing any number of potential combinations of data lists and algorithms to be created and examined. These include comparisons of a data list using different algorithms, or different data lists using a single algorithm. When multiple scenarios are to be compared, an animation panel is created and setup for each scenario (Figure 2). The animation panel, once created, is placed within the unified desktop. Scenarios can then be individually selected for parallel display. This feature supports the requirements of comparative algorithm analysis (R8) and multiple views (R9)

Figure 2: The play control managing an animation race



Since this prototype was developed as an instrument for an existing Data Structures module within the department there was no need to develop additional instructional material (R10).

Based on the scope of the initial project, it was decided not to support rewinding of animation (R2) as well as the provision of questions to predict algorithm behaviour (R4). In a follow up project which investigated the extensibility of the designed framework, these two features were successfully added [Moruge, 2006].

5. Conclusion

Algorithm animation can potentially be utilised to support students while learning and analysing algorithms. Literature studies showed that there was no unified and commonly accepted requirements framework for evaluating the effectiveness of algorithm animation systems in an algorithm course environment. The derived list of requirements (Table) is a theoretical contribution towards identifying instructionally effective features of algorithm animation systems.

The list of requirements can be used as a foundation for the creation of a unified algorithm animation system evaluation instrument. Tertiary educational institutions can also utilise the proposed requirements as a preliminary method for evaluating the suitability of algorithm animation systems in particular course environments, with the aim of integrating the systems to complement the institution's existing teaching strategies.

References

- AKINGBADE, A., FINLEY, T., JACKSON, D., PATEL, P. and RODGER, S.H. (2003): JAWAA: Easy Web-Based Animation from CS 0 to Advanced CS Courses. Proc. Technical Symposium on Computer Science Education, Reno, Nevada, USA. 35:162 - 166, ACM Press
- ANDERSON, J.M. and NAPS, T.L. (2000): A Context For The Assessment Of Algorithm Visualization Systems As Pedagogical Tools. Proc. First Program Visualization Workshop, Helsinki, Finland.
- BAECKER, R. and PRICE, B. (1998): The Early History of Software Visualization. In *Software Visualization: Programming as a Multimedia Experience*. 29 - 34. STASKO, J., DOMINGUE, J., BROWN, M.H. and PRICE, B.A. (eds). The MIT Press.
- BAZIK, J., TAMASSIA, R., REISS, S.P. and VAN DAM, A. (1998): Software Visualization in Teaching at Brown University. In *Software Visualization: Programming as a Multimedia Experience*. 383 - 398. STASKO, J., DOMINGUE, J., BROWN, M.H. and PRICE, B.A. (eds). The MIT Press.
- GLOOR, P.A. (1998): User Interface Issues For Algorithm Animation. In *Software Visualization: Programming as a Multimedia Experience*. 145 - 152. STASKO, J., DOMINGUE, J., BROWN, M.H. and PRICE, B.A. (eds). The MIT Press.
- GURKA, J.S. and CITRIN, W. (1996): Testing Effectiveness of Algorithm Animation. Proc. 1996 IEEE Symposium on Visual Languages. 182, IEEE Computer Society
- HÜBSCHER-YOUNGER, T. and NARAYANAN, N.H. (2002): Influence of Authority on Convergence in Collaborative Learning. Proc. Computer Support for Collaborative Learning Conference, Boulder, Colorado USA.
- HUNDHAUSEN, C.D. (2002): Integrating algorithm visualization technology into an undergraduate algorithms course: ethnographic studies of a social constructivist approach. *Computers & Education* 39(3):237 - 260. November 2002.
- MURUGE, K (2007) An Interactive Algorithm Animation Prototype. BCom Honours Dissertation. Nelson Mandela Metropolitan University, South Africa.
- NAPS, T.L., et al. (2003): Evaluating the Educational Impact of Visualization. Proc. ITiCSE 2003, Thessaloniki, Greece. ACM Press
- NAPS, T.L., EAGAN, J.R. and NORTON, L.L.(2000): JHAVE - An environment to Actively Engage Students in Web-based Algorithm Visualizations. Proc. Technical Symposium on Computer Science Education, Austin, TX, USA. 109 - 113, ACM Press

3

Multi-Scale Angiography Filters: Techniques Today

Fred N. Kinunuka

Vessel Enhancement and extraction in Angiography is still in developing state as many important problems still remain to be solved. The computation procedure of vessels in angiography is a very important due to limited computational resources. Computational procedure based on multi-scale has received considerable attention from scientists. Multi-scale approaches perform enhancement based on image resolutions and structure sizes. Many contributions have been made on the problem of vessel enhancement multiscale computing of volume data sets but there has never been a head to head evaluation of these approaches. In this paper the various multi-scale vessel enhancement approaches are put in perspective through a head to head comparison of algorithms of the existing research and a generic framework for linear multiscale is presented.

1. Introduction

Vessel Enhancement and extraction of blood vessels is one of the essential medical computing tools in clinical assessment of vascular diseases. Vascular diseases are one of the major source of deaths all over the world and therefore, developing reliable and robust vessel extraction methods has been a priority by many researchers. One of the aims of modern medical image processing is an automatic localization of suspected regions in 3D images of vasculature. This has various medical imaging applications including; surgical planning and treatment, stenosis and aneurysm quantification, disease monitoring and remission, segmentation of veins and arteries.

While many vessel enhancement techniques and algorithms have been developed, there is still no standard algorithm that has been am blessed fully and it's a developing state. This is because filtering of anatomical medical images is not only a daunting data-dependent task [Sherbondy et al., 2003],but also offer several unsolved challenges to enhancement algorithms like, the complicated structures found in these images, lack of global defining morphological characteristics, scanner noise,artifacts,incomplete or weak separation between voxels representing neighboring tissue, peculiar shapes of vessels that twist and turn throughout the anatomy making them very difficult to track along the crowded cross-sections of the volume, vessel density and small diameter, dynamic range of intensities, scan- ning limitation, vessel contrast complexity, [Suri et al., 2002], [Bullitt, 2002].

Multiscale vessel enhancement classification is actually a computation process. Multi-scale approaches perform segmentation task on different image resolutions and the size of the actual structures. The importance of analysing images at many scales arises from the nature of vessels themselves and any analysis procedure that is applied at a single scale may miss information at other scales, the vessel under investigation may possess important features at different scales and resolutions, the underlying vessels may have multi-resolution features or the vessels could be at several different resolutions as shown in Fig.1 using a desired criteria like based on area or volume. Multi-scale offer a basis for evaluating robustness and persistence across scales. In this paper the various multi-scale enhancement techniques are put in perspective through a head to head comparison of the existing algorithms.

The remainder of this article is organised as follows. Multi-scale Approaches are discussed in Section 2 inclusive is the generic theory and framework behind these approaches. Also a tabular summary of a head to head comparison is presented in this section. Section 3 gives the conclusions and future work.

2. Multi-scale Approaches

2.1 Linear Multi-scale approaches

There are a number of different types of multiscale approaches of which Linear multi-scale scale filters are the most popular and studied in literature. They are generally based on scale space theory by [Witkin, 1983, Lindeberg, 1994]. Scale-space representation generally involves generating a parameter class of derived signals in which the fine-scale information is sequentially suppressed. The structures at coarse scales in the multi-scale representation constitute simplifications of corresponding structures at finer scales. To ensure smoothing transformations in multiscale vessel analysis, the use of the convolution product of a Gaussian kernel and its derivatives has been shown to be the only way to ensure linearity, invariance under, translation, rotation, rescaling [Krissian et al., 2000], [Lindeberg, 1994]. This idea has been formalised in a variety of ways by different authors and can be summarised as in Fig.2. In scale space theory, for any N -dimensional signal; $f: \mathbb{R}^N \rightarrow \mathbb{R}$ its scale space is represented by:

$$S(x; t) = \int_{\tau \in \mathbb{R}^N} f(x - \tau) g(\tau) \partial \tau$$

Where

$$S: \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$$

$$\tau \in \mathbb{R}^N$$

and

$$g: \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$$

is the Gaussian kernel which can be computed. For linear space, the scale can be obtained using diffusion equation solution.

Given Initially

$$S(; t) = f \quad [1]$$

From Fig.2, Step 1: **Image Analytical descriptor**: The selection process of the image descriptor and its corresponding cross section is done. In selecting these descriptors invariance properties under certain transformations like, rotations, rescaling and affine or perspective deformations are taken into consideration.

Step 2: **Scale Space determination**: This is an important step which involves selection process of candidate pixels mostly dependent on the Hessian matrix. Multiscale response functions are computed to determine whether a pixel corresponds to different scales of vessels. The linear scale space is obtained from the convolution of Gaussian of the vessels and its derivatives.

Step 3: **Multiscale Local Integration/computation**. In this step, the scale range, minimum scale, and the number of scale levels are determined in most cases based on eigenvectors of Hessian Matrix. Filter responses tuned to different image descriptor scales can be combined in order to recover image structures of various sizes. Various methods of integrating multi-scale responses can be applied like normalize the filter responses of each scale and then select the maximum response among the multiple scales [Lindeberg, 1994]. This also enables rescaling.

Step 4: **Enhanced Volume**. This step is optional and visualization or skeleton construction of vessels.

Most of the linear multiscale approaches in literature tend to follow that framework. They deviate only at step 1 where the analytical computation is different. The descriptor and cross section function determines, the approach robustness, efficiency or complexity. For instance, The work by [Sato et al. 1997] considers vessel width with an elliptical cross section of vessels. There is discrimination of line structures from other structures and recovering line structures of various width. Multi-scale local integration is done by taking the maximum among single-scale filter responses, and its characteristics are examined to derive criteria for the selection of parameters in the formulation. A similar approach by [Frangi et al. 1998] focuses on local structure and examines the local second order ellipsoid. In other words this method is based on geometrical interpretation of eigenvalues of Hessian matrix. The only difference with [Sato et al. 1997] is it uses all eigenvalues simultaneously. [Lorenz et al. 1997] also present line structures method based on normalized first and second derivatives and on the eigenvector analysis of the hessian matrix. He uses a Taylor expansion to the second order of the image intensity [Koller et al. 1995] propose a multiscale response for linear structures. The response function uses eigenvectors of the Hessian matrix of the image to define at each point an orientation orthogonal to the axis of a potential vessel that goes through the point.

Similarly, [Descoteaux et al., 2004], [Poli, 1997], [Shikata et al., 2004], have similar image descriptors but only defer on analysis of Hessian matrix. Another approach by [Krissian et al., 2000] estimates the sensitivity of the image second order derivatives according to elliptical cross section, curvature of axis or partial volume effects in order to detect and extract centreline of vessels of different sizes based on scale. Importance is attached to the relationship between the radius of the structure and the scale it is detected at. They also use a simple model of cylindrical vessel with circular Gaussian cross-section guide detection. The work by Manniesing et al. [2006] can be said to be an extension of [Sato et al., 1997], [Krissian et al., 2000], [Frangi et al., 1998] and [Koller et al., 1995]. In this approach to enhance vascular structures within the framework of scale space theory combines a smooth vessel filter based on a geometrical analysis of the Hessians eigensystem, with a non-linear anisotropic diffusion scheme. The amount and orientation of diffusion depend on the local vessel likelihood. The scale space is generated by the diffusion equation or scheme with diffusion tensor preserving the features of interest. The tensor is a function of Hessian matrix with direct geometrical interpretation to measure vesselness.

Another approach by [C.Toumoulin et al., 2001] is based on geometrical moments and a local cylindrical approximation. An estimation of vessel and background intensity levels, position, orientation and diameter of the vessels with adaptive control of parameters, is provided during vessel tracking. [Bullitt 2002] method extracts centerlines of vessels using tubular object modeling. Multiscale extraction of vessel centerlines is done through dynamical scale ridge transversal and radii estimation. Dynamic ridge scale transversal is achieved through uncoupling width estimation process and using scales proportional to vessel width. While [Wink et al. 2004] proposes a method for determining the central axis of vessels for close proximity or crossing of vessels and varying vessel widths. This method is based on the minimum cost path search of a multiscale vessel enhancement filter that propagates the wave front through the filter response at a range of scales instead of through the maximum response of the filter, which provides scale selection. A vectorial multi-scale feature image for wave front propagation between two or more user defined points to retrieve the central axis of tubular objects is proposed. [M.Martinez et al. 1999] present a method to automatically segment retinal blood vessels based upon multiscale feature extraction. This method overcomes the problem of variations in contrast inherent in these images by using the first and second spatial derivatives of the intensity image that gives information about vessel topology. This approach also enables the detection of blood vessels of different widths, lengths and orientations. The local maxima over scales of the magnitude of the gradient and the maximum principal curvature of the Hessian tensor are used in a multiple pass region growing procedure.

The idea of morphologic multiscale can be summarised by [Van et al. 1995] as first downscale the image by a factor using an invertible scaling, then apply an image operator at a unit scale, and finally resize the image to its original scale. It is then required that the resulting one-parameter family of image operators satisfies some semigroup property.[Bangham et al., 1995, 1996] have done some useful work on non linear multiscale morphology filters. In these approaches image decomposition in multiscale is referred to as a sieve in n dimensions they highlight properties that an enhanced image under non linear multiscale should have;scale space causality, localization, piecewise smoothing and scale calibration and manipulation. The sieves can be obtained using the conventional openings and closing feature of morphology. In one-dimension the zones created by a sieve have a length implying that the scale parameter can be used for the precise measurement of features. For a two-dimensional image with regular pixelation the flat zones created by the sieve have a defined area and is referred to as an area decomposition. In three dimensions it is a decomposition by volume. [Meyer and Maragos 1999] describes the transformations properties from a finer scale to a coarser scale for multiscale of invariance by translation, invariance by rotation, invariance under a change of illumination that have to be satisfied. Similarly, Works by [Summers et al. 1997] use multi-resolution data structure to extract vascular morphology and local own parameters from phase contrast magnetic resonance angiograms.[Salembier, 1994], [Maragos, 1996], [Comer and Delp, 1999], [Meyer, 2004], [Bosworth and Acton, 2003], [Nacken, 1994], [Crespo and Serra, 1993], [Debayle and Pinoli, 2006], [Sun and Sang, 2008] follow the same principles and can be classified as non linear multiscale morphology based filters. For more about these approaches the reader should see the references and Table.I .

Table I. Summary of Multi-Scale Approaches

Algorithm	Model	Span	Features Preserved	Computational Cost	User Interaction
[Sato et al., 1997)	Linear	Local	Edges	slow	N/A
[Krissian et al., 2000)	Linear	Local	edges,junctions	slow	N/A
[Frangi et al., 1998)	Linear	Local	edges	fast	yes
[Lorenz et al., 1997)	Linear	Local	edges	N/A	N/A
[Koller et al., 1995)	Linear	Local	junctions	N/A	no
[Descoteaux et al., 2004)	Linear	Local	edges	N/A	N/A
[Poli, 1997)	Linear	Local	edges	N/A	N/A
[Shikata et al., 2004)	Linear	Local	edges,junctions	N/A	N/A
[Manniesing et al., 2006)	Linear	Local	edges	N/A	no

[C.Toumoulin et al., 2001)	Linear	Local	edges	N/A	N/A
[Bullitt, 2002) [Wink et al., 2004)	Linear	Local	nodules	fast	N/A
[M.Martinez et al., 1999)	Linear	Local	centerline	N/A	N/A
[Bangham et al., 1995)	Non linear	Local	edges,junctions	fast	no
[Meyer and Maragos, 1999)	Non Linear	local	centerlines	N/A	no
[Summers et al., 1997)	Non Linear	Local	edges	N/A	no
[Salembier, 1994)	Non Linear	Local	edges,junctions	fast	yes yes
[Maragos, 1996)	Non Linear	Local	edges,junctions	fast	yes yes
[Comer and Delp, 1999) [Meyer, 2004)	Non Linear	Local	edges	N/A	no
[Bosworth and Acton, 2003)	Non Linear	Local	edges edges	N/A	no
[Nacken, 1994)	Non Linear	Local	edges	N/A	N/A
[Crespo and Serra, 1993)	Non Linear	Local	edges	N/A	N/A
[Debayle and Pinoli, 2006)	Non Linear	local	edges	N/A	N/A
[Sun and Sang, 2008)	Non Linear	Global	edges	N/A	N/A

3. Conclusions and Future Work

Multiscale filters are considered as a set of sieves of different grades which allows details of certain size to pass while rejecting others. Linear multi scale are mostly based on Hessian matrix of intensity analysis and tend to have so much noise associated with them due to second order partial derivatives estimation of intensity matrix which leads to smoothing of data sets at different scales resulting into high computational cost. These filters also struggle in classifying complex structures like junctions and nodules. On the other hand, non linear multi scale require less computational cost as compared to linear ones and tend to do better on complex structures. Shape based filters rather than size based are now becoming prominent because of their ability to preserve the shape of objects without suppressing junctions, edges, nodule and less computation cost. Future work we intend to focus on exploring shape filters and possibly a combination of both size and shape based filters.

Reference

- BANGHAM, J. A., CHARDAIRE, P., PYE, C. J., AND LING, P. D. 1996. Multiscale non-linear decomposition: The sieve decomposition theorem. *IEEE Trans. Pattern Anal. Mach. Intell.* 18, 5, 529–539.
- BANGHAM, J. A., LING, P., AND HARVEY, R. 1995. Scale-space from nonlinear filters. In *ICCV '95: Proceedings of the Fifth International Conference on Computer Vision*. IEEE Computer Society, Washington, DC, USA, 163.
- BOSWORTH, J. H. AND ACTON, S. T. 2003. Morphological scale space in image processing. *Digital Signal Proc.* 13, 338–367.
- BULLITT, S. E. 2002. Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction.
- COMER, M. L. AND DELP, E. J. 1999. Morphological operations for color image processing. *Journal of Electronic Imaging* 8, 279–289.
- CRESPO, J. AND SERRA, J. C. 1993. Morphological pyramids for image coding. B. G. HASKELL AND H.-M. HANG, Eds. *Visual Communications and Image Processing '93* 2094, 1, 159–170.
- C.TOUMOULIN, BOLDAK, C., DILLENSEGER, J., COATRIEUX, J., AND ROLLAND, Y. 2001. Fast detection and characterization of vessels in very large data sets using geometrical moments. *IEEE Trans Biomed Eng* 48, 604606.
- DEBAYLE, J. AND PINOLI, J.-C. 2006. General adaptive neighborhood image processing. *J. Math. Imaging Vis.* 25, 2, 267–284.
- DESCOTEAUX, M., COLLINS, L., AND SIDDIQI, K. 2004. Geometric flows for segmenting vasculature in mri: Theory and validation. In *Medical Imaging Computing and Computer-Assisted Intervention*. 500–507.
- FRANGI, R. F., NIESSEN, W. J., VINCKEN, K. L., AND VIERGEVER, M. A. 1998. Multiscale vessel enhancement filtering. *Springer-Verlag*, 130–137.
- KOLLER, T. M., GERIG, G., AND DETTWILER, D. 1995. Multiscale detection of curvilinear structures. *IEEE Computer Society Press*, 864–869.
- KRISSIAN, K., MALANDAIN, G., AYACHE, N., VAILLANT, R., AND TROUSSET, Y. 2000. Model-based detection of tubular structures in 3d images. *Comput. Vis. Image Underst.* 80, 2, 130–171.
- LINDEBERG, T. 1994. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA.
- LORENZ, C., CARLSEN, I.-C., BUZUG, T. M., FASSNACHT, C., AND WEESE, J. 1997. Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2d and 3d medical images. In *CVRMed-MRCAS '97: Proceedings of the First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medial Robotics and Computer-Assisted Surgery*. Springer-Verlag, London, UK, 233–242.
- MANNIESING, R., VIERGEVER, M., AND NIESSEN, W. 2006. Vessel Enhancing Diffusion - A Scale Space Representation of Vessel Structures. *Medical Image Analysis* 10, 815–825.
- MARAGOS, P. 1996. Differential morphology and image processing. *IEEE Transactions on Image Processing* 5, 922–937.

- MEYER, F. 2004. Levelings, image simplification filters for segmentation. *J. Math. Imag. Vis.* 20, 1–2, 59–72.
- MEYER, F. AND MARAGOS, P. 1999. Morphological scale-space representation with levelings. In *SCALE-SPACE '99: Proceedings of the Second International Conference on Scale-Space Theories in Computer Vision*. Springer-Verlag, London, UK, 187–198.
- M.MARTINEZ, A.HUGHES, STANTON, A. V., THOM, S. A., BHARATH, A., AND PARKER, H. 1999. Retinal blood vessel segmentation by means of scale-space analysis and region growing. 6, 90–97.
- NACKEN, P. F. M. 1994. Image analysis methods based on hierarchies of graphs and multi-scale mathematical morphology. Ph.D. thesis, University of Amsterdam, Amsterdam, The Netherlands.
- POLI. 1997. An algorithm for real-time vessel enhancement and detection. *Computer Methods and Programs in Biomedicine* 52, 1, 1 – 22.
- SALEMBIER, P. 1994. Morphological multiscale segmentation for image coding. In *Proceedings of of the IEEE winter workshop on Nonlinear digital signal processing*. Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, 359–386.
- SATO, Y., NAKAJIMA, S., ATSUMI, H., KOLLER, T., GERIG, G., YOSHIDA, S., AND KIKINIS, R. 1997. 3d multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In *CVRMed-MRCAS '97: Proceedings of the First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medial Robotics and Computer-Assisted Surgery*. Springer-Verlag, London, UK, 213–222.
- SHERBONDY, A., HOUSTON, M., AND NAPEL, S. 2003. Fast volume segmentation with simultaneous visualization using programmable graphics hardware. In *VIS '03: Proceedings of the 14th IEEE Visualization 2003 [VIS'03]*. IEEE Computer Society, Washington, DC, USA, 23.
- SHIKATA, H., HOFFMAN, E. A., AND SONKA, M. 2004. Automated segmentation of pulmonary vascular tree from 3d ct images. In *Society of Photo-Optical Instrumentation Engineers [SPIE] Conference Series*, A. A. Amini and A. Manduca, Eds. Society of Photo-Optical Instrumentation Engineers [SPIE] Conference Series, vol. 5369. 107–116.
- SUMMERS, P., BHALERAO, A., AND HAWKES, D. 1997. Multi-resolution model based segmentation of mr angiograms. *Journal of Magnetic Resonance Imaging* 7, 950–957.
- SUN, K. AND SANG, N. 2008. Morphological enhancement of vascular angiogram with multiscale detected by Gabor filters. *IEEE Transactions on Medical Imaging* 44, 1, 86–87.
- SURI, J. S., LIU, K., REDEN, L., AND LAXMINARAYAN, S. 2002. A review on mr vascular image processing: Skeleton vs. non-skeleton approaches: Part-ii. *IEEE Transactions* 6, 324–337.
- VAN, R., BOOMGAARD, D., AND HEIJMANS. 1995. Morphological scale-space operators.
- WINK, O., NIESSEN, W., AND VIERGEVER, M. 2004. Multiscale vessel tracking. *IEEE Transactions on Medical Imaging* 23, 1, 130–133.
- WITKIN, A. P. 1983. Scale-space filtering. In *8th Int. Joint Conf. Artificial Intelligence*. Vol. 2. Karlsruhe, 1019–1022.

Fig. 1. Multi-scale filtering of time-of-flight MRA: [a] X-ray rendering of original volume; [b] filtered using desired criteria with threshold at $\lambda = 1.0$; [c] original filtered at $\lambda = 2.0$; [d] original filtered at $\lambda = 4.0$; [e] $\lambda = 6.0$; [f] $\lambda = 9.0$;

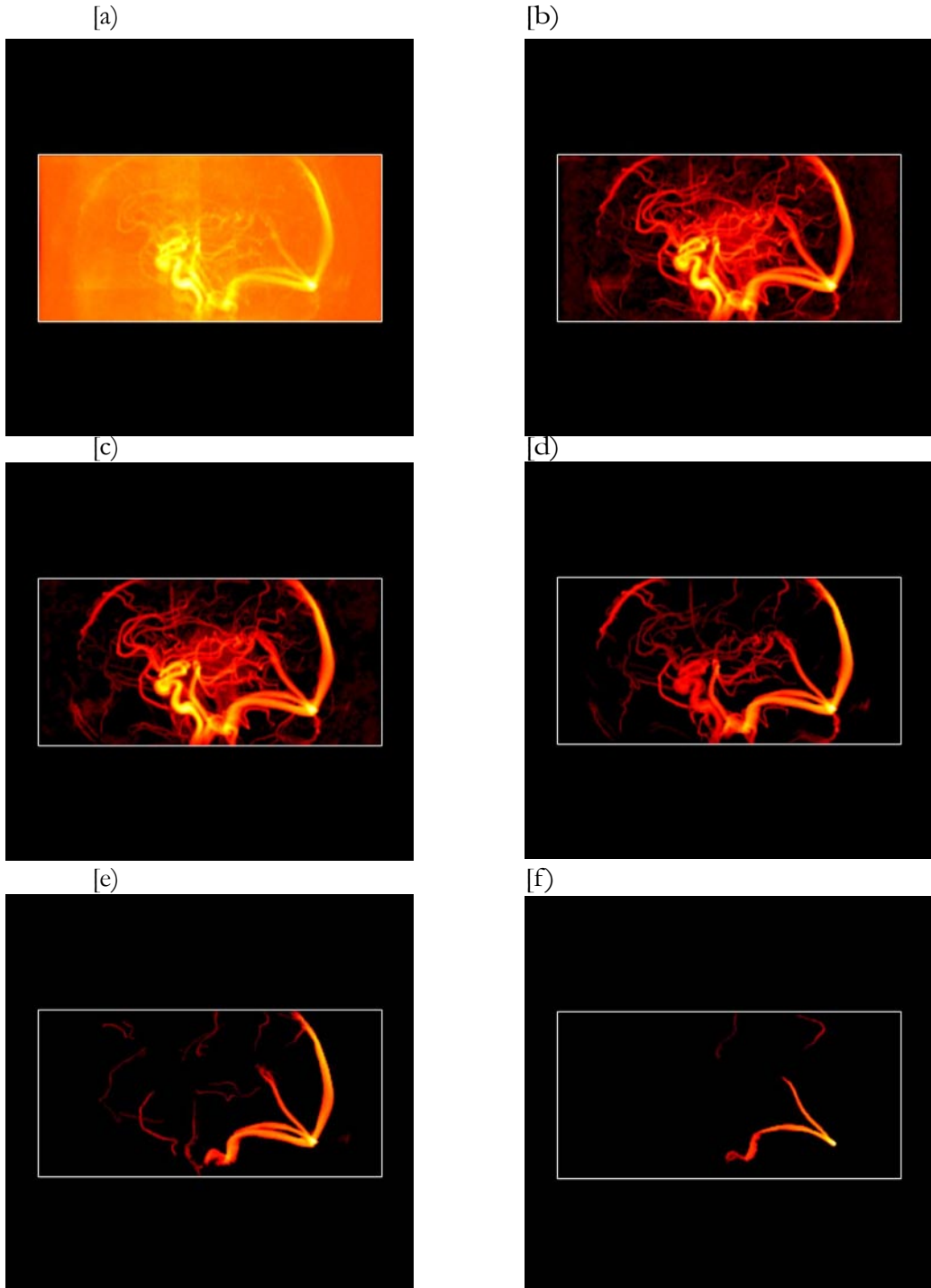
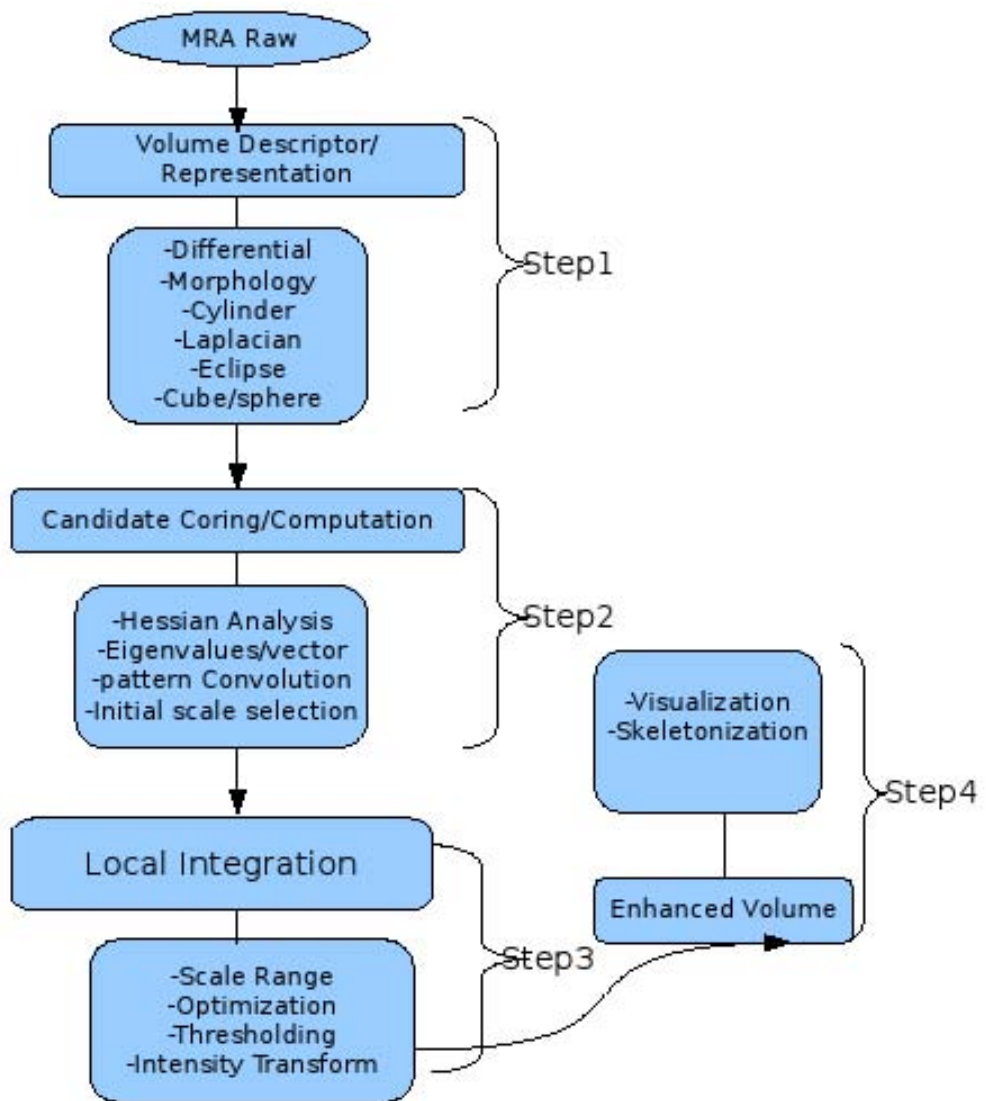


Fig. 2. Generic framework for linear multiscale vessel extraction



4

Security Analysis of an Agent-Mediated Book Trading Application

Richard Ssekibuule and Jose Ghislain Quenum

This paper presents a security analysis of an agent mediated application in an open distributed environment. We use a case study of a booktrading application that we implemented using AgentScape and JADE agent platforms. The paper analyzes whether security requirements, threats and countermeasures for an agent mediated application change when implemented on different types of agent platforms and presents countermeasures to generic and application specific threats.

1. Introduction

Research and development in Agent system has seen tremendous progress in recent years leading to the development of several agent platforms such as NOMADS[Suri et al. 2000], AgentScape[Overeinder and Brazier 2005], Havana[Mahmoud and Yu 2004], JADE[Bellifemine et al. 2007] and Aglets[Lange et al. 1997]. In order for agents to execute tasks that have been assigned to them, they have to interact with other agents in the open distributed environment whose intentions could be malicious. Several researchers [Li et al. 2004; Farmer et al. 1996; Jansen 2000] have suggested solutions to different types of threats that can be anticipated in agent-based applications and platforms. However, most of these solutions are based on generic analysis of threats and security countermeasures. Consequently, such solutions do not address specific application security requirements and threats. To address this problem, we perform security analysis using an agent-mediated case study and propose generic countermeasures to security threats based on a concrete application. The study also proposes an approach for performing systematic security analysis for agent-mediated application based on our results and experiences. The paper also investigates whether security requirements and threats change when an application is implemented using different types of agent platforms. Section 2 presents the booktrading application that is used as a case study, stakeholders and assets are presented in section 3, security requirements and threat modeling are presented in sections 4 and 5 respectively.

2. The Booktrading Application

The booktrading application comprises of booksellers and bookbuyer software agents that are ideally owned by separate individuals. The bookseller owners are responsible for maintaining bookstore operations which include stocking and setting prices for books. Bookbuyer software agents are responsible for buying books from bookseller agents on behalf of their owners. The implementation of the booktrading application was done in

AgentScape[Overeinder and Brazier 2005] (version 0.9.1) and Java Agent Development Environment[Bellifemine et al. 2007] (JADE version 3.6) agent platforms. The purpose of implementing the application in two agent platforms was to investigate whether threats, security requirements and threat countermeasures for the application would change when implemented in either of the platforms. The bookseller interface is used by the bookseller owner to store information about books available for sale. In addition to the book title and year of book publication, the bookseller also records the first price of the book and lowest price at which he/she can sell a book. The first price is the best reasonable price that a seller finds competitive in the market. It is in the best interest of booksellers to keep information about the lowest price of the book secret from buyers and other sellers. Keeping this information secret prevents the bookseller from being exploited by buyers during negotiation. Additionally, the bookseller might be helped in keeping pricing information secret from other booksellers to avoid competition that may arise from other booksellers in the environment setting their prices based on what they know about the seller. The bookbuyer interface is used by the bookbuyer owner to send out requests to booksellers. Using the bookbuyer interface, the bookbuyer owner informs the agent of the best price at which they wish to buy the book and the maximum amount they are willing to pay.

The bookbuyer agent also implements a protocol for negotiation with the bookseller in case the desired book is found to be at a price higher than the bookbuyer's best price. The current implementation of the negotiation protocol always proposes a price that is half the sum of the best and highest price. Ideally this strategy should not be known to the bookseller, otherwise a bookseller agent can exploit this information to sell at a higher price to the bookbuyer than they should have done in case they didn't know the buyer's negotiation strategy. Figure 3 below presents an interaction protocol between the bookseller and bookbuyer for the booktrading application. The current implementation of the booktrading application does not support mobility of either the bookseller or the bookbuyer. The choice of having the bookseller stationary was a design decision made to support functionality for storing books in a MySQL[MySQL, AB] database, while the choice of having the bookbuyer stationary was due to limitations in inter-platform migration of JADE agents. The current version (3.6) of JADE does not have proper support for inter-platform migration. The mobility addon was developed for an older version of JADE and has not been since updated. Similar design decisions were taken with AgentScape, but the implementation limitations were different from those experienced with JADE. Nevertheless, security considerations for having the bookbuyer mobile were considered. The requirement of having the bookbuyer mobile could arise when it is deemed necessary to have the buyer migrate to the platform that has the desired book. Such a necessity would arise in case it is considered computationally expensive for the buyer to perform all tasks from the bookbuyer owner's platform.

3. Stakeholders and Assets

This section presents stakeholders and assets in the booktrading application and agent platforms that were used in the implementation of the booktrading application. These

stakeholders and assets are partly derived from the functional requirements of the booktrading application.

3.1 Booktrading Application Stakeholders

(i) **Application creator:** This is the individual or organization that developed both the agent bookseller and agent bookbuyer. (ii) **Bookseller owner:** The individual or organization that owns the bookseller agent. (iii) **Bookbuyer owner:** The individual or organization that owns the bookbuyer agent. (iv) **Platform creator:** The individual or organization that developed the agent middleware (here in referred to as the agent platform). (v) **Platform owner:** This is the stakeholder category that owns or administers the operating system (host) on which the agent platform is installed.

3.2 Booktrading Application Assets

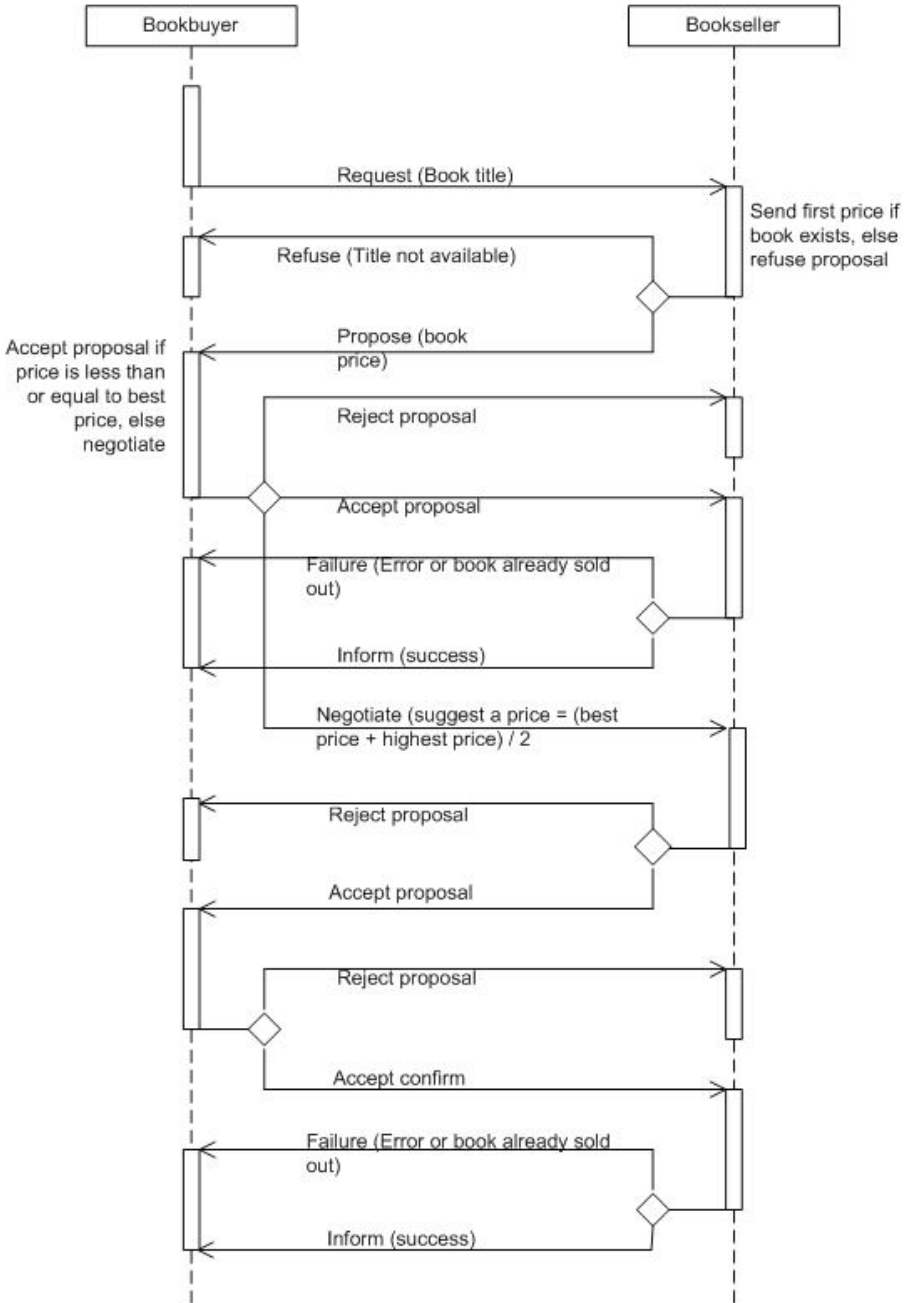
Bookseller agent: This is the agent code that is responsible for handling tasks related to storing books in the bookstore and interacting with bookbuyers.

Bookbuyer agent: The agent code that represents the human bookbuyer in booktrading tasks. The bookbuyer agents captures the title, the maximum age of the book, best and highest price that the agent owner would be willing to pay for the book.

Interaction protocol: The interaction protocol represents a set of rules through which messages exchanged between the agent buyer and agent seller are interpreted. In case the interaction protocol is is not followed, it is assumed that either party would not understand what they other is saying. The format of interaction messages were previous defined by Foundation of Intelligent Physical Agents (FIPA)[FIPA 2002]. The booktrading application extends the interaction protocol to include negotiation. The negotiation protocol is implemented by the bookseller and book-buyer agents to handle the logic through which they can agree on an alternative price that is different from the booksellers first price and the bookbuyers best price.

Agent platform: The agent platform provides the execution environment for both the bookseller and bookbuyer.

Fig. 1. Interaction protocol for booktrading application



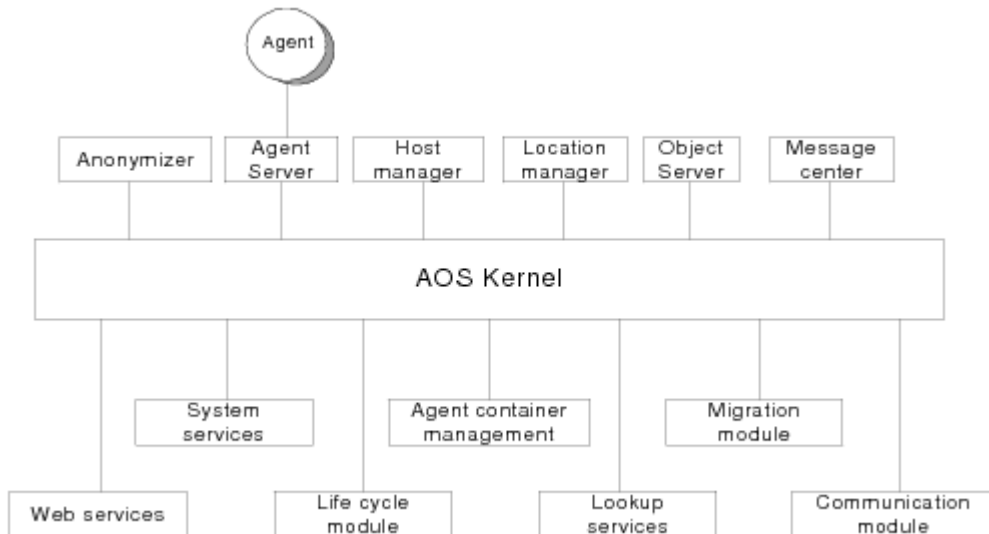
3.3 Platform Assets

This section presents platform assets for AgentScape and JADE agent platforms. Separating the assets and possible attacks to the agent platforms created a basis for thinking about generic countermeasures for attacks on agent platforms.

3.3.1 AgentScape Platform Assets.

AgentScape system services: The component represents core kernel services for AgentScape. System services ensure that the right classes are used in the communicator. AgentScape system services

Fig. 2. AgentScape Architecture



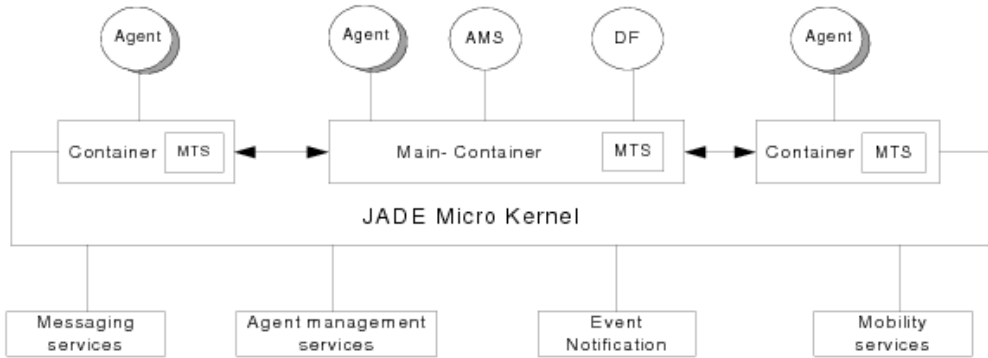
provide lookup services through which agents and services are registered and discovered. They also provide a communication module through which messages are exchanged between agents and remote agent management platforms.

Agent server: Provides services for loading of agent code from the agent container, startup and termination of the agent. Facilitates agent access to kernel services, provides mechanism for making negotiation calls. The agent server also uses an agentwrapper to provide interaction between a running agent and agentscape middleware.

Host manager: AgentScape host manager provides agent container management for mediating access to agent stores. The hostmanager additionally facilitates data handling during agent migration. The host manager provides agent life cycle management services such as migration, suspending, stopping and running of an agent.

Location manager: Provides an agent management module that facilitates inserting new agents into a location and handling of migration requests from local agents and remote location managers.

Fig. 3. JADE Platform Architecture



3.3.2 Jade Platform Assets.

JADE Core Base Service: Microkernel of the JADE system that provides a uniform mechanism for management and service discovery.

Agent Management Services: Agent management services define the agent, provide platform administration, status information for agents and a unique naming scheme.

Messaging Services: Microkernel service responsible for managing communication between agents and other entities in across agent platforms **Event Notification Services:** Represents events related to the agent life-cycle and configuration.

Agent Mobility Services: JADE microkernel provides an agent mobility service to facilitate migration within containers on the same platform and migration from one platform to another.

4. Security Requirements

This section covers security requirements for both the agent platform on which the agent-mediated application can be executed and security requirements for the booktrading application.

4.1 Application Security Requirements

We use the Confidentiality, Integrity, Authentication and Non-repudiation (CIAN) taxonomy to define security requirements for both the bookseller and bookbuyer agents. The Confidentiality, Integrity and Authentication (CIA) taxonomy has been reviewed by Howard et. al [Howard and Lipner 2006] in regard with risk analysis. This section presents a forth requirement of non-repudiation and brief definition of CIA components. **Confidentiality:** This is a security requirement that ensures information exchanged and stored in the system is accessed by only authorized users. **Integrity:** This requirement ensures that changes to information or application code are only done by authorized users. **Availability:** This is a security requirement that ensures that information and application resources can be accessed by all legitimate parties. The

legitimate parties may include users and software processes. **Non-repudiation:** This security requirement ensures that all parties can be held accountable for their actions. It ensures that all actions taken cannot be denied at a later time.

4.2 Bookseller Security Requirements

This subsection presents security requirements for the bookseller agent application.

Confidentiality: The bookseller should not release buyers private information to the public and the information should be protected from attackers.

Integrity: Bookseller should prevent book reviews from being manipulated or changed by people who did not write them.

Availability: (i) The bookseller should not prevent some books from being available to some buyers for one reason or another. (ii) The bookseller should allow the buyer to choose a book based on attributes such as price, publication date, quality and relevance among others. (iii) The bookseller should not deny bookbuyers a chance to negotiate when requested. (iv) The bookseller should not disable possibilities of writing reviews and posting them from genuinely critical reviewers and buyers. (v) The bookseller should not prevent potential buyers from reading book reviews.

Non-repudiation: The bookseller should not be able to deny any information that is exchanged with the bookbuyer such as having received payment from bookbuyers or purchase orders.

4.3 Bookbuyer Security Requirements

This subsection presents security requirements for the bookbuyer agent application.

Confidentiality: (i) Should be able to migrate from one platform to another and perform computational tasks from local and remote hosts/platforms without leaving information traces for attackers. Such information could include platforms/hosts they have previously visited and budget information for their shopping. (ii) The bookbuyer should have control on the type of information that they have to submit to the operation environment. Withholding some information for privacy reasons should not be a cause for denying them access to requested services. (iii) Bookbuyer should avoid forming a coalition in which it will be exploited. E.g exposing information it carries to malicious bookbuyers.

Integrity: (i) The bookbuyer should not be manipulated to buy a book that is not the best on offer. (ii) The information carried by the bookbuyer agent on behalf of the agent owner, should be protected from attackers that might want to change it.

Availability: (i) Bookbuyer should be able to find a book available for sale.

(ii) The bookbuyer agent should protect the information it carries and prevent it from being destroyed by attackers. (iii) Bookbuyer should not be prevented from forming a coalition with other bookbuyers for mutual benefit.

Non-repudiation: The bookbuyer should not be able to deny that they ever ordered for a book.

4.4 Platform Security Requirements

The agent platform security requirements are derived from the CIAN acronym that was defined in subsection 4.1.

Confidentiality: In case of agent migration, the agent platform needs to protect the migration path of the agent. The migration path might contain information concerning platforms the agent has previously visited and intended destinations.

Integrity: (i) The agent platform is expected to maintain integrity of agents (agent code) and protect them from malicious attackers in the environment. It is important to note that malicious platforms might try to do otherwise. (ii) The agent platform should provide a mechanism for detecting compromised or malicious agents.

Availability: (i) The agent platform needs to ensure availability of messaging service for agent communication. Termination of the messaging service would prevent agents from communicating, while a comprised messaging service could yield unexpected and undesired results for the intentions of the communication. For example, attacker could delay messages for the intended destination whose requests might have been time bound. (ii) The agent platform should have policies for regulating access to system resources to prevent starvation of some agents by others that may intentionally or otherwise over consume system resources.

Non-repudiation: The agent platform needs to provide a mechanism through which agents can be accountable for the actions they perform when visiting platforms.

5. THREAT MODELING

We use the STRIDE[Howard and Lipner 2006] taxonomy to identify possible threats faced by the booktrading application and the agent platform on which the application is executed.

Spoofing Identity: This is a form of attack in which someone or an entity pretends to be someone else of another entity. For example agent X pretending to be agent Y.

Tampering : Tampering attack refers to unauthorized changing of software code or information.

Repudiation: This refers to a circumstance in which a software process or an individual deny responsibility for their actions.

Information disclosure: This refers to unauthorized access to information.

Denial of Service: This is a form of attack that denies legitimate access to resources such as information, storage space, processor and communication channels.

Elevation of privileges: This refers to a form of attack in which an entity with lower privileges gains unauthorized higher privileges.

A more detailed explanation of the STRIDE components was presented by Howard et. al [Howard and Lipner 2006].

5.1 Application Threat Model

The book trading application is an agent mediated application in which one agent acts as a seller and another agent as a buyer. The bookseller agent provides a variety of books for sale in a manner similar to bookstores such as amazon, but in this setting the bookseller expects the buyers to be software agents. The bookbuyer agents are supposed to search in the books catalogue and compare prices and other attributes such as publication date, relevance and book ratings on behalf of their owners. The bookbuyer agent is expected to perform these tasks for a range of booksellers that have books available for sale. This section covers possible goals of the attacker against the bookseller and bookbuyer agents and suggests possible countermeasures to the identified threats.

5.1.1 Possible Goals of Attackers against Bookseller

- i. **Spoofing Identity:** An attacker could spoof the identity of a bookseller and requests payment from buyers for books that will not be delivered.
- ii. **Tampering:** (i) The attacker may want to change information in the booksellers catalogue so that book attributes such as price, publication date, quality and relevance are not correct. These changes could lead bookbuyers into choosing items that they were not supposed to buy. (ii) An attacker could change book reviews and rating so that consumers will not buy books from that particular bookseller.
- iii. **Repudiation:** A malicious bookbuyer could deny having requested or received a book from the bookseller.
- iv. **Information Disclosure:** (i) The attacker could compromise the negotiation logic implemented in the bookseller. E.g. if an attacker knows the price that a consumer wishes to pay for a product, the attacker could lower their prices to outcompete other sellers or simply to distract the buyer from making a genuine negotiation or purchase. (ii) An attacker may wish to access and log information concerning bookbuyers. The intention of this attack would be to compromise buyers privacy.
- v. **Denial of Service:** (i) An attacker could block the messaging service between the bookbuyer and the bookseller. (ii) The attacker could try to remove items from the booksellers catalogue of books so that books requested by the book-buyer are not available. (iii) An attacker could block buyers from writing and sending reviews on books. Such an attack could negatively affect the bookseller if buyers wish to know whether the books on sale are good and price worthy.

5.1.2 Possible Goals of Attackers against Bookbuyer:

- i. **Spoofing Identity:** An attacker could spoof the identity of a bookbuyer agent and purchases books that could be reputation damaging to the agent owner. This attack could be more severe in a general purpose e-commerce application where many types of products could be bought.
- ii. **Tampering:** (i) The attacker could alter message responses from the bookseller to indicate to the bookbuyer that the requested book is not available, even when it

is actually available. (ii) An attacker could change book reviews and rating so that consumers are lured into buying books that are not price worthy. (iii) An attacker could change information requests from the bookbuyer to indicate different requests to the bookseller from the ones submitted by the bookbuyer. Such attacks could lead the bookbuyer into getting invoices for books they did not order. Additionally, buyers could get false responses such as requested books not being available, even in circumstances where the books are available.

- iii. **Repudiation:** A malicious bookseller could deny having received payment for a book. In such a case, the bookbuyer would end up losing money.
- iv. **Information Disclosure:** (i) An attacker may have interest in accessing private information that is carried by a bookbuyer agent. (ii) An attacker could lure a bookbuyer into forming a coalition in which it would be exploited. E.g leaking information it carries to malicious bookbuyers.
- v. **Denial of Service:** (i) An attacker could lure a bookbuyer into forming a coalition in which it would be exploited. (ii) The attacker could block the messaging service between the bookbuyer and the bookseller. (iii) The bookbuyer could be denied a chance of forming a coalition with other buyers by withholding coalition formation information. This attack could also affect the bookseller by not making a needed sale. (iv) An attacker could prevent potential buyers from reading book reviews. (v) An attacker could create a malicious bookstore to prevent a bookbuyer from finding a genuine book to buy.

5.1.3 Application Specific Countermeasures.

This section presents countermeasures for the security challenges that are likely to be faced by the agent bookseller and bookbuyer. The countermeasures are meant to prevent attackers' goals that were identified in subsection 5.1.1. The countermeasures are combined for attacks on the bookseller and bookbuyer agents, because these attacks are similar in nature. We also assume that safe coding procedures were followed for both the application and agent platform. When software security flaws such as buffer overflows exist in software, then authentication and authorization schemes can be subverted.

- i. **Spoofing Identity:** Spoofing of an agent's identity can be prevented by providing an identity management system [de Groot and Brazier 2006] through which agents are assigned names (or identities) that are difficult to be changed by the agent or an attacker. Authentication and authorization systems such as kerberos [ste 1988] or message authentication codes [Kaliski and Robshaw 1995] can be used to authenticate agents' identity or their owners. In this setup an agent would be required to submit a message to the service provider indicating their identity and a small message encrypted by their private key. The service provider would then retrieve a public key (from the key-management system) that is needed to decrypt the short message. It is assumed with public-key cryptography that the private key of the agent is key secret.

- ii. **Tampering:** Two things need to be protected against tampering. That is the information or data carried by agents and the agent code. The countermeasures available against these attacks fall into categories of prevention and detection. Message authentication codes (MAC)[Kaliski and Robshaw 1995] and digital signatures on the agent code and data are used to detect any form of tampering on the agent code and data.
- iii. **Repudiation:** Public-key digital signatures can be used to prevent repudiation by either a malicious bookseller or bookbuyer. In order to prevent repudiation, digital signatures would be required on messages from either the bookbuyer or bookseller. When a given agent (A) encrypts messages using their private key, those messages can only be decrypted by a corresponding public key that certainly belongs to the sending agent. The main challenge to this kind of solution is that security depends on the secrecy of the secret key.
- iv. **Information Disclosure:** This countermeasure should protect information carried by the agent. Such information includes target book titles, best and highest price that the bookbuyer is willing to pay for the book. Confidentiality of this information can be provided by encrypting the information carried by agents.
- v. **Denial of Service:** The agent platform and agent execution environment needs to provide strong authentication [Stein 1988] and authorization for processes that access system resources. Authentication and authorization are useful in preventing non-authorized users and processes from accessing privileged resources that could be critical for correct functionality of the agent application. Apart from preventing users and processes from accessing privileged resources and services, authentication is useful for detecting users and processes that may breach the imposed restriction. In detecting which users or processes performed certain tasks, accountability can be achieved for all activities undertaken in the system. The concept of a trusted third party can be used to determine ratings of a bookseller before it can be considered by the bookbuyer for purchase of a book. A trusted third party would help in preventing malicious booksellers from participating in booktrading transactions.

5.2 Platform Threat Model

The platform threat model is based on the Java Agent Development Environment (JADE)[Bellifemine et al. 1999] and AgentScape[Overeinder and Brazier 2005] platforms whose assets were presented in subsections 3.3.2 and 3.3.1 respectively.

5.2.1 Possible Goals of Attackers against Agent Platform

- i. **Spoofing Identity:** An attacker could launch counterfeit agents using the agent platform to participate in a transaction they are not supposed to be involved. For example, an agent Z (representing an attacker) could spoof the identity of agent X in order to perform actions privileged to X.
- ii. **Tampering:** (i) An attacker could be interested in altering agent code through the agent platform so that the agent does not do what it is supposed to do. (ii) The attacker could use a weakness in the agent platform to reach and subvert the communication

channel for agents. (iii) The attacker could use the platform to migrate an agent from a trusted platform to a compromised one.

- iii. **Information Disclosure:** In circumstances where agent platforms keep a non-repudiable log of agent events, an attacker could be interested in knowing about action of a particular agent. An attacker accessing this information could violate agent's confidentiality requirements. Additionally, the attacker could use this information to launch other forms of attacks against the agent.
- iv. **Denial of Service:** An attacker could be interested in subverting the Agent Management System (AMS), Directory Facilitator (DF) and Management services.
- v. **Elevation of Privileges:** Exploiting a flaw in the agent platform to grant higher privileges to malicious agents. Such malicious agents could be interested in using free resources on platform hosts, or even overconsuming system resources in order to deny services to legitimate users

5.2.2 Platform Specific Countermeasures.

This section presents security techniques that are needed by the platform to achieve the security requirements indicated in section 4.4 and to prevent attackers from achieving their objectives stated out in subsection 5.2.1.

- i. **Spoofing Identity:** Provide identity management: Only authenticated and registered agent owners are allowed to launch booksellers into the environment to prevent scenarios of malicious booksellers. A trusted third party can be used to verify and certify credentials agents.
- ii. **Tampering:** As indicated in subsection 5.2.1 the attacker may want to change the agent code or migrate an agent to a malicious platform. The action of changing or tampering with agent code can be prevented by code signing [Jansen 2000] with a digital signature. Agent migration to malicious platforms can be prevented by use of security policies to authorize sensitive actions to be executed by only trusted parties.
- iii. **Repudiation:** Combined with identity management, a non repudiable log kept by the agent platforms would be useful in tracking actions that were performed by various agents.
- iv. **Information Disclosure:** Confidentiality of the agent logs resident on the agent platform can be achieved through encryption. However, encryption has to be applied with consideration for low resource platforms like mobile devices.
- v. **Denial of Service:** Use of security policies for authentication and authorization of users and processes to key assets for the agent platform would prevent denial of service attacks.
- vi. **Elevation of Privileges:** This attack can be stopped by preventing software security flaws such as buffer overflows and SQL injection. Language based security mechanisms such as static source code analysis to enforce safety properties of the programming language, sandboxing and proof carrying code can be used.

6. Conclusions And Remarks

In this paper, we present a systematic approach for performing a security analysis of an agent mediated application. We have presented the Confidentiality, Integrity, Availability and Non-repudiation (CIAN) framework through which security requirements for an application can be derived and combined it with STRIDE[Howard and Lipner 2006] to obtain possible attacker goals. The proposed countermeasures (presented in subsections 5.2.2 and 5.1.3) are clearly generic for any agent application and indicate generic assets that need to be protected in an agent mediated application.

Our results indicate that security requirements for the agent application did not change significantly when implemented with either AgentScape or JADE. In reference to subsection 3.3, the assets of the agent platforms fall in the categories of (i) *Agent Management Services*, (ii) *Directory Facilitator*, (iii) *Messaging Services*, (iv) *Mobility Services* and (v) *Event Notification Services*. This implies that these assets need to be protected irrespective of the agent platform against all possible forms of attacks for the agent environment to be considered secure. Furthermore, for any application to be considered secure, its security requirements have to be catered for in the implementation and assets protected.

6.1 Implementation Experiences and Issues

This section presents some key issues and experiences that were encountered during the implementation phase of the booktrading application on JADE and AgentScape platform.

The implementation challenges were different for the platforms (AgentScape and JADE) that were used. Intra-platform migration challenges were faced with JADE (version 3.5 and 3.6) mainly because the mobility add-on was implemented for lower version of JADE. In AgentScape, implementation challenges were faced with service discovery facility due to the nature of requirements that were imposed by the application. It is worthy noting that these middleware platforms (mostly especially AgentScape) are still under heavy development and most problems are being solved as they are reported by users.

7. Future Work

The information disclosure countermeasure needs to protect the bookbuyer against traffic analysis by an intelligent bookseller. In this example booktrading application, the bookbuyer negotiates by sending a second price that is half the sum of the best and highest price. Using traffic analysis, an intelligent bookseller could be able to generate these two (best and highest) prices of the bookbuyer. The bookseller knowing these prices puts the bookbuyer in a poor negotiation position. Furthermore, the countermeasure should protect the negotiation protocol from revealing negotiation strategy. Additionally, some malicious participants could start the negotiation protocol and then terminate it with the intentions of stealing information concerning pricing and introducing annoyance attacks.

References

1988. KERBEROS: An Authentication Service for Open Network Systems. Proc. *Winter USENIX Conference*.
- BELLIFEMINE, F., CAIRE, G., AND GREENWOOD, D. 2007. *Developing Multi-agent Systems with JADE*. Springer.
- BELLIFEMINE, F., POGGI, A., AND RIMASSA, G. 1999. JADE—A FIPA-compliant agent framework. *Proceedings of PAAM 99*, 97–108.
- DE GROOT, D. AND BRAZIER, F. 2006. Identity Management in Agent Systems. *Proceedings of the First International Workshop on Privacy and Security in Agent-based Collaborative Environments (PSACE)*, 23–34.
- FARMER, W., GUTTMAN, J., AND SWARUP, V. 1996. Security for Mobile Agents: Issues and Requirements. *Proceedings of the 19th National Information Systems Security Conference 2*, 591–597.
- FIPA, F. 2002. Contract Net Interaction Protocol Specification.
- HOWARD, M. AND LIPNER, S. 2006. The Security Development Lifecycle. *Microsoft Press Redmond, WA, USA, Chapter Risk Analysis*, 114–115.
- JANSEN, W. 2000. Countermeasures for mobile agent security. *Computer Communications* 23, 17, 1667–1676.
- KALISKI, B. AND ROBSHAW, M. 1995. Message authentication with MD5. *CryptoBytes (RSA Labs Technical Newsletter) 1, 1*.
- LANGE, D., OSHIMA, M., KARGOTH, G., AND KOSAKA, K. 1997. Aglets: Programming Mobile Agents in Java. *Lecture Notes in Computer Science*, 253–266.
- LI, X., ZHANG, A., SUN, J., AND YIN, Z. 2004. The Research of Mobile Agent Security. *Lecture Notes in Computer Science*, 187–190.
- Mahmoud, Q. and Yu, L. 2004. Havana: a mobile agent platform for seamless integration with the existing Web infrastructure. *Electrical and Computer Engineering, 2004. Canadian Conference*.
- MySQL, AB. MySQL Database Server. *Internet WWW page, at URL: <http://www.mysql.com> (last accessed 5/21/2008)*.
- OVEREINDER, B. AND BRAZIER, F. 2005. Scalable Middleware Environment for Agent-Based Internet Applications. *Lecture Notes in Computer Science 3732*, 675.
- SURI, N., BRADSHAW, J., BREEDY, M., GROTH, P., HILL, G., JEFFERS, R., MITROVICH, T., POULIOT, B., AND SMITH, D. 2000. NOMADS: toward a strong and safe mobile agent system. *Proceedings of the fourth international conference on Autonomous agents*, 163–164.