Advances in Systems Modelling and ICT Applications

Editors

Joseph Kizza
Janet Aisbett
Andrew Vince
Tom Wanyama

Volume II

FOUNTAIN PUBLISHERS

Kampala

2006
# Contents

Series Information ........................................................................................................ iii
Preface .......................................................................................................................... viii
Editors ........................................................................................................................ viii
About Contributing Authors ...................................................................................... ix
Acknowledgements .................................................................................................... xviii
Introduction ................................................................................................................ xix

## Part One: Advanced System Modelling ................................................................. 1

1. Indexing a Discrete Global Grid  
   *Andrew Vince* ........................................................................................................ 2

2. Information Quality and Information Systems Success  
   *Janet Aisbett* ......................................................................................................... 18

3. A System Dynamics Tool for Evaluating IT Investment Projects  
   *Paul Ssemaluulu* and *Ddembe Williams* ............................................................. 29

4. The Influence of Job Physical Characteristics on their Schedulability in Multi-cluster Systems  
   *John Ngubiri* and *Mario van Vliet* .................................................................. 39

5. On Solving Large-scale Linear Systems Arising from Interior Point Methods for Linear Programming  
   *Venansius Baryamureeba* ................................................................................... 50

6. Adaptive and Optimisation Predictive Text Entry for Short Message Service (Sms)  
   *Peter Waiganjo Wagacha*, *Dennis Chege* .......................................................... 65

   *Ddembe Williams* ................................................................................................ 77

8. A Fundamental View on the Act of Modelling  
   *H.A. (Erik) Proper*, *P. van Bommel*, *S.J.B.A. Hoppenbrouwers* and  
   *Th.P. van der Weide* .......................................................................................... 97
# Part Two: Natural Language Processing

9. Kinyarwanda Speech Recognition for Automatic Voice Dialling Systems  
   Jackson Muhirwe .......................................................... 114

10. Computational Modelling in Bantu Language  
    Katherine W. Getao and Evans K. Miriti .......................... 128

---

# Part Three: Computer Networks and Applications

11. On Network Measurement and Monitoring of End-to-End Paths Used for E-VLBI  
    Julianne Sansa, Arpad Szomoru and J.M. van der Hulst .................. 140

12. Multiagent Systems for Distributed Resource Allocation  
    Eric Ayienga, Elijah Opiyo, William Okello-Odongo, Kate Getao,  
    Bernard Manderick and Anne Nowé ...................................... 154

13. Efficient Dynamic Admission Control for End-to-End Delay Guarantees in IP Networks  
    J. Kasigwa, Venansius Baryamureeba and Ddembe Williams ................. 167

14. Towards Web Design Frameworks (Wdfs)  
    Rehema Baguma, Patrick Ogao and Tom Wanyama ................................ 178

15. The Future of Intelligent Networks in Developing Countries  
    Narcis T. Rwangoga and Martin Ngobye .................................... 185

---

# Part Four: ICT and Legal Applications

16. Identifying Sensitive Knowledge for Law Enforcement Agencies  
    Habibu Atib and John Zeleznikow ......................................... 196

17. Using JAD to Bridge the Design-Reality Gaps; a Major Cause of IS Projects’ Failures in the Developing Countries  
    Euphrath Muthoni Masinde .................................................. 205

18. Survey of Data Mining Methods for Crime Analysis and Visualisation  
    Fredrick R. Okwangale and Patrick Ogao .................................. 221

19. Informing Regional ICT Policy: Case Study of Trends in ICT Indicators of OECD, EU, COMESA and EAC  
    Joseph Muliaro Wafula, Anthony J. Rodrigues and Nick G. Wanjobi .......... 227
Part Five: E-Commerce .................................................................................................................. 247

20 Bridging Africa’s Digital Divide: Building Sustainable ICT Infrastructures
Joseph M. Kizza ......................................................................................................................... 248

21 An Organisational Climate Awareness Toolkit for Nurturing the Effectiveness of Team/Group Interactions
Shushma Patel and Dilip Patel ................................................................................................. 254

22 ICTs in Developing Countries: Contexts, Challenges and Interventions
Anthony J. Rodrigues ................................................................................................................. 261

23 Architectural Building Blocks For Reusable Service Oriented Architecture With Integrated Transaction Processing
Benjamin Kanagwa and Ezra K, Mugiza .................................................................................... 277

24 Information and Communication Technologies (ICTs) in Small and Medium Enterprises (SMEs): Findings from Uganda
Ali Ndiwalana, Steve Esselaar, F.F. Tusubira and Christoph Stork ........................................ 283

Part Six: ICT and Education ........................................................................................................ 299

25 An Exploration of the Factors That Affect Quality Assurance Decision-making in Higher Education
Benedict Oyo and Ddembe Williams ......................................................................................... 300

26 Using Developmental Research to Design a Web Instructional Design Subsystem
J.K. Njenga and A.J. Bytheway .................................................................................................... 313

27 Survey of Data Mining Methods for Crime Analysis and Visualisation
Fredrick R. Okwangale and Patrick Ogao ................................................................................... 322

28 ICT for Educational Use: Cases of Implementing a Partnership Approach Model for Reaching Target Groups
Elijah I. Omwenga ..................................................................................................................... 328

29 The Impact of ICT on Universities: Classroom/Lecture Theatre Design and Curriculum Delivery
Collins Gyavira Rubanju Asifiwe .............................................................................................. 340

30 African Virtual Environment Collaborative (Afro-velab)
Madara Ogot ............................................................................................................................... 349
Part Seven: ICT and Gender ................................................................. 361
31 ICT Liberalisation and Changing Gender Relations in Contemporary Uganda: A Research Agenda
   Aramanzan Madanda ........................................................................ 362

Part Eight: Bio-Informatics ................................................................. 375
32 Towards a Reusable Ontology Framework for Biological Information Integration
   Gilbert Maiga and Ddembe Williams ............................................. 376
Preface

The future of Africa, its peoples, economy, and overall development will ultimately depend on the development of its infrastructure by her indigenous peoples. Perhaps unlike in other developed countries on other continents, the African case is a complex one, needing the concerted effort and sustained determination of all Africans and development partners of Africa. Building the infrastructure essential for African development is a formidable task that requires individual, national, and multinational effort.

It is against this background that the Makerere University Faculty of Computing and Information Technology (CIT) started a conference on the information communications technology (ICT) infrastructure and capacity-building in developing countries. The goal was to create an international forum for discussion of ideas, proven solutions and best practices to Africa’s development problems that can be easily adapted and utilised to build and sustain an ICT environment and infrastructure for Africa’s development.

In its second year, the Makerere University Faculty of Computing and Information Technology’s International Conference on Sustainable ICT Capacity brought together interested academicians, developers and practitioners for four days at the Makerere University campus to discuss, share ideas and experiences, and develop and recommend strategies for ICT-based development.

The conference attracted more than 100 participants with over thirty scholarly papers presented covering a wide array of topics with relevance to ICT-based development in four tracks: Computer Networks, Information Technology, Information Systems, and Computer Science and Software Engineering.

Organising a conference like this one requires considerable effort and sacrifice. This is exactly what all the people on the organising committee did.

We are grateful for their dedication and hard work, without which the conference and book would not have been possible. We are also grateful to the publishing team at Fountain Publishers, Kampala. Lastly special thanks to Jackson Muhirwe, Secretary to the conference and his team for their dedication, which made the conference the success that it was.

Joseph M. Kizza
Tennessee, USA.
Editors

Kizza, Joseph Migga received a BSc in Mathematics and Computer Science from Makerere University, Kampala, Uganda, an MSc in Computer Science from California State University, USA, an MA in Mathematics from the University of Toledo, Ohio, USA and a PhD in Computer Science from the University of Nebraska-Lincoln, Nebraska, USA. Dr Kizza is currently a professor in the department of Computer Science and electrical engineering at the University of Tennessee at 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.

Aisbett, Janet has a PhD in Pure Mathematics from the University of Western Ontario. She worked in signal and image processing before becoming interested in problems associated with information storage and retrieval. For the last decade her theoretical research has concerned the effects of representation on information exchange and cognition. Her applied research has focused on the integration and evaluation of health systems and electronic commerce. She has held various administrative posts within the University of Newcastle, including being Deputy President of Academic Senate and a member of the University's Council, and within regional and national groups is involved in encouraging IT skills development.

Vince, Andrew received his BSc from Stanford University, his MA from Harvard University and his PhD in Mathematics from the University of Michigan in 1981. Prof. Vince has researched extensively and won several grants. He has been full Professor of Mathematics at the University of Florida since 1992. His current areas of research interest include: Combinatorics and Graph Theory, Polytopes and Tilings, Combinatorial algorithms, and Discrete Geometry.

Wanyama, Tom received his BSc and MSc degrees in Electrical Engineering from Makerere University, Uganda, and the University of New South Wales, Australia in 1993 and 1995 respectively. In addition, he received a postgraduate certificate in university teaching in 2005, and a PhD in Electrical and Computer Engineering (majoring in Software Engineering, with minors in Artificial Intelligence and in Communications) in 2006 from the University of Calgary, Canada. He has over ten years of university teaching experience in Electrical, Communications and Computer Engineering. Moreover, he has carried out research and worked in various areas of Electrical and Computer Engineering for over ten years. Dr Wanyama is currently a Lecturer in the Faculty of Technology and a Research Advisor in the Faculty of Computing and Information Technology at Makerere University. His research interests include: Agent-based Software Engineering, Software Decision Support Systems, Component-based Software Development, Intelligent Software Systems, Data Mining and Knowledge Discovery, Data Warehousing, Machine Learning and Computer Networks. Dr Wanyama has published extensively in the various areas of his current research interest.
About Contributing Authors

Aisbett, Janet has a PhD in Pure Mathematics from the University of Western Ontario. She worked in signal and image processing before becoming interested in problems associated with information storage and retrieval. For the last decade her theoretical research has concerned the effects of representation on information exchange and cognition. Her applied research has focused on the integration and evaluation of health systems and electronic commerce. She has held various administrative posts within the University of Newcastle, including being Deputy President of Academic Senate and a member of the University’s Council, and within regional and national groups is involved in encouraging IT skills development.

Atib, Habibu is a PhD Student at the School of Information Systems, Victoria University. Mr Habibu’s research activities focus on the areas of Knowledge Discovery of Databases, Data Mining, Information Security and Policies, Distributed Systems, Client Server Technology, Databases (Distributed, OO, Relational), E-Technologies (E-Learning, E-Health, E-Commerce, E-Government). Mr Habibu spent 6 month’s with the IBM Java Security Group in Germany and has taught Master of Technology (computing) in RMIT University as well as related subjects at the Faculty of Computing and IT, Makerere University, School of Computer Science and IT, RMIT University, Australia, the School of Information Systems, Victoria University, Australia and Information Systems, Nkozi University.

Ayienga, Eric is a member of academic staff at the School of Computing and Informatics, University of Nairobi. His interests include multi-agent systems, quality of service assurance and grid computing networks.

Baryamureeba, Venansius is the Dean and Senior Lecturer in Computer Science at the Faculty of Computing and Information Technology, Makerere University, Uganda. Baryamureeba holds a PhD in Computer Science. He is currently the Executive President of the Uganda National Academy of Computing and Information Technology, an association that brings together all computing faculties/ institutes in all public and private universities to address common objectives and challenges. Baryamureeba is also Chairman and Managing Director of ICT Consults Ltd (http://www.ict.co.ug), one of the prestigious ICT consultancy firms in Africa. He is a distinguished lecturer, researcher, administrator and manager. Baryamureeba lectures courses in Computing at all levels, including doctoral level. He has supervised several students at master’s and PhD levels in the area of Computing. Baryamureeba has published extensively in refereed journals and books.

Bytheway, Andy has degrees in Electrical Engineering and Computer Science. He qualified as a Chartered Engineer in the United Kingdom and he is a Fellow of the IMIS. He has written and travelled extensively, and has held visiting lectureships at Victoria University of Wellington (New Zealand), the University of the Witwatersrand (South Africa), Rhodes University (South Africa), HEC Lausanne (Switzerland), Zayed
University (UAE) and Kingston University (United Kingdom). Andy Bytheway moved to South Africa in 1998 to take up the “Old Mutual Chair in Information Systems” at the University of the Western Cape, following ten years with the Information Systems Group at the Cranfield School of Management (UK). At UWC he established and led a new Department of Information Systems that developed rapidly in terms of undergraduate student numbers, new postgraduate programmes, and outreach activity. He became preoccupied with “information management” as an issue that affects everyone, and also with the wider societal aspects of information technology – what we now term “Community Informatics”. In order to develop and contribute to this new area he joined the Cape Peninsula University of Technology in 2004. He has since worked with his colleagues there to initiate new postgraduate programmes in community informatics and new research into related subjects.

Esselaar, Steve is a researcher at the Link Centre at Witswaterstrand University in Johannesburg, South Africa. He holds an MBA (GIBS) from the University of Pretoria and a BA Hons (Joint) (Rhodes) in Political Science and Philosophy. His main research interests include Telecommunications Policy, Management and Regulation, Competition Theory, and Structure of Markets.

Getao, Katherine is the Director of the School of Computing and Informatics, University of Nairobi. Her interests include artificial intelligence and natural language applications for Bantu languages. She supervises a number of undergraduate and postgraduate students. She is a joint supervisor of the first author.

Kanagwa, Benjamin is an Assistant Lecturer and PhD student at the Faculty of Computing and Information Technology, Makerere University. His research interests are in the area of software architectures, service-oriented software engineering, components and software reuse.

Kizza, Joseph Migga received a BSc in Mathematics and Computer Science from Makerere University, Kampala, Uganda, an MSc in Computer Science from California State University, USA, an MA in Mathematics from the University of Toledo, Ohio, USA and a PhD in Computer Science from the University of Nebraska-Lincoln, Nebraska, USA. Dr Kizza is currently a professor in the department of Computer Science and electrical engineering at the University of Tennessee at 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.

Madanda, Aramanzan is an assistant lecturer in the Department of Women and Gender Studies, Makerere University. From 2002 he has been ICT Coordinator. He holds a Bachelor of Arts with Education (History & Economics) degree a postgraduate diploma in Computer Science, and a Master of Arts in Women Studies, from Makerere University. He is currently a PhD candidate by research on the topic “ICT Liberalisation, Changing Gender Relations and Ideology in Contemporary Uganda”. 
Maiga, Gilbert holds the following academic qualifications: MSc Computer Science (MAK), MSc Animal Physiology (UoN), Bachelor of Veterinary Science (MAK), Diploma in Computer Systems Design (WLC) Certificate in Computer Programming (WLC). Networking Basics Certificate (CISCO). The MSc Computer Science project was on “Information Systems Integration in HIV-AIDS Management: A Web-based Database Approach”. He registered for a PhD in the Department of Information Systems, Faculty of Computing and IT, Makerere University in August 2005, where he is also an Assistant Lecturer. Bioinformatics is his main area of interest for research. His PhD research topic: “An Ontology-based Framework for Integrating Large-Scale Biological and Clinical Information”.

Manderick, Bernard is the Director of the Computational and Modelling Laboratoy (COMO), Free University of Brussels, Belgium. His interests include artificial intelligence, machine learning and evolutionay computational modelling. He is a joint supervisor of both first and second authors.

Miriti, Evans is a Tutorial Fellow and Masters degree student at the School of Computing and Informatics, University of Nairobi, Kenya. He has research interests in natural language engineering and mobile computing.

Mugisa, Ezra K., is a visiting senior lecturer at the Faculty of Computing & IT, Makerere University. He holds a PhD from Imperial College, London. His research interest include Component-based software engineering, formal methods, software architectures, software reuse, e-learning and ICTs for development.

Muhirwe, Jackson holds a BSc (Mathematics, Psychology) degree and a MSc in Computer Science degree from Makerere University. Muhirwe is a registered PhD student in the Department of Networks, Faculty of Computing and IT, Makerere University. His main areas of research interest are computational linguistics and open source software. Muhirwe is employed in the Faculty of Computing and IT, Makerere University, as a Research Coordinator and his main duty is to coordinate and manage research projects for (post) graduate students. He teaches courses on the undergraduate degree programmes and has strong programming skills. He has very good communication, writing and interpersonal skills.

Muliaro Wafula, Joseph obtained a BSc (Kenyatta University), an MSc (University of Nairobi) and MPhil (University of Cambridge UK). He is currently a lecturer at the Institute of Computer Science and Information Technology at Jomo Kenyatta University of Agriculture and Technology, Kenya and a PhD student researching on ICT policies and e-strategies.

Muthoni Masinde, Euphraith (MS) holds a Master’s as well as a bachelor’s degree in Computer Science. She has five years’ teaching experience as a lecturer at the School of Computing and Informatics, University of Nairobi, Kenya where she specialises in databases and programming courses. M Masinde is also a member of the Distributed Systems and Computer Architecture Research Group at the School and her research interests are in middleware for heterogeneous distributed databases. Emphasis is on how the Common Object Request Broker (CORBA) approach can be used to develop
middleware for such databases. Prior to this, she worked as a software developer in the Management Information Systems (MIS) Department of the University of Nairobi ICT Centre where she held positions of Senior Analyst/Programmer and then as Chief Analyst/Programmer and Team Leader of the University’s Students Management Information System (SMIS). Ms Masinde is an active member of IEEE and at the moment is registered as a PhD student at the School.

Ndiwalana, Ali is a PhD student researcher and works for the Directorate for ICT Support at Makerere University, Uganda. He holds an MSc in Computer Science from Virginia Tech. in USA and a B Arch from Makerere University. His main research interests are in how we can better leverage information and communication technologies (ICTs) for development. Specific areas of focus include the small business enterprises (SMEs) and business support organisations as well as education and educational institutions all in the context of developing nations with.

Ngobye, Martin obtained his master’s degree in Computer Science & IT in 2005 from Hunan University, China. Mr Ngobye also holds a postgraduate diploma in Computer Science from Makerere University and a bachelor’s degree in Education (Mathematics and Physics) of Makerere University. He is currently employed as a part time lecturer in the Department of Information Technology, Faculty of Computing & IT, Makerere University. His research interests are in software testing with object-oriented programming languages.

Ngubiri, John is an Assistant Lecturer in the Department of Computer Science, Faculty of Computing and Information Technology, Makerere University. His research interests are in performance evaluation, system optimisation and parallel and distributed processing. He is currently a PhD student at Radboud University Nijmegen, The Netherlands. He is in the Information Retrieval and Information Systems (IRIS) research group of the Nijmegen Institute for Informatics and Information Science (NIII). Faculty of Science, Mathematics, and Computing Science. His promoter is Prof. Mario van Vliet.

Njenga, James is a graduate student in the Department of Information Systems, University of the Western Cape. His research interests are on the use of ICTs in education. His master’s study dwelt on the area of instructional design for web-based Learning Management Systems. This study was inspired by his previous experiences at the University of Nairobi, Kenya, where he was first employed and worked for nearly two years after graduating with a Bachelor of Science in Computer Science degree from the same university. He is currently pursuing his doctoral studies and at the same time working for the newly formed E-learning Division at the University of the Western Cape.

Nowé, Ann is Deputy Director of the Computational and Modelling Laboratory (COMO), Free University of Brussels, Belgium. Her interests include artificial intelligence, machine learning and multi-agent systems. She is a joint supervisor of both first and second authors.
Ogao, Patrick has been involved in visualisation research since 1996. He has a PhD (Utrecht University, the Netherlands) and an Msc (ITC Enschede, The Netherlands) focusing on visualisation in information systems and dynamic geo-phenomena. Prior to joining Makerere University, he worked with a research group in scientific visualization and computer graphics in the department of computing and mathematics of the University of Groningen, The Netherlands. He is currently a visiting senior lecturer and acting Head of Information Systems Department in the Faculty of Computing and Information Technology, at Makerere University. He is the Long Term Expert (on secondment by the University of Groningen) in a NUFFIC/NPT project in ICT training capacity-building in four public universities in Uganda. His research interests are in visualisation applications in bio-informatics, geo-informatics, and software engineering and in developing e-strategies for developing countries.

Ogot, Madara is currently an Associate Professor of Engineering Design and Mechanical Engineering at Pennsylvania State University. After completing his BSE in Mechanical Engineering at Princeton University in 1987, and his PhD at Pennsylvania State in 1991, he joined the Mechanical and Aerospace Engineering Department at Rutgers University in 1991, where he remained until moving to Penn State in 2003. He received an MBA from Rutgers in 1997. His research interests include optimisation methods, design under uncertainty, innovative design, and engineering design education. Dr Ogot is also the founding Editor-in-Chief of the Journal of TRIZ in Engineering Design.

Okello-Odongo, William is a senior member of academic staff at the School of Computing and Informatics, University of Nairobi. His interests include distributed systems and high performance computing. He is a joint supervisor of the second author.

Okwangale, Fredrick Richard is an Assistant Lecturer since 2001 at the Faculty of Computing and Information Technology (CIT), Makerere University (MAK) after having been a Teaching Assistant at the then Institute of Computer Science since 1993. He lectures and supervises students at both undergraduate and postgraduate levels. He is currently pursuing a doctorate degree at Makerere University in the area of information systems. The doctoral dissertation he is working on is entitled ‘The GIS-based Crime Forecasting and Prediction Model’. His research interest is in the area of data mining and geographic information systems.

Omwenga, Elijah is a Computer Science lecturer at the University of Nairobi, School of Computing and Informatics. He is currently serving as the Acting Director of the University ICT Centre. Dr Omwenga is a leading consultant for UNESCO in its ICT for Science and Engineering Education for Africa programme. He is actively engaged in e-learning research and is a seasoned software engineer with over 15 years’ experience in software design and development.

Opiyo, Elisha T.O. is a member of academic staff at the School of Computing and Informatics, University of Nairobi. His interests include applications of multi-agent systems, scheduling and grid computing.
Oyo, Benedict is a PhD student in Information Systems at Makerere University, Faculty of Computing and Information Technology, Uganda. His research area is Quality Assurance in Higher Education using System Dynamics Modelling. Mr Oyo has an MSc in Computer Science from Makerere University.

Patel, Dilip holds the chair of Information Systems and is the head of the Centre for Information Management and E-Business, which consists 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 of 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.

Patel, Shushma is a principal lecturer in the Faculty of Business, Computing and Information Management at London South Bank University. Her research interests include knowledge management and organisational decision-making. Dr Shushma Patel is on the editorial board of two international journals: The International Journal of Cognitive Informatics and Natural Intelligence (IJCiNi) and The International Journal of Computing and ICT Research. She has worked on a number of research projects funded by the European Union, DTI and industry.

R.G. Collins, Asifiwe’s career in education includes secondary teaching in government. His advocacy for computers in the classroom began in 2005; as a result, a short time after enrolling for the master’s degree course in Information and Communication Technology he has carried out training sessions in ICT in various areas, especially in secondary schools. He is currently a student at the University of Makerere doing research as the final part of his master’s degree in Educational Information and Communication Technology. Mr. Asifiwe has had a teaching experience of more than five years in a number of schools and for the past three years has been in active research with a number of research consultancies, both academic and socio-economic. In his advocacy, he is committed to preparing young people to succeed in a rapidly changing world and sees the development of ICT for staff and students as fundamental in creating schools and institutions for the future. His wish is to help develop staff capabilities such as network infrastructure and a technical support infrastructure as precursors to inspiring teacher enthusiasm for integrating ICT into the curriculum.

Rehema, Baguma (Ms) holds a Bachelor of Library and Information Science degree from Makerere University and an MSc in Computer Science & Applications from Huazhong University of Science and Technology, China. Rehema is registered for a PhD in the Department of Information Systems, Faculty of Computing and IT, Makerere University in August 2005. Her main areas of research interest are in software frameworks, e-strategies and open source software. MS Baguma is employed in the Faculty of Computing and IT, Makerere University as an Assistant Lecturer teaching Client/Server and Distributed Computing, Database Systems and Information Security Policy, Privacy and Ethics.
Rodrigues, Anthony J. obtained his BSc, MSc and PhD degrees from the University of Manchester. He is a Professor of Computer Science at the University of Nairobi, Kenya. His research interests include ICT policies and e-strategies that support good governance and sustainable development.

Rwangoga, Narcis T. obtained his master’s degree in Computer Science in 2004 from Central South University, Hunan Province, China. Mr Rwangoga also holds a postgraduate diploma in Management from the Uganda Management Institute (2000) and a bachelor's degree in Agricultural Economics from Makerere University (1994). Mr Rwangoga is currently employed Assistant Lecturer in the Computer Science Department at the Faculty of Computing and IT, Makerere University. He has been involved in teaching and research supervision for students in areas of Artificial Intelligence and Expert Systems, IT Strategic Planning and Management, Networked Systems and Information Security Management. He has professional certifications in different computer application fields, especially in databases and networking technologies. Mr Rwangoga’s current research interests lie in the areas of distributed systems (mobile agents technology, distributed databases), and artificial intelligence (expert systems, neural networks and genetic algorithms).

Sansa, Julianne holds a BSc (Maths, Computer Science) and MSc (Computer Science) from Makerere University. Since 2001 she has worked with the Faculty of Computing and IT of Makerere University in various capacities. She is currently registered as a PhD student at the University of Groningen, in the Netherlands, and her research interests are quality of service and network protocols.

Ssemaluulu, Paul obtained a FTC in Telecommunications from Uganda Polytechnic Kyambogo in 1985. He also holds a BBA from Makerere University in 2001, an MSc in Computer Science in 2006 and is a PhD student in Information Systems at the Faculty of Computing and IT, Makerere University.

Stork, Christoph is a senior researcher at the Link Centre at Witswatersrand University in Johannesburg, South Africa and Associate Director of the research programme in Financial Economics and ICT at the Namibian Economic Policy Research Unit. He holds a PhD in Financial Economics from London Guildhall University, UK, a Diplom Kaufmann (MA) from the University of Paderborn, Germany, and a BA Economics from the University of Nottingham, Trent, UK. His main research interests are the economics of ICT, financial economics and private sector development. He also writes for Insight Magazine, a Namibian current affairs magazine on policy issues in information and communication technology.

Szomoru, Arpad obtained a PhD in astronomy from the University of Groningen, the Netherlands. He has worked at (JIVE) the Joint Institute for VLBI in Europe for several years as a senior software scientist. During this period he was heavily involved in the deployment and integration of disk-based recording systems and the early development of e-VLBI. Since 2006 he has been the head of data processor research and development at JIVE.

Tusubira, F. is the Director, Directorate for ICT Support, Makerere University and an Associate Professor in the Electrical Engineering Department. He obtained his PhD
from the University of Southampton, England; his MSc E from the University of New Brunswick, Canada; and his BSc (Eng.) (1st Class Honours) from Makerere University. He is a telecommunications professional whose research interests include: policy formulation and management; project management from conception to implementation; regulation and regulatory issues; information and communication technology (ICT) infrastructure design and implementation; ICT as a tool for development.

**Van der Hulst, Thijs** got his PhD from Groningen in 1977 on a study of neutral hydrogen emission from interacting galaxies. He spent 5 years in the USA at the National Radio Astronomy Observatory and the University of Minnesota before returning to the Netherlands, where he joined the staff of ASTRON, the Netherlands Foundation for Radio Astronomy. In 1982 he moved to the Kapteyn Astronomical Institute of the University of Groningen, of which he now is the director. His main field of interest is structure and evolution of galaxies.

**van Vliet, Mario** is a professor in software project management at the Radboud University Nijmegen, the Netherlands, and a vice president at Capgemini. At Capgemini, he advises clients on optimising their business processes and related IT systems. He holds a PhD in Operations Research. Mario’s research interests lie in machine scheduling, queueing networks, information management, software project management, supply chain management and business process improvement. He has published in leading operations research and management science journals.

**Vince, Andrew** received his BSc from Stanford University, his MA from Harvard University and his PhD in Mathematics from the University of Michigan in 1981. Prof. Vince has researched extensively and won several grants. He has been full Professor of Mathematics at the University of Florida since 1992. His current areas of research interest include: Combinatorics and Graph Theory, Polytopes and Tilings, Combinatorial algorithms, and Discrete Geometry.

**Wagacha, Peter W.** is a lecturer and researcher at the School of Computing & Informatics, University of Nairobi. His main interests are Machine Learning, Natural Language Processing (NLP) and Mobile Telephony. He is currently researching on local languages and the application of Machine Learning in NLP.

**Wanjohi, Nick G.** obtained his BA, MA and PhD degrees from the University of Nairobi. He is an Associate Professor of Political Science and currently the Vice Chancellor of Jomo Kenyatta University of Agriculture and Technology, Kenya.

**Wanyama, Tom** received his BSc and MSc degrees in Electrical Engineering from Makerere University, Uganda, and the University of New South Wales, Australia in 1993 and 1995 respectively. In addition, he received a postgraduate certificate in university teaching in 2005, and a PhD in Electrical and Computer Engineering (majoring in Software Engineering, with minors in Artificial Intelligence and in Communications) in 2006 from the University of Calgary, Canada. He has over ten years of university teaching experience in Electrical, Communications and Computer Engineering. Moreover, he has carried out research and worked in various areas of Electrical and Computer Engineering for over ten years. Dr Wanyama is currently a Lecturer in the...
Faculty of Technology and a Research Advisor in the Faculty of Computing and Information Technology at Makerere University. His research interests include: Agent-based Software Engineering, Software Decision Support Systems, Component-based Software Development, Intelligent Software Systems, Data Mining and Knowledge Discovery, Data Warehousing, Machine Learning and Computer Networks. Dr Wanyama has published extensively in the various areas of his current research interest.

**Williams, Ddembe** is a Visiting Senior Lecturer in Information Systems at the Faculty of Computing and Information Technology (CIT), Makerere University, Uganda. He is also the Deputy Dean for Graduate Studies and Research. For the last seven years, Ddembe has been a Senior Lecturer in Decision Sciences and Head of Centre for Advanced Systems Modelling and Simulation within the faculty of Business, Computing and Information Management (BCIM) at London South Bank University (LSBU) in the United Kingdom. Ddembe has an MSc in Advanced Information Technology Management and a PhD in Computer Science/System Dynamics from London South Bank University. Ddembe’s practical and professional work in teaching and consultancy focuses on applied computing to systems requirements engineering, process analysis, problem structuring and model-based decision support systems. His theoretical/research work centres on the contribution that system dynamics/information systems can make to the advancement of systems modelling and simulation. Ddembe has published over 20 refereed conference and journal papers.

**Zeleznikow, John** is a full Professor in the School of Information Systems, Victoria University of Technology. Professor Zeleznikow has a Bachelor of Science (Hons) and Doctor of Philosophy from Monash University as well as a Graduate Diploma in Computer Science from the University of Melbourne. Professor Zeleznikow has written 2 books, 40 journal articles, 100 conference papers, workshop articles, and 8 book chapters, mainly in the area of Information Systems. He has tremendously contributed to the field of Information Systems by giving seven invited talks at international conferences and has presented eighteen half day tutorials. He is on the editorial board of numerous journals including *Artificial Intelligence and Law, Internet Research and Information and Computer Security*. He has twice been chairperson of the International Conference on Artificial Intelligence and Law and founded the Australian Database Conference (in 1990). He has successfully supervised ten PhD students, has been an expert commentator on Information Technology and Law and has worked with many industry partners, including legal firms (Phillips and Wilkins and Allan Moore and Company), non-governmental organisations (Victoria Legal Aid Relationships Australia, Software Engineering Australia and Therapeutic Guidelines) and Police Forces (Lothian and Borders and West Midlands in the UK). He has conducted teaching and research in France, The Netherlands, Israel, the United Kingdom and the United States of America.
Acknowledgements

We are grateful to those who contributed peer reviewed and edited papers to this book. The papers were accepted for presentation at the second International Conference on Sustainable ICT Capacity in Developing Countries (SREC 06), with the theme “Strengthening Research Networks in Developing Countries”, held on 6-9 August 2006, at the Faculty of Computing and Information Technology, Makerere University, Kampala Uganda.

The editors would like to recognise and also give special thanks to the following:

- The different researchers around the world who reviewed the papers anonymously;
- The Conference Chair and the Chairpersons of the conference technical committees;
- Members of the conference organising committee and conference technical committees;
- The Conference Secretariat for supporting the conference activities and putting together this book;
- Makerere University Management and the University of Groningen for their overwhelming support;
- The Netherlands Organisation for International Cooperation in Higher Education for providing financial support for the Conference and publication of this book.
- Lastly but not least our publishers, Fountain Publishers, for a job well done.

Editors
Prof. Joseph Kizza
Prof. Janet Aisbett
Prof. Andrew Vince
Dr Tom Wanyama Tom
Introduction

Joseph M. Kizza, Department of Computer Science, The University of Tennessee-Chattanooga, TN 37403

Computers in particular and information communication technology in general have, in the last fifty years, driven development in every country. There is documented evidence that computers and information technology indeed speed up and enhance development. However, the rate of growth has been uneven. In rich industrialised countries, where resources and capacity are abundant, such growth and the rate of growth have been phenomenal. Developing countries, however, have not been so lucky as to benefit substantially. There are observed and documented complex problems that need to be overcome before any meaningful development can be registered.

The papers in this book all focus on the theme of ICT and development. “The Second International Conference on Sustainable ICT Capacity in Developing Countries – SREC06”, out of which this book grew, was a platform and a forum where ICT-based development was discussed over a course of three days. Papers were presented in four tracks: Computer Networks, Information Technology, Information Systems, and Computer Science and Software Engineering. There were seven keynote papers to headline the conference. Joseph M. Kizza in “Bridging Africa’s Digital Divide: Building Sustainable ICT Infrastructures” takes note of the changing fortunes of Africa with the acquisition of computing technologies, especially wireless, resulting from the miniaturisation and the plummeting prices and how these will help Africa to quickly leapfrog into the 21st Century with the kind of rapid development not seen in generations. Janet Aisbett in “Information Quality and Information Systems Success” reviews key models of IS success in the context of the prevailing IS. She identifies information quality as a constant factor in the models, and poor content quality as a continuing contributor to perceptions of failure. She concludes that fundamental research into information quality is still needed, and provides an example of such investigations concerning new forms of representation. Dilip Patel in “An Organisational Climate Awareness Toolkit for Nurturing the Effectiveness of Team/Group Interactions” discusses the findings of a study of organisational climate issues in several commercial organisations and outlines an organisational awareness toolkit.

Anthony J. Rodrigues in “ICTs in Developing Countries: Contexts, Challenges And Interventions” discusses the barriers that Africa must overcome, including lack of: information coordination, coordination of physical connections and technical personnel cooperation. He notes that this is exacerbated by telecommunications monopolies and obsolete regulatory frameworks, limited involvement of research institutions in network building and diffusion, and language barriers. There is a need to lower the overall costs of creating a competitive workforce and to do this continuously on a 24/7 basis for people, especially those in rural areas. In “A Fundamental View on the Act of Modelling”, H.A. (Erik) Proper and P. van Bommel discuss the role
of models and modelling in the information system development life-cycle as part of their ongoing research effort to better understand the act of modelling. They describe a formal framework by which the process of modelling can be regarded as involving the selection of more and more refined interpretations in terms of the underlying metamodel of the modelling language used. The resulting framework will be used to create a laboratory set-up in which to consequently and closely study (and support) modelling processes. Madara in “African Virtual Environment Collaborative (Afrovelab)” describes the African virtual environment collaborative that aims to solve the failure of African science and technology institutions to take advantage of the immense collaborative synergies the information age has brought about. He proposes a medium where African S&T researchers can readily form research teams – based on areas of expertise – and take on sophisticated collaborative research projects that each could never attempt by going it alone.

Lastly Andrew Vince in “Indexing a Discrete Global Grid” discusses a method for computer representation and manipulation of global data based on a multi-resolution sub-division of a regular polyhedron. In particular he considers the problem of efficiently indexing the cells of such a discrete global grid.

Presenters in each of the four tracks discussed in depth the latest research and developments in the area protocols and best practices that are suitable for development and the how-to in the implementation of some of these notable solutions and best practices.

In Computer Networks, five papers were presented. In “Multi-agent Systems for Distributed Resource Allocation”, Eric Ayienga et al. discuss the additional challenges introduced by wireless grids that are above and beyond those already existing in resources allocation in the wired grid environment. Since wireless grids are considered as complex systems with economies where multiple applications (consumers) compete for resources and services from network providers (suppliers), optimisation techniques can be done on network QoS parameters to solve problems in wireless networks. In “On Network Measurement and Monitoring of End-to-End Paths Used for e-VLBI”, Julianne Sansa et al. present and discuss the bottlenecks currently limiting the transfer speeds of astronomical data from radio telescopes across the world over high speed links to the central processing centre in the Netherlands. The processed data is used to produce images, which are used by radio astronomers. They image the sensitivity of continuum observations scales with the data rate, so the higher the data rate, the more sensitive these observations will be. They observed that while the required high data rates should be attainable on these network links, it is not, however, the case. So they discuss how to uncover the prevailing bottlenecks and propose solutions to address them. Narcis T. Rwangoga and Martin Ngobye in “The Future of Intelligent Networks in Developing Countries” discuss the growing demand for network-based services in developing countries and how wireless technologies for cellular and personal communications have extended these network services to most areas in developing countries. They explore how Intelligent Networks (IN) are a promise that countries, which are implementing telecommunication networks, can use to deliver network-based
services. Habibu Atib and John Zeleznikow in “Identifying Sensitive Knowledge for Law Enforcement Agencies” take a hard look at how law enforcement and intelligence agencies can effectively manage knowledge to further their goals of crime detection and prevention. Techniques such as computer profiling, link analysis and knowledge discovery from databases are discussed. Finally Kasigwa, Baryamureeba and Williams in “An Efficient Dynamic Admission Control for End-to-End Delay Guarantees in IP Networks” describe and classify a broad set of proposed admission control algorithms, evaluate the accuracy of these algorithms via experiments using both on-off sources and long traces of compressed video, compare the admissible regions and QoS parameters predicted by their implementations of the algorithms with those obtained from trace-driven simulations, and identify the key aspects of an admission control algorithm necessary for achieving a high degree of accuracy and end-to-end QoS of a computer network.

In Information Technology, five papers were presented. Aramanzan Madanda in “ICT Liberalisation and Changing Gender Relations in Contemporary Uganda: A Research Agenda”, identifies research needs in the area of gender and ICT in Uganda, especially in relation to adoption/non-adoption and changing gender relations at individual and community levels. Madanda takes up a gender perspective in this discussion of the current Ugandan ICT policy framework and ICT diffusion and adoption, drawing on arguments advanced by the liberal economic theory of technology diffusion and adoption and the feminist theory to illuminate key gender and ICT research issues in Uganda. Elijah Omwenga in “ICT for Educational Use: Cases of Implementing a Partnership Approach Model for Reaching Target Groups” discusses how institutions of higher learning can use ICT to further their mission. Two phases to achieve this are advanced: phase one of awareness and advocacy through seminars and workshops, sensitisation on the relationship between ICTs and wealth-creation and training of stakeholders and trainers; phase two of taking the programmes to the people – the students and target groups. Omwenga also proposes a partnership model that can be applied to help realise the goals of effectively reaching the target groups. The model recognises a three-stage pyramidal communication model involving target groups and partners, networking and collaboration and finally policy strategising on thematic areas. Joseph Wafula et al. in “Informing Regional ICT Policy: Case Study of Trends in ICT Indicators of OECD, EU, COMESA and EAC”, discuss trends and analyses of ICT indicators such as mobile phone subscribers, mobile communications revenue, annual telecommunications investment, international bandwidth and mass media usage in relation to ICT policy, e-strategy and the enabling environment for OECD, EU, COMESA and EAC based on the 2005 ITU and DHS databases. Correlation analysis among indicators is also done in an attempt to establish emerging commonalities and differences among these regions and the accompanying lessons. Asifiwe Rubanju in “The Impact of ICT on Universities: Classroom/Lecture Theatre Design and Curriculum Delivery”, describes a research done to determine the impact of information and communication technologies on class/lecture room design and curriculum delivery. Finally P. J. Ogao in “Beyond the Computer Graphics Course: A Case for African Universities”, proposes a knowledge
continuity strategy to the computer graphics course by outlining a visualisation course content that supports science and society.

In Information Systems, ten papers were presented. In “Information and Communication Technologies (ICTs) in Small and Medium Enterprises (SMEs): Findings From Uganda” Ndiwalana Ali et al. report findings from a research carried out on 351 SMEs in Uganda with a varying degree of formality to help understand the potential of ICTs on SMEs, by understanding the entrepreneurs behind the SMEs, their information practices, the needs that dictate the way they operate as well as the environments in which they operate. Benedict Oyo and Ddembe Williams in “An Exploration of the Factors that Affect Quality Assurance Decision-making in Higher Education” argue that the system dynamics modelling approach is a valuable alternative for quality assurance research in higher education. This approach is emphasised owing to the fact that a greater portion of university quality problems and their solutions do not have quantitative foundations mostly because such systems involve qualitative elements that are difficult to quantify and model using other approaches. A framework is used to demonstrate the theoretical value of system dynamics modelling in combining diverse factors responsible for quality assurance in higher education and arriving at a consistent decision about the quality of awards. Euphraith Masinde in “Using JAD to Bridge the Design-Reality Gaps: A Major Cause of IS Projects’ Failures in the Developing Countries” explains how Joint Application Development (JAD), a software development methodology that will involve the stakeholders in the entire process of IS implementation, can be used to eradicate most of the causes of IS projects’ failures in developing countries using the University of Nairobi case study. The CHAOS Ten Success factors have been employed to analyse data for nine IS projects. In “Using Developmental Research to Design a Web Instructional Design Subsystem”, J.K. Njenga and A.J. Bytheway report on a process and findings of using development research design and methodology to design an instructional design subsystem. The development research used involves a review of the literature with the aim of exploring, analysing, integrating and synthesising the broad field of learning and instructional theories, paradigms and best practices with the design of the instructional design subsystem. They also discuss how the subsystem can be improved by rigorous attention to the formulation of theory. Rehema Baguma et al. in “Towards Web Design Frameworks (WDFs)” discuss Web Information Systems, the increasing rate of internet/web usage calls for efficiency and effectiveness in the development and deployment of web systems to provide high quality systems in the shortest time possible, the concept of web design frameworks, and their importance as a formal method of web application development. Ddembe Williams in “Revisiting Dynamic Synthesis Methodology: A Reference Theoretical Framework and Tool for Business Process Modelling and Analysis” provides a useful and systematic reference point for researchers who wish to work in the business process modelling and generally to encourage careful work on the conceptualisation and execution of the Dynamic Synthesis methodology in business processes and to a wider process-modelling field. Ddembe also suggests that the potential usefulness of the Dynamic Synthesis Methodology is in aiding researchers and managers in improving both building and
testing theories in understanding business processes and strategic modelling and analysis. Gilbert Maiga and Ddembe Williams in “Towards a Reusable Ontology Framework for Biological Information Integration” provide the research background and approach to an ongoing study that aims to develop a reusable framework for the integration of biological and clinical research data. A theoretical basis for the reusable framework is given and the proposed approach for developing and validation of the framework using the Protégé ontology development environment is also outlined. In “Survey of Data Mining Methods for Crime Analysis and Visualisation”, Fredrick R. Okwangale and Patrick Ogao review the applicability of various data mining methods and Geographic Information Systems in crime analysis and visualisation in mainly poorly planned settings characterised by missing electronic data, a common phenomenon in developing countries like Uganda, focusing on criminality of places rather than the tracing of individual criminals. And finally Paul Ssemaluulu and Ddembe Williams in “A System Dynamics Tool for Evaluating IT Investment Projects”, investigate five different methodologies, taking into account the suitability or goodness of the framework, bias, focus and complexity. The simulation tool was used to analyse how different variables interact to affect the total benefits of an information system. It was observed that only a strong interaction of people, information, and technology can improve business performance, and consequently lead to information systems success.

In Computer Science and Software Engineering, seven papers were presented. Venansius Baryamureeba in “On Solving Large Scale Linear Systems Arising from Interior Point Methods for Linear Programming” proposes a strategy for solving the linear systems arising from interior point methods for linear programming. Furthermore, Baryamureeba proposes how to construct a preconditioner for the iterative approach for solving linear systems. In “Architectural Building Blocks for Reusable Service-Oriented Architecture with Integrated Transaction Processing”, Benjamin Kanagwa proposes a generic high-level architecture that specifies the structural elements for transaction centered service oriented architecture. Elisha T. O. Opiyo et al. in “Multi-Agent Systems Scheduler and Open Grid Computing Standards” propose a scheduling model based on multi-agent systems and review market-based approaches to the resource allocation problem for decentralized contexts. They interestingly note that increased reliance on agents and Web services may one day lead to a situation where the grid computer is the Internet and the Internet is the grid computer. Peter Waiganjo Wagacha and Dennis Chege in “Adaptive and Optimisation Predictive Text Entry for Short Message Service (SMS)” describe the development of a generic mobile phone predictive text application they experimented with using local Kenyan languages. They also present an adaptive learning model for improving text entry speed that incorporates a specific user’s word usage habits. In “Kinyarwanda Speech Recognition for Automatic Voice Dialling Systems”, Jackson Muhirwe describes the design and training of a Kinyarwanda language speech recognition system that could be used by developers to create applications that will help those who want to use and speak Kinyarwanda. Katherine W. Getao and Evans K. Miriti in “Computational Modelling in Bantu Language” propose a derivational approach to discovering the structure of compound Bantu words that does
not depend on large amounts of prior knowledge. This technique is amenable to machine learning methods such as reinforcement learning. Through this they demonstrate a new approach to natural language processing that may be of specific application to Bantu languages and to other language groups with a strong inflectional strategy. Finally, in “The Influence of Job Physical Characteristics on their Schedulability in Multi-cluster Systems”, John Ngubiri and Mario van Vliet investigate the influence of jobs’ physical characteristics on their schedulability in a multi-cluster system using the Fit Processor First Served (FPFS) scheduler and their relative strength in determining schedulability as well as their sensitivity to scheduler parameters.

This book will be a valuable tool for teaching and research for scholars, students and practitioners of ICT-based development as well as an enriching experience for all interested in ICT-related issues.
Part One

Advanced System Modelling
In an increasingly interconnected world, the development of computer friendly systems for the display and analysis of global data is an important issue. This paper concerns a method for computer representation and manipulation of global data based on multi-resolution subdivisions of regular polyhedra. In particular, the problem of efficiently indexing the cells of such a discrete global grid is addressed.

1. Digital Earth

The subject of this report, the representation and analysis of global data, has a history that dates back several millenia. Maps, possibly the earliest depiction of geographical information, are easily understood and appreciated regardless of language or culture. The oldest known maps are preserved on Babylonian clay tablets from about 2300 B.C. That the Earth is spherical was known by Greek philosophers by the time of Aristotle (about 350 B.C.). Ptolemy’s map (about 85-165 A.D.) depicted the Old World from about 60N to 30S latitudes. The first whole world maps began to appear in the early 16th century, following voyages by Columbus and others to the New World. In 1569 Mercator published a map of the world intended as an aid to navigation, using a projection method now known by Mercator’s name. Buckminster Fuller invented the geodesic dome in the late 1940’s. Geographic information systems (GIS) emerged in the 1970-80s. The emphasis over the past few decades has been on the computer display and analysis of georeferenced information and remotely sensed data about the Earth, collected by organizations and institutions. As more global datasets are acquired and as the world becomes more interconnected, this endeavor becomes increasingly important.

This paper concerns such a computer representation of global data, called a discrete global grid, that is based on cellular subdivisions of regular polyhedra. An array of data consists of one data element for each grid cell. The user has the flexibility to define the meaning of grid cell values according to the application at hand. Traditional digital image processing in the plane is carried out on a rectangular grid. For some applications, however, hexagonal grids are advantageous. Hexagonal grids have a higher packing density, approximate circular regions, and each cell has equal distance from its six immediate neighbors. The most commonly used grids on the sphere are those based on latitude-longitude coordinates. On the sphere, however, an (almost) hexagonal grid is an even more natural choice than in the plane. The hexagons have almost the same shape and size as compared with a lat/long grid.
There is a substantial recent literature on the subject of spherical grids based on tessellations of regular polyhedra, including [1, 3, 8, 13]. One commonly mentioned is an aperture 3, multi-resolution tessellation of the sphere into mainly hexagons. Multiresolution means that there is not just a single tessellation, but a hierarchical sequence of progressively finer tessellations. Going further in the sequence zooms in on smaller areas. Aperture 3 refers to the approximate ratio between the areas of hexagons at successive tessellations in the sequence. In fact, this small ratio is one of the features that makes an aperture 3 tessellation appealing. Such a tessellation of the sphere will be referred to as an aperture 3 hexagonal discrete global grid (A3H). A mathematical construction is given in Section 2 of this paper. Figure 1 shows a few levels of resolution of such an A3H.

Of the numerous research challenges in the field, this paper concentrates on the problem of efficiently addressing or indexing the cells of A3H. While Section 2 describes the geometry and basic data structure of A3H, the rest of the paper provides two methods of indexing its cells, which we call barycentric indexing and radix indexing. The first method, based on an investigation of the barycenters of the cells in A3H, was developed in [18]. The second is based on a generalization of the concept of positional number systems (radix systems) for the integers [17]. The particular radix system discussed here for A3H was developed in conjunction with Canadian based company the PYXIS innovation [12]. Each of the two methods has advantages in providing efficient algorithms for the relevant applications. Indexing based on quad tree data structures is appropriate for certain discrete global grids not discussed in this paper [2, 5, 7, 11, 15]. Proofs of theorems appearing in this paper have been referenced but not included.

2. The geometry of A3H

A3H is obtained by tessellating a regular polyhedron, then projecting the tessellated polyhedron onto the surface of the sphere. For this paper the relevant regular polyhedra are the octahedron and the icosahedron. Before giving the construction, the following basic geometric operations are introduced.

The barycenter $\beta(X)$ of a set $X = \{x_1, x_2, \ldots x_n\}$ of points in $\mathbb{R}^3$ is given by Figure 1: A3H.
By a *tessellation* $T$ of a polyhedron or of the sphere, we mean a collection of closed, non-overlapping cells that cover the surface. On a polyhedron the edges are straight lines and, on the sphere, arcs of great circles. For a tessellation $T$, the notation $V(T)$ and $E(T)$ denote the set of vertices and edges, respectively. For a tessellation $T$ let:

$$\beta(X) = \frac{1}{n} \sum_{i=1}^{n} x_i.$$ 

denote the set of barycenters of its cells. (Here $t$ is considered as the set of its three vertices.) Two basic operations on tessellations are the dual and centroid subdivision. They are defined as follows and illustrated in Figure 2.

1. **The dual.** For a tessellation $T$ with vertex set $V$, the dual tessellation $D(T)$ has vertex set $\beta(T)$. Two vertices of $D(T)$ are joined by an edge if and only if the corresponding cells of $T$ share an edge.

2. **Centroid subdivision.** The centroid subdivision $C(H)$ of tessellation $H$ has vertex set $V(H) \cup \beta(H)$. The edge set of $C(H)$ is the union of $E(H)$ and the set of edges formed by joining $\beta(h)$ to each vertex of $h$ for all $h \in H$. 

Fig 1: A3H.
Construction

Let $P$ be either the regular octahedron or icosahedron. Define a sequence $(T_n, H_n)$ of dual pairs of tessellations of $P$ as follows. First, $T_0$ is the polyhedron itself centered at the origin of $\mathbb{R}^3$. The sequence is then defined recursively in terms of the dual and centroid subdivision:

$$H_n = D(T_n)$$
$$T_{n+1} = C(H_n).$$

The sequence $H_n, n \geq 0$, of tessellations is the one called the A3H. We leave open until the next section whether A3H is based on the octahedron or icosahedron, i.e., whether $T_0$ is the octahedron or icosahedron. The number $n$ is called the resolution of A3H. The dual tessellations $T_n$ and $H_n$ are shown, in part, in Figure 3. In fact, Figure 3 shows a patch of two successive subdivisions $T_n$, $T_{n+1}$ and $H_n$, $H_{n+1}$. Note that $V_n := V(T_n)$ is the set of cell centers of the tessellation $H_n$.

Properties

The following properties of A3H are either obvious or easily proved by induction.

1. The tessellation $T_n$ is a triangulation for each $n$, i.e., the faces are triangles. Moreover, the triangulation $T_n$ contains exactly $c \cdot 3^n$ triangles where $c = 8$ or 20 depending on whether $T_0$ is an octahedron or icosahedron.

2. The $n^{\text{th}}$ resolution $H_n$ of A3H contains exactly $\frac{c}{2}(3^n - 1)$ hexagons and either 6 squares or 12 pentagons, depending on whether $T_0$ is an octahedron or icosahedron.

3. The sets of barycenters of cells of the A3H are nested: $V_0 \subset V_1 \subset V_2 \subset \cdots$.

4. The ratio of the area of a hexagon in $H_n$ to the area of a hexagon in $H_{n+1}$ is 3.
Figure 3: Two successive subdivisions.

5. Each hexagonal (pentagonal) cell $x$ of $H_n$ intersects exactly 7 cells (6 cells) of $H_{n+1}$, a centroid cell, with the same barycenter as $x$, and 6 vertex cells (5 vertex cells) whose centers are the vertices of $x$. See Figure 3.

6. Each cell $x \in H_n$ intersects either 1 or 3 cells of $H_{n-1}$, one cell in the case that $x$ is a centroid cell, three cells in the case that $x$ is a vertex cell.

**Projection onto the Sphere**

There are various methods for projecting A3H from the tessellated polyhedron onto the surface of the 2-dimensional sphere, for example by central projection from the origin. No projection method can be both equal area (areas related by a constant scaling factor) and conformal (angles preserved). One of the most common equal area projections, due to Snyder, is well suited to the icosahedron and is called the icosahedral Snyder equal area projection (ISEA3H). The projection formulas can be found in [14].

### 3. Indexing A3H

Recall that the $n^{th}$ resolution of A3H is denoted $H_n$. The set of barycenters of the cells in $H_n$ is denoted $V_n$ and, when no confusion arises, we will identify a cell with its barycenter. Indexing is a means to reference or address the cells in $H_n$ (points in $V_n$). So an index function is an injection

$$I : V_n \rightarrow W_n,$$

where $W_n$ is referred to as the index set. A good index function is one that allows for efficient implementation of algorithms for the relevant applications. This requires the efficient handling of relevant data structures and fast algebraic operations on indices corresponding to local vector operations on the cell centers. In Sections 4 and 5 two methods are introduced for indexing A3H.
4. Barycentric Indexing

Barycentric indexing of the cells of A3H is based on the tessellations of the regular octahedron as described in Section 2. Thus T0 is the octahedron centered at the origin of R^3, and H0, the 0^th resolution of A3H, is the cube. The first resolution H1 is the truncated octahedron. As will be explained soon, the index set W_N consists of all strings of n + 3 digits, each digit taken from the set \{-1, 0, 1\} of “trits”. Thus the index set W_N has size 27 \cdot 3^n while H_N has 4 \cdot 3^n + 2 cells.

To define the index function, first place the vertices of the octahedron T0 at the six points (±1, 0, 0), (0, ±1, 0), (0, 0, ±1) in R^3. Theorem 1 gives the Cartesian coordinates of the set V_N of cell centers of A3H on the surface of the octahedron (prior to projection onto the sphere). The proofs of Theorems 1-3 appear in [17]. All congruences in this section are modulo 3.

**Theorem 1**

The set V_N of barycenters of cells of the octahedral A3H at resolution n is given by

\[
V_n = \begin{cases} 
\{ \frac{1}{3^{\frac{n}{2}}} (a, b, c) : a, b, c \in \mathbb{Z}, |a| + |b| + |c| = 3^{\frac{n}{2}} \} & \text{if } n \text{ is even} \\
\{ \frac{1}{3^{\frac{n+1}{2}}} (a, b, c) : a, b, c \in \mathbb{Z}, |a| + |b| + |c| = 3^{\frac{n+1}{2}}, a \equiv b \equiv c \} & \text{if } n \text{ is odd.}
\end{cases}
\]

Define the A3-coordinates of a point in V_N, or the corresponding cell in H_N, as the ordered triple (a, b, c) of integers as given in Theorem 1. In other words the A3-coordinates of the cell with center \((1/3^{\frac{n}{2}}) (a, b, c), n \text{ even, or } (1/3^{\frac{n+1}{2}}) (a, b, c), n \text{ odd, is simply } (a, b, c)\). Hence the A3-coordinates of a cell of H_N is an ordered triple (a, b, c) of integers such that

\[
|a| + |b| + |c| = 3^{\frac{n}{2}} \quad \text{if } n \text{ is even}
\]

\[
|a| + |b| + |c| = 3^{\frac{n+1}{2}}, a \equiv b \equiv c \mod 3 \quad \text{if } n \text{ is odd.}
\]

Given the A3-coordinates of a cell \(\star\), the location of \(\star\) in terms of the Cartesian coordinates of its center is immediately known via Theorem 1. Each cell \(\star \in H_n\) is adjacent to either 5 or 6 cells called neighbors of \(\star\). The centroid and vertex children of \(\star\) are the centroid and vertex cells, respectively, of \(\star\) as defined by Properties 5 and 6 in Section 2. Call the cells at resolution \(n - 1\) that overlap \(\star\) the **parents** of \(\star\). Theorems 2 and 3 give simple rules, in terms of A3-coordinates, for the following basic procedures that are essential for global grid applications. Given the A3-coordinates of an arbitrary cell \(\star\) at any resolution,

- determine the neighbors, children, and parents of \(\star\); and
- perform local algebraic operations in the vicinity of the cell.
Theorem 2

\( A^3 \) addressing satisfies the following properties.

1. The neighbors of a cell in \( H_n \) with \( A^3 \)-coordinates \((a_1, a_2, a_3)\) have \( A^3 \)-coordinates \((b_1, b_2, b_3)\), where, for \( i = 1, 2, 3, \)

\[
|a_i - b_i| \leq \begin{cases} 1 & \text{if } n \text{ is even} \\ 2 & \text{if } n \text{ is odd.} \end{cases}
\]

2. The centroid child in \( H_{n+1} \) of \((a, b, c) \in H_n\) is

\[
\begin{cases} 3(a, b, c) & \text{if } n \text{ is even} \\ (a, b, c) & \text{if } n \text{ is odd.} \end{cases}
\]

3. Cell \((a, b, c) \in H_n\) is a centroid child if and only if

\[
\begin{align*}
& a \equiv b \equiv c & \text{if } n \text{ is even} \\
& a \equiv b \equiv c \equiv 0 & \text{if } n \text{ is odd.}
\end{align*}
\]

4. The vertex children of \((a, b, c)\) are, in either the even or odd case, the neighbors of the centroid child as given by item (1).

5. The parent in \( H_{n-1} \) of a centroid child \((a, b, c) \in H_n\) is

\[
\begin{cases} (a, b, c) & \text{if } n \text{ is even} \\ \frac{1}{3}(a, b, c) & \text{if } n \text{ is odd.} \end{cases}
\]

6. Let \( x = (a, b, c) \in H_n \) be a vertex child.

For \( n \) even, the cell \( x \) has exactly three neighbors \((d, e, f)\) with the property that \( d \equiv e \equiv f \). These three are the \( A^3 \)-coordinates of the parents of \((a, b, c)\) in \( H_{n-1} \). For \( n \) odd, the cell \( x \) has exactly three neighbors \((d, e, f)\) with the property that \( d \equiv e \equiv f \equiv 0 \). For these three the triples \( \frac{1}{3}(d, e, f) \) are the \( A^3 \)-coordinates of the parents of \((a, b, c)\).

Using the \( A^3 \)-coordinates, a local vector arithmetic can be efficiently implemented as follows. By “local” we mean centered at any cell of \( H_n \) and restricted to a single face of the octahedron. By “vector arithmetic” we mean usual vector addition and multiplication by scalars for vectors contained on a single face of the octahedron.

The octant of a triple \((a, b, c)\) of integers is the ordered triple of signs (+ or −) of the three entries. The octant of \((5, -2, 2)\), for example, is \((+, -, +)\). Two points of \( V_n \) lie on the same face of the octahedron if and only if their \( A^3 \)-coordinates are in the same octant. Let \( x_0 \) be the \( A^3 \)-coordinates of a fixed cell \( x_0 \in H_n \). Let \( x_1 \) and \( x_2 \) be the \( A^3 \)-coordinates of two other cells \( x_1 \in H_n \) and \( x_2 \in H_n \) in the same octant as \( x_0 \). Let \( v_1 \) denote the vector pointing from the center of \( x_0 \) to the center of \( x_1 \); similarly \( v_2 \) the vector pointing from the center of \( x_0 \) to \( x_2 \).
Theorem 3

1. With notation as given above, the vector sum \( v_1 + v_2 \) points from \( x_0 \) to \( x_1 + x_2 - x_0 \), where the sum of the \( x_i \) in the above formula is the usual addition in \( \mathbb{R}^3 \). The formula is valid as long as \( x_1 + x_2 - x_0 \) lies in the same octant as \( x_0 \).

2. For an integer \( k \), the scalar product \( k v_1 \) points from \( x_0 \) to \( k x_1 + (1 - k) x_0 \).

We have shown that each A3H cell can be represented as an ordered triple of integers called the A3-coordinates. An indexing scheme for A3H is now introduced for referencing the cells at any level of resolution \( n \) using a string of \( n + 3 \) “trits.” The balanced ternary is a base 3 positional number system using the digit set \( \mathbb{D} = \{-1, 0, 1\} \), with \( \mathbb{D} \) often referred to as the set of trits. The integer 8, for example, has the balanced ternary representation \( 8 = (1)(0)(-1) \). Relevant to our application are the following properties of the balanced ternary. Further information can be found in Knuth’s “The Art of Computer Programming” [9].

1. Every integer, positive or negative, has a unique representation in the balanced ternary. Moreover, every integer between \( -3^n / 2 \) and \( 3^n / 2 \) has a unique representation of the form

\[
\sum_{k=0}^{n-1} d_k 3^k,
\]

where \( d_k \in \mathbb{D} \).

2. The negative of an integer in balanced ternary is obtained by merely changing the sign of each digit.

3. Two integers in balanced ternary are congruent modulo 3 if and only if they have the same last digit. In particular, an integer is divisible by 3 if and only if the last digit in its balanced ternary representation is 0.

Given A3-coordinates \( (a, b, c) \), encode this triple as a string \( S \) of \( n + 3 \) trits as follows. The cases \( n \) even and odd are considered separately.

Even \( n = 2k \). The first \( k + 1 \) trits in \( S \) represent the integer \( a \). The second \( k + 1 \) trits represent the integer \( b \).

The third integer \( c \) is given by the formula \( c = \pm (3^k - |a| - |b|) \), where the \( \pm \) is the last trit in \( S \).

Odd \( n = 2k - 1 \). The first \( k + 1 \) trits in \( S \) represent the integer \( a \). The second \( k + 1 \) trits represent the integer \( b \).

The third integer \( c \) is given by the formula \( c = \pm (3^k - |a| - |b|) \), where the \( \pm \) is chosen to make \( a \equiv b \equiv c \mod 3 \).

The index function \( I : V_n \rightarrow W_n \) for A3H is now defined as follows. For a point \( x \in V_n \), let \( (a, b, c) \) be its A3-coordinates, and let \( S \) be the balanced ternary string encoding of \( (a, b, c) \) as explained above. Then define \( I(x) = S \).
It should now be apparent, from Theorems 1, 2, and 3 and the stated properties of the balanced ternary, that the essential operations on the cells of the octahedral A3H can be performed using elementary base 3 arithmetic applied to the indices.

5. Radix Indexing

The indexing function for A3H introduced in this section is based on tessellations of the regular icosahedron. Thus $T_0$ is the regular icosahedron; $H_0$ is the dodecahedron; and the first resolution $H1$ is the truncated icosahedron. The indexing is first defined on planar “tiles.” Thirty two such tiles are then assembled to form the icosahedral A3H. This indexing scheme is based on a 2-dimensional generalization of positional number systems for the integers. It is a fundamental result in number theory that if $b$ is an integer greater than 1, then every positive integer has a unique representation of the form

$$\sum_{i=0}^{N} b^i \cdot d_i,$$

where the digits $d_i$ belong to a set $D = \{0, 1, 2, \ldots, b - 1\}$. It was remarked in the previous section that every integer, positive or negative, has a unique representation of the above form if $b = 3$ and $D = \{-1, 0, 1\}$. This is the balanced ternary system. The set of integers is a 1-dimensional lattice. Analogous positional number systems exist for higher dimensional lattices. A 2-dimensional lattice is the set of all integer linear combinations of two linearly independent vectors in $\mathbb{R}^2$. The hexagonal lattice is the lattice generated by the vectors $(1, 0), (\frac{1}{2}, \frac{1}{2})$. Call a triple $(L, B, D)$ a 2-dimensional radix system if $L$ is a 2-dimensional lattice, $B$ is linear map ($2 \times 2$ matrix) such that $B(L) \subset L$, and $D$ is a finite subset of $L$ that includes the origin. The set $D$ is called the digit set and $B$ the base. A natural, but difficult, question is for which radix systems does there exist a unique representation of each lattice point $x \in L$ in the form

$$x = \sum_{i=0}^{N} B^i (d_i),$$

where $d_i \in D$. See [16] for partial answers and references. As is standard for the base 10 radix system $(\mathbb{Z}, (10), \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\})$, for any radix system we will use he abbreviated notation

$$x = d_N d_{N-1} \cdots d_1 d_0$$

instead of the form (1).

For the purposes of this paper, $L$ is the hexagonal lattice. The base is 3, more precisely $B$ is the linear transformation given by multiplication by 3. The hexagonal lattice $L$ is the set of centers of the hexagonal grid and, if no confusion arises, we will identify each lattice point in $L$ with the corresponding hexagon of which it is the center. For the set of digits, initially take

$$D' = \{0, 1, \omega, \omega^2, \omega^3, \omega^4, \omega^5, \omega + \omega^2, \omega^4 + \omega^5\},$$
these 9 lattice points shown as shaded hexagons in Figure 4, where in complex number notation \( \omega = \frac{1}{2} + \frac{1}{2} \sqrt{3} i \). Although the proof is omitted, this radix system \((L, B, \mathcal{D}^0)\) has the property that every point in \(L\) has a unique representation in the form given in equation (1) above.

Figure 4: Digit set

Next expand the digit set \(\mathcal{D}^0\) with 9 elements to the digit set
\[
\mathcal{D} = \{0, \omega^k, \omega^{k-1} + \omega^k : 1 \leq k \leq 6\}
\]
with 13 elements as shown in Figure 4. To simplify notation, for \(1 \leq k \leq 6\), make the following replacements:
\[
\begin{align*}
\omega^k & \leftarrow 0k; \\
\omega^k + \omega^{k+1} & \leftarrow k0.
\end{align*}
\]
Hence in the radix system \((L, B, \mathcal{D})\), each lattice point \(x \in L\) can be expressed as
\[
x = e_{2N+1} e_{2N} \cdots e_3 e_2 e_1 e_0,
\]
where each \(e_i \in \{0, 1, 2, 3, 4, 5, 6\}\). In the radix system \((L, B, \mathcal{D})\) the representation of lattice points in the form of equation (1), equivalently (3), is no longer unique. For example the lattice point \(1 + \omega\) (see Figure 4) can be expressed as either 0140 or 60 because \(0140 = (01)(40) = 3\omega + (\omega^4 + \omega^5) = 1 + \omega = \omega^6 + \omega^7 = 60\). We have used the identities \(\omega^6 = 1\) and \(1 + \omega^2 + \omega^4 = 0\), which can also be used to prove the following theorem.

Theorem 4

In the hexagonal radix system \((L, B, \mathcal{D})\) every element of \(L\) can be written uniquely in the form (3) with the restriction that there exist no two consecutive non-zero digits, i.e., between any two non-zero digits there is at least one zero.
In analogy to decimals, next introduce the notation \( .e_1 e_2 \cdots e_{2N} - 1 \) \( e_{2N} \), where \( e_i \in \{0,1,2,3,4,5,6\} \) with no two consecutive non-zero digits. Before the replacement (2) this is \( .d_1 d_2 \cdots d_N \), where \( d_i \in D \), which stands for

\[
x = \sum_{i=1}^{N} 3^{-i} \cdot d_i, \quad d_i \in D.
\]

Define \( P_n \) to be the set all points in the plane of the form \( .e_1 e_2 \cdots e_n \) where \( e_i \in \{0,1,2,3,4,5,6\} \) with no two consecutive non-zero digits. If \( n \) is odd, the implicit \((n + 1)^{\text{st}}\) digit is 0. For ease of notation, the decimal point will be dropped. For example, 105 represents the point in the plane:

\[
(10) \frac{1}{3} + (50) \frac{1}{9} = (\omega + \omega^2) \frac{1}{3} + (\omega^5 + 1) \frac{1}{9} = \frac{1}{9} (5 \omega - 1).
\]

The set \( P_n \) is a subset of points of an hexagonal lattice, the larger the \( n \), the finer the lattice (points closer together). The integer \( n \) will be referred to as the resolution of \( P_n \). Figure 5 shows two resolutions \( P_1 \) and \( P_3 \). (For clarity, the set \( P_2 \) is omitted from the figure, but each hexagon of \( P_2 \) has area in the ratios 1/3 and 3 to the hexagons in \( P_1 \) and \( P_3 \), respectively, and each hexagon of \( P_2 \) is rotated by 30° relative to those in \( P_1 \) and \( P_3 \).) For a given hexagon (or corresponding lattice point), the string \( e_1 e_2 \cdots e_n \) will be referred to as its index. If \( W_n \) is the set of all strings of length \( n \) with elements from the set \( \{0,1,2,3,4,5,6\} \), no two consecutive non-zero, then define

\[
I : P_n \rightarrow W_n
\]

by \( I(x) = e_1 e_2 \cdots e_n \) for \( x \in P_n \). The indices are also shown in Figure 5. The next result lists properties of the sets \( P_n, n \geq 0 \). (Note that \( P_0 \) consists just of the origin.)

**Theorem 5**

The set \( P_n, n \geq 0 \), satisfy the following properties.

1. The \( P_n \) are nested: \( P_0 \subset P_1 \subset P_2 \subset \cdots \).

2. The ratio of the area of a \( P_n \) hexagon to the area of a \( P_{n+1} \) hexagon is 3.

3. The hexagons in \( P_n, n \) even, are oriented a 30° rotation to the hexagons in \( P_n, n \) odd.

4. The number of hexagons in \( P_n \) is \( \frac{1}{2} (3^{n+2} - (-2)^{n+2}) \).

5. \( P_n \) has 6-fold rotational symmetry about the origin.

**Tree Data Structure**

The data structure underlying the sets \( P_n, n \geq 0 \), has the form of a tree \( T \). The set of nodes of \( T \) is \( \{P_n : n \geq 0\} \). The nodes at depth \( n \) are the cells at resolution \( n \). Each node \( x \) at depth \( n \) has 7 children, one at level \( n + 1 \) and six at depth \( n + 2 \). The
child $x_0$ at depth $n + 1$ is the centroid cell (as defined by Property 5 in Section 2) of $x$ while the six children at depth $n + 2$ are the vertex cells of $x_0$. Note that the children and parents of a cell as defined in Section 4 do not correspond exactly to the children and parent of a cell in $T$. The basic idea here is that each cell $x$ at resolution $n$ generates

- one centroid child $x_0$ at resolution $n + 1$ with the same center as $x$, and
- five or six vertex children at resolution $n + 2$ centered at the vertices of $x_0$.

It can be shown that the indexing is closely related to the data structure. If $\alpha \in W_n$ is the index of a hexagon of $P_n$ at resolution $n$, then

- $\alpha 0 \in W_{n+1}$ is the index of its centroid child, and
- $\alpha 0k \in W_{n+2}$, $1 \leq k \leq 6$, are the indices of its 6 vertex children.

**Face tiles and Vertex Tiles**

The sets $P_n$ will be referred to as a face tiles. The construction is now altered slightly to obtain a vertex tile. For each $1 \leq k \leq 6$ construct a set $P_{n,k}$ as follows. The set $P_n$ can be partitioned into 6 subsets each one approximately a $60^\circ$ wedge of $P_n$. More precisely, for $1 \leq k \leq 6$, let $P_n^{(k)}$

$$P'_{n,k} = P_n \setminus P_n^{(k)}.$$
Given an index $\alpha$, let $\alpha^+$ denote the result of adding 1 (modulo 6) to every non-zero digit in $\alpha$. Now construct $P_{n,k}^*$ as follows. Remove the edge in $P_{n,k}^*$ that is contained only on the hexagon centered at the origin. The hexagon at the origin now becomes a pentagon. Identify, in pairs, the other edges in $P_{n,k}^*$ that are contained on exactly one hexagon of $P_{n,k}^*$ as follows. Identify each such edge common to hexagons with indices $\alpha$ and $\beta$ in $P_n$ with the edge common to hexagons $\alpha^+$ and $\beta^+$. Whereas each $P_n$ possesses 6-fold rotational symmetry about the origin, each $P_{n,k}$ possesses 5-fold symmetry. Geometrically the six sets $P_{n,k}$, $1 \leq k \leq 6$, are identical, but their indexing is not. With the indexing ignored, denote each $P_{n,k}$ simply by $P_1^*$. While $P_n$ is planar, $P_n^*$ is not. The title $P_n^*$ called a vertex tile.

**Indexing the Icosahedral A3H**

We are now in a position to define the radix indexing on the isocahedral A3H. It is based on the following theorem, which is illustrated in Figure 6 depicting the 20 faces of the icosahedron flattened onto the plane.

**Theorem 6**

For each $n \geq 1$, the $(n+1)^{st}$ resolution $H_{n+1}$ of A3H is the non-overlapping union of 20 copies of face tile $P_{n-1}$, each such tile centered at the barycenter of a triangular face of the icosahedron, and 12 copies of vertex tile $P_n^*$, each such tile centered at a vertex of the icosahedron.

The indexing at resolution $n+1$ on A3H is inherited from the indexing on copies of tiles $P_{n-1}$ and $P_n^*$. Label the vertices of the icosahedron $U_k, U'_k$, $1 \leq k \leq 6$, where $U_k$ and $U'_k$ are antipodal to each other. Then at vertices $U_k$ and $U'_k$ of $H_{n+1}$ provide the indexing of patch $P_{n,k}$. Label the faces of the icosahedron $F_k$, $1 \leq k \leq 20$. At face $F_k$ of $H_{n+1}$ provide the indexing of patch $P_{n,k}$. There still remains some ambiguity because of the possible rotations of the patches about their centers (6-fold symmetry of face tiles and 5-fold symmetry of vertex tiles). In practice the vertex tiles are labeled A-T and the face tiles 1-12, their relative positions given by a table.

**6. Conclusion**

This paper concerned the display and processing of global data on a discrete global grid, in particular, on the aperture 3 hexagonal discrete global grid A3H. After a brief geometric description, the focus was on the problem of indexing the cells of A3H. Two indexing schemes were introduced, barycentric indexing on the octahedral A3H and radix indexing on the icosahedral A3H, both leading to efficient processing.
Figure 6: Unfolded Icosahedron

References


Information Quality and Information Systems Success

Janet Aisbett, Faculty of Science and Information Technology, The University of Newcastle, Industrial Drive, Newcastle, 2300, Australia

Information systems (IS) failure has been a concern of IS practitioners and researchers for more than 40 years. Those years have seen continuing change in the technology and application of computerized IS. The question arises as to whether failure to achieve expectations of stakeholders is because, or in spite, of the changing environment. This paper reviews key models of IS success in the context of the prevailing IS. It identifies information quality as a constant factor in the models, and poor content quality as a continuing contributor to perceptions of failure. We conclude that fundamental research into information quality is still needed, and provide an example of such investigations concerning new forms of representation.

Introduction

High levels of dissatisfaction amongst information systems (IS) users and owners have been reported in the hundreds of academic studies into IS success and failure published over the last forty years (eg Lyytinen & Hirschheim, 1987; Byrd et al, 2006). Over this time, a range of approaches have been taken to identifying the factors which contribute to success of operational IS, including descriptive case studies (Sauer, 1993), surveys (Byrd et al, 2006), literature reviews (van der Meijden et al, 2003) and theory building (Delone & McLean, 1992). IS evaluation remains an active research area around the world, with, for example, 30 journal articles published on success or failure in the last 5 years (Web of Science, 2006).

The history of computerized IS has been one of innovation and rapid growth in scope and complexity. High perceptions of failure may have persisted because, as existing problem areas are solved, new problems emerge associated with new features enabled by technology; for example, local area networking leading to document sharing and problems with information consistency, or browser technology and the internet leading to business websites and problems with information currency. The new functionality causing the problems may have motivated extensions to success models, and the primary cause of perceived failure of IS may be sourced to these extensions.

It is therefore useful to identify factors that have not changed in key models in successful use of IS over the years, as well as those that have. If the constant factors are significant contributors to IS failures, then failure cannot be attributed only to the...
pace of change. This paper outlines the historical development of IS success models in the context of the prevailing IS, identifies information quality as a constant factor and considers the implications for IS researchers.

The paper is organized as follows. Sections 2 and 3 review the development of models of successful IS use, from their initial accounting focus to their incorporation of, first, the responses of the users of the systems, then the goals of the organizations that funded them, then broader notions of user and business benefit. Section 4 observes that content quality—the quality of the information accessible from the system—has been a stable factor in IS success models, and a persistent source of failure. An underlying problem in providing quality information is ambiguity in structured and unstructured data, and therefore this section touches on our investigations into more expressive representational forms.

**Evolution of IS and the Effect on IS Success Models**

Twelve thousand years ago the Sumerians used clay tokens of varying shapes and sizes to denote amounts of specific products, such as number of jars of oil (Robson, 1992). These tokens provided records of goods received and exchanged, and were used for taxation as well as trading purposes. With time and increasing sophistication of both the traded goods and the taxation systems, tokens eventually were replaced by clay tablets containing records of multiple transactions. The legacy of these tablets remains in the tabular databases which continue to be a major part of IS.

Foreign trade eventually led to another important development in accounting practice, double entry bookkeeping, which was promoted in Renaissance Italy to protect joint ventures partnerships conducting foreign trade and to protect overseas branches of Italian banks against fraud (Robson, 1992:692). Double entry accounting allowed the commercial activity of a firm to be reconciled over time, and provided a technique to reduce arithmetic errors in accounts. The resulting bookkeeping required skilled workers to perform the arithmetic calculations, and, from Pascal’s inventions in the mid 1600s, savings in time and cost motivated development of mechanical and, later, electronic adding machines.

The dominant criterion for success in early computerized systems was therefore return on investment (ROI). ROI could be calculated simply as the savings in the cost of employing human “computers” less the cost of installing and running the machines (Chase, 1984; Lay, 1986), assuming a known time to write-off the system. However, mechanical computers suffered from reliability problems and then, at the start of the modern electronic era, programming errors (Gigerenzer, 2004), so information quality was also a recognised factor in their successful use.

Expanded functionality was incorporated into computer systems as peripherals and compute power improved. In the bespoke systems of this time, users of the so-called “data processing” systems had a major role in defining what functionality was needed, as part of the requirements analysis phase enshrined in IS development methodologies. Logically, the views of users should therefore be important in IS evaluation. Bailey and Pearson’s methodology for measuring what they called User Information Satisfaction
(UIS) (Bailey & Pearson, 1983) was a landmark in IS success models. Their evaluation tool asked users to record their reactions to a large set of factors contributing to success, including attitudes of support staff and management, but also covering classical information quality parameters such as accuracy, timeliness, precision and relevance.

Despite criticism that Bailey and Pearson’s instrument mixed causes as well as consequences amongst the factors (Ives, Olson, & Baroudi, 1983; Doll & Torkzadeh, 1988), the concept of UIS was incorporated into many success models and evaluation methodologies. See Zviran and Erlich (2003) for a review and McHaney, Hightower and Pearson (2002) for an example from outside the OECD.

The commitment of senior management to the systems had long been identified as a key factor in success (eg. Micallef, 1996), not least because the business case for new systems often had to be made to senior decision makers to get funding approval. Such a case could initially be made in terms of savings, in the traditional ROI framework. But as functionality was introduced that extended rather than replaced existing functions, the business case had to be made against intangibles such as new opportunities for the firm, customer perception or staff morale (Edwards, Ward & Bytheway, 1995). The need to evaluate an operational system in terms of its success in achieving the business goals used in making the initial case for its development was recognized in some of the debate arising from Bailey and Pearson’s UIS, and achievement of business benefits became explicit in discussion of IS success (Strassman, 1990). As typical examples, Saarinen (1996) and Mirani and Lederer (1998) developed evaluation tools that required estimation of the impact of a new IS on the organization, through estimating the impact on aspects such as efficiency, profitability, decision making, work processes and competitive advantage.

Figure 1. IS success model, from Delone and McLean (1992). See text.

Information quality remained an explicit factor in the success models. For example, Mirani and Lederer saw it as being part of the informational benefits that formed organizational benefits, and Saarinen saw it as part of IS quality.

This accords with survey results on IS use failures; for example, Lyytinen (1988) identified failure in IS use as due in part to difficulties in maintaining error free information and the users’ difficulties in interpreting it.
An influential causal model of success was proposed by Delone and McLean (1992) (summarized in Figure 1) as a synthesis and reduction of existing models. This model culminated in a dependent variable which the authors called organizational impact. System quality referred to features of the supporting software which contribute to how the system was used and satisfaction with that use. The individual impacts were things like effectiveness on the job, decision making performance and quality of work. Information quality appeared in this model as an independent variable contributing alongside system quality to the use, and satisfaction with the use, of the system.

Changing Nature Of Is In The 1990s

Despite increasing sophistication in the models of IS success, and in the associated pre- and post-development evaluation approaches, the literature reporting dissatisfaction with IS continued (e.g. Beynon-Davies, 1999, Olsen, 1999, Poon & Wagner, 2001). The 1990s was an era of rapid changes in the nature of IS and the environment in which they operated. For a start, it was an era of globalisation. More firms were looking for suppliers and markets across the globe, and used joint ventures, partnerships or mergers to tap local knowledge. Corporate structures were more fluid than they had been in earlier decades. Efficiency, including elimination of duplication and delay, was seen as critical for business success in a highly competitive environment (eg. Porter, 1998).

An initial response to the desire for efficiency was to develop monolithic IS internal to the organisation. Enterprise Resource Planning (ERP) systems grew from merely being integrated IS for manufacturing into, first, systems that integrated operational data such as inventories with accounting and human resource data, and then systems with specialized software suites for various industry sectors (Sprott, 2000). As the software involved was very complex, a few large vendors such as SAP and Peoplesoft emerged to dominate the market. Unfortunately, high failure rates continued to be reported. While some ERP failures were attributed to the costs and complexities of implementation, they were also due to on-going problems with lack of the flexibility needed to meet changing organizational requirements, and quality problems with the information supplied, including representational misfits (eg Soh et al, 2000; Olsen, 1999).

A supply chain emphasis meant that data that had traditionally been quarantined within an organization was increasingly shared with buyers and/or suppliers. Such changes affected PC-based IS in smaller firms which had to provide compatible data to their larger suppliers or buyers. Once-only data entry, which was designed to reduce duplication of effort and improve information quality through fewer errors and better currency and consistency, raised issues of ownership, quality (because data could be supplied externally to the organization) and interpretation (eg. Nicholson, 1994). Nevertheless, linking along supply chains represented an important change in IS, as no longer could the user of the system be assumed to be an internal member of the firm or organisation.

As the original ERP market of large organizations became saturated, vendors turned to smaller sized businesses, and were faced with the need to reduce implementation
costs and development time (Kumar & Van Hillegersberg, 2000). For those taught
the conventional IS approach, the solution again was radical: change the organization
that purchased the software to fit the software, rather than change the software to suit
the organization. (Of course, this had been happening for a long time with smaller
software applications, especially for personal computers.) Vendors justified the need for
the organization to change by claiming that the software enforced business processes
that were “best practice”. However, local factors such as cultural differences made the
notion of a universal best practice solution questionable, even without taking into
account environmental changes over time that affect what is the best practice (Kumar

ERP systems evolved to accommodate the Internet, and to become less monolithic in
their data structures (Sprott, 2000). The introduction of the internet along with browser
technologies altered all systems, not just ERP systems, improving communication
amongst internal users as well as further opening up organizational IS to users who were
not members of the organization. The potential to connect a broad customer base to
organizational information, including databases, and also to collect information from
them opened the way for retail engagement with an organization's IS.

Development of vocabularies and basic domain models to support sharing of
structured data across IS has become a major research efforts in many domains -- health,
environment, business, scientific subdisciplines and so on. The term “ontologies” has
been corrupted to describe these information structures describing domains of interest
(Zuniga, 2001), and their development has been assisted by the rapid penetration of
XML as an efficient and easy to understand description language for the structures.
Ontology development, however, faces the same problems underlying the monolithic
ERP approach. Specifically, there are problems of mis-matched modeling of concepts
and terminology at the periphery of domains (for example, when nursing applications
use an ontology developed for medical domain); and there are problems with change
over time. Multiple ontologies may be available to describe the required subject matter,
but while all may cover some of the domain of interest (in different ways) none may
be complete. The evaluation of ontologies and of ontology builders is a subset of IS
evaluation focusing on quality of information, but research in this area has tended to be
independent (eg. Duineveld et al, 2000).

These changes in the use of IS further changed the way that IS success should be
viewed. DeLone and McLean (2003) produced an update of their success model, which
took account of the voluntary aspect of use of IS by customers with an attitudinal
factor intention to use, alongside the behavioural use factor which described how the
system was actually used. The new model also substituted the concept of net benefits
to the owner or sponsor of the system for the complex concept individual impact
on which organizational benefits had been dependent. Net benefits were described
very broadly by the authors, as including impacts on customers, employers, suppliers,
markets, industries and so on. Finally, service quality, meaning the overall support
from the service provider, was added as a base factor alongside information quality and
system quality.
Information Quality as a prerequisite for IS success, and the limitations of Data Formats

Information quality has been a factor in IS success models since their inception. Despite this, and despite being an important driver of developments such as ERP systems, information quality has been relatively neglected by IS researchers. A notable exception over the last 15 years has been the research due to Richard Wang and colleagues at the Sloan School.

Figure 2. A decomposition of the complex concept of Information Quality, in which information is viewed as both a product and a service. From Kahn, Strong and Wang (2002)

<table>
<thead>
<tr>
<th></th>
<th>Conforms to Specifications</th>
<th>Meets of Exceeds Consumer Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Product Quality</strong></td>
<td>Sound Information</td>
<td>Useful Information</td>
</tr>
<tr>
<td></td>
<td>• Free-of-Error</td>
<td>• Appropriate Amount</td>
</tr>
<tr>
<td></td>
<td>• Concise Representation</td>
<td>• Relevancy</td>
</tr>
<tr>
<td></td>
<td>• Completeness</td>
<td>• Understandability</td>
</tr>
<tr>
<td></td>
<td>• Consistent Representation</td>
<td>• Interpretability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Objectivity</td>
</tr>
<tr>
<td><strong>Service Quality</strong></td>
<td>Dependable Information</td>
<td>Usable Information</td>
</tr>
<tr>
<td></td>
<td>• Timeliness</td>
<td>• Believability</td>
</tr>
<tr>
<td></td>
<td>• Security</td>
<td>• Accessibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Ease of manipulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reputation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Value-added</td>
</tr>
</tbody>
</table>

Typically, information quality is represented as a multidimensional concept, with dimensions such as reliability, timeliness, accuracy, completeness and so on, some of which are intrinsic to the information and some task-dependent (Wang 1998) Figure 2, reproduced from Kahn, Strong and Wang (2002), structures such a list of attributes in a 2x2 framework. First, the structure differentiates between information as a product (where data is in some sense viewed as the output of a process that “manufactures” it from raw input) and as a service, where the focus is on how data are obtained and used as information -- which brings in computer and communications system characteristics such as reliability and speed (Zeithaml, Berry & Parasuraman, 1990).

According to Kahn, Strong and Wang, when information conforms to a priori specifications to do with its characteristics as a product, it is sound, and when it conforms to specifications of a service, i.e. to do with its “consumption” for a particular purpose, then it is dependable. Sound information, depicted in the top square of Figure 2, covers the accuracy, consistency and completeness sought by doubly entry bookkeeping,
which, improved information quality on its introduction to accounting in the 14th and 15th centuries. The right hand column in this figure identifies factors to do with what the user of the information expects: this clearly influences their later satisfaction with the IS that holds and supplies the information. Key to quality information as product is how well the content of the data suits users’ purposes. As a service, users’ expectations concern how well the data are delivered to suit their current purposes.

The traditional form of the information product is structured, and most structured information is still presented in the tabular format employed in pre-computerised bookkeeping where trading activities and relevant stocks, including cash and inventory items, were modeled as alphanumeric vectors (records). This form of representation was not unique to commerce, of course. Descriptions of complex concepts using sets of attributes taking numeric or alphanumeric values are common in science, and, as Robson (1992) suggests, derive from Platonian belief in the representational power of numbers.

However, limitations of the structured representations of early computer systems are evident in the rapid uptake of systems providing unstructured information and easy interaction which started to become available in the 1980s with software such as Lotus Notes supporting document exchange, annotations, hypertext and later hypermedia (eg. Tung et al, 2000). Technology has continued to improve access and delivery mechanisms (through search engines, broadband networks and so on), and enhanced the service view of information by raising users’ expectations.

The product view has been the traditional view, and therefore most pertinent to our discussion of constant factors in IS success. From this perspective, consider the example of a person wanting to purchase a pen over the Internet. They may be able to interrogate conventional record descriptors of the make, colour, cost and so on of the various pens on offer; however, an ink colour described as “red” might well vary from reddish-violet to reddish-orange. Prospective customers may also view pictures of pens, even three dimensional models of them, and possibly look at examples of the pens’ writing which reproduce ink colour and other characteristics (distorted possible by the viewer’s display device). But how can concepts be conveyed such as the feel of the pen or the flow of the ink, which might be critical criteria in the purchase decision? In structured representations, feel might be selected from a coarse grained value set such as “heavy weight”, “medium weight” and “light weight”, which are very open to ambiguity. Even verbal descriptions are ambiguous when conveying subtle meanings such as these.

The quality of the information available fails in its soundness, and in expectations of the virtual shopping experience being able to provide the same decision information as is available in a conventional shopping experience. In terms of the attributes summarized in Figure 2, the information lacks completeness (having the depth and breadth to convey concepts of feel and so on) and relevance (being helpful for the purchasing task). Lack of completeness leads to ambiguity.

The problem of producing an information product which can capture more of the properties of a real-world object such as a pen is a long term research question. Our research in this area (eg Aisbett & Gibbon, 2005, Aisbett &
Gibbon 2001) has argued that information structures which assume a finite number of dimensions (no matter how large that number) are likely to fail to convey the subtlety of concept instances, especially sensory or emotional ones. We proposed that many attributes conventionally represented by numerical values or on Likert scales are better represented by numerical functions defined on a continuum, for example, as a real-valued function defined on a plane, or equivalently, as grey scale images on a subset of the plane.

As illustration, consider the representation of two real world objects, a red cube and a green cone made of plasticine, depicted in the small images in the left of Figure 3. In conventional IS these objects would be described in tables of attribute values, possibly also by photographs such as the ones at the left edge of the figure, and possibly also by textual description. In an experiment reported in Aisbett and Gibbon (2003), people were asked to describe these objects using attributes prescribed by the experimenters, namely, \textit{texture \\& feel}, \textit{size \\& weight}, \textit{hue}, and \textit{shade}. Figure 3 shows the sets of attribute values chosen by two different people for the same two objects. The set of values had to be chosen from a family of functions (images), of the types shown in the larger images to the right of the object photographs in this figure. The four values selected by a person to describe each object were arranged as quadrants of a composite image, with \textit{size \\& weight} and \textit{hue} on the top row, and \textit{shade} and \textit{texture \\& feel} on the bottom row. The outline of the hand in the depiction of \textit{texture \\& feel} was used to help record the \textit{feel} of the physical objects, which the people handled prior to making their selection, while \textit{texture} was coded through patterns on the “pressure points” depicted as dark blobs within the hand silhouette.

This is early work into enriched representation, which we have been applying to automated recognition systems. The work has potential application in extending conventional IS, including use of such systems by people who cannot read, or even by animals which can be trained to recognize patterns, such as pigeons.
Conclusion

This paper has looked at the continuing issue of successful IS use, against the background of the changing commercial and technological environment. We first looked at pre-computerised and early computer systems to explain why IS success models were initially based on accounting principles, with an emphasis on ROI and error free information. The rise of user involvement in design was reflected in the incorporation of user satisfaction in IS success models. As business benefits brought by systems included more intangibles and less direct ROI, achievement of the goals of the organization that funded the system became an explicit part of the models. When systems started to deliver information to users outside the organizational boundaries, IS success models had to account for more complicated notions of user satisfaction and business benefit.

Innovation in technology has led to many changes in IS, notably affecting the availability of unstructured information (documents) and the choice and power of delivery mechanisms. Case studies, surveys and literature reviews into both successful and unsuccessful applications of the new technologies have been necessary to develop useful success models which can help organizations effectively employ the new technologies. A large body of IS research investigating electronic commerce applications in many countries has resulted from these approaches.

We observed that the quality of the content delivered by IS, that is, information quality, was a constant of IS success models. Moreover, this factor contributed to recorded failures. We conclude that IS failure cannot only be attributed to changing technology and usage. Therefore, although IS research is needed into the effects of new technology and new applications such as web sites in electronic commerce, there remains a need for fundamental research. An underlying problem in providing quality information is lack of completeness which leads to ambiguity in structured and unstructured data. Our investigations into more expressive representational forms are examples of the long term research still needed in IS.

References


A System Dynamics Tool for Evaluating IT Investment Projects

Paul Ssemahulu and Ddembe Williams Faculty of Computing and Information Technology, Department of Information Systems, Makerere University

Systems dynamics modeling is the technique of constructing and running a model of an abstract system in order to study its behavior without disrupting the environment of the real system. The process simulated in this study, that is, evaluation of IT investment projects, is one of increasing importance as it has been observed that 90% of all senior IS managers have no idea how to determine the value of information systems. In an era where cost overruns and cancelled projects cost millions of dollars, a simulation tool which can rapidly calculate the benefits to be derived from an information system can be very useful. The existing literature identifies noticeable gaps between academic theories, commercially available methodologies and actual evaluation practices promoted by organizational rules and structures, informal practices implemented by stakeholders and academic recommendations which are not used in practice. Problems relating to evaluation of IT investment projects were established. Understanding these problems would in the long run reduce losses due to failed IT investments. In this study, five different methodologies were investigated taking into account the suitability or goodness of the framework, bias, focus and complexity. The System Dynamics Methodology was found to be the best as the others had serious shortcomings. A model derived from earlier work by Seddon et al., as well as Delone and Mclean was used to construct a dynamic hypothesis that helps to realize the interrelationships between the critical variables. A causal loop diagram derived from the dynamic hypothesis was also constructed. A simulation tool for evaluating IT investment projects was developed to help managers cut down on time spent debating investment decisions, cut down on costs, reduce information overload and help researchers evaluate related problems. The simulation tool was used to analyze how different variables interact to affect the total benefits of an information system. It was observed that only a strong interaction of people, information, and technology can improve business performance, and consequently lead to Information Systems success.

Introduction

Increased competition and the global economic aspects have forced companies to cut costs significantly, through the empowerment of lower-level staff and the removal of middle management. These trends are increasingly supported by developments in information technology (IT) (Serafeimidis and Smithson, 1998) The existing literature (Willcocks and Lester, 1994; Ballantine et al., 1995; Ward et al., 1996; Farbey et al.,
1999; Serafeimidis and Smithson, 2000 identifies noticeable gaps between academic theories, commercially available methodologies and actual evaluation practices promoted by organizational rules and structures, informal practices implemented by stakeholders and academic recommendations which are not used in practice.

Many organizations are experiencing difficulties with evaluating information systems. It is further indicated that more than half of the organizations do not apply any formal evaluation at all (Willcocks, 1996; Dhillon, 2000). And it has been shown that 90% of all senior IS managers have no idea how to determine the value of information systems (Dhillon, 2000).

Many researchers have stated problems and suggested solutions related to evaluation of IT investments and IS projects, notably (Hirschheim and Smithon, 1988; Orlikowski and Crash, 1994; Farbey et al., 1995; Dhillon and Backhouse, 1996; Irani and Love, 2001), but they are not appropriate for a developing country like Uganda, where managers are not comfortable with information Systems (Dhillon, 2000). Although there are already other studies in IT investment evaluation, they are not customised to support an individual manager at the desktop. Others are overly complex in nature and require experts to operate them. It is impossible to design a perfect Decision Support system that would be both highly pre-customised and highly customizable (Gachet and Haettenschwiler, 2003). The growing dependence of organisations on IT is viewed by many as a source of uncertainty. It is difficult to identify and measure the potential benefits and costs of an IT investment (Serafeimidis and Smithson, 1998). It is also true that IT evaluation is complex and elusive (Hirschheim and Smithson, 1988; Dhillon, 2000). According to Dhillon (2000), a phenomenal amount of money is lost because of inability of organisations to realise IS/IT benefits. Figures coming from the United States suggest nearly $ 59 million in cost overruns and some $ 81 million in cancelled IS/IT projects (Johnson, 1995). In Uganda, IT investments seem to have run into difficulties with the latest being the Electoral Commission Voters’ Register System. It was hatched in 1995 but to date, its benefits have not been realised. Executives often face information overload when they have to decide whether to invest in new IT projects or when management demand an audit of a completed project. The existence of a simple and easy to use tool for evaluation would greatly reduce this information overload.

**IT Investment Evaluation**

The essential questions are how and when to evaluate IT investments. There are three interrelated questions. How does information technology (IT) improve business performance? How do we decide the IT projects in which to invest? How do we assess the performance of systems after their implementation? The use of different evaluation techniques to answer this question varies from organisation to organisation. Research into the use of these techniques and their value to different organisations provides varying responses (Serafeimidis and Smithson, 2003; [69], Berghout, 1997; [10]). Earl (1989), stresses that not all organisations face an identical challenge, their business sectors differ, the competitive forces they combat vary, their histories are not alike and
they make different strategic choices. In addition, organisations must evaluate where in their evolution of IT developments they stand so as to ensure that they are able to make and manage the appropriate degree of strategic change. Farbey et al.,(1993) argue that the search for a single technique for evaluating investments in IT is fruitless. The range of circumstances that one technique would have to be applied to is so wide that no single technique is likely to be applicable. Berghout (1997) concludes that a mixture of both qualitative and quantitative methods should be used.

Seddon et al- IS Effectiveness Matrix (Seddon, 1998) was a clear breakthrough. In their presentation, they argued that a large number of IS effectiveness measures can be found in the IS literature. What is not clear in the literature is what measures are appropriate in a particular context.

The traditional evaluation methods and their shortcomings are discussed below (Berghout, 1997; Andresen (1999); (Kennedy, 1999; Adler, 2000):

To guide management decision making in the Information Systems Investment Appraisal process a number of “traditional” investment appraisal techniques, based on financial management techniques are normally employed. As commonly used these “traditional” investment appraisal techniques such as Payback, Accounting Rate of Return [ARR], Net Present Value [NPV] and Internal Rate of Return [IRR], are not able to measure many of the benefits offered by IS investments that are intended to gain tactical or strategic business advantages. This is a particular problem with those projects designed to achieve a ‘transformation’ of the business processes

We may conclude that each of these methods does have a shortcoming as far as ease of use, bias and objectivity are concerned. This shows that a common platform is needed for managers to have confidence in the decisions they make (Berghout, 1997; Andresen 1999; Adler, 2000)

Methodology

Nowadays, more modeling and analysis techniques often based on simulation or work flow techniques are used in predicting the effects of change. System dynamics is a method for analysing the behaviour of any kind of system: biological, physical, sociological, economic, and others. It provides a high level view of the system emphasising the interactions between its constituent parts, as well as the impact of time on its dynamic behaviour (Hustache et al., 2001). It is also a technique of constructing and running a model of an abstract system in order to study its behaviour without disrupting the environment of the real system. Simulation is the process of forming an abstract model from a real situation in order to understand the impact of modification and the effect of introducing various strategies on the situation (Williams, 2004).
The study followed the methodology as espoused by Williams (Williams, 2004), Fig.2; where he advocates simulation as one of the main strategies for this kind of research.

Model Subsystems
The simulation tool had three subsystems, that is the Information System, Users and Benefits.

Information System
The key variables for the Information System are: Rate of Information Quality, User Satisfaction, Rate of System Quality, and Service Quality. The information system sector provides intention to use, user satisfaction, information quality as well as system and service quality to the users sector.

User Interaction with the Tool
The user opens the tool by double clicking on it. When the tool opens, the graphical interface for running the simulation also opens. The user can click on the question marks of each sector to view the documentation about that particular sector. By using the navigation keys, she can navigate between the interface and the stock and flow part. In case the user wants to change the variables to run in the experiment, she simply double clicks the graphical interface, which allows her to change them. Clicking on the Run button on the tool starts the experiment.
Users
The key variables for the Users sector are: Individual Benefits, Organisational Benefits, and Society Benefits. The inputs to the Benefits Sector are the Individual Benefits, the Organisational Benefits, and Society Benefits. The users generate individual benefits, society benefits, and organisational benefits to the benefits sector.

Benefits
Benefits are fed back to the information and users sectors as feedback, thus bringing in complexity.

The Dynamic Hypothesis

The total benefits change over time as a result of variations in system quality and information quality. These affect the users’ intention to use the system and customer satisfaction. The result is that individual, society and organisation benefits are influenced which again affects total benefits in another feedback loop. A change in total benefits has a direct effect on customer satisfaction and users’ intention to use that system. This is what is captured in our dynamic hypothesis in Fig. 3.

The initial IT investment feedback structure (Fig. 3) contains seven dominant feedback loops all of which are Reinforcing loops (R). An increase in Information Quality leads to an increase in System Quality, giving rise to an increased Intention to use the system by the users. On the other hand, an increase in System Quality gives rise to increased Customer Satisfaction, which in turn increases the Intention to use the system by Users.
Increased Customer satisfaction increases both Society benefits and Organizational benefits. These serve to increase the Total Benefits giving rise to an increased Intention to Use the system by the Users and increased Customer Satisfaction.

**Simulation Experiments**

A number of simulation experiments were carried out as seen from two of them in Fig. 4 and Fig 5. As the graph below shows (Figure 4), both the total benefits and user satisfaction rise slowly up to around 10 months where the two intersect and user satisfaction increases at a higher rate than total benefits. Intention to use increases slowly up to 15 months where it is affected by changes in individual benefits. On the other hand, individual benefits are sporadic but low up to 15 months when they rise and increase sporadically. From Figure 4 we can infer that the changes in user satisfaction and total benefits are affected by feedback that affects the changes in individual benefits.

**Fig 4.** Total Benefits as a function of User Satisfaction, Intention to Use and Individual Benefits.
As the graph below shows (Figure 5), total benefits and user satisfaction rise at a slow rate giving an almost straight line from the start up to 13 months, as opposed to the other variables. It is observed that these are spurred on by an initial rate of system quality, which also increases gracefully up to 18 months. Around 18 months, the rate of system quality suffers from sporadic increases, which could be the reason why the rate of system quality suffers from some instability at this time. Information system success in this experiment is very stable and rises predictably. From Figure 5 we can infer that this scenario captures the most important elements related to information system success.

**Discussion Of Findings And Contribution**

Shortcomings in evaluation of information systems were identified. Critical variables in information system evaluation were identified for inclusion in a simulation-based tool. Five different research methodologies were investigated taking into account the suitability or goodness of the framework, bias, focus and complexity. The Systems Dynamics Methodology was found to be the best as the others had serious shortcomings.

The literature review highlighted the fact that depending on the background of the manager, they would use a different evaluation method. This means that there is no common platform or communication language for managers to communicate their decisions and the reasons as to why they arrived at particular decisions. This study made an attempt to provide that common platform by presenting a Systems Dynamics Tool for evaluation of information systems.
The Simulation tool runs with a maximum of five variables. When the model is run it shows us a peep into the future of what our policies and actions can do without taking the real risk of trying them out in the real world. These can be changed until a satisfactory mix of policies and actions are arrived at.

This research makes a significant contribution to the literature in terms of bringing together disparate areas of IT evaluation in a coherent and systematic way. The System Dynamics model of IT investment evaluation constitutes a novel source of new knowledge and provides an understanding of the area to both researchers and managers. The model helps in understanding the patterns of change or dynamics that a system exhibits over time and identifies the conditions that cause these patterns to be stable or unstable. This knowledge of the system can then suggest what kinds of prescriptions for governing it will work and what kinds may not.

Conclusions and future work

The main contribution of this paper is a tool that will help managers cut down on time spent debating investment decisions, cut down on costs, reduce information overload and help researchers evaluate related problems. The simulation model was used as a tool for evaluating the benefits derived from an information system using different variables like rate of information quality, rate of system quality, intention to use, user satisfaction, individual benefits, society benefits, organisation benefits, as well as total benefits. It was possible to calculate and observe the total benefits, user satisfaction, rate of information quality, rate of system quality, organizational benefits, societal benefits as well as intention to use. It was observed that the combination of rate of system quality, rate of information quality, and user satisfaction provided the best combination for information system success. System dynamics demonstrates how most of our own decision-making policies are the cause of the problems that we usually blame on others, and how to identify policies we can follow to improve our situation (Morecroft, 1999). The purpose in applying Systems Dynamics is to facilitate understanding of the relationship between the behavior of the system overtime and its underlying structure and strategic policies or decision rules (Caulfield and Maj, 2002). Taking a close look at the variables at play, it is evident that only a strong interaction of people, information, and technology can improve business performance, and consequently lead to Information Systems success.

Further research will be necessary to improve the tool in order to assess the theoretical content of the model. Another area that could be investigated could be the inclusion of the cost of the resources like equipment, software and people as a variable to be modelled along with the original variables in the model.

References


The Influence of Job Physical Characteristics on their Schedulability in Multi-cluster Systems

John Ngubiri and Mario van Vliet, Nijmegen Institute for Informatics and Information Sciences, Radboud University Nijmegen, Toernooiveld 1, 6525 ED, Nijmegen, The Netherlands

Performance (and sensitivity) studies in parallel job scheduling mostly use average values of the measurement metrics over the entire job stream. This does not give an idea of relative job performance (hence starvation) and sensitivity to scheduler parameters. Some jobs can therefore be easily starved without being detected. We investigate the influence of jobs’ physical characteristics on their schedulability in a multi-cluster system using the Fit Processor First Served (FPFS) scheduler. We also investigate their relative strength in determining schedulability as well as their sensitivity to scheduler parameters. We deduce the implication of the findings in a practical scheduling scenario.

1. Introduction

Clusters and Distributed Memory Multiprocessors (DMM) are taking a central place in high performance computing. Clusters constitute over 50% of the top 500 supercomputers in the world today (Top 500, 2005). This is due to fault tolerance, scalability, low cost and ability to be gradually set up. Recently, more interest has been put on combining clusters to form larger computing infrastructures called multi-cluster systems (mini grids). Unlike in the Grid (Grid Forum site, 2006), multi-cluster systems are comparatively small and more homogeneous.

In multi cluster systems, clusters in different geographical locations work together to process the jobs. A job submitted in one of the clusters may be processed in any of the clusters. Jobs need resources like memory, processors, software and data to execute (processors only considered in this research). Some of the resources (like software and data) may limit the clusters where a certain job may be processed.

Each job needs a specific number of processors (called size) to execute. A job may be made up of multiple components each of which is to be processed in a different cluster. Components of the same job however have to start executing at the same time (co-allocation). Despite the relatively slow wide area communication introduced, co-allocation improves performance (Bucur and Epema, 2001; Jones, 2005).
2. Research Model

Our research is based on the Distributed ASCII Supercomputer (DAS) (DAS site, 2006). DAS is a five cluster system located in five universities in The Netherlands (Delft, Utrecht, Leiden, Amsterdam and The Free University). It was developed by the Advanced School of Computing and Imaging (ASCII) to facilitate research in parallel computing. One of the clusters (Free University) has 72 nodes while the rest have 32 nodes each. The nodes are connected by Myrnet LANs while inter cluster communication is facilitated by a fast ATM backbone.

3. Related Work

Parallel job scheduling techniques, workload modeling, and performance measurements have been an active field of research in the past decade (Feitelson et al, 2004). For a scheduling scheme, both job based (like Average Waiting Time (AWT), job slowdown and throughput) and system based (like utilization and capacity loss) metrics are used to measure the effectiveness (Zhang et al, 2000). The effectiveness of these parameters in representing the relative performance of scheduling instances is not only contentious but has also attracted less attention in research. Short jobs for example make a large statistical influence on average job slowdown leading to misleading interpretation (Schwiegelshohn et al, 1997; Mu’alem and Feitelson, 2001). Improving it to bounded job slowdown (considering only jobs with runtimes above a certain threshold) makes it more realistic. A scheduling algorithm can starve some job types which may lead to a misleading numeric value depending on the measurement metric considered. For example, the Shortest Job First (SJF) scheduler starves long jobs. This leads to an impressive throughput and AWT values which does not imply the starvation creating a practical evaluation shortfall. If however the metrics are recorded separately for long and short jobs, the differences in values can easily imply the relative starvation. Moreira et al (1999) established that approximately 80% of the system load is from the largest 30% of the jobs. This implies that starvation of large jobs is equivalent to starving an exceptionally large portion of the workload.

Concentrating at the metric values can lead to wrong deductions (Feitelson (2002); Feitelson (2005)). Likewise, sometimes the difference in performance is due to the job mix rather than the philosophy of the scheduler (Srivasan and Kettimuthu, 2002).

The variation of AWT with utilization has been widely used in online scheduling performance studies. We also use it in this work. A scheduling instance $A$ is considered to have a better performance than $B$ if at a similar utilization $A$ has a lower AWT value.

Like in some of the previous work on multi cluster scheduling (Bucur, 2004, Jones 2005), we do not assume prior knowledge of job runtimes. This is because users are unable to accurately estimate the duration of their jobs (Lee et al, 2004). We use FPFS scheduling algorithm for our investigations.

In FPFS, jobs are placed in the queue in their arrival order. When searching for the next job to process, we start from the head search deeper into the queue for the first job that fits into the system. In case we get none, we wait for a job to finish execution.
or a job arrives and we search again. This may however lead to starvation of some jobs as they are continuously jumped. This is avoided by limiting the number of times \( (\text{maxJumps}) \) a job can be jumped while at the head of the queue. After being jumped \( \text{maxJumps} \) times, no other job is processed until when enough space has been created for the job at the head of the queue to start processing. We use FPFS(\( x \)) to represent FPFS at \( \text{maxJumps} \) value of \( x \).

4. Experimental Set Up

Using CSIM (Mesquite site, 2005), we simulated a system of five (homogeneous) clusters of 32 nodes each processing by pure space slicing. The system is served with one queue and one scheduler. The scheduler uses FPFS scheduling policy and Worst Fit placement policy. We do not model inter/ intra cluster communication.

A job can have up to four components. We consider three job streams where the respective composition of one, two, three and four component jobs is 10\%, 20\%, 30\% and 40\% (10-20-30-40), 25\%, 25\%, 25\% and 25\% (25-25-25-25) and 40\%, 30\%, 20\% and 10\% (40-30-20-10). Jobs execution times are exponentially distributed with mean 45. We continuously reduce the mean job inter arrival times (also exponentially distributed) until when we achieve system saturation.

The component width distribution \( D(q) \) is defined by the probability density function \( p_i = 3q^i/Q \) if \( i \) is a power of 2 and \( p_i = q^i/Q \) if \( i \) is not a power of 2. \( p_i \) is the probability that a component has width \( i \), \( q \) is a value used to vary the average component width and \( Q \) is in such a way that \( p_i \) sums up to 1. The distribution is defined over an interval \([n_1, n_2]\) (0 < \( n_1 < n_2 \)). It favors small jobs and those whose size is power of 2 which makes it a realistic choice. It has been used in previous studies (Bucur, 2004). We consider \( D(0.645839) \) over \([1, 18]\). All measurements are recorded at a maximum relative error of 0.08 and 95\% confidence interval.

5. Results

We classify jobs by the number of components they have, their size (some of all component widths), and by the width of the widest component. A job is narrow if it has a size less than 5 (otherwise it is wide) and it has a narrow widest component if the widest component has width less than 4 (otherwise it has a wide widest component).
The values 5 and 4 are chosen because they are the integral approximations of the medians for the job sizes and component widths for a 25-25-25-25 job stream. We consider single and mixed criteria classifications.

1.1 Single Criterion Classification

FIG 1: Performance of 1, 2, 3 and 4 component jobs on FPFS (10) and FPFS (50) for a 10-20-30-40 (top), 25-25-25-25 (middle) and 40-30-20-10 (bottom) job streams.
Fig 2: Performance of jobs grouped by size (left) and width of the widest component (right) for FPFS (10) and FPFS (50) on a 10-20-30-40(top), 25-25-25-25 (middle) and 40-30-20-10 (bottom) job streams.
From Fig 1 and Fig 2, we observe that jobs physical characteristics have a big contribution on their performance (schedulability). We observe that increasing the components, size and width of the widest component reduces the individual job's performance. This is largely independent of the (component wise) composition of the job stream. The difference in performance is observed to be large with large jobs performing up to over 10 times worse than the small jobs at high utilization values. We also observe that like in the work of Bucur (2004) (though for the entire job stream in
her case), increasing the value of \( \text{maxJumps} \) improves the performance of both small and large jobs. This implies that allowing jobs more chances of jumping to get scheduled does not significantly disadvantage large jobs. However, we observe that small jobs get a higher relative improvement compared to large jobs when \( \text{maxJumps} \) is increased.

5.2 Multiple Criteria Classification

We also grouped jobs by a combination of criteria and compared their performance. In component wise classification, we limit ourselves on one and four component jobs (since they are the extremes). We use a 25-25-25-25- stream since the job stream does not significantly affect the relative performance.

Fig 3: Performance of jobs classified by number of components and size (top) and number of components and width of widest component (bottom) for FPFS (10) (left) and FPFS (50) (right) on a 25-25-25-25 job stream.
Fig 4: Performance of jobs classified by size and width of the widest component for FPFS (10) (left) and FPFS (50) (right) on a 25-25-25-25 job stream.

From Fig 3 and Fig 4, we observe that (i) irrespective of the number of components, wide jobs are harder to schedule compared to narrow jobs. This implies that job size has a higher influence on schedulability than the number of components; (ii) irrespective of the width of the widest component, jobs with more components are harder to schedule compared to jobs with fewer components. This implies that the number of components has a higher influence on schedulability compared to the width of the widest component (we however observe that the width of the widest component gets more influential at large values of $\text{maxJumps}$) and (iii) irrespective of the width of the widest component, wide jobs are harder to schedule compared to narrow jobs implying that the job size has a higher influence on schedulability compared to the width of the widest component.

5.3 Job And Load Distribution

To illustrate the practical impact of the big deviation in relative performance, we summarized at the numerical and load (processor hours) distribution of the jobs in the job streams. We use the physical characteristics investigated in the research to do the classifications.

Table 1: Percentage numerical and load contribution for jobs classified by the number of components.

<table>
<thead>
<tr>
<th>Job stream</th>
<th>1 - COMP</th>
<th>2 - COMP</th>
<th>3 - COMP</th>
<th>4 - COMP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jobs</td>
<td>Load</td>
<td>Jobs</td>
<td>Load</td>
</tr>
<tr>
<td>10-20-30-40</td>
<td>10.8</td>
<td>3.4</td>
<td>19.5</td>
<td>13.3</td>
</tr>
<tr>
<td>25-25-25-25</td>
<td>24.9</td>
<td>9.4</td>
<td>25.0</td>
<td>20.6</td>
</tr>
<tr>
<td>40-30-20-10</td>
<td>40.2</td>
<td>20.1</td>
<td>30.0</td>
<td>30.2</td>
</tr>
</tbody>
</table>
Table 2: Percentage numerical and load contribution for jobs classified by width of widest component.

<table>
<thead>
<tr>
<th>Job stream</th>
<th>NWC</th>
<th>WWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jobs</td>
<td>Load</td>
<td>Jobs</td>
</tr>
<tr>
<td>10-20-30-40</td>
<td>52.2</td>
<td>31.8</td>
</tr>
<tr>
<td>25-25-25-25</td>
<td>58.6</td>
<td>34.4</td>
</tr>
<tr>
<td>40-30-20-10</td>
<td>66.0</td>
<td>40.5</td>
</tr>
</tbody>
</table>

Table 3: Percentage numerical and load contribution for jobs classified by size.

<table>
<thead>
<tr>
<th>Job stream</th>
<th>Narrow</th>
<th>Wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jobs</td>
<td>Load</td>
<td>Jobs</td>
</tr>
<tr>
<td>10-20-30-40</td>
<td>38.6</td>
<td>12.4</td>
</tr>
<tr>
<td>25-25-25-25</td>
<td>51.4</td>
<td>19.8</td>
</tr>
<tr>
<td>40-30-20-10</td>
<td>65.9</td>
<td>31.9</td>
</tr>
</tbody>
</table>

6. Discussion Of Results

We observe that there is a large difference between jobs of different types classified by any of the parameters: - size, number of components and width of the widest component. This implies that FPFS is largely an unfair scheduler. The use of \textit{maxJumps} to avoid selective starvation still leaves a large performance deviation (large jobs waiting up to over 10 times the small jobs at high utilization).

The job physical characteristics, greatly affect the schedulability of the job. The extent of influence is in the (reducing) order of size, number of components and width of the widest component.

The large difference in performance implies a practical setback where the AWT is not an accurate measure of when a job can be expected to wait in the queue. While a user with a small job waits for far less time, the user with a large job waits for a far larger time compared to the average. The average therefore is not a good guide to a user as to what he expects from the system. The difference also implies that by delaying the large jobs, the system is effectively delaying a larger portion of the load while quickly processing the low load jobs. These small jobs, due to their numerical majority highly influence the AWT parameter hence making appear to be a better scheduler. This is synonymous to the impact of short jobs on the job slow down metric studied by Schwiegelshohn et al (1997) and the impact of job mix on the average performance parameter studied by Srivasan and Kettimuthu (2002).
7. Conclusions And Future Work

Using simulation, we investigated the impact of job physical characteristics on their schedulability. The constraint to for job scheduling in our case is the availability of processors in the clusters that can accommodate the job. Generally, we realize that to get a better understanding of the performance, we need to go beyond the overall average and group the jobs using the characteristics related to the constraining parameter. For example, if we were to consider memory as well, it would be useful to classify jobs by the memory the jobs need since now memory is a constraint. In a broader grid environment, several resources are scarce (like bandwidth, memory, processors, software, etc). Grouping by these parameters, on top of the average gives a better practical picture on how restrictive a certain resource is.

The research also creates more opportunities for future work. While FPFS is known to be a good scheduler especially in cases where the job duration is not known, we observe that it is highly ‘unfair’. Fixing $\text{maxJumps}$ does not adequately handle the starvation problem. We therefore need to look at ways of reducing inter-job performance gap beyond the $\text{maxJumps}$ condition.

References


DAS Website: www.sc.vu.nl/das2. (Accessed Jan 2006)


On solving large scale Linear Systems arising from Interior Point Methods for Linear Programming

Venansius Baryamureeba, Department of Computer Science, Faculty of Computing and Information Technology, Makerere University

Every Newton step in an interior point method for linear programming requires a solution of a symmetric indefinite system of linear equations. The most expensive part of an interior point method is computing the search direction by solving one or more of the linear systems. Such systems are indefinite and can be written in symmetric form, which is known as an augmented system. A common approach in interior point solvers for linear programming reduces the augmented system to a smaller positive definite one, called a normal equation system. Both direct methods and iterative methods have been used in the solution of the linear systems. The success of implementations using iterative methods depends on how to choose the appropriate preconditioner since the matrix system usually becomes ill-conditioned as the optimal solution of the problem is approached. In this paper we propose a strategy for solving the linear systems arising from interior point methods for linear programming. We also propose how to construct a preconditioner for the iterative approach for solving linear systems.

1. Introduction

This paper concerns solving linear systems of equations arising from interior-point algorithms for linear programming by preconditioned conjugate gradient methods combined with a direct method. The central issue is to propose an efficient preconditioner for the normal equations linear systems.

1.1 Interior-Point Methods for Linear Programming

Consider the standard form primal linear program

Minimize \( c^T x \)
Subject to: \( Ax = b \)
\( x \geq 0 \)

and its dual program

Maximize \( b^T y \)
Subject to: \( A^T y + z = c \)
\( z \geq 0 \)
where $A \in \mathbb{R}^{m \times n}$ ($m < n$), $b \in \mathbb{R}^{m}$ and $c \in \mathbb{R}^{n}$. For convenience, we will assume that $A$ has full rank $m$.

The optimality (Karush-Kuhn-Tucker or KKT) conditions for the above linear program pair can be written as a linear-quadratic system of equations plus nonnegativity constraints,

$$F(x, y, z) = \begin{pmatrix} \begin{bmatrix} b^T & b \\ A^T & y + z - c \\ XZ & e \end{bmatrix} \end{pmatrix} = 0, (x, z) \geq 0,$$

where $X = \text{diag}(x), Z = \text{diag}(z)$ and $e$ is the vector of all ones in $\mathbb{R}^{n}$.

We consider primal-dual interior-point algorithms that have proven to be the most efficient for linear programming, although the results in this paper can be applied to primal or dual interior-point algorithms as well. The basic framework of primal-dual interior-point methods can be viewed as applying Newton's method to the so-called perturbed KKT conditions,

$$F_\mu(x, y, z) = \begin{pmatrix} A x - b \\ A^T y + z - c \\ XZ \righte - \mu e \end{pmatrix} = 0, (x, z) \geq 0,$$

for a sequence of positive $\mu$ values that decreases to zero, while keeping $x$ and $z$ positive. The parameter $\mu$ is called the centering (or barrier) parameter, and the set of triples $(x, y, z)$ that satisfy (1) for all $\mu > 0$ is called the (primal-dual) central path. There exists several references (El-Bakry et al., 1996 [11]; Kojima et al., 1989 [20]; Wright, 1997 [28]) on primal-dual interior-point methods.

At a given iterate $(x, y, z)$, where $x, z > 0$, and for a fixed value of $\mu$, the linear equation that defines a Newton step for (1) is

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ Z & 0 & X \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = - \begin{pmatrix} A x - b \\ A^T y + z - c \\ XZ \righte - \mu e \end{pmatrix}$$

The next iterate is obtained by the update $(x, y, z) \leftarrow (x, y, z) + \alpha (\Delta x, \Delta y, \Delta z)$.
where the step length $\alpha$ is chosen to ensure that the new $\mathbf{x}$ and $\mathbf{z}$ remain positive.

The most widely-used approach for solving (2) is to use a block Gaussian elimination to reduce it to a smaller system

$$\begin{align*}
(AGA^T)\mathbf{y} &= AG(\mathbf{A}^T - \mathbf{A}^T \mathbf{y} - \mu X^{-1} \mathbf{e}) + (\mathbf{b} - \mathbf{Ax}) \\
(3)
\end{align*}$$

where $\mathbf{G} = Z^{-1}X$ is a positive definite diagonal matrix. After $\Delta \mathbf{y}$ is determined from solving the above linear system, $\Delta \mathbf{x}$ and $\Delta \mathbf{z}$ can then be easily computed. Hence, in this approach, the primary computation at each iteration is to solve (3).

It is worth noting that whenever there holds $\mathbf{b} = A\mathbf{\bar{x}}$ for some $\mathbf{\bar{x}}$, system (3) can be written into

$$\begin{align*}
(AGA^T)\mathbf{y} &= \mathbf{v} \\
(4)
\end{align*}$$

where $\mathbf{v} = AGh$ and $h = (\mathbf{c} - \mathbf{A}^T \mathbf{y} - \mu X^{-1} \mathbf{e}) + G^{-1}(\mathbf{\bar{x}} - \mathbf{x})$.

Let $\mathbf{A} = [\mathbf{A}_1 - \mathbf{A}_2]$ be the matrix whose columns have been reordered into two block matrices $\mathbf{A}_1$ and $\mathbf{A}_2$. After reordering of $\mathbf{A}$, let $\mathcal{I}_1 = \{1, 2, \ldots, r\}$ and $\mathcal{I}_2 = \{r + 1, r + 2, \ldots, n\}$ be column indices of $\mathbf{A}$ corresponding to $\mathbf{A}_1$ and $\mathbf{A}_2$ respectively.

Let $\mathbf{G}$ be partitioned such that $\mathbf{G}_1$ and $\mathbf{G}_2$ are block (square) submatrices corresponding to $\mathcal{I}_1$ and $\mathcal{I}_2$ respectively. Then

$$\begin{align*}
AGA^T &= A_1 G_1 A_1^T + A_2 G_2 A_2^T \\
(5)
\end{align*}$$

In this notation, (4) becomes

$$\begin{align*}
(A_1 G_1 A_1^T + A_2 G_2 A_2^T)\mathbf{y} &= \mathbf{v} \\
(6)
\end{align*}$$

Let $\mathbf{H} \in \mathbb{R}^{n \times n}$ be a positive semi-definite diagonal matrix. Likewise, we partition $\mathbf{H}$ such that

$$\begin{align*}
AH\mathbf{A}^T &= A_1 H_1 A_1^T + A_2 H_2 A_2^T \\
(7)
\end{align*}$$

In this paper, we propose how to construct a preconditioner for the linear system (4). The preconditioner is based on a low-rank correction of a matrix $A_1 H_1 A_1^T$ where
is a positive definite diagonal matrix from a previous iteration and is not "vastly different" from \( G \); (ii) the Cholesky factorization of \( A H A^T \) is already available. We will use the bounds on the condition number of the preconditioned coefficient matrix derived by Baryamureeba (2002) \cite{Baryamureeba2002} and provide guidelines for the choice of low-rank corrections. We will also provide guidelines on choosing the approximate linear system to be solved.

### 1.2 Motivation

Many problems in engineering applications involve the solution of large sparse linear programs when an interior point method is used. For example, linear programs occurring in the area of VLSI layout optimization (Hu and Shing, 1985 \cite{Hu1985}; Vannelli, 1991 \cite{Vannelli1991}; Vannelli, 1998 \cite{Vannelli1998}; Weis and Mlynski, 1987 \cite{Weis1987}). Before solving these large sparse linear programs, they are usually reduced to smaller linear systems, namely the augmented linear system (or the normal equations linear system). The coefficient matrix of the normal equations linear system corresponding to these problems is of the form (\ref{eq:5}). In particular, the matrix \( A \) is large, sparse, and usually with some dense columns.

### 1.3 Organization and Notation

The outline of the paper is as follows: In Section 2, we introduce the construction of the preconditioners based on low-rank corrections and give theoretical results to guide the choice of low-rank corrections. Section 3 discusses computation with the preconditioner and Section 4 gives concluding remarks.

Throughout this paper we use the following notation. The symbol \( \min_i \) or \( \max_i \) is for all \( i \) for which the argument is defined. For any matrix \( A \), \( A_{ij} \) is the element in the \( i \)-th row and \( j \)-th column, \( A \) is the \( j \)-th column. The vector norm \( \| \cdot \| \) is the Euclidean norm \( \| x \|^2 = x^T x \). The notation \( x > 0 (x \geq 0) \) means that all components of the vector \( x \) are positive (nonnegative). The symbol \( 0 \) will be used to denote the number zero, the zero vector, and the zero matrix. The symbol \( I \) is used to denote the (square) identity matrix; its size will always be apparent from the context. For any square matrix \( X \) with real eigenvalues, \( \lambda_i(X) \) are the eigenvalues of \( X \) arranged in nondecreasing order \( \lambda_{\min}(X) \) and \( \lambda_{\max}(X) \) denote the smallest and largest eigenvalues of \( X \) respectively; i.e.,

\[
\lambda_{\min}(X^{-1}) = \lambda_1(X) \leq \lambda_2(X) \leq \ldots \leq \lambda_m(X) = \lambda_{\max}(X).
\]

If \( X \) is symmetric and positive definite, then the above arrangement gives

\[
\lambda_i(X^{-1}) = 1/\lambda_{m-i+1}(X).
\]
In addition, we will denote the spectral condition number of \( X \) by \( \kappa(X) \) where by definition
\[
\kappa(X) = \frac{\lambda_{\text{max}}(X)}{\lambda_{\text{min}}(X)}.
\]
The letters \( L \) and \( R \) represent unit lower triangular factors of a symmetric, positive definite matrix.

The matrices \( D \) and \( T \) are positive definite and diagonal. For any set \( I \) and \( |I| \) is the number of elements in \( I \).

2. The Preconditioner

In most implementations of primal-dual interior-point algorithms for linear programming, the direct method of Cholesky factorization is used to solve the system (4). The use of iterative methods of conjugate-gradient type (Golub and van Loan, 1996) [16] have also been considered (Carpenter and Shanno, 1993 [10]; Karmarkar and Ramakrishnan, 1991 [19]; Mehrotra, 1992 [21]). However, the success of iterative methods has been, at best, very limited for interior-point algorithms due to the difficulty in constructing general-purpose, effective preconditioners.

It is well known that effective preconditioners are critical for accelerating the convergence rate of conjugate-gradient type methods. In search for a preconditioner, one has to balance two conflicting goals: (i) to minimize the condition number of the preconditioned coefficient matrix, and (ii) to minimize the cost of solving linear systems with the preconditioner as the coefficient matrix.

To attain rapid convergence for conjugate gradient type methods we require that either the spectral condition number of the preconditioned matrix be close to one in order for the error bound based on the Chebyshev polynomial to be small, or the preconditioned matrix have great clustering of eigenvalues. From the computational point of view, we require that the linear systems with the preconditioner as coefficient matrix be easier to solve, and the construction cost of the preconditioner be low.

2.1 Class of Preconditioners

We investigate a mixed strategy for solving linear systems of the form (6), in which a direct solver and an iterative solver are combined in some manner. Let \( k \) denote the computation step. At the initial computation step \( k = 0 \) we use a direct solver. Let \( A_{1}, H_{1}, A_{1}^{T} \) be a matrix from the previous computation step with a known factorization.

For a given index set
\[
Q_{1} \subseteq \{ j : j \in I_{1} \text{ and } G_{jj} \neq H_{jj} \},
\]
let the $r \times r$ diagonal matrices $D_1$ and $K_1$ be given by

$$D_{1_{jj}} = \begin{cases} G_{jj} - H_{jj} & \text{if } j \in Q_1 \\ 0 & \text{if } j \in \bar{J} \setminus Q_1 \end{cases}$$

and

$$K_1 = H_1 + D_1. \quad (9)$$

Let $A_1 \in \mathbb{R}^{n \times q_1}$, where $q_1 = |Q_1|$, consist of all columns $A_j$ such that $j \in Q_1$ and let $D_1 \in \mathbb{R}^{q_1 \times q_1}$ be the diagonal matrix corresponding to the nonzero diagonal elements of $D_1$. In this notation,

$$A_1K_1A_1^T = A_1(H_1 + D_1)A_1^T = A_1H_1A_1^T + A_1D_1A_1^T,$$

namely, $A_1K_1A_1^T$ is a rank $q_1$ correction of $A_1H_1A_1^T$.

Given $Q_2 \subseteq \{ j : j \in \bar{J}_2 \}$

let the $(n - r) \times (n - r)$ diagonal matrix $K_2$ and the $(n \times n)$ diagonal matrix $K$ be given by

$$K_{2_{jj}} = \begin{cases} G_{jj} & \text{if } j \in Q_2 \\ 0 & \text{if } j \in \bar{J}_2 \setminus Q_2 \end{cases}$$

and

$$K = \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix}, \quad (10)$$

where $K_1$ is defined in (9) and $s = j - r$. Let $A_2 \in \mathbb{R}^{n \times q_2}$, where $q_2 = |Q_2|$, consist of all columns $A_j$ such that $j \in Q_2$ and let $G_2 \in \mathbb{R}^{q_2 \times q_2}$ be the diagonal matrix corresponding to the nonzero diagonal elements of $K_2$. In this notation,

$$A_2K_2A_2^T = A_2G_2A_2^T.$$
Thus the general class of preconditioners is given by

\[ AKA^T = \left( A_1 H_1 A_1 + \overline{A_1 D_1 A_1^T} \right) + A_2 G_2 A_2^T. \]  

(11)

Let \( Q = Q_1 \cup Q_2 \) and \( q = q_1 + q_2 \). The elements in the class of preconditioners (11) are determined by the choice of the index set \( Q \).

### 2.2 Bounds on eigenvalues of the preconditioned matrix

Our interest is in bounding the spectral condition number of the preconditioned matrix. By the definition of the spectral condition number, this is equivalent to establishing a lower bound for the smallest eigenvalue and an upper bound for the largest eigenvalue of the preconditioned matrix. Once we have the bounds on the eigenvalues, the upper bound on the spectral condition number can be easily computed.

#### Lemma 2.1: Baryamureeba et al. (1999) [7]:

Let \( G, K \in \mathbb{R}^{n \times n} \) be positive definite diagonal matrices.

Then

\[
\min_j \left\{ \frac{G_{jj}}{K_{jj}} \right\} \leq \lambda_i \left( (A_1 K_1 A_1^T)^T A G A^T \right) \leq \max_j \left\{ \frac{G_{jj}}{K_{jj}} \right\} .
\]

#### Theorem 2.1: Baryamureeba (2002) [5]:

Let \( G, H \in \mathbb{R}^{n \times n} \) be positive definite diagonal matrices. Let \( AGA^T \) and \( AHA^T \) be partitioned as in (5) and (7) respectively. Let \( K_1 \) and \( K \) be defined in (9) and (10) respectively. Assume that \( A_1 \) has full rank \( m \). Then

\[
(\text{i}) \min_{j \in \Phi_1} \left\{ \frac{G_{jj}}{H_{jj}} \right\} \leq \lambda_i ((A_1 K_1 A_1^T)^{-1} A_1 G_1 A_1^T) \leq \max_{j \in \Phi_1} \left\{ \frac{G_{jj}}{H_{jj}} \right\}
\]

and

\[
(\text{ii}) \min_{j \in \Phi_2} \left\{ \frac{G_{jj}}{H_{jj}} \right\} \leq \lambda_i ((AKA^T)^{-1} AGA^T) \leq \max_{j \in \Phi_2} \left\{ \frac{G_{jj}}{H_{jj}} \right\} + \Psi,
\]

where

\[
\Psi = \sum_{j \in \Phi_2 \setminus \Phi_1} G_{jj} \left\| A_j \right\|_2^2 / \lambda_{\min} (AKA^T).
\]
2.2.1 Related work

Bocanegra et al. [3] propose a hybrid preconditioner to solve the normal equations system where the Controlled Cholesky Factorization Preconditioner (Campos, 1995) [9] is used in earlier iterations (phase one) and the split preconditioner (Oliveria, 1997) [9] is used in last iterations (phase two).

Adler et al. (1989) [2] and Gill et al. (1986) [14] propose computing the factorization $RTR^T = A_1G_1A_1^T$; and then use this factorization as a preconditioner for conjugate gradient algorithm to solve (6). As we will show later in this paper, the sparse part $A_1G_1A_1^T$ is not always a good preconditioner for $AGA^T$. As an alternative method to handle dense columns Gill et al. (1988) [13] suggest the Schur complement which factors a large sparse matrix $A_1G_1A_1^T$ and a small dense matrix but it is not iterative in character.

The idea of using low-rank updates is not new. In his seminal paper on the first polynomial-time interior-point method for linear programming, Karmarkar (1984) [18] already proposed to use low-rank updates to decrease the theoretical complexity bounds. His idea has been pursued by many later papers (for example, Goldfarb and Liu (1991) [15]) for the similar reason. Moreover, a combination of a direct method and an iterative method was reported by Karmarkar and Ramakrishnan (1991) [19].

Wang and O’Leary (2000) [25] propose a strategy of preconditioning the normal equation system based on low-rank corrections. In their implementation, they choose the index set $Q$ to consist of the indices corresponding to the largest values of $|G_{i'} - H_{i'}|$. 

Baryamureeba et al. (1999) [7] suggest to choose $Q$ to consist of $j \in I$ corresponding to the largest $G_{i'} / H_{i'}$ and/or the smallest $G_{i'} / H_{i'}$ such that $\kappa(K^{-1}G)$ (for $K$ positive definite) is minimized. However, these papers (Baryamureeba et al. 1999 [7]; Wang and O’Leary, 2000 [25]) do not discuss the case when $A$ contains some dense columns.

Baryamureeba (2002) [5] extends the result by Baryamureeba et al. (1999) [7] and suggests a low-rank correction preconditioner for the case when the problem matrix $A$ contains some dense columns. In this case $Q_1$ is chosen such that $\kappa(K_1^{-1}G_1)$ (for $K_1$ positive definite) is minimized and $Q_2$ consists of the indices corresponding to the largest values of $G_{i'} / A_{i'}$ for $j \in I_2$. This implies that $Q_1$ should consist of indices $j \in I_1$ corresponding to the largest $G_{i'} / H_{i'}$ and/or the smallest $G_{i'} / H_{i'}$ such that $\kappa(K_1^{-1}G_1)$ is minimized. In this case, the matrices $A_1$ and $A_2$ consist of the sparse columns in $A$ and the dense columns in $A$ respectively. The matrix $A_1A_1^T$ is the sparse part of the coefficient matrix of the normal equations (or the sparse part of the preconditioner) from the previous computation step with known factorization.
In this paper we suggest improvements to the preconditioner proposed in Baryamureeba (2002) [5] that reduces the computational cost of the preconditioner by eliminating columns from the preconditioner corresponding to nonbasic variables.

2.3 Choosing the index set \( Q \)

The idea is to choose the index set \( Q \) such that the upper bound on the spectral condition number of the preconditioned matrix \( (AKA^T)^{-1}AGA^T \) is minimized. We also want to minimize the computation cost of the preconditioner.

At the direct step, before doing the Cholesky factorization, we propose that \( A_1 \) comprises of the sparse part of \( A \) excluding the sparse columns corresponding to the smallest \( G_{jj} / H_{jj} < 1 \). This implies that \( A_2 \) will comprise of all the dense columns in \( A \) and the sparse columns not in \( A_1 \). We factorize \( A_1G_1A_1^T = RTR^T \) during the direct step. Thus the corresponding matrix from the previous iteration with the known factorization \( A_1H_1A_1^T \) corresponds to \( A_1G_1A_1^T \).

The Simplex Method approach looks for an \( m \times m \) basis matrix that gives an optimal solution by eliminating \( n - m \) columns of \( A \) and \( G \) that correspond to slack variables. We use this idea in eliminating sparse columns from \( A_1 \) corresponding to the smallest \( G_{jj} / H_{jj} < 1 \) where \( G_{jj} < 1 \), i.e. corresponding to slack variables. Thus \( A_2 \) will comprise of columns corresponding to slack variables and dense columns. The sparse columns in \( A_2 \) can only get back to \( A_1 \), i.e. join the basis matrix if and only if the corresponding \( G_{jj} \) is large.

Using the result in Theorem 2.1, we choose \( Q_1 \) such that the upper bound on \( \kappa((AKA_1^T)^{-1}A_1G_1A_1^T) \) is minimized and \( Q_2 \) such that \( \psi \) is minimized. This implies that \( Q_1 \) should consist of indices

- \( j \in \mathbb{I}_1 \) corresponding to the largest \( G_{jj} / H_{jj} > 1 \)
- \( j \in \mathbb{I}_1 \) corresponding to the smallest \( G_{jj} / H_{jj} < 1 \)

such that the upper bound on \( \kappa((A_1K_1A_1^T)^{-1}A_1G_1A_1^T) \) is minimized. Due to the preprocessing at the direct step \( Q_1 \) may not consist of any \( j \in \mathbb{I}_1 \) corresponding to the smallest \( G_{jj} / H_{jj} < 1 \). This decreases the cost of computing the low rank preconditioner.

Similarly, \( Q_2 \) consists of indices \( j \in \mathbb{I}_2 \) corresponding to the largest \( G_{jj} \| A_j \|_2^2 \). Thus the required index set \( Q = Q_1 \cup Q_2 \). In this case, the matrix \( A_1 \) consists of the sparse columns in \( A \) excluding sparse columns corresponding to the smallest
Similarly the \( A_2 \) consists of the columns not in \( A_1 \) including the dense columns. The matrix \( A_1 H_1 A_1^T \) is the sparse part of the preconditioner from the previous commputation step with a known factorization.

### 2.4 Interpretation of the Results

Baryamureeba (2002) [5] and Baryamureeba et al. (1999) [7] proposed formation of a low rank preconditioner involving low rank updates \( G_j A_j A_j^T \) corresponding to the smallest \( G_j / H_j < 1 \). Under the simplex method these are updates corresponding to nonbasic variables, which essentially should have been ignored from the basis matrix. So in this paper, we propose a technique for avoiding the columns corresponding to the nonbasic variables from the preconditioner, i.e. from the matrix to be factorized which will later form the factorized part of the preconditioner at later iterations.

Our assumption is that for slack variable \( x_j \) the value of \( X_j Z_j^{-1} \) will tend towards 0 as we approach the solution since \( x_j \rightarrow 0 \).

Let \( A = [A_2, A_2, A_3] \) be the matrix whose columns have been reordered into three block matrices \( A_1, A_2, \) and \( A_3 \). After reordering of \( A \), let \( \bar{1}_1 = \{1, 2, ..., r\}, \bar{2}_1 = \{r + 1, r + 2, ..., s\}, \) and \( \bar{3}_1 = \{s + 1, s + 2, ..., n\} \) be column indices of \( A \) corresponding to \( A_1, A_2 \) and \( A_3 \) respectively.

Let \( G \) be partitioned such that \( G_1 G_2 \) and \( G_3 \) are block (square) submatrices corresponding to \( \bar{1}_1 \bar{2}_1 \) and \( \bar{1}_1 \bar{3}_1 \) respectively. Then

\[
AGA^T = A_1 G_1 A_1^T + A_2 G_2 A_2^T + A_3 G_3 A_3^T.
\] (12)

In this notation, (4) becomes

\[
(A_1 G_1 A_1^T + A_2 G_2 A_2^T + A_3 G_3 A_3^T) \Delta y = \nu.
\] (13)

Let \( H \in \mathbb{R}^{m \times m} \) be a positive semi-definite diagonal matrix. Likewise, we partition \( H \) such that

\[
AH A^T = A_1 H_1 A_1^T + A_2 H_2 A_2^T + A_3 H_3 A_3^T.
\] (14)
The matrix $A_3 G_3 A_3^T$ corresponds to the smallest $G_{jj} / H_{jj} < 1$. Thus the index set $J_3$ corresponds to the prospective slack variables. We suggest solving a reduced system

\[(A_1 G_1 A_1^T + A_2 G_2 A_2^T) \Delta y = \nu\]  

(15)

at the direct step where $A_1$ comprises of the sparse columns and $A_2$ the dense columns.

The computational cost of using the preconditioner in Baryamureeba (2002) [5] and Baryamureeba et al. (1999) [7] is proportional to the number of updates and downdates. Also the value of $\|A_3 G_3 A_3^T\|^2_2$ in (13) is usually very small. Thus at the direct step solving (15) gives a good approximate solution to the linear system in (13). The most recent preconditioner of the type (11) was suggested in Baryamureeba (2002) [5]. It easily follows that the preconditioner suggested in this paper is more efficient than the preconditioners suggested in Baryamureeba (2002) [5] and Baryamureeba et al. (1999) [7].

3. Computing with the Preconditioner

In this paper, we investigate a mixed strategy for solving linear systems of the form (13), in which a direct method and an iterative method are combined in some manner. Specifically, we will combine Cholesky factorization (or LDL^T) with a preconditioned conjugate-gradient method. A simple, but not necessarily effective, combination would be that the Cholesky factorization is used at every even iteration and the preconditioned conjugate-gradient method at every odd iteration.

Suppose that we have calculated the Cholesky factorization of $A_1 H_1 A_1^T$ at a previous iteration and we are to solve a new linear system with coefficient matrix $A_1 G_1 A_1^T + A_2 G_2 A_2^T$ where $G_1$ is different but not “too far away” from $H_1$. Then it is advantageous to utilize the existing Cholesky factorization of $A_1 H_1 A_1^T$ in the construction of a preconditioner for the conjugate-gradient method. Moreover, a good candidate for such a preconditioner is a low-rank correction (update) to $A_1 H_1 A_1^T$ because of two reasons. Firstly, a good correction may be able to at least partially account for the changes from $H_1$ to $G_1$. Secondly, given the availability of the Cholesky factors, a low-rank update will keep the cost low for solving linear systems with the preconditioner as the coefficient matrix.

Initially ($k = 0$) we solve (6) by a direct solver. For every computation step where a direct solver is used, we compute the factorization

\[RTR^T = A_1 G_1 A_1^T,\]  

(16)
and solve for $y$ by applying the Sherman-Morrison-Woodbury formula approach (Baryamureeba et al., 1999) [7] to the linear system

$$(RTR^T + A_2G_2A_2^T)y = v.$$ 

The factorization (16) can be computed by directly computing the Cholesky factors. However, when the matrix $A_1^T A_1$ is very ill-conditioned, we are most likely to compute less accurate Cholesky factors because of loss of precision inherent in forming $A_1^T A_1$. Thus in this case we suggest computing the factors $R$ and $T$ indirectly from the QR factorization of $G_{1/2}^{1/2} A_1^T$, see (Björck, 1996[8]; Wang and O’Leary, 2000[25]). This results in more accurate factors, but it is very expensive compared to direct computation of the Cholesky factors.

When the rank of $A_1$ is less than $m$ then the matrix $A_1^T A_1$ is singular. In this case computing the standard Cholesky factorization of $A_1^T A_1$ is numerically unstable. An alternative is to compute the Cholesky-infinity factorization (Zhang, 1998) [29] for example. For other approaches on handling rank deficiency, see Andersen (1996) [1] and references there in.

Let $LDL^T = A_1 H_1 A_1^T$ be the given factorization of the sparse part of the preconditioner (or sparse part of the coefficient matrix of the normal equations) from the previous computation step. Then the linear system with the preconditioner as coefficient matrix is of the form

$$(17)$$

Given the sparse low-rank correction $A_1 D_1 A_1^T$ and the dense update $A_2 D_2 A_2^T$, in the remaining part of this section we show how to solve for $z$ in (17).

### 3.1. Sherman-Morrison-Woodbury formula approach

In the case of odd-even alternation (using a direct solver for every even step and an iterative solver for every odd step), we directly apply Sherman-Morrison-Woodbury formula approach (Baryamureeba et al., 1999) [7] to (17) in the form

$$(18)$$

### 3.2. Modifying the triangular factors

We modify the triangular factors when the solver is chosen adaptively, see Wang and O’Leary (2000) [25] and Algorithm 4.1. in Baryamureeba (2002) [5]. In particular, we consider (17) in the form

$$(19)$$
At the current step we first modify the $LDL^T$ factorization with a sparse matrix of low-rank to get $RTR^T$ (Bartels and Kaufman, 1989 [4]; Baryamureeba and Steihaug, 2000 [6]; Gill et al 1974 [12]) i.e., we compute $R$ and $T$ such that $RTR^T = LDL^T + A_1D_1A_1^T$. Thus the preconditioner (11) reduces to

$$AKA^T = RTR^T + A_2G_2A_2^T.$$ 

To solve the linear system (18) for $z$, we apply Sherman-Morrison-Woodbury formula approach [7] to

$$\left(RTR^T + A_2G_2A_2^T\right)z = v$$

### 4. Concluding Remarks

In this paper we have proposed a strategy for solving the linear systems arising from interior point methods for linear programming. We also proposed how to construct a preconditioner for the iterative approach for solving linear systems. This approach will reduce on the computational cost (cost of updates) of the preconditioner since it eliminates the columns in the problem matrix corresponding to slack variables from the matrix to be factorized at the direct step.

### References

- S. Bocanegra, EF. Campos, A.R.L.Oliveira, Using Hybrid preconditioner for solving large-scale linear systems arising from interior point methods, To appear in the Special Issue of Comput. Optimization and Applications on LINEAR ALGEBRA ISSUES ARISING IN INTERIOR POINT METHODS


A.R.L. Oliveria, A new class of preconditioners for large-scale linear systems from interior point methods for linear programming, PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1997.


Adaptive and Optimization Predictive Text Entry for Short Message Service (Sms)

Peter Waiganjo Wagacha and Dennis Chege, School of Computing and Informatics, University of Nairobi, Kenya

The ability to enter text on phones is crucial for utilizing the Short Message Service (SMS) of digital cellular networks. Standard 12-button phone keypads are not naturally suited for text entry as each button on them encodes three or more alphabetic letters. Various techniques have been proposed and implemented to improve text entry tasks. One such technique is predictive text entry. In this paper we describe the development of a generic mobile phone predictive text application. We have tried it on some local Kenyan languages. It is implemented as a word weighting system that is based on the sigmoid function. It initially relies on the natural word frequency of a given language. We also present an adaptive learning model for improving text entry speed that incorporates a specific user's word usage habits. Our results indicate an improvement in the performance of the developed system as evaluated and compared to other current predictive text techniques.

1. Introduction

Over the past few decades, wireless devices have revolutionized the communication industry giving rise to unprecedented level of mobile phone usage across the population spectrum in Kenya. Because of this it is increasingly becoming apparent that mobile phone usage is poised to grow significantly. For users who are more comfortable using local languages, a few fundamental technical challenges need to be addressed before they can efficiently input text messages in a language of their choice.

The main challenge is convenient and effortless composing of text messages in local languages. The number of taps for entering a single character as well as composing whole words and sentences in local languages could be impractically high thereby discouraging widespread use. The demand for the input of text or alphanumeric information will become even stronger as accessing the Internet from a mobile device becomes more common. This makes research into efficient text entry methods all the more important.

This work is, therefore, an attempt to enable practical and efficient composing of text messages in local languages. To improve efficiency even further, we introduce an adaptive prediction algorithm was developed so that the system automatically learns a user's word usage habits and suggests the most probable word for the case when a key sequence results in more than one possible word (disambiguation). This is a an improvement over current predictive text systems which are largely static and do not adapt to user’s word usage.
1.1. Short Message Service

Short Message Service (SMS) allows people to send or receive text messages of up to 160 characters from mobile phones. SMS messaging has gained popularity among people, especially the younger generation, in recent years. The various uses of SMS extend from keeping in contact with one another to communicating vital business information to mobile executives, sales representatives, field technicians and customers. According to the GSM association, the number of SMS messages sent globally increased from approximately 4 billion in Jan 2000 to 24 billion in May 2002 (GSM, 2005). Because of this almost exponential growth, it is essential to investigate and find ways to improve the efficiency of the text entry methods currently available.

1.2. Predictive Text

The mobile phone keypad is small in size, with all the keys supporting both numeric and sharing of alphanumeric characters, as well as other functions. This is illustrated in Figure 1.

For the predictive text method, the user presses the key that corresponds to each letter of a word once. The system uses a dictionary of words to determine which of the words the key sequence matches. As many words may share the same key sequence (ambiguous), users often have to press a NEXT key to scroll among the alternatives. On many mobile phones, the NEXT key is typically the asterisk-key (*).

Figure 1: Standard 12-keypad

Likewise, the predictive method is imperfect in some ways. Some commonly used words in SMS messages may require a higher number of NEXT key presses than those used rarely. Also, it is impossible to achieve 1 keystroke per character because some words require disambiguation. For example, the word “home” which shares the same key sequence as “good” requires one NEXT key press but the less common word “gown” does not require any NEXT key presses. In addition, a user can only input words in the dictionary using the predictive method. If a user requires to input a word not in the dictionary, one will have to revert to using the multi-tap method or insert the word into the dictionary. Once a word is inserted into the dictionary, the user can enter the word as per normal predictive method.

Currently, the T9 application is the most widely used implementation of this text entry method. However T9 does not implement word completion, whereby it would
complete a word before the entire word is spelt out. Other implementations such as eZiText by Zi Corporation and iTap by Motorola feature word completion.

1.3. Optimized Predictive Text

A notable shortcoming of the three text entry methods mentioned above is that all of them are static and do not adapt to a user’s word usage habits. Having a system that adapts to word usage would help improve text entry efficiency further because it would result in a significant reduction of the number of times a user presses the NEXT key to scroll to another word particularly when an ambiguous key sequence was input. For ambiguous key sequences, the system uses usage statistics to predict the most common possible word first. The optimized adaptive predictive text entry technique is the focus of this paper and is discussed further in Section 4.

For example, if a user wishes to form the word “gone”, using an English language predictive text enabled mobile phone, the key combination for this is 4663. This key combination also produces good, home, hood, hoof, hone and goof. To obtain the correct word, the users would have to enter the above key sequence and scroll twice.

2. Literature Review

2.1. Text Entry

An interesting technology utilizing prediction to aid in text input is POBox. POBox allows users to enter part of a word and then search for similar words by spelling, pronunciation, or shape (for pictograph-based languages) (Matsui (1998), Matsui (1999)). It uses a static database for the searching functionality. Reactive Keyboard, which is another predictive input technology monitors what a user enters and presents text predictions for the user to choose from (Darragh et al (1990), Darragh & Witten (1991). The predictions are generated by finding the longest matching sub-strings in the previously entered text. This technology is not only predictive, but it also adapts to the user’s input and hence is not limited to a static set of words or phrases.

Another predictive text input technology, T9 that has been licensed widely and is currently available on PDAs and cellular telephones. The T9 input technology is widely used in cellular telephones and uses the ubiquitous telephone keypad with the letter arrangement as defined by an international standard (ISO (1994), Tegic). As the user types, the text is not completely deterministic, because each keypress does not map to a single letter, but, instead, to one of three or four letters on each key. To overcome the ambiguity, word possibilities are presented as the user enters text. Although the disambiguating algorithm is proprietary, it likely uses character and word frequencies. Three evaluations of T9 exist (Silfverberg et. al, 2000, Bohan et. al, (1999), James & Reischel (2001)).

Predictive text input technologies rely heavily upon statistical language models and their ultimate performance is always limited by the quality of the model.
2.2. Evaluation Techniques

Tremendous interest in developing mobile text input techniques that are optimized or otherwise enhanced have brought about the need to have a mechanism to determine which text input technique is better. As a result, several authors have proposed standard metrics for evaluating text input techniques (MacKenzie, 2000), Soukoreff & MacKenzie (2003)).

Scott MacKenzie proposed a measure to characterize text entry techniques MacKenzie, (2000). It is calculated apriori, using a language model and a keystroke-level description of the technique. The measure is used to characterize and compare methods at the design stage, thus facilitating analyses prior to labour-intensive implementations and evaluations. This KeyStrokes Per Character (KSPC) is a metric for comparing the performance of text entry methods, KSPC is the number of keystrokes, on average, required to generate each character of text in a given language using a given text entry technique. MacKenzie systematically describes the calculation of KSPC and provides examples across a variety of text entry techniques. KSPC is useful for apriori analyses, thereby supporting the characterization and comparison of text entry methods before labour-intensive implementations and evaluations.

Using this technique the “Multitap” was used for English. It is based on the standard mobile phone key layout that meant one key was tapped several times to select the required character. Multitap was evaluated at 2.0342 KSPC. The “Qwerty” keyboard has a KSPC of 1.0000, since on this keyboard there is a key for each character for any given English word.

Word prediction is used to enhance this score, consequently reducing it to less than 1.0000 because entering all letters in a word is not necessary. Given a sufficient portion of a word to identify it in a dictionary, an interaction technique is invoked to select the full word and deliver it to the application.

3. Text Corpus Used

A text corpus is a collection of linguistic data, or written texts. Typically, corpora have to be quite large to be of any linguistic use. An important part of this research was to obtain a text corpus from which the dictionaries were developed.

Part of this work focused on local Kenyan languages. In the work discussed here, we report our findings based on Swahili only, which is a more widely spoken language in Kenya and in the east and central parts of Africa. We were unable to obtain a freely available text corpus, and collected our own. The corpus was analyzed to obtain an important statistic required to develop the dictionary. This statistic was the natural occurring frequency of a given word, based on the number of times the word appeared in the text corpus.

In developing the corpus, a conscious attempt was made to obtain text from various domains so as to have a corpus representative of the languages under study. These included the Holy Bible, Quran, stories, poems, news articles, government documents etc.
3.1 Building the data for the dictionary

Working with the text corpus collected for Swahili, we created a dictionary for this language. The word frequency was also extracted using our own Java program. Each unique word found was stored in a separate file along with its frequency. The resultant file was then examined and the less frequent words were removed from the wordlist.

4.0 Adaptive Predictive Text

As described earlier, predictive text input attempts to tackle word disambiguation given a key sequence. In Swahili for example lipia (pay on-behalf), kisha (thereafter) and lisha (feed) are all typed by pressing the key sequence 54742. Such an ambiguous sequence will therefore require the user to press the NEXT key to scroll through the ambiguous words until the required word is obtained.

As mentioned earlier, focus of this work is a user optimized adaptive predictive text entry. This is a move away from the rather static frequency based text prediction to prediction based on the most probable word given a user’s word usage habit. The system will therefore be able to adapt and learn a user’s word usage by adapting the natural word frequency.

This adaptability will result in a significant reduction in the number of times a user presses the NEXT in cases when the first word suggested is incorrect. The system does this by assigning weight values to all ambiguous words and updating the weights accordingly each time the user selects a word from the ambiguous list. The word with the highest weight corresponds to the most probable word for a given user and will be automatically suggested the next time the ambiguous key sequence is input.

Only ambiguous words with a given threshold weight will be suggested. The weights are updated using the sigmoid function, so named because it produces a sigmoid curve — a curve having an “S” shape. Often, sigmoid function refers to the special case of the logistic function shown in Figure 2 and defined by the formula:

\[ f(x) = \frac{1}{1+e^x} \]

The weight will be changing by \( f(x) \) where \( x \) represents the number of times NEXT key was pressed to get to the given word.

i.e., \( \text{newWeight} = \text{oldWeight} + \frac{1}{1 + e^{\text{keytaps}}} \)

\( f(x) \) is inversely proportional to the value of \( x \) i.e. the smaller the \( x \) value, the higher the value of \( f(x) \). keytaps is an integer representing the number of times the NEXT key was pressed to scroll through. keytaps decreases with the number of key presses. The implication of this is that the more ‘inaccessible’ words will always be having their weights updated by a larger factor.
In general, a sigmoid function is real-valued and differentiable, having a non-negative or non-positive first derivative, one local minimum, and one local maximum and are often used in neural networks to introduce non-linearity in the model and/or to make sure that certain signals remain within a specified range. This function was used to update the weights because it ensures that the weights are bounded (between 0 and 1) and that the change in weight is gradual (smooth).

The weight update algorithm used in adaptive predictive text is represented in pseudo code as follows:

```plaintext
For each keyseq
    Possible_words = All words in Dictionary corresponding to keyseq;
    IF Possible_words > 1
        Add words corresponding to keyseq to Ambiguous list
        Sort Ambiguous list by weight, max to min
        Integer K = 0;
        WHILE keyPressed == NEXT
            Suggest word K in Ambiguous;
            K = K + 1;
        End WHILE
        IF keyPressed == SPACE
            Search for word Ambiguous [K] in Dictionary;
            // Update weight for Ambiguous [K];
            NewWeight [K] = oldWeight [K] + 1/ (1 + e^keyTaps)
        End IF
    End IF
End For
```

An advantage of this word weight update scheme is that its overheads are minimal. An operation such as updating the word frequency is completely avoided as it would require more resources.
5.0. Design

As the interface was set by the current mobile phone systems, its layout was relatively constrained, along with the buttons that could be used. There were few design choices to be made for the traditional multi-tap system, as this was a control from which evaluation results could be measured, and therefore had to copy the current system. However, choices needed to be made for the dictionary-based system. There had to be methods to perform word selection (changing the suggested word), and word completion, as well as handling punctuation.

5.1. Word Selection

Word selection involved the system suggesting a word, and the user accepting it by entering a SPACE or punctuation mark, or telling the system it is the incorrect word, by pressing the STAR key. The space or punctuation marks signify the word is correct as the user has now entered the next character in the message, by entering a character that finishes a word (an apostrophised word can be treated as two separate words separated by an apostrophe). Pressing the STAR key notifies the system that the suggested word is incorrect and at the same time suggests the next most probable word. (The criteria used for this may be frequency, weight values or simply alphabetic order depending on the system).

5.2. Word Completion

Pressing the SPACE button after a partially input word is auto-completed performs the Word Completion operation. This was chosen as the user is again confirming a suggestion and saying that the suggested word is the correct word.

5.3. Auto capitalization

The developed system takes advantage of the fact that the first character of the first word in every sentence is always capitalized. The system performs automatic capitalization on the word immediately a terminal punctuation (full stop, question mark, exclamation mark, full colon etc.) is noted. This is expected to reduce the number of keystrokes even further. However, this automatic capitalization dictated by a terminal punctuation is not always correct and may lead to unnecessary capitalization.

5.4. Design

In order to realize the above design choices various software modules were conceptualized to perform specific functions. The conceptual analysis models were then transformed into a specific design, which is physically realisable. In this section we review the various basic design issues of the system. The methodology used in the design was Object Oriented Design (OOD). This methodology entailed the transformation of the conceptual models into Classes, Relationship diagrams and Object Interaction diagrams. The development platform used J2ME (Java2 Mobile Edition) which is fully object-oriented hence the choice of methodology.
6.0. Results

For the purpose of this work, three different implementations of text prediction were developed. These are:

- Basic Predictive Text
- Frequency Based Prediction
- Adaptive Predictive Text

This section gives a detailed account of the findings obtained after comparing efficiency of the three input methods.

The KSPC of each of the methods was used as the basis of comparison. Since KSPC is an average, a more accurate value would be obtained by analyzing as much text as possible. SMS messages are short (usually less than 160 characters) and it is therefore impractical to type in 100 pages of text, for example, just to calculate the number of keystrokes required and hence compute the KSPC.

A more practical way of doing this is through simulation. With this in mind, a program was developed that would compute the required metrics given any amount of text in a text file. The program parses through a file and outputs the number of words in the file, number of characters, the number of key taps that would be required to type in the next, and most importantly the program computes the KSPC of the input method used. In addition, the program outputs a word-frequency file, which is required by the Frequency Based technique. Frequency based prediction makes use of words natural frequency as the basis of disambiguation and therefore a word-frequency file is of utmost importance. Compared to actual typing, simulation is much faster as well as more efficient and allows more comprehensive analysis to be done.

6.1. Analysis

The results obtained from the simulation were validated by comparing with those obtained manually from typing in a small piece of text (the Kenya National Anthem, Swahili). Ideally, an SMS corpus would have been better to use, but this was not available. The result from the two compared fairly well hence the simulation model was validated. The model was then used to perform text input simulation for three text input methods. A Swahili poem was used as the input text. The poem had a total of 159 words but only 96 of them were unique. The repeated words would form a good model of ambiguous words that require weight update.

The simulation was repeated several times with the same text until no significant change in KSPC could be observed. The cumulative results for 17 iterations are summarized in Table 2.
<table>
<thead>
<tr>
<th>Iteration Number</th>
<th>Number of Words</th>
<th>Multi-Tap Method</th>
<th>Frequency-Based Predictive Text</th>
<th>Adaptive Predictive Text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No of Keytaps</td>
<td>KSPC</td>
<td>No of Keytaps</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00000</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>159</td>
<td>2076</td>
<td>2.14685</td>
<td>1286</td>
</tr>
<tr>
<td>2</td>
<td>318</td>
<td>4152</td>
<td>2.14685</td>
<td>2572</td>
</tr>
<tr>
<td>3</td>
<td>477</td>
<td>6228</td>
<td>2.14685</td>
<td>3858</td>
</tr>
<tr>
<td>4</td>
<td>636</td>
<td>8304</td>
<td>2.14685</td>
<td>5144</td>
</tr>
<tr>
<td>5</td>
<td>795</td>
<td>10380</td>
<td>2.14685</td>
<td>6430</td>
</tr>
<tr>
<td>6</td>
<td>954</td>
<td>12456</td>
<td>2.14685</td>
<td>7716</td>
</tr>
<tr>
<td>7</td>
<td>1113</td>
<td>14532</td>
<td>2.14685</td>
<td>9002</td>
</tr>
<tr>
<td>8</td>
<td>1272</td>
<td>16608</td>
<td>2.14685</td>
<td>10288</td>
</tr>
<tr>
<td>9</td>
<td>1431</td>
<td>18684</td>
<td>2.14685</td>
<td>11574</td>
</tr>
<tr>
<td>10</td>
<td>1590</td>
<td>20760</td>
<td>2.14685</td>
<td>12860</td>
</tr>
<tr>
<td>11</td>
<td>1749</td>
<td>22836</td>
<td>2.14685</td>
<td>14146</td>
</tr>
<tr>
<td>12</td>
<td>1908</td>
<td>24912</td>
<td>2.14685</td>
<td>15432</td>
</tr>
<tr>
<td>13</td>
<td>2067</td>
<td>26988</td>
<td>2.14685</td>
<td>16718</td>
</tr>
<tr>
<td>14</td>
<td>2226</td>
<td>29064</td>
<td>2.14685</td>
<td>18004</td>
</tr>
<tr>
<td>15</td>
<td>2385</td>
<td>31140</td>
<td>2.14685</td>
<td>19290</td>
</tr>
<tr>
<td>16</td>
<td>2544</td>
<td>33216</td>
<td>2.14685</td>
<td>20576</td>
</tr>
<tr>
<td>17</td>
<td>2703</td>
<td>35292</td>
<td>2.14685</td>
<td>21862</td>
</tr>
</tbody>
</table>

As can be observed in Table 1, the KSPC value for Multi-tap method remained constant at 2.15 across the 17 iterations. Similarly, the KSPC value for Frequency-based predictive text was constant at 1.33. However, the KSPC value for adaptive predictive text tends towards 1 across the iterations (see Figure 3). Note that even in the very first iteration, the KSPC value for adaptive prediction is slightly less than that of Frequency-based prediction. This is because the text used had a lot of repeating words as mentioned earlier (only 96 unique words out of 159). The results in Table 1 are graphically represented in Figure 3.

An interesting and surprising observation was that Swahili had a lesser disambiguation problem as compared to English.
7.0 Conclusion

The adaptive and optimized predictive text system provided in this work was successful. An evaluation has been carried out with the delay-based system acting as a control, and the results have suggested a number of useful features of the system provided:

- Repeated key presses are mainly avoided.
- Delays waiting for the cursor to move on are mainly avoided.

The complete word features can reduce the time taken and the effort required to type a message.

Figure 3: Comparison of text entry techniques based on the number of words.

<table>
<thead>
<tr>
<th>Interaction Technique</th>
<th>KSPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-Tap</td>
<td>2.15</td>
</tr>
<tr>
<td>Frequency Based Predictive Text</td>
<td>1.33</td>
</tr>
<tr>
<td>Adaptive Prediction without Auto-complete</td>
<td>1.16</td>
</tr>
<tr>
<td>Adaptive Prediction with Auto-complete</td>
<td>0.89</td>
</tr>
</tbody>
</table>
From comparison of the three systems developed, it can be seen that the adaptive-based system outperforms the other two. It relies on words already being in the dictionary. However, some users may require extra help in using the dictionary-based system, so it does not seem to be a good system for absolute novices.

7.1. Future Work

Below are suggestions for future work:

1. Improvement on the prediction using the concept of bi-grams (words pairs), and even n-grams. Instead of developing a bi-gram dictionary (i.e. a dictionary containing all words along with the next most probable word) we would do an analysis of SMS messages and look for word clusters. If a user incorporates several words in a SMS, one could weigh the other words more heavily. The use of this clustering concept is necessitated by the fact that a mobile phone may not have adequate memory to store a n-gram dictionary. This may improve the prediction further by eliminating suggestions of similar ambiguous words that are not in the bi-gram list.

2. An interesting technique worth exploring is context-based disambiguation i.e. disambiguating words based on the context.

3. Implement this technique for other local languages.

8.0. References


Revisiting Dynamic Synthesis Methodology: A reference theoretical framework and tool for business process modelling and analysis

Ddembe Williams, Department of Computer Science, Faculty of Computing and Information Technology, Makerere University

The case for integrating research methods generally, and more specifically that for combining system dynamics and case study research approaches in business process modelling is strong. Yet, research designs that extensively combine both system dynamics modelling and case study in process modelling are rare. This paper aims to provide a useful and systematic reference point for researchers who wish to work in the business process modelling and generally to encourage careful work on the conceptualisation and execution of Dynamic Synthesis methodology in business process and to a wider process-modelling field. The paper suggests that the potential usefulness of the Dynamic Synthesis Methodology is in aiding researchers and managers in improving both building and testing theories in the understanding business processes and strategic modelling and analysis.

1. Introduction

This paper proposes a model-based business process modelling and analysis methodology that brings together two complementary methods using system dynamics modelling approach. Simulation modelling and case study are powerful research methods, for modelling and analysis whose added advantages can complement each other in terms of theory building, testing and theory extension. The paper addresses the philosophical and theoretical issues concerning the nature of combining the case study and process simulation research approaches, and methodological issues on conduct and reporting of this type of research design in business process problem solving methodology. Triangulation of methods while not new, has not been applied in business process modelling and analysis research. Problem solving, in this paper is used loosely to refer to the process of investigating variables ranging from specific, short-term and well-defined issues to more general long-term and ill-defined issues in attempt to generate some improvement (Keys, 1983). Studying a real world sequence of events in the phenomenon of interest provides a foundation for understanding and explanation of how particular outcomes in the business processes modelling are arrived at. In order to aid such an understanding of the business processes, this paper proposes a problem solving model-based paradigm – Dynamic Synthesis Methodology (DSM). The paper suggests that the potential usefulness of the Dynamic Synthesis Methodology is in aiding researchers in improving both building and testing theories in the understanding business processes and strategic modelling and analysis.
The rest of this paper is divided into five sections. Section 2, discusses the philosophical underpinning of the Dynamic Synthesis Methodology. Section 3 introduces the basis for triangulating case study and simulation research methods, while section 4, introduces the detailed reference study. Section five discusses the potential contributions of Dynamic Synthesis methodology as a research and practice-based methodology.

2. Philosophical Basis for Dynamic Synthesis Methodology

Dynamic Synthesis Methodology refers to the integration of theoretical concepts and structuring of parts and elements of a research process over time in such a manner to form a formal functional entity, underpinned by synthesis as philosophy of science. Within Dynamic Synthesis Methodology case study research method is complemented with system dynamics modelling to form a framework for aiding empirical explanation and prediction of the behaviour of complex systems often observed in business process performance.

In business process modelling, and requirements engineering process in particular, synthesis may be considered in the specification activities when checking for logical correctness and at the problem design stage where quantitative analysis is carried out concurrently with qualitative analysis (Williams and Kennedy, 2000). Business process modelling and analysis research, particularly for large-scale projects, is a complex process. For example a great deal of work has been carried out on process-based approaches to requirements engineering (Pohl, 1993), but very little has been done in utilising dynamic process-based tools. This paper contributes to developing such an understanding by proposing a methodology that facilitates dynamic modelling and analysis of business process. Both simulation and case study research methods have been used in the development of theories in the business process modelling of the software development process. However, triangulation of both methods to study the requirements engineering and its business process modelling and analysis phenomenon has not been used before (Finkelstein, 2000). The next section discusses philosophical issues of triangulating simulation modelling and case study research methods for studying in the business process-modelling field.

2.1. Applied Research Design

In model-based systems modelling, traditional research approaches and pure science generally take the format of: Observation, Experimentation, and Generalisation. Notable researchers have proposed variations of applied research methods like (D’Abro, 1951 pp.3):

- observation stage
- experimentation stage
- the theoretical and mathematical stage (in Physics).

While Ackoff (1962, pp.26) proposes expansion of the three stages into ‘cyclic’ six phases:
1. Formulating the problem,
2. Constructing the model,
3. Testing the model,
4. Deriving a solution from the model,
5. Testing and controlling the solution, and
6. Implementing the solution.

Early discussion on primacy of observation or theory was reported in Churchman, (1961). Modern philosophical analysis shows that observation always presupposes a criterion of relevance. Ackoff (1962) observes that, “some modern philosophies of science, assert the cyclic and interdependent characteristic of these stages of scientific method” (pp.26). Table 1 presents a comparison of pure and applied research approaches to problem solving as examined by D’Abro (1951) and Ackoff (1962).

<table>
<thead>
<tr>
<th>Pure Research</th>
<th>Applied Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>Formulating the Problem</td>
</tr>
<tr>
<td>Generalisation</td>
<td>Model Construction</td>
</tr>
<tr>
<td>Experimentation</td>
<td>Derivation of the Solution</td>
</tr>
<tr>
<td></td>
<td>Testing the Model</td>
</tr>
<tr>
<td></td>
<td>Implementing the Solution</td>
</tr>
</tbody>
</table>

Ackoff contends that “the control phase of applied research corresponds with efforts to further generalise the results of pure research, and implementation corresponds with efforts to use the results of a piece of pure research in another” (pp.26). This paper extends Ackoff’s (1962) applied methodological research design strategy, modified by integrating Yin’s (1984) case study research as a basis for theory development. The applied methodological research approach described by Ackoff (1962) is similar to that proposed in this paper for dealing with systems-based problems (Richardson, 1981). The advantage of modelling in the system dynamics tools’ environment is its iterative nature that yields understanding, which forms a basis for further analysis, theory testing and extension (Meadows, 1982).

The advantage of a systemic paradigm is that it can be used to explore (description) and explain (prescription) both living and those categorized as human activity systems is an important aspect that may shed light on the treatment of causality in System dynamics (Philips, 1987). This is a positivist position that uses an applied science method to investigate existence of causal relationship between business processes of interest. (Williams, Hall and Kennedy, 2000). The above concept underpins the importance of DSM in solving differing views of reality in business processes as part of systems requirements engineering. This is an important characteristic of the DSM
as it makes its philosophical objectives very explicit, as the focus of the methodology is not necessarily computerisation but also improvement in business process modelling, which in practice may not require computerised changes. The overwhelming objective is to provide improved understanding of social theories, literature and stakeholder concerns. The DSM proposed in this paper extends Ackoff’s (1962) applied research approach where he argues that the study of scientific method has been a major part of the philosophy of science. Ackoff (1962) makes a claim that this branch of philosophy has at least three other types of science including:

- conceptual analysis: the attempt to define concepts or problem areas in such a way amenable to scientific study;
- examination of assumptions: concerning the reality that ‘underlies’ science;
- synthesis: the attempt to fuse the findings of various branches of science into one consistent view of reality, “weltanschauung” (pp. 27).

It is the third branch of philosophy of science (synthesis) that is adopted in this paper to explain the emergent properties of the whole business process modelling and requirements engineering process improvement based on systems theory, servomechanisms theory and measurement theory as theoretical anchors. Synthesis as a method of performance evaluation identifies the behaviour inherent of a system of which the emergent properties to be explained is part. Table 2 compares two modes of designing problems, analysis and synthesis in aiding the understanding of the emergent patterns meaningful process modelling and improvement. Ackoff (1981) uses Descartes’ principles of 1637 to raise the concept of reductionism as fundamental to problem solving. In systems analysis, of which business process modelling is a major activity as illustrated in Table 2, parts are put together by synthesis in order to understand why things operate the way they do (1981).

**Table 2: Understanding facilitates explanation of the Whole [adapted from Ackoff, 1981]**

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Synthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focuses on structure, it reveals how things work</td>
<td>Focuses on functions, it reveals why things operate as they do</td>
</tr>
<tr>
<td>Yields Knowledge</td>
<td>Yields understanding</td>
</tr>
<tr>
<td>Enables description</td>
<td>Enables explanation</td>
</tr>
<tr>
<td>Looks into things</td>
<td>Looks out of things</td>
</tr>
</tbody>
</table>

Table 2 illustrates that gaining an understanding of the social theory and the factors that lead to its effective explanations is the prerequisite for improving the our literature, methods and concerns. The advantage of dynamic synthesis methodology is in its ability to incorporate the results of earlier research (Churchman and Ackoff, 1950).
An effective business process modelling research methodology and tool should be able to capture both informal and fuzzy concepts often relevant to social theory or models of the social process involved. A good modelling methodology should lend itself to automation, the pseudocode generated by SD software provides a basis for formalising these often-diverse concepts. The concept of DSM as a business process modelling–based research methodology is underpinned by the relevant theoretical anchors, including: General Systems Theory, Servomechanisms Theory, and Measurement Theory.

The theoretical concepts discussed in this section aid the positioning data collection in its contextual setting. Van Maanen (1983) asserts that one may not describe the observed behaviour of a phenomenon until they have developed a description of the context in which the behaviour takes place and have attempted to see the behaviour from the position of the problem owner. Theories that anchor the DSM conceptualised in this section serves as a useful foundation to stimulate and organise research efforts in business process modelling. In order to position this paper in the epistemological theory of knowledge (Churchman and Ackoff, 1950), dynamic synthesis methodology has potential to play an important role in increasing the acquisition of knowledge, so as to bring together a breadth of techniques to facilitate understanding of business process modelling problems. The domains of these philosophical and theoretical characteristics in the applicability of DSM as a model-based problem solving, for business process modelling and analysis as described in the reference study.

3. Triangulation of case study and system dynamics methods
The case for combining methods generally, and more specifically that of combining qualitative (case study) and quantitative (System dynamics modelling) is strong (Klein et al, 1993, Wynekoop, 1992; Galliers; 1984; Visala, 1991; Gable 1994; Kaplan and Ducheon, 1988 and Orlwowski and Baroudi, 1991; Ives et al, 1980). Wynekoop (1992) suggests that ‘micro level’ quantitative analysis should be integrated with qualitative ‘macro-level’ analyses, in order to understand the ways in which individual variables’ behaviour have an impact on organisational phenomenon. Gable (1988) takes the same position by adding that “the ways in which macro phenomenon has effects through individual variable may be explicated” (pp. 115).

The notion of combining qualitative and quantitative research methods has been said to increase the robustness of results. Findings can be strengthened through cross validation achieved when different kinds and sources of data converge and are found to be congruent (Kaplan & Ducheon, 1988; Gable, 1988). Klein et al (1991) and Galliers (1984), calls for tolerance of methodological pluralism and recognition of method and personal bias. Orlikowski and Baroudi (1991) claim that given human limitations, individuals must specialise in a limited number of methods. For example, three methods have tended historically to dominate information system research: survey, laboratory experimentation and case study.

Simulation modelling is a form of laboratory experimentation with very high levels of constraints. Visala (1991) contributes to the debate by proposing a conceptual
framework to help overcome the gap between positivist and interpretative research approaches. Visala’s work is adapted from the information systems research framework proposed by Ives et al (1980). Visala (1991) cross-references epistemological approaches (causal model, technological explanations, hermeneutics, dynamic structure models, formal methods and phenomenology) with classes of variable of interest in Information Systems research proposed by Ives et al (1980). However Visala does not explore how specific research methods, for example simulation and case study, ought to be operationalised in combinations. Williams, 2000, 2001 urged the combination of case study and simulation methods in Requirements Engineering process modelling and analysis, in order to improve the stakeholder’s understanding of how requirements changes over time.

3.1. Case Study Research Method

A case study is an empirical investigation that probes and examines responses of convenient influences within the real operational environment of the task, user, and system. The case study approach generally refers to group methods, which emphasise qualitative analysis (Yin, 1984; Gable 1988), although some case studies are quantitative in nature. In the SD literature quantitative case studies have been used to validate SD simulation models (Senge and Forester, 1980; Graham, Morecroft, Senge and Sterman, 1982). However, the unit of analysis may differ between case study and Process Simulation modelling, for example in a RE process at simulation level, and information systems development project or IS use at Case Study level.

Data collection from Software Development Organisations (SDOs) thorough techniques such as surveys, interviews, observation and documentation analysis may vary. The case study method seeks to understand the problem being investigated. Gable (1988) articulates that the word understand is used in the phenomenological or hermeneutic sense, where “understanding” the meaning held by a subject or group is contrasted to ‘explanation’ produced by a scientific observation (pp 113). This position is consistent with that of Graham, Morecroft, Senge and Sterman (1992). The proponent and advocate of qualitative methods, Yin (1984), suggest that case study is appropriate where the objective is to study contemporary events, and where it is not necessary to control behavioural events or variables. This lack of experimental control in case studies makes it difficult to minimise possible confounding of variables effects than in a formal experiment. Perhaps most importantly Yin (1984) suggests that single case studies are appropriate, if the objective of the research is to explore a previously not researched subject vehicle but multiple case studies are desirable for description, theory building, or theory testing. Sage (1991) suggests “there are no really good methodological frameworks for conduct of case studies or use of case study research” (pp193). The qualitative bias of case studies may be contrasted with the quantitative bias of System dynamics modelling.

3.2. System Dynamics Modelling

System dynamics modelling is the technique of constructing and running a model of an abstract system in order to study its behaviour without disrupting the environment
of the real system. Simulation is the process of forming an abstract model from a real situation in order to understand the impact of modification and the effect of introducing various strategies on the situation. Pidd (1992) describes simulation as an approach, which involves experimentation on a computer based-model in a trial and error way. It is a process of imitating important aspects of the behaviour of a system in real, compressed time or expanded time by constructing and experimenting with a model of the system. Simulation is one of the main strategies used in this research. The main objective of the simulation program is to aid the user, most often an operations research specialist or systems analyst, in projecting what will happen to a given situation under given assumptions. It allows the user to experiment with real and proposed situations otherwise impossible or impractical.

Hill (1996) describes the process of simulation as carried out by creating an abstract of a real system (model) to evolve in real time in order to assist the understanding of the functioning and behaviour of this system and to understand certain of its dynamic characteristics, and with the aim of evaluating different decisions. This technique therefore allows simulation of the operation of existing or non-existing systems, for a given work load (studies of transient operations, testing of different strategies). The assumptions made in order to simulate some process tend to be a key to the accuracy of the results. The more simplified the results, the less reliable the results, whereas the more details included, generally the better the results. No matter how detailed the program, the prediction obtained will be inaccurate if the initial assumptions are incorrect. In this case, the simulation program becomes invalid as a prediction tool. Simulation experiments have a series of advantages that can be used in various applications areas. Advances in hardware and software technology have greatly reduced the two major weaknesses of simulation, cost and the unavailability of the data necessary for building and validating the model. The integration of simulation modelling and case study research methods provides a conceptual framework for a dynamic synthesis research methodology.

In Information Systems (IS) Janvenpaa (1988) and Galliers and Land (1988) suggest the use of alternative IS research approaches in the process of theory building, testing and extension. The conceptual synthesis proposed here (Figure 1), extends Vogel and Wetherbe’s (1980) criteria of parsimony and completeness. It also draws on the application areas reported in Galliers and Land (1987) as a foundation for a research method, being a paradigm suitable in context of System dynamics. The Dynamic Synthesis Methodology is grounded on well-tested and developed theoretical anchors and builds on an existing epistemological philosophy of science in the acquisition of knowledge (Churchman and Ackoff, 1950), as a basis for theory building and extension in the field of System dynamics.

4. The Reference Study

This paper proposes revised research strategy that combines System dynamics modelling (Forrester, 1961; Richardson and Pugh, 1981) and case study research method. (Galliers, 1984; Mason and Mitroff, 1973, Yin, 1984). The Case study research
method (Yin, 1984) involves identifying key factors that may affect the outcome of an activity and then to document the activity’s inputs, constraints, resources, and outputs. Simulation experiments (Law and Kelton, 1988; Coyle, 1996) are rigorous controlled activities, where key factors are identified and manipulated to document their effects on the outcome. The usefulness of the system dynamics (SD) modelling in DSM is its ability to capture ‘hard’ and ‘soft’ concepts into a formal model, thus bringing together theoretical constructs that impact on the research phenomenon of the system dynamicists (Bell and Bell, 1980; Lane, 2001). System dynamics currently does not guide researchers how to carefully collect data that is relevant to the phenomenon of interest. In order to attain the specified mechanism, a research design strategy for collecting and synthesising data and information must be adopted. As Popper (1972) puts it “the activity of understanding is essentially the same as that of problem solving”. This concept reflects to the domain of business process modelling particularly where paradigms that support social and natural sciences may play common roles for effectiveness of research methodology. Kuhn (1970) proposes a need to change our conception of science to “more appropriately” in terms of problem or puzzle solving. This is in line with Ackoff’s (1962) applied research approach and extended in the DSM proposed in this paper.

In the SD literature, researchers have proposed the use of either case study or purely problem simulation modelling, however researchers have not stated the advantages of combining the two research methods (Williams, 2000; 2001a/b). Jick (1983) on the other hand underscores the desirability of mixing methods given the strengths and weakness found in each single method design. While other researchers have examined the relative merits of quantitative (simulation modelling) and qualitative (case study) debate.

Combining case study and System dynamics modelling makes Dynamic Synthesis Methodology a powerful empirical research method that potentially makes useful contribution to body of System Dynamics. 134 successful PhD research dissertations in Information System, between 1971 and 1973 by Manson and Mitroff, (1973) and Ives et al (1980) support combinations of this sort. Figure 1 presents the overall reference study for the DSM proposed in this paper. The DSM includes six iterative research process phases, namely: Problem statement, Field Studies, SD Model Building, Case Studies, Simulation Experiments and Model Use and Theory Extension.
The case study (Yin, 1984; Munford et al, 1985; Benbasat et al 1987; Lee, 1989; Smith, 1990) and Simulation (Graham, Morecroft, Senge and Sterman 1992; Pidd 1992; March et al, 1992; Roberts 1983; Hill 1996; Meadows 1983; Law and Kenton, 1988; Morecroft, 1979, 1988) methods have seen extensive application in software and systems research. Based on figure 1, the next section discusses each of the six phases of the Dynamic Synthesis Methodology.

4.1. The Problem Statement

The word “problem statement” is used in preference to research question(s) because modelling and analysis being part of System dynamic requires solving problems rather than answering questions. System dynamics modelling refers to a ‘problem’ to be solved (Richardson and Pugh, 1981), while other qualitative methods like case studies refer to the “research question” (Lane, 2001; Checkland and Scholes, 1990. The statement of the problem is an important early phase in modelling and analysis. Kerlinger (1986, pp.16-17) lists several characteristics of a good problem statement:
• The problem statement should state the expected relationships between variables (in experimentation this is a causal relationship).
• The statement of the problem must at least imply the possibility of an empirical test of the problem (hypothesis).
• The problem should be stated in form of a question.
• Reviewing the state of the art relevant to the problem statement refines the problem. Field studies may be necessary, through a pilot case study, in order to understand important variables.

4.2. Field Studies

Field studies and supporting data collection methods provide invaluable insights and discoveries during the System dynamics modelling. Field study is a term that applies to a variety of research methods, ranging from low to high constraints. These methods share a focus on observing naturally occurring behaviour under largely natural conditions. Curtis et al (1988) have used field studies to study large-scale systems in software development. Like field study research falls near the low-constraint end of Graziano and Raulin's (1996) classification presented in table 3.

Table 3: Low-High Constraints Categories of Field Research [After Graziano and Raulin, 1996]

<table>
<thead>
<tr>
<th>Naturalistic Observation</th>
<th>Direct observation of events as they occur in natural world</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archival research</td>
<td>Studying information from already existing records made in natural settings.</td>
</tr>
<tr>
<td>Survey</td>
<td>Asking direct questions of persons in natural settings.</td>
</tr>
<tr>
<td>Case Study</td>
<td>Making extensive observation of a single group or person.</td>
</tr>
<tr>
<td>Program Evaluation</td>
<td>Conducting evaluation of applied procedures in natural settings</td>
</tr>
<tr>
<td>Field experiments</td>
<td>Conducting experiments in neutral settings where casual inferences are sought.</td>
</tr>
</tbody>
</table>

As illustrated in Table 3, the process modelling field studies may take different forms, but all low-constraint field research constitute naturalistic observations, archival research, case studies, and surveys, while high-constraint include programme evaluation and field experiments. These low-constraint field research methods are not necessarily inferior to higher-constraint research. The appropriate level of constraint in any research and indeed in a business processes depends upon a number of factors, the most important of which is the nature of the process research questions or problem statement.
Gable (1988) urges that fieldwork is a poor method for objectively verifying hypotheses. However Attwell and Rennle (1991) suggest that the use of data collection techniques like the survey is strong in areas where field methods are weak. The strengths of field studies are the collection of data, and description of the phenomenon in its natural settings. Surveys, semi-structured interviews, participant observation and document analysis are data collection techniques used in this paper and recommended in the DYNASIS Descriptive Framework. Field studies are used to collect on site information on the current systems; process owners and required proposed system are gathered to facilitate identification of user and specification of system requirements, and constraints. Input and output information to activities identified in a Descriptive process model resulting from field studies are used to identify activities, resources and products used by the process. Data on processes, resources and product are used to develop a generic system dynamics model.

4.3. System Dynamics Model Building

System dynamics model development is a system stage process that begins and ends with understanding. The result of field studies should provide a descriptive model, on which SD conceptual feedback structure can be developed. The feedback structural model is developed with the help of a causal loop diagram. The next stage is the conversion of the cause loop diagram into stock and flow diagrams, which is a formal quantitative model of the problem in question. In order to simulate the model, we must define the mathematical relationship between and among variables. Pugh and Richardson (1981) suggest that a system dynamics modelling effort begins and ends with understanding. Simulations can then be run on the important variables. Once confidence is gained, through validation and ownership by RE process stakeholders, then the model is available to test hypotheses or policies of interest. In some cases a developed model can be validated, using a post-mortem analysis or cases studies in RE processes, for prediction or prescriptive purposes.

4.4. Case Studies

Case Study is an exploratory (single in-depth study) or explanatory (cross-case analysis) research strategy, which involves an empirical investigation of a particular contemporary phenomenon within its real life context using multiple sources of evidence. Apart from case study, other sources of information may facilitate the practical grounding of theories in the RE process domain. To attain reliable and high quality data of high-level requirements, RE project data need to be captured. The forms specifically elicits information on RE projects’ resources, processes and products, their cost and time expended. Example of these RE process metrics can be: number of changes, the decisions taken and why, number of reviews, errors discovered, total number of requirements, requirements traced, rejected, frozen, documentation size, number of pages, document production cost, cost per page. The duration of each activity within the RE process, and any other factors that the requirement team thinks might have an influence on the way the projects are managed, are collected and documented for analysis and simulation experimentation.
4.5. Simulation Experiments

Simulation models are abstractions of the real worldview of a system or problem being solved. Shannon (1998) defines simulation as “the process of designing a model of a real system and conducting experiments with this model for the purpose of understanding the behavior of the system and/or evaluating various strategies for the operation of the system”. During the course of simulation, the model mimics important elements of what is being simulated. The model is used as a vehicle for experimentation in a “trial and error” way to demonstrate the likely effects of various policies. Those policies, which produce the best result in the model, will be implemented in real life. It is sometimes necessary to study the behaviour of a system in order to find answers to a problem or predict the possible outcome of policies adopted but it may be impossible or impractical to experiment on the real system. In such situations, simulation can be an effective, powerful and universal approach to problem solving in different areas of application (Matko et al, 1992), to extend existing theories or identify new problems (Williams, Hall and Kennedy, 1999).

4.6. Model Use and Theory Extension

The System dynamics modelling approach takes a philosophical position that feedback structures are responsible for the changing patterns of behaviour we experience in complex problems (Richardson and Pugh, 1981). System Thinking/ System Dynamics notation can be used to model and test different hypotheses propounded about RE process modelling and analysis. A dynamic hypothesis, can be tested verbally, or as a causal loop diagram or as a stock and flow diagram. In view of SD’s capacity to deal with the dynamic complexity created by interdependencies, feedback, time delays and non-linearity. Steer (1992) suggests that it should be used to complement traditional scheduling and project management tools. Figure 2, provides a detailed description of DSM and how the resulting products of each research phase contribute to theory building and testing. Earlier sections discussed the merits and objectives of the dynamic synthesis methodology presented and overall reference research design. The main objective of Figure 2 is to depict input and outputs of each phase and their relationships and outputs and to identify data requirements and data collection methods of each phase, and validation and verification points within the methodology.

As a process modelling and analysis methodology, DSM can be used to study a range of problems in both fields of science (systems engineering, operational research, control science) and social science (i.e. behavioural science, political science and management science). However, intended problem solvers need to acquire the richest possible understanding of organisational problems if they are to become effective problem solvers. (Jayaratna, 1984). An analysis of the strength and weakness of case study and process simulation methods suggests complementary sources of evidence and ideas appropriate to business process modelling and analysis. While case studies can be used to capture reality description, simulation models can be used to build theories or test them.
Figure 2 aids different stakeholders to view requirements engineering process problems from shared contexts to facilitate shared understanding and common ownership of problems.

A Descriptive framework is a set of theories and associated experimental evidence and field studies concerned with how actual business process stakeholders perform the process modelling and analysis. The dominant features of the descriptive theory are the limitations in cognitive abilities of decision makers, (Kleindorfer et al, 1993).

**Figure 3: Detailed Reference Model for Dynamic Synthesis Methodology**

DYNASIS Prescriptive framework is an integrative holistic approach aimed at improvement and not necessarily optimisation of RE process stakeholders’ decisions as depicted in Figure 4. Figure 4 aids different stakeholders to view requirements engineering process problems from shared contexts to facilitate shared understanding and common ownership of problems.

DSM as an applied research methodology has to demonstrate its usability in practice (Jayaratna, 1994). While customers may understand the problems that need to be solved, problem solvers or researchers may refine those problem statements depending on known problems. The Dynamic Synthesis Methodology uses a general model, (Figure 1) to illustrate the relationship and interconnections between different phases and artefacts as illustrated in Figure 4. The phases, their relationships and resulting artefacts are dynamic. The degree of connections depends not only on time and space but also on
the process stakeholder. This means different people perceive different connections and process problems differently and would make different choices to solution approaches (Kleindorfer et al, 1993). As depicted in the reference study in Figure 3 and figure 4, the key support for DSM is the DYNAMıc syntheSIS (DYNASIS) Tool.

Figure 4: Realisation of DYNASIS Tool to Support business process modelling and analysis

As illustrated in Figure 4, Increasingly, methodologies are becoming automated through the use of tools and models. A good process modelling methodology should lend itself to automation. The DYNASIS Tool to support business process modelling in the context of activities, technology and stakeholders (people), transforming customers’ needs, into the finished business process model and its likely performance. In respect to the proposed DYNASIS tool, credibility is as important as validation in terms of actual implementation of simulation results, although credibility has not been discussed in great detail in simulation literature (Carson, 1986; Law and Kelton, 1991). In respect to DYNASIS tool, animation of the business process is an important aspect of establishing credibility, “since animation is an effective way for the researcher and other RE process stakeholders to communicate the essence of the model to the manager” (Law and Kelton, 1991, p. 299). STELLA one of the system dynamics tools has the capability to animate the stock-and-flow diagrams (HPS, 1996-2000).

The benefits of DSM over case study or process simulation research methods are due to the fact that it is well structured for effective data collection purposes. Yin (1984), suggest that case studies are appropriate where the objective is to study
contemporary events, and where it is not necessary to control behavioural events or variable. Case study as a qualitative research method is difficult to manipulate variables independently and have a high risk of interpretation (Kerlinger, 1986). On the other hand, simulation does not generate optimised solutions, such as those obtained by linear programming or other analytic methods and provides statistical estimates, not exact results. Each simulation model is applicable to a particular situation and transferring the results to other problems is generally not difficult (Law and Kelton, 1991). DSM is easy to use, quick to draw and made changes, therefore information collected can be easily maintained and updated. Also data collected can be objectively represented in the SD model as mathematical formulas of numerical value of the various parameters (both soft and hard aspects). As a research methodology this forms grounds for originality and significant contribution to scientific knowledge in the business process management research field.

4.7. Model Verification, Validation and Credibility

Building valid and credible process models is an important aspect of a researcher’s representation of the actual system being studied. In determining the accuracy of a given model, three important terms are used, namely verification, validation and credibility.

Verification is the process of determining that a simulation computer program (or model) performs as intended. Verification checks the translation of the conceptual model (e.g. influence diagrams, flowcharts and assumptions) into a correctly working program or pseudocode in the Dynamic Synthesis methodology.

Validation is the process of determining whether the conceptual model (as opposed to the computer program or model) is an accurate representation of the system under study. Law and Kelton (1991) suggest that if the model is “valid”, then the decisions made with the model should be similar to those that would be made by physically experimenting with the system (pp. 299). This position is similar to that of Fishman and Kiviat (1968). A model is said to be credible when a simulation model and its results are accepted by managers/customers as being valid, and used as an aid (tool) in making decisions. In the DSM approach credibility is established by process owners i.e managers, who sponsor the RE process. In the absence of the sponsors, to establish credibility of the model can be performed by a focus group as surrogate stakeholders (Weigers, 1999).

Many researchers in the general simulation literature have suggested important approaches to validation, verification and establishing credibility in simulation models. Law and Kelton (1991) suggest that validation should be contrasted with output analysis, which is concerned with determining (estimating) a simulation model’s true measures of performance. In the system dynamics literature, perhaps the most sighted approach to validate models is that proposed by Senge and Forrester (1980). The strengths of combined process simulation and case study methods. This combination compensate for weaknesses in each individual methods. In many cases triangulation or the use of multiple methods (both qualitative and quantitative) to crosscheck each other, is
desirable and can enhance confidence in findings. This symbiotic interaction between
deductive and inductive approaches, theory building and testing, and exploratory and
explanatory research is the strength of the DSM. Other researchers (McGrath, 1979;
Babbie, 1989) highlight the benefits of combining inductive and deductive research
approaches. Ackoff (1962) sums up this concept by stating that “is probably the best
representation of the scientific research cycle” (pp 62).

5. Potential Contribution of DSM as a Practice Methodology

The benefits of DSM over case study or process simulation are in its theoretical
grounding and well structured methods for effective data collection. The data collected
can be objectively represented in the SD model as mathematical formulas with numerical
value of various parameters from both the soft and hard aspects of the RE process.
This is a significant contribution to scientific knowledge in the RE process management
field. The requirements engineering process, as a social process, needs approaches or
problem solving paradigms that can captures both the quantitative and qualitative issues
commonly found in complex systems, as opposed to traditional techniques that support
either hard or soft issues only (Williams and Kennedy, 1997). The issue raised here is
the need to reorganise the role qualitative sociological process inputs play in facilitating
exploration and understanding of a system’s RE process. Although there is insufficient
literature on the use of sociological qualitative methods (Macaulay, 1996; Hirscheim,
1984; Earneson, 1991; and Eastely-Smith, 1991), the scientific paradigm adopted by the
natural sciences is appropriate to the social-technical information systems only in so far
as it is appropriate for the social sciences (Hirscheim, 1984). These tools and techniques
can be used in the RE process modelling to improve its effectiveness. Methodological
pluralism or paradigms that support both hard and soft should be welcomed, regardless
of inherent differences in epistemological and ontological issues (Galliers, 1984),
particularly when researching systems requirements engineering process.

This paper makes key contributions to the business process modelling debate and
to the literature on model-based systems and requirements engineering by proposing
the DSM. Viewing business process problems from a dynamic feedback control
viewpoint provides an enhanced understanding enshrined in the Dynamic Synthesis
Methodology – a major contribution of this paper. Although the heading may not
be mutually exclusive with other methodologies, there are obvious inter-relationships
between DSM and other problem-solving methodologies. For example, aspects of
philosophical background and practice are reflected to some extent in all other problem
solving methodologies (Avison and Fitzgerald, 1996). DSM is still in its infancy, but
has a lot of potential in developing into a practice-based RE process modelling and
analysis. Further academic evaluations and practice-based validation will be necessary in
order for the DSM to be accepted both in industry and academia as a general problem
solving methodology. The DSM helps to provide a basis for collecting data about a
phenomenon and then experimentation to predict the behaviour with what if dynamic
analysis.
Application of Dynamic Synthesis methodology and the proposed DYNASIS tool can help researchers and managers address a broad range of strategic business process management problems. The DSM can be used to understand the various business process stakeholders' qualitative meaning of concepts while objectively simulating quantitative behaviour of variables that help to explain the patterns akin to business process analysts and managers. The solution to many of the problems of the RE process management and improvement are found in the definition of DSM. The DSM does not only integrate case study and process simulation modelling methods but it also integrates concepts from the general systems theory, servo mechanisms theory, measurement theory, underpinned by synthesis as a philosophy of science. The value of the research process model in theory building, testing and extension was highlighted in the paper. The paper suggests that the model has the potential to provide a framework for building a body of knowledge on business process modelling. There are practical implications of the framework for both theory and practice. The paper therefore calls for a comprehensive research programme that tests the strengths and weakness of the Dynamic Synthesis Framework on business process modelling and analysis.

References


van Maanen J (1983), Qualitative Methodology Sage Beverly Hills, California.


A Fundamental View on the Act of Modeling

H.A. (Erik) Proper, P. van Bommel, S.J.B.A. Hoppenbrouwers and
Th.P. van der Weide, Institute for Computing and Information Sciences Radboud University
Nijmegen Toernooiveld 1, 6525 ED Nijmegen, The Netherlands, EU E.Proper@cs.ru.nl

This paper is part of an ongoing research effort to better understand the role of models and modeling in the information system development life-cycle. During this life-cycle, several models are produced, ranging from high level sketches, via conceptual models to source code. This paper is part of an ongoing research effort to better understand the act of modeling. We describe a formal framework by which the process of modeling can be regarded as involving the selection of more and more refined interpretations in terms of the underlying meta-model of the modeling language used. The resulting framework will be used to create a laboratory setup in which we can consequently more closely study (and support) modeling processes.

1. Introduction

Modeling is at the core of information systems engineering. In (Mylopoulos, 1998) a distinction is made between usage world, subject world, system world and development world, when producing deliverables during information systems engineering. Understanding each of these worlds require considerable modeling efforts, be it to define the requirements on the system, or be it to produce the design of a system.

The work reported in this paper is part of an ongoing effort to better understand the act of modeling (Hoppenbrouwers et al., 2005c,f,g,d; Proper et al., 2005c; Hoppenbrouwers et al., 2005b; Proper et al., 2005a; Proper and Weide, 2005; Proper et al., 2005b; Bommel et al., 2006) in the context of information system engineering. One of our longer term goals is to turn the art of modeling into a science of modeling.

This research effort is one of three focal areas in our research:

1. Syntax and semantics of modeling languages (Bommel et al., 1991; Hofstede and Weide, 1993; Hofstede et al., 1993; Proper and Weide, 1994; Bronts et al., 1995; Campbell et al., 1996; Creasy and Proper, 1996; Hoppenbrouwers et al., 1997; Bommel et al., 1996; Hoppenbrouwers et al., 2005e).

2. The process of modeling (Derksen et al., 1996; Frederiks and Weide, 2004; Bosman and Weide, 2004a,b; Bleeker et al., 2004; Hoppenbrouwers et al., 2005a; Proper et al., 2004; Proper and Hoppenbrouwers, 2004; Hoppenbrouwers et al., 2005c,f; Proper and Weide, 2005; Proper et al., 2005b; Bommel et al., 2006; Hoppenbrouwers et al., 2005g).
3. The use of models in information systems engineering (Hoppenbrouwers et al., 2005d; Hoppenbrouwers and Proper, 2004; Proper et al., 2005c; Veldhuijzen van Zanten et al., 2004; Hoppenbrouwers et al., 2005b; Proper et al., 2005a).

In the past our focus was mainly on the formal definition of syntax and semantics of modeling languages. We have recently expanded this focus to include the process of modeling and the usage of models in information systems engineering. This expansion was inspired by a desire to better understand the modeling process itself, as well as the requirements on the languages used to express these models by the context in which they are to be used (Proper et al., 2005c).

**Figure 1: Refinement of models and meta-models**

The primary concern of this paper is therefore a further elaboration of a hypothesis put forward in (Proper et al., 2005b). We argue that one can observe how many modeling techniques are in use to model several aspects of domains, such as processes, objects, information being processed, the flow of information, the flow of control, etc. Scholars and practitioners have produced numerous modeling techniques (Bubenko, 1986; Avison and Wood–Harper, 1991; Avison, 1995; Bernus et al., 1998). The resulting plethora of techniques has, in the past, already been referred to as “a methodology jungle” (Avison, 1995). Each of these modeling techniques focuses on specific aspects of a domain, and is especially geared towards the representation, study, analysis or design of such aspects. Nevertheless, all of these techniques deal with facts about a domain describing how (from the perspective of a specific aspect) concepts in the domain relate to each other. Put more operationally, we argue that any activity model, sequence diagram, information model, etc. has an accompanying domain model (Bleeker et al., 2004; Proper et al., 2004) of the underlying concepts and their relations. Such a domain model could be expressed in terms of a general purpose domain modeling language such as ORM (Halpin, 2001), but also using ontology modeling languages such as OWL (McGuiness and Harmelen, 2003) and KL-ONE (Woods and Schmolze, 1992).

This leads to the situation as depicted in Figure 1. On the right hand side we find the meta-models of the modeling techniques used, while on the left hand side we find the actual models. The ‘XXX’ represents an aspect of the domain that is being modeled.
The ‘XXX’ model is a re-interpretation of the original model in terms of the refined ‘XXX’ meta-model. To illustrate this point, consider the example depicted in Figure 2. In this example, we have used the ORM domain modeling technique (Halpin, 2001) to represent a general domain model of a small sample domain dealing with involvement of people with a University department. The involvement starts with candidature, then might move on to the coworkership level, and will typically end in the alumnus status. In the example, we have (partially) re-interpreted the underlying domain model into two directions: an UML class diagram focusing on the core concepts in the domain, and a state-transition diagram focusing on the state changes of the involvement of people with departments.

As another example, consider the compacted version, as depicted in Figure 3, of the case study used in (Proper et al., 2005b). This example focuses on workflow modeling and shows two interpretation steps. The first step, moving from A) to B), requires modelers to select which object types are really actor and actand types. The second interpretation step, from B) to C), can actually be done automatically given a pre-defined mapping between the meta-models of the modeling techniques involved. The modeler does not need to add additional information to the model. Note that the situation depicted in A) is not a static view on the domain. The arrows from fills in form to examines, etc, show a temporal dependency between states, thus providing a flow of states and activities.

**Figure 2: Example interpretations**
In each of the interpretation steps, modelers need to make a choice of how to re-interpret (if at all!) specific concepts in the general domain model in terms of the modeling concepts in the refined meta-model. We argue that modeling can be regarded as a process of (iteratively!) refining one’s view on the world in terms of more and more refined modeling concepts (the types in the meta-model). This process is driven by the motivations for producing the model in the first place.

Using the framework presented, one could actually experiment with situations in which the metamodel is defined during the modeling process versus situations in which the meta-models are pre-defined and standards-based.

One may also argue that in practice, modelers will quite often directly produce UML class diagrams, workflow diagrams, etc. In our view, doing so leaves implicit numerous interpretive decisions about the domain. If one were to first produce a domain model as depicted in Figure 3 A), one could argue that the understanding of the domain being modeled would be deeper, providing a better base from which to then produce model C) via B). Note that it is not our goal not to cast judgement on how to best model. Our goal is rather to better understand the actual act of modeling, and as such, we do want to study how modelers implicitly or explicitly move from A) to C). The resulting framework will be integrated with the logbook perspective (Bommel et al., 1996; Bosman and Weide, 2004b; Hoppenbrouwers et al., 2005c) on the modeling process to create a system that will allow us to conduct modeling experiments in a laboratory setting.

We have structured the remainder of this paper as follows. In Section 2 we briefly explore the notion of subjectivity in relation to modeling. Section 3 then focuses on hierarchies of modeling languages, i.e. meta-model hierarchies. Given such a hierarchy, Section 4 shows how hierarchies of models as depicted in Figure 3 can be represented formally.
2. Subjectivity in Modeling

The aim of this section is to define more precisely what we mean by the modeling of a domain, in other words, our fundamental way of thinking about modeling. In doing so, we will start by introducing a framework describing the essential processes that take place when an observer observes a domain.

It is our assumption, based on the work of C.S. Peirce (Peirce, 1969), that observers perceive a universe and then produce a conception of that part they deem relevant. The conceptions harbored by an observer are impossible to communicate and discuss with other observers unless they are articulated somehow (the need for this ability in the context of information systems engineering is evident). In other words, a conception needs to be represented. Peirce argues that both the perception and conception of an observer are strongly influenced by their interest in the observed universe. This leads to the following set of definitions (also inspired by the ones provided in (Falkenberg et al., 1998), which are based on the work by Peirce as well):
Universe – the ‘world’ around the observer.

Observer – an actor perceiving and conceiving the universe, using their senses.

Perception – that what results, in the mind of an observer, when they observe the universe, using their senses.

Conception – that what results, in the mind of an observer, when they interpret a perception of the universe.

Observers may zoom in on a particular part of the universe they observe, or to state it more precisely, they may zoom in on a particular part of their conception of the universe:

Domain of interest – any ‘part’ or ‘aspect’ of a conception of the universe, an observer may zoom in on.

Note that when observers zoom in on a domain of interest, they produce yet another conception.

In the context of information systems engineering, observers may have different domains of interest depending on their concern with regards to the information system being engineered. For example, the operators who will be required to maintain a planned information system, will regard this system in terms of costs of keeping the system up and running, costs and efforts involved in implementing the system, etc. Future users of the same planned system, however, will be more interested in the impact/support the system is likely to have on their work related tasks. In our effort to obtain a fundamental understanding of the act of modeling, we initially focus on situations where we only have one specific concern and associated domain of interest. In line with (Falkenberg et al., 1998) we define a model to be a specific kind of conception:

Model – a purposely abstracted and unambiguous conception of a domain of interest.

Conceptions that are harbored by an observer are impossible to communicate and discuss with other observers, unless they are articulated somehow. In other words, the conception needs to be represented:

Representation – the result of an observer representing a conception, using some language to express themselves.

The resulting situation is illustrated in Figure 4 showing how an observer in observing the universe has a conception, which may be represented in terms of a representation.

We are now also in a position to define more precisely what we mean by modeling:
Figure 4: An observer observing a universe

**Modeling** – The act of purposely forming a model from (what is conceived to be) a part of the universe, and representing the resulting model by means of some language and medium.

The same domain of interest may be regarded by different observers, which is bound to lead to different conceptions, depending on the specific observers. The fact that when referring to the same universe, people are likely to refer to different models is, as reported in e.g. (Falkenberg et al., 1998), one serious cause for the current confusion in the development of information systems. People, tend to think about a system as something that can be objectively determined (Falkenberg et al., 1998). An assumption that is bound to lead to serious ‘accidents’. However, at present our focus is on better understanding the act of modeling when only one observer is involved, which is difficult enough as even one observer is not likely to behave like a monotonic function when modeling.

In the context of information systems engineering, observers will approach a domain with the aim of expressing the domain in terms of some set of modeling constructs, such as classes, activity (types), event (types), constraints, etc. The set of modeling constructs a observer is used to employ (or trained to use) when modeling a domain, will strongly influence his/her conceptions. For example, when viewing a domain of interest from the perspective of UML class diagrams, this is bound to lead to a different model than when the same domain is viewed from the perspective of UML sequence diagrams. To make this explicit, we therefore presume that when observers model a domain, they do so from a certain perspective; their Weltanschauung (Wood–Harper et al., 1985). Figure 5 also illustrates how an observer observes (a domain of interest within) a universe from the perspective of different meta-models (M1, . . . , Mn), leading to equally many models (m1, . . . , mn) and model representations (r1, . . . , rn).
The remainder of this paper is primarily concerned with the development of a precise understanding of the relationships between these meta-models, the corresponding models (or rather their representations), as well as their evolution during a modeling process. Here we will operate under the hypothesis that modeling can be viewed as an iterative process of:

- Defining an (unspecified) model of a domain using some suitable (suitable; not necessarily the) generic meta-model, focusing on domain concepts and their relationships in a general sense.
  
  In the examples of the previous section, we used ORM (with temporal extensions) as an example of such a generic meta-model.

- Selecting more specific interpretations of the concepts identified in the initial model, using more refined meta-models.
  
  In the previous section we showed examples of interpretations in terms of a UML class diagram, a state-transition diagram, and a workflow model.

  The latter step, selection of interpretation, is an essential aspect of our way of thinking with regards to modeling.

3. Meta-model Hierarchies

The foundation of our modeling framework is formed by a hierarchy of meta-models. The concept of a meta-model hierarchy is not new. It was already introduced in (Oei et al., 1992; Falkenberg and Oei, 1994) as a way of comparing modeling techniques, and to some extent refined further in (Falkenberg et al., 1998). Our goals of viewing the act of modeling as a process of stepwise selection of interpretations over a hierarchy of meta-models is a way to operationalise the ‘old’ notion of a meta-model hierarchy.

A meta-model is seen as a formal system (Mendelson, 1987). Such a system consists of (1) a signature that specifies its concepts, providing a base for the definition of well-formed formulae, and (2) a set of such well-formed formulae (also referred to as axioms) that are assumed/referred to hold for concrete systems that realize the formal system. In this context we shall refer to the concepts of the formal system as the (modeling) types of the meta-model. We will denote a metamodel by its signature and its axioms. We will use $hT,Ai$ to denote the system with signature $T$ and axioms $A$. 
Let MT be the set of all meta-types from some class of modeling techniques, MA be the set of all axioms, and MM-MT × MA the set of all meta-models. We focus on meta-models that satisfy the following rules. Each meta-model is consistent, meaning that the axioms are not contradictory.

[M1] If \( hT, Ai 2 MM \), then \( A \) is a consistent set of well-formed formula’s over \( T \).

Each meta-model is required to have different modeling types.

[M2] If \( M_1 = (T_1, A_1) \) and \( M_2 = (T_2, A_2) \), such that \( M_1, M_2 2 MM \), then:
\[
M_1 \neq M_2 \Rightarrow T_1 \cap T_2 = \emptyset;
\]
This latter requirement is added to allow us to study relations between modeling concepts in more depth.

A model is regarded as an instantiation of a formal system; the associated meta-model. This model thus contains instantiations of the meta-types contained in that meta-model. Let EL be the set of all those instantiations, which are referred to as model elements. We define the possible interpretation of these elements in terms of the metatypes: \( \text{IN} = \text{EL} \times \text{MT} \). In other words, an interpretation is the combination of a model element and a meta-type. Since meta-models may contain sub-types, elements may be associated to multiple meta-types.

If \( m \) is a model with associated meta-model \( M \), we will also say that \( m \) is an \( M \)-model. An \( M \) model \( m \) can be regarded as a set of interpretations \( m 2 \text{IN} \) that meet the axioms of meta-model \( M \). The set of valid \( M \)-models for a given meta-model \( M = (T, A) \) is therefore defined as:
\[
\text{M}(M) \triangleq \{ m 2 \text{EL} \times T \mid m \models A \}
\]

The set of interpretations fitting a meta-model is defined as: \( \text{I}(M) \triangleq \cup \text{M}(M) \).

The next step is to introduce hierarchies of meta-models. Such a hierarchy is composed of refinement relations between meta-models. Let RF be the set of possible refinement relations for the considered class of meta-models and let From, To : RF \( \rightarrow \) MM be functions returning the start and destination meta-model of a refinement respectively. Then RF, From and To together span a space in which we will be able to identify meta-model hierarchies to be used in modeling. We do require RF to be acyclic:

[M3] The graph spanned over MM by From and To is acyclic. A specific meta-model hierarchy is a set of refinements, so we can define the set of possible metamodel hierarchies as \( \text{MH} \subseteq \mathcal{O}(\text{RF}) \), where we do require:

[M4] If \( R 2 \text{MH} \) then \( R \) is a tree. Let \( \text{Top}(R) \) denote the top of such a tree. We will write \( \text{R}^{\text{MM}} \) as an abbreviation for the set of meta-models involved in \( R \).

To really capture the notion of refinement between meta-models, we must be able to map models upward in the hierarchy. We therefore need a function that is able to ground models stated in a refined meta-model in terms of the more general meta-model:

\[
\text{Ground} : \text{RF} \rightarrow (\mathcal{O}(\text{IN}) \rightarrow \mathcal{O}(\text{IN}))
\]

In terms of the example shown in Figure 3 the grounding function would have to map any actor type and actand type in a workflow model onto an object type in an ORM model, and each activity type onto an ORM relationship type. The working of the grounding function is illustrated in Figure 6. Models are grounded by grounding the
interpretations they are made of. Multiple models conform a refined meta-model may be grounded onto the same generalized model. For example, in Figure 3 we might have selected a person being examined to be an actand (i.e. passive) in the examination, rather than considering it to be an actor as well (as is currently shown in B). In either case, the grounding of model B) would still be the model shown in A).

**Figure 6: Grounding of models and interpretations**

![Grounding of models and interpretations](image)

For a given refinement \( r \), the grounding function should limit itself to interpretations associated to the meta-models involved in the refinement:

\[ [M5] \times 2 \text{ dom}(\text{Ground}(r)) x - I(\text{To}(r)) \quad \text{and} \quad y \in \text{ran}(\text{Ground}(r)) y - I(\text{From}(r)) \]

Empty models have an empty grounding:

\[ [M6] \text{Ground}(;) = ; \]

Even more, the grounding function should behave strict monotonous in terms of inclusion of sets of interpretations:

\[ [M7] m_1 \subseteq m_2 \rightarrow \text{Ground}(m_1) \subseteq \text{Ground}(m_2) \]

where \( \subseteq \) is used as a proper subset. This allows us to ground any non-empty fragment of a re-interpreted model back to (a non-empty) fragment at the more generic level:

Corollary 3.1 \( m \neq ; \Rightarrow \text{Ground}(m) \neq ; \)

### 4 Model Hierarchies

In this section we extend the meta-model hierarchy of the previous section to a hierarchy of models over such meta-model hierarchies. First we follow the interpretation of a single model element in a hierarchy. When modeling, decisions are made pertaining to interpretations of the domain.

These modeling decisions are almost as important as the resulting model. Let MV be a carrier set for motivations of such decisions, then we can define an interpretation hierarchy as a partial function: \( h : MM \rightarrow +(IN) \times MV \) where \( +(X) = \{X\}-\{\} \). Let \( IH \) be the set of all such interpretation hierarchies.

\[ IH, MM \rightarrow +(IN) \times MV \]
If we are only interested in the set of interpretations, we will use:

\[ h!(M), I \text{ such that } h(M) = hI, vi \]

An interpretation hierarchy should follow a meta-model hierarchy. This is laid down in three rules. We consider \( h \) to be an interpretation hierarchy fitting a meta-model hierarchy \( R \), written as \( h \in I(R), i \):

1. The first condition requires that an interpretation hierarchy can only contain interpretations for the meta-models present in \( R \). Formally: \( \text{dom}(h) - RMM \).
2. The second condition requires the top of the interpretation hierarchy to contain one interpretation only; the root. Formally: \( |h!(\text{Top}(R))| = 1 \).
3. The third condition requires the interpretation hierarchy to obey the grounding function. Formally this is enforced by:

\[ 8r \in R [\text{Ground}_r(h!(\text{To}(r))) - h!(\text{From}(r))] \]

Note that in a refinement step, one is allowed to exclude elements from the original model. If we would want to forbid this, the third condition would have to read:

\[ 8r \in R [\text{Ground}_r(h!(\text{To}(r))) = h!(\text{From}(r))] \]

Two interpretation hierarchies are disjoint if they do not overlap for any meta-model:

\[ h \perp i, 8M \in \text{MM} [h!(M) \not= i!(M)] \]

A model hierarchy is a set \( H \) of interpretation hierarchies. The set of possible model hierarchies is therefore given as:

\[ MH, } (IH) \]

If \( H \) is a model hierarchy, then for any meta-model \( M \), the complete model is defined as the union of the interpretations in the interpretation hierarchies (as illustrated in Figure 7):

\[ H!(M), [h2H \bigcup h!(M)] \]

For a given meta-model hierarchy \( R \), the set of valid model hierarchies consists of those interpretation hierarchies \( H \) such that:

\[ 8M \in \text{dom}(H) [H!(M) \not= \bigcup 8h,i \in H [h \not= i] \text{ and } h \not= i] \]

The first condition requires that all models in the hierarchy conform to their respective metamodels, while the second condition requires interpretation hierarchies to not overlap.
5. Conclusion

In this paper we have discussed a framework to study the act of modeling, where a modeling process is regarded as involving the selection of more and more refined interpretations in terms of the underlying meta-model of the modeling language used. The resulting framework will be used, in conjunction with the logbook system, to create a laboratory environment in which modeling experiments can be conducted.

The logbook system (Hoppenbrouwers et al., 2005c) takes the view that a modeling process is a (controlled) dialogue between a domain expert, a modeling mediator and a model builder. This process is regarded as a questioning & answering process involving these three roles. When combined with the theory as presented in this paper, the goal of such a questioning & answering process can be made explicit as the creation of a model hierarchy on top of a pre-determined (dictated by the modeling goals at hand (Proper et al., 2005c,a)) meta-model hierarchy.

In future versions of our framework we also intend to refine it such that we are able to deal with multiple views and concerns, as well as multiple (contradicting!) observers. In the latter case we would like to be able to even log the negotiation that may have to take place in reconciling different views held by different observers of the same domain.

References


Part Two

Natural Language Processing
Kinyarwanda Speech Recognition for Automatic Voice Dialling Systems

Muhirwe Jackson, Department of Computer Science, Faculty of Computing and IT, Makerere University, P.O Box 7062 Kampala, jmuhirwe@cit.mak.ac.ug

Speech recognition tools exist for English, Spanish, Portuguese, Italian and many languages of Eastern Europe and Asia. There are barely no tools for African languages and other third world countries’ languages. In this paper we present the process of building a speech recognizer for Kinyarwanda language. Two different corpora were collected of audio recordings of indigenous Kinyarwanda language speakers, in which subjects read aloud numeric digits. One of the collected corpora contained the training data and the other the testing data.

The system was implemented using the HMM toolkit HTK by training HMMs of the words making the vocabulary on the training data. The trained system was tested on data other than the training data and results revealed that 94.47% of the tested data were correctly recognized.

The developed system can be used by developers and researchers interested in speech recognition for Kinyarwanda language and any other related African language. The findings of the study can be generalized to cater for large vocabularies and for continuous speech recognition.

Introduction

Speech recognition can be defined as the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words (Zue et al., 1996 [10]; Mengjie, 2001 [3]). Automatic speech recognition (ASR) is one of the fastest developing fields in the framework of speech science and engineering. As the new generation of computing technology, it comes as the next major innovation in man-machine interaction, after functionality of text-to-speech (TTS), supporting interactive voice response (IVR) systems. In ASR systems acoustic information is sampled as a signal suitable for processing by computers and fed into a recognition process. The output of the system is a hypothesis transcription of the utterances.

The list of applications of automatic speech recognition is so long and is growing; some of known applications include virtual reality, Multimedia searches, auto-attendants, travel information and reservation, translators, natural language understanding and many more applications (Scansoft, 2004 [7]; Robertson, 1998 [6]).

As the use of ICT tools, especially the computer, is becoming inevitable, there are many Rwandans who are left out due to inadequate human computer interface (HCI) design considerations. A case in point is the many Rwandans who are left out due to
language barrier (Earth trends, 2003) [1]. These people can only read and write in their mother-tongue, Kinyarwanda language, making it impossible for them to use ICT conventional tools that are built in the two International languages, English and French used in Rwanda.

The purpose of this project was therefore to design and train a speech recognition system that could be used by application developers to develop application that will take indigenous Kinyarwanda language speakers aboard the current information and communication technologies to fast-track the benefits of ICT.

Kinyarwanda Language

Kinyarwanda Language, is the national language of Rwanda and is the second largest spoken language in the Bantu group after Swahili(Kimenyi, 2005) [2]. It is a sister dialect of Kirundi, the national language of Burundi and Giha, another dialect spoken in Tanzania. Despite genocide which took place taking lives of more than one million Rwandans, its speakers are perhaps more than 20 million people spread in the great lakes region. (Kimenyi, 2005) [2].

Much as Kinyarwanda is spoken by over 20 million compared to Walpiri spoken by just 3000 people it has not yet received the attention it needs.

Several papers and books have been written on how to construct an automatic speech recogniser for most languages in Europe and Asia. Nevertheless, each language presents its own unique challenges. In this paper, we describe the process we went through to develop a Kinyarwanda language automatic speech recognition for voice dialling systems. Though this process was used for Kinyarwanda language, it can be used for any other language with minimal modification.

Kinyarwanda speech recogniser construction process

The project was limited to only isolated whole words and trained and tested on only one (1) word sentences consisting of the numeric digit 0 to 9 that could be used on operating a voice operated telephone system.

The goal of the project was to build a robust whole word recognizer. That means it should be able to generalise both from speaker specific properties and its training should be more than just instance based learning. In the HMM paradigm this is supposed to be the case, but the researcher intended to put this into practice.

For pragmatic reasons and also to be able to focus on more specific issues than HMM in general, the Hidden Markov Model toolkit (HTK) was used. HTK is a toolkit for building Hidden Markov Models (HMMs). HMMs can be used to model any time series and the core of HTK is similarly general-purpose. However, HTK is primarily designed for building HMM-based speech processing tools, in particular recognisers (Young S.et al., 2002) [9].

Secondly to reduce the difficulties of the task, a very limited language model was used. Future research can be directed to more extensive language models.

There are several different approaches for the implementation of an ASR but in this paper the four major processing steps as suggested by HTK (Young et al., 2002) [9] are considered namely data preparation, training, Recognition/testing and analysis.
Data Preparation

The first stage of any recogniser development project is preparing data both for training and testing. In the system built here, all of this speech was recorded from scratch. The training data is used during the development of the system and testing data provides the reference transcriptions against which the recogniser’s performance can be measured. In the case of the training data, the prompt scripts were be used in conjunction with a pronunciation dictionary to provide the initial phone level transcriptions needed to start the HMM training process.

It follows from above that before the data can be recorded, a phone set must be defined, a dictionary must be constructed to cover both training and testing and a task grammar must be defined.

The Task Grammar

The task grammar defines constraints on what the recogniser can expect as input. As the system built provides a voice operated interface for phone dialling, it handles digit strings. For the limited scope of this project, only a the digits 0, 1, 9 making toy grammar were needed. The grammar was defined in BN-form, as follows: $variable$ defines a phrase as anything between the subsequent $=$ sign and the semicolon, where $|$ stands for a logical or. Brackets have the usual grouping function and square brackets denote optionality.

The used toy grammar was:

```
#
#Task grammar
#
$digit=RIMWE|KABIRI|GATATU|KANE|GATANU|GATANDATU|KARI
NDWI|UMUNANI|ICYENDA|ZERO;
(SENT-START[$digit] SENT-END)
```

The above grammar can be depicted as a network but this is simply a high level representation of a task grammar is provided for user convenience. The HTK recogniser actually requires a word network to be defined using a low level notation called HTK Standard Lattice Format (SLF) in which each word instance and each word-to-word transition is listed explicitly.

A Pronunciation Dictionary

The dictionary provides an association between words used in the task grammar and the acoustic models which may be composed of sub word (phonetic, syllabic etc.,) units. Since this project provides a voice operated interface the dictionary could have been constructed by hand but the researcher wanted to try a different method which could be used to construct a dictionary for a large vocabulary ASR system. In order to train the HMM network, a large pronunciation dictionary is needed.

Since we are using whole-word models in this assignment, the dictionary has a simple structure. A file called 'lexicon' was created with the following structure:
A file named wdlist.txt was created containing all the words that make up the vocabulary:

GATANDATU gatandatu
GATANU gatanu
GATATU gatatu
ICYENDA icyenda
KABIRI kabiri
KANE kane
KARINDWI karindwi
RIMWE rimwe
SENT-END [] sil
SENT-START [] sil
UMUNANI umunani
ZERO zero

The dictionary was created finally by using HDman as follows:

HDman -m -w wdlist.txt -n models1 -l dlog dict lexicon. This creates a new dictionary called dict by searching the source dictionary(s) lexicon to find pronunciations for each word in wdlist.txt. Here, the wdlist.txt in question needs only to be a sorted list of the words appearing in the task grammar given above. Once training and test data has been recorded, an HMM will be estimated for each of these words.

Recording

In order to train and test the recognizer on the domain and on the voices of some selected people, 10 sentences were automatically generated from the grammar with HTK’s HSGen. Speech data of six (6) different speakers 3 males and 3 females of different age groups was recorded. Due to my lack of access to a recording studio, the recordings were done in an room but with minimal noise. As the toolkit does not require phoneme duration information for the training sentences, the (differences in) timing in the pronunciation of the training sentences is not important. The toolkit learns to recognise the words through fitting the word transcriptions on the training set. These transcriptions are used for all realisations of the same sentence, even though there might be variation between speakers relative to the transcription.

The speakers were given a list with sentences which they had to read aloud. After about 5 sentences they took a short break, and drank a glass of water. The training corpus consisting of 150 sentences were recorded and labelled using the HTK tool HSLab.

After recording and labelling the training sentences, a test corpus was also created the same way as the training corpus but in this case 70 sentences were used for training. The differences noted in pronunciation between speakers (and their consequences) can be categorised as articulation variation E.g., some speakers had a rolling ‘r’, others not in, example,
‘kabiri’, ’rimwe’. Phonetic change degrades the quality of the training set, since the same phonetic transcription was used for all speakers. These phonetic changes problems were solved by using isolated whole word models and having many different sentences such at the end of the day I created a speaker independent system.

Articulation variation on the other hand is of course a problem for recognition but if there was no articulation variation the task of recognising would become an instance based learning problem.

**Phonetic Transcription**

For training, we need to tell the recognizer which files correspond to what digit. HTK uses the so-called Master Label Files (MLF) to store information associated to speech. What makes things a bit confusing is the fact that there are two things an MLF can contain: words and phonemes. In the tutorial the usages of various HTK tools are shown that can convert lists of sentences into lists of words and then lists of phonemes, the last two in an MLF. Since the objective of this project was to create an isolated word recognizer, a file called source.mlf was created associating each recorded and labelled speech data with a word. #!MLF!# ”data/train/rimwe01.lab”

RIMWE
. 
"data/train/rimwe02.lab”RIMWE
. Etc..

It is assumed that rimwe01.WAV contains the utterance ’rimwe’, and so on. Next, the model transcriptions must be obtained. For this, an HTK edit script called ’mkphones0.led’ was created containing the following:

EX
IS sil sil
DE sp

The HTK tool HLed was used to the word transcriptions into model transcriptions (models0.mlf): HLed -d dict –I
models0.mlf mkphones0.led source.mlf

**Encoding the Data**

Speech recognition tools cannot process directly on speech waveforms. These have to be represented in a more compact and efficient way. This step is called ”acoustical analysis”:

The signal is segmented in successive frames (whose length is chosen between 20ms and 40ms, typically), overlapping with each other. Each frame is multiplied by a windowing function (e.g. Hamming function).

A vector of acoustical coefficients (giving a compact representation of the spectral properties of the frame) is extracted from each windowed frame.

In order to specify to HTK the nature of the audio data (format, sample rate, etc.) and feature extraction parameters

(type of feature, window length, pre-emphasis, etc.), a configuration file (config.txt) was created as follows:
# Coding parameters

SOURCEKIND = waveform  SOURCEFORMAT = HTK
SOURCE_RATE = 625

TARGETKIND = MFCC 0 D  A TARGETRATE = 100000.0
SAVECOMPRESSED = T  SAVEWITHCRC = T  WINDOWSIZE = 250000.0
USEHAMMING = T 4
PREEMCOEF = 0.97
NUMCHANS = 26
CEPLIFTER = 22
NUMCEPS = 12
ENORMALISE = F

To run a HCopy a list of each source file and its corresponding output file was created. The first few lines look like:

```
data/train/rimwe01.SIG  data/MFC/rimwe01.MFC
data/train/rimwe02.SIG  data/MFC/rimwe02.MFC
data/train/rimwe03.SIG  data/MFC/rimwe03.MFC
data/train/sil10.SIG  data/F/mfc/sil10.MFC
```

One line for each file in the training set. This file tells HTK to extract features from each audio file in the first column and save them to the corresponding feature file in the second column. The command used is:

```
HCopy -T 1 -C config.txt -S hcopy.scp
```

**Parameter Estimation (Training)**

Defining the structure and overall form of a set of HMMs is the first step towards building a recognizer. The second step is to estimate the parameters of the HMMs from examples of the data sequences that they are intended to model. This process of parameter estimation is usually called training. The topology for each of the hmm to be trained is built by writing a prototype definition. HTK allows HMMs to be built with any desired topology. HMM definitions can be stored externally as simple text files and hence it is possible to edit them with any convenient text editor. With the exception of the transition probabilities, all of the HMM parameters given in the prototype definition are ignored. The purpose of the prototype definition is only to specify the overall characteristics and topology of the HMM. The actual parameters will be computed later by the training tools. Sensible values for the transition probabilities must be given but the training process is very insensitive to these. An acceptable and simple strategy for choosing these probabilities is to make all of the transitions out of any state equally likely. In principle the HMM should be tested on a large corpus containing wide range of word pronunciations. For this purpose 150 sentences were recorded and labelled as stated above see the training corpus CD for training data.

**Training Strategy**

HTK offers two different approaches to training speech data. Since this project we were interested in isolated whole word the following strategy was used as described below:
Firstly, an initial set of models must be created. If there is some speech data available for which the location of the word boundaries have been marked, then this can be used as bootstrap data. In this case, the tools HInit and HRest provide isolated word style training using the fully labeled bootstrap data. Each of the required HMMs is generated individually. HInit reads in all of the bootstrap training data and cuts out all of the examples of the required phone. It then iteratively computes an initial set of parameter values using a segmental k-means procedure.

On the first cycle, the training data is uniformly segmented, each model state is matched with the corresponding data segments and then means and variances are estimated. If mixture Gaussian models are being trained, then a modified form of k-means clustering is used. On the second and successive cycles, the uniform segmentation is replaced by Viterbi alignment. The initial parameter values computed by HInit are then further re-estimated by HRest.

In the absence of marked data, the tool HCompV is used. In this project since all the data was labelled, then HInit and HRest were used for training purposes.

**HMM Training**

The first step in HMM training is to define a prototype model. The purpose of the prototype is to define a model topology on which all the other models can be based. A prototype model was defined for each HMM refer to HTK book on details about HMM prototype definition (Young S.et al., 2002,) [9].

**Initialisation**

The HTK tool HInit was used to initialize the data as given below

```
HInit -A -D -T 1 -S train.scp -M model/hmm0 -H hmmfile -l label -L label dir nameofhmm where:
nameofhmm is the name of the HMM to initialise (here: yes, no, or sil). hmmfile is
```
a description file containing the prototype of the HMM called nameofhmm (here: hmm rimwe, hmm kabiri, etc).

trainlist.txt gives the complete list of the .mfcc files forming the training corpus (stored in directory data/train/mfc). label dir is the directory where the label files (.lab) corresponding to the training corpus (here: data/train/lab/).

label indicates which labelled segment must be used within the training corpus (here: rimwe, kabiri, etc.. model/hmm0 is the name of the directory (must be created before) where the resulting initialised HMM description will be output.

This procedure has to be repeated for each model (hmm rimwe, hmm kabiri, hmm gatatu etc..). The HMM file output by HInit has the same name as the input prototype. E.g

HInit -A -D -T 1 -S train.scp -M model/hmm0 -H hmm 1.txt -l rimwe -L data/train/rimwe This process was repeated for all the models. The HTK tool HCompV was used to initialize the models to the training data as follows.

HCompV -C config.txt -f 0.01 -m -S train.scp -M hmm0 proto.txt HCompv was not used to initialise the models (it was already done with HInit). HCompv is only used here because it outputs, along with the initialised model, an interesting file called vFloors, which contains the global variance vector multiplied by a factor 0.01. The values stored in varFloor1 (called the ”variance floor macro”) are to be used later during the training process as floor values for the estimated variance vectors. This results in the creation of two files - proto and vFloors - in the directory hmm0.

These files were edited in the following way: An error occurs at this point which rearranges the order of the parts of the MFCC 0 D A label as MFCC D A 0. This was corrected. The first three lines of proto were cut and pasted into vFloors, this was then saved as macros.

Training

The following command line was used to perform one re-estimation iteration with HTK tool HRest, estimating the optimal values for the HMM parameters (transition probabilities, plus mean and variance vectors of each observation function):

HRest -A -D -T 1 -S train.scp -M model/hmm1 -H vFloors -H model/hmm0/hmm 1.txt -l rimwe -L data/train rimwe. train.scp gives the complete list of the .mfcc files forming the training corpus (stored in directory data/train/mfc). Model/hmm1, the output directory, indicates the index of the current iteration. vFloors is the file containing the variance floor macro obtained with HCompv. Hmm 1.txt is the description file of the HMM called rimwe. It is stored in a directory whose name indicates the index of the last iteration (here model/hmm0). -l rimwe is an option that indicates the label to use within the training data (rimwe, kabiri, etc).

Data/train is the directory where the label files (.lab) corresponding to the training corpus. rimwe is the name of the HMM to train. This procedure has to be repeated several times for each of the HMMs( kabiri. Gatatu. Kane. Sil) to train. Each time, the HRest iterations (i.e. iterations within the current re-estimation iteration) are displayed on screen, indicating the convergence through the change measure. As soon as this
measure do not decrease (in absolute value) from one HRest iteration to another, it's time to stop the process. In this project 3 re-estimation iterations were used. The final word HMMs are then: hmm3/hmm 1, hmm3/hmm 0, and hmm3/hmm sil etc.. A file called hmmdefs.txt was created by combining all the hmms into one file which was consequently named hmmdefs.txt.

After each iteration an error occurred which rearranges the order of the parts of the MFCC_0 D A label as MFCC_ D A 0 which was consequently corrected after each iteration.

**Recognition**

The recognizer is now complete and its performance can be evaluated. The recognition network and dictionary have already been constructed, and test data has been recorded. Thus, all that is necessary is to run the recognizer. The recognition process can be summarized as in the figure below:

**Figure 2: Speech recognition process**

An input speech signal is first transformed into a series of ”acoustical vectors” (here MFCs) using the HTK tool HCopy, in the same way as what was done with the training data. The result was stored in a file known as test.scp (often called the acoustical observation).

The input observation was then processed by a Viterbi algorithm, which matches it against the recogniser’s Markov models using the HTK tool HVite: As follows

HVite -A -D -T 1 -H model/hmm3/hmmdefs.txt -i recout.mlf -w wdnnet dict hmmlist.txt -S test.scp. Where:

hmmdefs.txt contains the definition of the HMMs. It is possible to repeat the -H
option and list the different HMM definition files, in this case: -H model/hmm3/hmm0.txt -H model/hmm3/hmm1.txt etc.. but it is more convenient (especially when there are more than 3 models) to gather every definitions in a single file called a Master Macro File. For this project this file was obtained by copying each definition after the other in a single file, without repeating the header information.

The output is stored in a file (recout.mlf) which contains the transcription of the input. recout.mlf is the output recognition transcription file.

Wdnet is the task network.
dict is the task dictionary.
hmmlist.txt lists the names of the models to use (rimwe, kabiri, etc..). Each element is separated by a new line character.

Test scp is the input data to be recognised.

**Running the Recognizer Live**

The built recogniser was tested with live input. To do this the configuration variables parameters were altered as given below:

```
# Waveform capture
SOURCERA TE=625.0
SOURCEKIND=HAUDIO SOURCEFORMAT=HTK ENORMALISE=F
USESILDET=T MEASURESIL=F OUTSILWARN=T
```

These indicate that the source is direct audio with sample period 62.5 secs. The silence detector is enabled and a measurement of the background speech/silence levels was made at start-up. The final line makes sure that a warning is printed when this silence measurement is being made. Once the configuration file had been set-up for direct audio input, the HTK tool HVite was again used to recognize the live input using a microphone.

**Results**

The recognition performance evaluation of an ASR system must be measured on a corpus of data different from the training corpus. A separate test corpus, with new Kinyarwanda language digits records, was created as it was previously done with the training corpus. The test corpus was made of 50 recorded and labelled data which were later converted into MFC. In order to test for speaker independency of the system, some of the subjects who participated in creation of the testing corpus had not participated in creation of the training corpus.

**Performance Test**

Evaluation of the performance of the speech recognition system was done by using the HTK tool HResults. On running and testing the tool against the testing data, the following performance statistics were obtained:
Figure 3: Speech recognition results

![Image](image.png)

Performance Analysis

The first line (SENT) gives the sentence recognition rate (\%Correct=92.00), the second one (WORD) gives the word recognition rate (\%Corr=94.87). The first line (SENT) should be considered here. H=46 gives the number of test data correctly recognized, S=4 the number of substitution errors and N=50 the total number of test data. These results imply that of the 50 sentences making the testing corpus only 46 were correctly recognized which is equivalent to 92.00% and four (4) sentences were substituted by other sentences. The statistics given on the second line (WORD) only make sense with more sophisticated types of recognition systems (e.g. connected words recognition tasks). Nevertheless, there were 6 deletion errors (D), 2 substitution errors (S) and 0 insertion errors (I). N=156 gives the total number of words making the test data and of these 148 were correctly recognized leading to a 94.87% recognition. The accuracy figure (Acc) of 94.87% is the same as the percentage correct (Cor) because it takes account of the insertion errors, which the latter does not but in this case the insertion errors are zero. These results indicate that the training of the system was successful and that the developed system is speaker independent.

Testing the System on Live Data

To further test the system on live data and also again test its speaker independency, the system was tested by running it live. Four (4) different speakers who never participated in the creation of the training corpus helped in testing the system live. Subjects read loudly the Kinyarwanda language numeric digits and the table below gives a summary of the results. These results show that the system is speaker independent with a few errors which can be reduced by training the system on a larger training data and also including recordings from speakers from different regions of the great lakes region who speak Kinyarwanda.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Words Correctly recognized</th>
<th>Substitution errors</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject1</td>
<td>9</td>
<td>1</td>
<td>90%</td>
</tr>
<tr>
<td>Subject2</td>
<td>8</td>
<td>2</td>
<td>80%</td>
</tr>
<tr>
<td>Subject3</td>
<td>8</td>
<td>2</td>
<td>80%</td>
</tr>
<tr>
<td>Subject4</td>
<td>8</td>
<td>2</td>
<td>80%</td>
</tr>
</tbody>
</table>
Discussion

Before attempting building a speech recogniser for a new language it is always advisable to start by building one language which is already tested and in this case the researcher first constructed an English Yes and No recogniser which paved the way for the new language speech recognisers.

The Cambridge University Hidden Markov Models toolkit (HTK) was used for the implementation of the recogniser. HTK was used because it is free and has been used by many researchers all over the world. HTK supports both isolated whole word recognition and sub-word or phone based recognition.

Although the research in the area of automatic speech recognition has been pursued for the last three decades, only whole-word based speech recognition systems have found practical use and have become commercial successes (Rabiner et al., 1981 [5]; Wilpon et al., 1988 [8]). Two important reasons for this success are that the effects of context-dependence and co-articulation within the word are implicitly built into the word models and that there is no necessity of lexical decoding.

Isolated word recognition was considered for this project because it proved to be much easier because the pauses between the words make it easy to detect the start and end making it possible to detect each word at a time.

A limited grammar and dictionary were constructed to be used by the recognizer. The Speech data was recorded and labeled from 6 different speakers making the training and the testing corpus.

Since the researcher had labeled training data, the HTK tools HInit and HRest were used during the initialization and training processing. The results obtained from the system showed that the system can automatically recognize 94.87 percent words of any Kinyarwanda language speaker. The system was also tested on live data and it performed well. Four different speakers participated in the testing of the system on live data and performance was very good as seen in figure 6. There were some cases where the word kane was substituted with the word karindwi. This problem was mainly observed with some specific speakers not all.

Conclusion

The objective of this study was mainly to build a speech recognizer for Kinyarwanda language. In order to meet this objective a limited word grammar was constructed, a dictionary created and data from different Kinyarwanda language speakers was recorded and trained thereafter.

The system was tested using testing corpus data and live data and the system scored 92.00% sentence recognition and 94.87% word recognition. This implies that the objective of creating a system that can recognize spoken Kinyarwanda language was achieved.

The study is however not all conclusive as it has catered for only a voice operated phone dialing system. As much as it has created a basis for research, this study can be expanded to cater for more extensive language models and larger vocabularies.
Areas for Further Study

In spite of the successes of the whole word model speech recognizers which is also exemplified in the success of this project, they suffer from two problems:

- Co-articulation effects across the word boundaries. This problem has been reasonably well solved and connected word recognition systems with good performance have been reported in the literature (Rabiner et al., 1981 [5]; Wilpon et al., 1988 [8]).
- Amount of training data. It is extremely difficult to obtain good whole word reference models from a limited amount of speech data available for training. This training problem becomes even worse for large vocabulary speech recognition systems.

It is because of the above reasons that I therefore recommend for future research to be taken in large vocabulary Kinyarwanda language speech recognition, using sub-words (phonemes) which solve the above mentioned problems. A sub-word based approach is a viable alternative to the whole-word based approach because here, the word models are built from a small inventory of sub-word units. Phoneme HMMs are generalisable (trainable) both towards larger vocabulary and towards different speakers.

References


The inability to communicate with human beings in their own language may be one of the biggest barriers to the adoption of information and communication technology in the third world and the bridging of the digital divide. At the same time natural language processing is one of the most complex computational problems that have faced computer scientists. A number of approaches to addressing natural language understanding have been adopted over the years. While all these approaches have merit, it is clear that they have been most useful for languages that have a large knowledge base in terms of syntactic information, combined with extensive textual corpora. In the case of most Bantu languages these resources are not available now, and are unlikely to become available in the near future. Yet these languages are spoken by over 100 million people spread over most of Sub-Saharan Africa. In this research we note that Bantu languages have common logical forms and computational structures, and which have utility in semantic processing of natural language. We propose a derivational approach to discovering the structure of compound Bantu words that does not depend on large amounts of prior knowledge. This technique is amenable to machine learning methods such as reinforcement learning. Through this we demonstrate a new approach to natural language processing that may be of specific application to Bantu languages and to other language groups with a strong inflectional strategy.

Research Motivation
The traditional scientific study of human languages has become even more urgent in the age of digitization. Languages with good models are candidates for the development of speech to text systems, text to speech systems, machine translation and other computerized processing that will enable these languages to more easily adapt to the world wide web and other digitized media. Conversely, languages that are spoken by smaller populations that have few digitized or even textual materials, and are perceived as having little ‘economic’ value are severely threatened with extinction. Yet there are a number of reasons why it is important to preserve languages including preservation of knowledge about human origins, protection of indigenous knowledge particularly knowledge about biodiversity, loss of human wisdom and even basic human rights (Nettle D.; Romaine S., 2000).

Sub-Saharan Africa is dominated by the Bantu language group that is estimated to have between three and seven hundred variants (depending on how a variant is defined.)
These languages are rich in Linguistic form, probably encode a large part of indigenous African knowledge and are good candidates for preservation. However, their diversity and their shortage of textual and digitized forms may soon result in some of the languages becoming endangered.

The research motivated by these concerns has the following questions:

- Can recurring structures in Bantu languages be processed using computational approaches?
- Can computational pre-processing help Bantu languages make use of machine-learning approaches to build linguistic knowledge-bases?

In this paper we consider the case of one Bantu language, Gikũyũ, to demonstrate that a computational approach can be used to pre-process highly compound word forms using a machine learning approach such as reinforcement learning.

**Overview of Natural Language Processing Methodology**

Since about 1960 (generally the age when computerized processing of natural language became both desirable and possible) there have been two major paradigms for the processing of natural language: the rational and the empirical. (Manning C.D.; Schütze H., 1999)

The rationalists (exemplified by early research by Noam Chomsky) had difficulty in accepting that a phenomenon as complex as human language could be achieved by children purely through the simple and often inaccurate method of learning from examples. They therefore preferred a genetic, pre-programmed view of language acquisition where children were born with the rules and structure of language in place. An Artificial Intelligence researcher who sought to develop a natural language understanding system guided by this paradigm would have to discover the rules of language and embody them into a program which would then have the ability to process language in a human-like manner. Failures in this program might be caused by the inability of the human to enumerate, discover and correctly code all the rules present. The rationalists therefore seek a complete understanding of language.

Conversely, the empiricists assumed that humans begin with a few, simple, language-oriented abilities and then gain the bulk of their language ability through employing association, pattern recognition and generalization methods to a rich variety of sensory experience. The work of the empirical language researcher thus can be described using the following algorithm:

- **select** initial general language model
- **while** language examples are present
- **apply** pattern-recognizer to language example to obtain pattern
- **induce** additional language information using derived patterns
- **store** additional language information
- **extend** language model using stored language information

The empiricists, therefore, seek a utilitarian understanding of language that can deal competently with the most common types of language use.
The field of natural language processing has not reached a stage where we can place an automated natural language understanding device in normal everyday situations so most ‘language examples’ used by such system are drawn from a corpus – a collection of usually textual, usually digitized, examples crafted by a human being (or automatically generated e.g. by collecting digitized documents through the world wide web. (Getao K.W.; Miriti E.K., 2005)

It may be easier to select empirical approaches for the study of Bantu languages because, in most cases, they are not documented to the extent required by rationalists (an exception might be Kiswahili, see for example (Hurskainen A., 1992), (Hurskainen A., 2004).) However, the choice of an empirical approach does not preclude the two requirements of such a study: the choice of the initial general language model and the construction of a corpus. Both of these initial efforts are challenging for the reasons given below:

Choosing a language model is to a large extent subjectively dependent on the world view, background knowledge and motivations (political / social / technical / …) of the language modeler. For example, spoken sentences generally form a continuous whole so the method of splitting a sentence into classes of words such as verbs and nouns, or even selecting those classes is at the discretion of the linguist.

To develop a corpus a number of factors need to be considered:

• Some languages have very few written texts;
• Written texts may be specialized into a certain domain, for example over 90% of the Kiswahili texts discovered on the web had a religious context so that, for example, the word ‘God’ occurred disproportionately in our corpus;
• It is expensive to digitize a large number of texts;
• Written language is often markedly different in form from spoken language;
• It is expensive to transcribe speech.

The matter of both language model and corpus building will thus remain important and challenging research tasks for Bantu natural language. In the meantime Computer Scientists can assist the effort by identifying computational methods for processing natural language that do not require a complex language model and / or a large corpus.

Computational Modeling of Bantu Natural Language

From very early studies of Bantu languages that:

“The Bantu languages are attractive to the explorer not only from the harmonious adjustment of vowels and consonants, but from the logic of their grammatical structure, which, in the majority of these tongues, provides for a wide range and a nice discrimination in the expression of ideas.” (Johnstone Sir H.H., 1919) p.15.

Researchers in Bantu languages have identified syntagmatic features that occur in most Bantu languages, which include nominal classes and concords. Attempts to explain
the grammatical function of these features have proved problematic. For example, although the noun class that describes humans is fairly regular and consistent in most Bantu languages, attempts to explain the classification strategy for other nominal classes have (incompletely) relied on such diverse categorizations as shape, size, affect and force of nature. Indeed (Mugane J.M., 1997) notes that “the grouping of referential entities into noun classes has clear semantic criteria for some of the classes, while for others the semantics are not clear cut.”

As for the concords – word parts that are thought to provide agreement between other word types such as verbs and adjectives and the noun they are associated with – no one has explained why such seemingly redundant concords should exist at all. For example the English language uses proximity to associate an adjective with a noun e.g. “nice chair” while Kiswahili uses both proximity and concord “kiti kizuri.” There are cases where proximity is not used, for example “viti vile alinunua ni vizuri” but such cases do not enjoy superior comprehensibility because of concord. We can see this by comparing them to the English translation “the chairs which he/she bought are nice.”

The researcher’s attention is therefore drawn to the question of why the most common structures, nominal class and concord, should persist when they seem to be redundant. Are they ‘syntactic sugar’ or pronunciation features or are they computational features that are important to the processing of Bantu linguistic forms?

In seeking to use computational methods to describe natural languages we use the example of (Fokker J., 1995) who used functional programs to define algebraic language. We will later demonstrate that this approach has profound implications for the nature of the basic language model and the learning methods that can be applied to a Bantu linguistic corpus based on this language model. We would ideally like to choose a formalism that generates a language model that learns robustly from a reduced number of examples (taking into account the scarcity of suitable Bantu corpora.) If the language model can also be generalized to other Bantu languages then this would be a bonus.

**Inflections and concords as functions**

An inflection is a change in the form of a word to indicate a change in its grammatical function. For example in English the addition of an inflexion ‘d’ changes the verb ‘pose’ to it’s past form ‘posed.’

Bantu languages generally form their noun phrases using agreement concords drawn from the referential noun class and inflectional prefixes for tense. Thus, for example,

\[ \text{a-li-mu-ona he / she-past-him / her-see} \]

In the structural analysis of this phrase ‘a’ is a concord for the person class (third person singular) that serves for the subject of the sentence, ‘li’ is an inflexion indicating past tense, ‘mu’ is a concord from the person class (third person singular) that serves as the object of the sentence, and ‘ona’ is a verb particle meaning ‘see.’ The referents of ‘a’ and ‘mu’ must be drawn from the surrounding context, either within or outside the
sentence. Thus ‘Alimuona’ is a complete sentence, but it could also have been more specific ‘Juma alimuona Fatuma’ with Juma as the subject and Fatuma as the object. A less common passive construction is ‘Alimuona Fatuma, Juma.’

In structural analysis these concords and inflexions are viewed as grammatical features of the Bantu sentence. However, in this research we investigate the possibility of using them as computational features, by creating a set of concatenation functions that affix various concords and inflexions to a class root to create new compound words. A class root is the most basic ‘atomic’ particle that exists in the language.

The concatenation functions belong to function classes. The function class Tense contains several functions for concatenating the inflexions for present, past and future and so on. The function class Source also has several functions relating to the concatenation of suitable inflexions to represent singular and plural, first, second and third person. In Bantu linguistics the noun classes are also critical to the selection of the correct person function and this may be an important facility in the selection of information from the Reference. The function class Target has similar properties to the function class Source. Finally the function class Deed defines all the class roots of the language. The basic language model consists of the definitions of these functions. The application of these functions for a fragment of language produces a semantic model of the fragment. The analysis of these functions and their interaction with the Reference can be a means of building a refined language model using a corpus.

Abstraction and specialization of nouns by higher-order functions

Another language feature that occurs in some Bantu languages is noun derivation, where suffixes and prefixes are used to subtly alter the meaning of nouns or even to change a word from one word class to another (e.g. nouns can be derived from verbs as in English run → runner.) The Gikũyũ language in Kenya is an example of a Bantu language that demonstrates prolific noun derivation strategies. Diagram 1 below shows some of these strategies

It is noted that the derivation strategies are regular (and their semantics are also generally regular.) Many of the derivation strategies are also compositional in nature. One derivation, for example an agentive verbal suffixation can be applied immediately after another derivation such as a nominal base prefixation to produce a perfectly valid agent (an agent is a noun representing a person.)
Diagram 1: Derivation of nouns in Gikũyũ language (Mugane J.M., 1997)

It is this regularity of application that causes one to speculate that a functional approach to exploratory compound word derivation may be a useful tool for discovering the structure of Bantu compound words.

For example concords drawn from a specific noun class can be fixed into a general class function using an argument fixing functions such as:

(define nominal-base-fix
  (lambda (prefix)
    (lambda (class-root)
      (string-concatenate prefix verb-root)
  )
)

Similarly verbs can be fixed into a general class of agentive verbs by composing another nominal derivation functions (such as the one above) with a string transformation function. Thus a functional interpretation of the noun derivation operations would add considerable power to Bantu language processing.

**Compound words in Gikũyũ**

A common quality of Bantu languages is the compound word. Compound words exist in many languages, for example, in English we have a set of related words (run, runs, ran, runner, runners, running) which are clearly related. They consist of the root ‘run’ with affixes which subtly alter its meaning. During linguistic processing we often pre-process linguistic expressions to ensure that ‘run’ and ‘runs’ are not recognized as separate words but rather as different forms of the same word.
In a Bantu language such as Gĩkũyũ the compounding can be quite complex (see for example (Ng’ang’a P. M., 1996) page 1.) For example, using the root verb ‘ciar’ roughly meaning ‘born’ many words can be formed by the addition of affixes including: <ciar>a (verb – give birth), agĩ<ciar>ũo or ara<ciar>ũo (a verb phrase – he / she was born), gũ<ciar>ũo (the infinitive – to be born), gũgũ<cia>rũo (a prepositional phrase – after the birth has taken place), mũ<ciar>i (noun – parent, the-one-who-has-given-birth-to. These many forms are related but would require a powerful pattern-matching algorithm in order to access the common <ciar> in all these compound words. This is complicated by the occurrence of ‘ciar’ in other forms such as <ciar>a (noun – fingers.) It would require a large corpus to access comparative compounds that would help us to derive the key component from each compound, for example:

1. gũgũ<ciar>ũo
2. gũgũ<thi>o
3. gũgũ<tham>ũo

The situation is further complicated by changes that take place in the structure of the compound word based on pronunciation—for example 1,3 share the common suffix ‘ũo’ whereas 2 has suffix ‘o’ based on issues of pronunciation. The recursive form of compounding in Gikũyũ means that regular expressions are not suitable for representing the diversity of structure that occurs in these compound forms.

Since there is limited written material in the language it is difficult to achieve a corpus of the size needed to use machine learning techniques such as clustering or neural networks to capture the essential structure of the common forms of these compounds. In this research we thus propose a generative method to enable the use of a trial and error machine learning approach such as reinforcement learning as an alternative method of gathering linguistic information from a resource-scarce Bantu language.

Derivational Method for Gikũyũ Compounds

We shall draw from the work of (Mugane J.M., 1997) in naming the minimal particle that can be achieved from a compound the class root. We shall assume that the class root is modified by various affixes to form a compound. The process of modification that is carried out using the recursive rules of the derivational method is successful if a derived compound matches a compound word that occurs in the language. (The word that is used to validate the derived compound may be obtained by any means, either from a text corpus, or from a speech corpus, or by recognition by a fluent speaker of the language.)

Drawing on the work of (Steedman M., 2004) we assume that the derivation is compositional. As he notes: Composition is one of the most primitive combinators, or operations combining functions, which (Curry H.B., F. R., 1958) call B, writing the above sequence a; b as Bab, where:

\[
Bab = I( a(b(t)))
\]
However, in this case the composition is not absolutely free because not all \(a\) and \(b\) functions can be composed into a function that will derive a valid Gikũyũ compound. The task is to find a utilitarian set of compositions that can be used to derive valid Gikũyũ compounds.

In the above equation for composition, \(r\), in the derivational calculus is always the class root. A series of functions: \(a, b, c, d, e, f\) is defined that are concatenation functions that prefix, suffix or infix \(r\) with a modifier. Concatenation functions can be developed analytically or they can be discovered using a machine learning algorithm such as clustering (or even a simple algorithm that splits words into syllable since many Bantu inflexions appear to be syllables.) Examples of concatenation functions are:

- Functions that prefix a compound with a noun class concord (for example the verbal compound ‘igua’ → <mũ>igua by prefixing with the class concord for the person class.
- Functions that suffix a compound with a relational verb prefix (for example the class root ‘igu’ → igu<irthania> by prefixing with a transitive verb prefix.

Each time an additional function is applied to the current composition the derived compound is tested for occurrence.

Composition is thus an infinite function that is only terminated when a valid word in recognized. Once a valid word is recognized the composed function is stored for testing with other roots. Since the function may not be valid for all roots, each function is associated with a list of roots for which it has been accepted as valid. In a machine learning approach such as reinforcement learning, the agent is rewarded for finding the most general compositions, that is, those that are valid for the greatest number of class roots. Eventually the compositional functions may be used to tag a real corpus.

**Reinforcement learning**

The definitive text on reinforcement learning (RL) defines it as “Reinforcement learning is learning what to do – how to map situations to actions – so as to maximize a numerical reward signal. The learner is not told what actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These two characteristics – trial and error search and delayed reward – are the two most important distinguishing features of reinforcement learning.” (Sutton R.S.; Barto A.G., 1998)

(Harmon M.E.; Harmon S.S., 2000) characterize RL as the marriage of dynamic programming, a mathematical method traditionally used to solve problems of optimization and control, and supervised learning a generalized method for training a function approximator (e.g. neural network) to represent functions. RL performs well in large complex environments where the correct answers (training examples) are not known.
The definition of an RL problem has the following parameters:

- A complex, dynamic environment;
- A reinforcement function usually represented as a mapping from state/action pairs to reinforcements, which is applied after each state change in the environment.

There are a number of strategies for representing complex dynamic environments and reinforcement functions in the literature. For example, reinforcement functions can be represented using formally justified techniques such as dynamic programming or learning automata, or ad hoc techniques such as greedy strategies (Kaelbling L.P.; Littman M.L.; Moore A.W., 1996).

**Learning the compositional functions for Gikũyũ compounds**

The complex dynamic environment for the reinforcement learner consists of:

1. A composition function, $C$, that recursively composes functions from the set of functions (consisting of the base set of concatenation functions and any other functions composed during the interaction), $S$.
2. A finite set of concatenation functions. Some of the concatenation functions are organized into function classes. Functions that belong to the same class are interchangeable in any composed function, $F \subseteq S$.
3. A class root set, $R$, from which strings defined as class roots may be drawn as starting function arguments.
4. A reinforcement function, $Q$, that rewards the reinforcement learner for composing more general functions.
5. A set of valid (compound) words, $V$, that consists of all those words which have so far been found to exist in the language.
6. An ordered set of composed functions, $W$, that consist of those functions that have been found to derive valid words, each being associated with a set of class roots, $T_n$, that are those class roots that are valid arguments for the function. The composed function with the largest $T_n$ has precedence in the ordered set.

The reinforcement learning model consists of:

- a discrete set of environment states, $S$. The state of the environment is represented by the set $W$;
- a discrete set of agent actions, $A$. The actions are the choice, composition and application of two functions from the set $S$ on a word from the set $R$; and
- a set of scalar reinforcement signals representing the reinforcement function, $Q$, (in this case a real number representing the length of the list of words $T_n$ from the validation set, $V$, that can be generated using the latest composed functions.)

The goal of the reinforcement learner is to learn a policy, $p$, that composes the most
general functions for generating compound words in Gikũyũ. This algorithm generates one composition per cycle. As with all reinforcement learners the learner incorporates an exploration strategy that causes it to periodically compose a random function to avoid becoming stuck in a local minimum (in this case, we wish to prevent the learner from continuously re-using the most successful composition functions.) Several researchers have developed reinforcement learning algorithms that are pertinent to this kind of an approach for example (Moriarty D.E.; Schultz A.C.; Grefenstette J.J., 1999) or (Dixon K.R.; Malak R.J.; Khosla K.P., 2000).

Conclusion

In this paper we began from the premise that BNL is a worthwhile topic for natural language processing research because we want to preserve the languages that are spoken by so many African voices, and that contain so much valuable knowledge. We went on to argue that current corpus-based techniques may not be adequate for languages that are disadvantaged in the quantity and quality of digitized texts available for inclusion in the corpus. Such languages may also have language models of variable quality.

We studied Bantu language structures from the novel perspective of viewing the concords and inflections that abound in Bantu languages from a functional viewpoint as the products of compositional concatenation functions derived using a reinforcement learner.

In this way we have demonstrated that it is possible to use a computational approach to derive linguistic knowledge from a small amount of knowledge about a Bantu language. We have demonstrated that a derivational, language-engineering approach can be posed as a reinforcement learning problem.

The exploratory and knowledge-efficient methodology posed in this paper holds promise for addressing the problem of digitizing a resource-scarce language such as Gikũyũ.

References


Mugane J.m. (1997) A Paradigmatic Grammar Of Gikûyu, Center For The Study Of Language And Information.


Part Three

Computer Networks and Applications
11

On Network Measurement and Monitoring of end-to-end paths used for e-VLBI.

Julianne Sansa, Makerere University, Faculty of Computing and IT, Kampala, Uganda, sansa@cit.mak.ac.ug; Arpad Szomoru, Joint Institute for VLBI in Europe, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands, szomoru@jive.nl and J.M. van der Hulst, Kapteyn Astronomical Institute, Postbus 800, 9700 AV Groningen, The Netherlands, vdhulst@astro.rug.nl

This paper presents the bottlenecks currently limiting the transfer speeds of astronomical data from radio telescopes across the world over high speed links to the central processing centre in the Netherlands. The processed data is used to produce images which are used by radio astronomers. The image sensitivity of continuum observations scales with the data rate, so the higher the data rate, the more sensitive these observations will be. The required high data rates should be attainable on these network links, however it is not the case. The aim of our study was therefore to uncover the prevailing bottlenecks and propose solutions to address them. The bottlenecks were identified by conducting data transfer tests over these links under various network condition. We also provide an explanation for the existing bottlenecks as well as potential solutions to overcome them.

1. Introduction

Very long baseline interferometry (VLBI) is a technique that enables astronomers to observe cosmic objects at radio wavelengths with an extremely high angular resolution. This is done by combining the signals of widely separated radio telescopes using a correlator, a custom-built supercomputer. In this way a telescope with the size of a continent, and even of the earth, can be simulated. The telescopes used in the European VLBI Network (EVN) produce data at rates of up to 1Gbps each. Traditionally, these data streams were recorded on tapes (nowadays hard disk drives) and shipped to the correlator located at JIVE, the Joint Institute for VLBI in Europe, in Dwingeloo, the Netherlands. During the last few years JIVE, in collaboration with the European National Research Networks and the pan-European Research Network GEANT, have worked on a proof-of-concept project to connect several telescopes across Europe in real time to the correlator via the Internet (electronic VLBI or e-VLBI). This project has led to an EC sponsored project called EXPReS, which over the next few years will transform the EVN to a fully functional real-time e-VLBI network.
During the PoC project it became clear that in spite of the vast capacity of the connecting networks, the actual transport of large amounts of data poses quite a challenge especially for real time correlation and in utilizing all the physically available bandwidth.

In order to investigate the throughput and examine what limits the data flow we carried out a number of network transfer tests under varying conditions. In this paper we report how the tests were set up and evaluate the results in terms of what limits the network performance.

The Mark5 [1] application that handles e-VLBI data uses the Transport Control Protocol (TCP). By the nature of e-VLBI, huge amounts of data have to be transported via the Internet over long distances from geographically dispersed telescopes to one central correlator. Internet measurement tools are classified into two classes, passive and active [2]. Active tools inject traffic into the network and use it to gather network statistics, while passive tools use already existing traffic to compute the status of the network. We monitored the network paths used for e-VLBI in the Netherlands with web100 [3, 4] (a mix of active and passive tools) and tcpdump [5] (a passive tool) to gather statistics on the congestion window, Round Trip Time (RTT), packet loss and the resulting throughput. The aim was to clarify the relation between these variables and the achieved throughput.

1.1 Background

When a TCP [6, 7] connection is established between sender and receiver, a congestion window (CWND) is negotiated. This is the amount of data that a sender can send before receiving any feedback from the receiver. A TCP receiver maintains a receive window (RWND), for the purpose of informing the sender how much data it is willing to accept in one go without needing to generate acknowledgements. It is thus in the best interest of the TCP connection that CWND and RWND are close to one another, otherwise the lesser value will be the effective value for both. TCP senders update the CWND in response to acknowledgments of received packets and the detection of congestion. Standard TCP interprets packet loss as congestion. This packet loss is detected through 1) the timeout of an unacknowledged packet, 2) the receipt of several duplicate acknowledgments, which emphasize that a previously acknowledged packet is being re-acknowledged implying to the sender that the next packet whose turn it was to be acknowledged has been lost, or 3) through selective acknowledgments (SACK), which indicate specific packets received among a block that was sent implying that the rest of the packets within the block were lost. A TCP sender can be in two states, the slow-start or the congestion avoidance state. A sender which has just established a TCP connection will be in the slow-start state where the initial value of CWND is very small (one packet for standard TCP) and increases exponentially as it senses the network congestion status with the aim of using the unused bandwidth as quickly as possible. As the sender's WND continues to grow it will cross the so-called slow-start threshold and the connection will move into the congestion avoidance state. In this state, the increased chance of congestion causes the increase of the CWND to slow down.
In these two states the sender will adjust the size of the CWND according to the following equations (1) - (4) depending on whether the previous packet was acknowledged or lost:

**Slow-Start:**

On acknowledgement:

\[ \text{CWND}_{\text{new}} = \text{CWND}_{\text{old}} + c \]  

(1)

On packet loss:

\[ \text{CWND}_{\text{new}} = \text{CWND}_{\text{initial}} \]  

(2)

**Congestion Avoidance:**

On acknowledgement:

\[ \text{CWND}_{\text{new}} = \text{CWND}_{\text{old}} + \frac{a}{\text{CWND}_{\text{old}}} \]  

(3)

On packet loss:

\[ \text{CWND}_{\text{new}} = \text{CWND}_{\text{old}} - b \times \text{CWND}_{\text{old}} \]  

(4)

where \( a = 1, b = 0.5, c = 1 \)

On high speed links the TCP connection quickly passes through the slow-start state into the congestion avoidance state and as a result most of the connection’s lifetime is spent in the congestion avoidance state. From Floyd [8] the following relationships are taken:

\[ \text{CWND}_{\text{optimal}} = \text{Bandwidth} \times \text{RTT} \]  

(5)

\[ \text{Throughput} = \frac{\text{CWND}_{\text{average}}}{\text{RTT}} \]  

(6)

\[ \text{CWND}_{\text{average}} = \frac{1.2}{\sqrt{p}} \]  

(7)

Where RTT is the Round Trip Time, and \( p \) is the packet loss rate. From equations (6) and (7), we note that the TCP throughput is directly proportional to the CWND and inversely proportional to the RTT and that the CWND in turn is inversely proportional to the square root of packet loss rate. Making long distance connections over high speed links as is the case for e-VLBI, implies that the optimal CWND required to ensure high
throughput has to be large, as suggested by equation (5). The TCP congestion avoidance algorithm, in equations (3) and (4), however exhibits a weakness in maintaining a large CWND in the presence of packet loss.

1.2. Questions to address

The main question we want to address is what information we need to predict network performance. We can specify this further with the following questions:

i) What is the available bandwidth for a connection that TCP is about to make.

ii) If the available bandwidth is less than the theoretically available bandwidth, what is the limiting bottleneck (sender, network or receiver).

iii) How much packet loss and RTT occurs in the end-to-end path of the established TCP connection.

iv) What is the stability of the TCP connection established.

v) How can we minimize the impact of the identified bottleneck(s).

In this paper we will address questions i) - iv) and provide potential answers to v).

2. Methodology

In this section we describe the network topology over which we conducted the network tests as well the hardware configuration and the tools we used to gather network statistics.

2.1. Setup

Tests were conducted between Mark5 units located at JIVE interconnected via Amsterdam through Netherlight [9] as shown in Figure 1.

Each of the links from the router AR5 to each of the Mark5’s is 1 Gbps. Statistics were gathered for iperf [10] flows or Mark5 disk2net-net2disk transfers from one Mark5 unit through the router AR5 to another Mark5 unit. Iperf flows are generated from the memory of the sending machine to the memory of the receiving machine and are able to measure the maximum bandwidth. The Mark5 application has several data transfer modes. The disk2net-net2disk mode which we used gets already recorded VLBI data on a disk and sends it to the network on the sending side and on the receiving side data.
is got from the network and written to the disk. Some tests were conducted through Amsterdam to Manchester, UK, via UKlight [11] where the end-to-end path has a minimum of 1 Gbps.

Figure 1: Network topology over which tests were conducted

2.2. Hardware configuration

The hardware configuration of the Mark5 units is as follows:

- Motherboard: Supermicro P3TDLE, FSB 133 MHz
- Processor: Pentium III 1.26 GHz, FSB 133 MHz, 512K L2
- Chipset: ServerWorks Serverset III LE
- Memory: 256M PC133 SDRAM
- Operating System: RedHat Release 9.0, linux kernel 2.4.20-6

2.3. High Performance Networking Options

In order to ensure maximum throughput, the following well known high performance networking options [12] were tuned. Maximum Transmission Unit (MTU) of 8192 bytes was supported for all tests in the Netherlands, while that of 4470 bytes was supported
for tests involving Manchester, compared to the default 1500 bytes. TCP buffers were set to 4 Mbytes compared to the 64 Kbytes default while txqueuelen was set from a default of 100 to 20000, which has been proved to offer good performance [13]. The default values are too small to support high speed data transfers.

2.4. Measurements with Web100 & TCPdump

We use Web100 [3, 4], a set of tools that together provide an advanced management interface for TCP. It is a passive tool in that it uses ordinary TCP traffic to calculate TCP event statistics per connection from the received acknowledgments. However, since statistics are computed for TCP connections from TCP acknowledgments, web100 is also an active tool. The diagnostics provided by Web100 are extremely useful for evaluating the link performance.

During e-VLBI transfers we gather statistics similar to those reported with Web100 using TCPdump [5]. TCPdump generates no traffic, it merely keeps track of all the traffic going through a particular network interface, making it a passive tool.

3. Tests

In this section we present the results of our tests in four subsections: observed CWND & RWND, packet loss, RTT and TCP throughput.

3.1. Observed CWND & RWND

As mentioned in Section 1.1, CWND and RWND should be nearly equal and large enough to allow full bandwidth utilisation. The links have a bandwidth of 1 Gbps, the RTT for tests in the Netherlands is ~ 4 ms, while the RTT for tests involving Manchester is ~ 16 ms. From equation (5) in Section 1.1 the optimal CWND and RWND for tests within the Netherlands are 0.5 Mbyte, while for tests to Manchester they are 2 Mbyte. In Figure 2A we show the results for a single data flow between JIVE and Manchester. Both CWND and RWND are in close proximity of each other with an average of \(1.58 \times 10^6\) bytes, which is 79% of the optimal CWND. When producing two parallel competing data flows as shown in Figure 2B we begin to see unfair allocation of the CWND for each flow, one with an average of \(1.48 \times 10^6\) bytes (which is 74% of the optimal CWND) and the other with \(1.19 \times 10^6\) bytes (which is 59.5% of the optimal CWND). The same situation is seen in Figure 2C for five parallel flows where we see that all the five values of the CWND range from \(5.29 \times 10^5\) bytes to \(5.76 \times 10^5\) bytes (which is ~25% of the optimal CWND). For multiple flows Figures (2B and 2C), we do not show the corresponding RWND plots because RWND values are always higher than the CWND values, making CWND the effectively used variable. In both Figures 2B and 2C the effective combined CWND of the multiple flows is greater than 100% of the optimal CWND, thus the implied throughput that could be achieved is close to the peak value of 1 Gbps.
3.2. Packet loss

From our measured maximum CWND we calculate a steady state packet loss of $7.49 \times 10^{-6}$ [8]. A rate larger than this steady state rate will cause a decrease of the CWND, while a smaller rate will cause an increase. We observed zero packet loss rate for single flows of both iperf and the Mark5 application and variable average packet loss of $1.18 \times 10^{-9}$ and $2.81 \times 10^{-8}$ for two parallel flows, while for five parallel flows it ranged from $3.03 \times 10^{-8}$ to $5.02 \times 10^{-8}$ as shown in Figure 3A and Figure 3B for 2 and 5 flows respectively. As the observed packet loss rates are much smaller than the steady state packet loss rate, the CWND should be increasing. This however is not the case. We will return to this in Section 4.

Figure 2: Congestion Window and Receive Window for A: a single memory-to-memory data flow (CWND and RWND are nearly equal), B: two parallel memory-to-memory data flows (RWND is not plotted as it is much larger than CWND, making CWND the effective limiting value), C: five parallel memory-to-memory data flows (here too RWND is not plotted as it is much larger than CWND) and D: disk2net-net2disk e-VLBI data transfer (RWND smaller than CWND).
3.3. RTT

In iperf tests from Dwingeloo to Manchester we observed an alternating RTT between two points 10 ms and 20 ms for a single flow, alternating RTTs between three points 10 ms, 20 ms and 30 ms for two and five parallel flows. During e-VLBI data transfers between two hosts both located at JIVE in the Netherlands, we noted an average RTT of 3.8 ms with a few spikes to approximately 100 ms. These results are shown in Figure 4. The general observation is that RTT values are quite stable even when multiple flows are generated, making RTT ideal to use as a sign of congestion.
Figure 4: Round trip time for A: a single memory-to-memory data flow, B: two parallel memory-to-memory data flows, C: five parallel memory-to-memory data flows between a Mark5 in the Netherlands and a machine in Manchester and D: a disk2net-net2disk e-VLBI data flow between two Mark5’s both in the Netherlands

3.4. TCP throughput

An average TCP throughput of 810.7Mbps was achieved for a single TCP flow as shown in Figure 5A. For two parallel flows the aggregate TCP throughput was 973.4 Mbps, which is more than what was achieved by a single flow. There was however unfair sharing of this throughput with one flow having average 578.6 Mbps and the other having 394.8 Mbps. This is illustrated in Figure 5B. For five parallel flows the aggregate TCP throughput is 977.8 Mbps, which is closest to linespeed, with the various flows having 203.8 Mbps, 174.7 Mbps, 204.2 Mbps, 192.8 Mbps and 202.3 Mbps. The unfair sharing of the available bandwidth is shown in Figure 5C. For the disk2net-net2disk e-VLBI transfer we observed an average of 366.9 Mbps, although performance was
affected by the tcpdump running at the same time by a factor of between 20% and 40%, which implies this average throughput would have been a value between 440.3Mbps and 513.7 Mbps without TCPdump.

Figure 5: Achieved throughput for: A - a single memory-to-memory data flow (a high stable throughput is achieved), B - two parallel memory-to-memory data flows (unequal throughput is attained by each flow, which slightly fluctuates in each case), C - five parallel memory-to-memory data flows (unequal throughput, which fluctuates significantly and tends towards stable value for each flow) and D - disk2net-net2disk e-VLBI data flow (throughput fluctuates at the beginning of the flow and stabilises a short while later).
Summary of Test Results

Table 1 shows a summary of the average CWND, RWND, packet loss rate, RTT and throughput observed for each data flow under the different scenarios.

4 Discussion of Results

In this section we discuss the results obtained from the tests presented in the Section 3 as well as the bottlenecks we found and their likely solutions.

Table 1: Average CWND, RWND, packet loss rate, RTT and throughput for each data flow under the different scenarios (some values have no error bound indicated because it is unmeasurably close to zero due to the fact that such values mostly maintain a constant value)

<table>
<thead>
<tr>
<th>Data Flow category</th>
<th>CWND (MBytes)</th>
<th>RWND (MBytes)</th>
<th>Packet loss rate (×10⁻⁹)</th>
<th>RTT (ms)</th>
<th>Throughput (Mbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single memory-to-memory data flow</td>
<td>1.6</td>
<td>1.6</td>
<td>0</td>
<td>16 ± 5</td>
<td>811 ± 1</td>
</tr>
<tr>
<td>Two parallel memory-to-memory data flows</td>
<td>1.5 ± 0.2</td>
<td>1.6</td>
<td>28 ± 1</td>
<td>21 ± 9</td>
<td>579 ± 8</td>
</tr>
<tr>
<td></td>
<td>1.2 ± 0.1</td>
<td>1.0</td>
<td>1 ± 3</td>
<td>22 ± 6</td>
<td>395 ± 4</td>
</tr>
<tr>
<td>Five parallel memory-to-memory data flows</td>
<td>0.6 ± 0.3</td>
<td>1.6</td>
<td>30 ± 5</td>
<td>23 ± 8</td>
<td>204 ± 20</td>
</tr>
<tr>
<td></td>
<td>0.5 ± 0.3</td>
<td>1.6</td>
<td>50 ± 32</td>
<td>23 ± 8</td>
<td>175 ± 9</td>
</tr>
<tr>
<td></td>
<td>0.6 ± 0.3</td>
<td>1.6</td>
<td>30 ± 3</td>
<td>23 ± 8</td>
<td>204 ± 11</td>
</tr>
<tr>
<td></td>
<td>0.6 ± 0.3</td>
<td>1.6</td>
<td>38 ± 3</td>
<td>23 ± 8</td>
<td>193 ± 7</td>
</tr>
<tr>
<td></td>
<td>0.6 ± 0.3</td>
<td>1.6</td>
<td>36 ± 24</td>
<td>23 ± 8</td>
<td>202 ± 8</td>
</tr>
<tr>
<td>Disk2net2disk e-VLBI data flow</td>
<td>0.06 ± 0.01</td>
<td>0.05 ± 0.003</td>
<td>0</td>
<td>4 ± 3</td>
<td>367 ± 106</td>
</tr>
</tbody>
</table>

4.1. Single Flows

In the case of one TCP flow (both iperf and e-VLBI) we see the same CWND sustained for a period of time and yet with both slow-start and congestion avoidance algorithms we should be seeing CWND either increase or decrease. This can be explained as idle connection window validation [14], in which the same constant CWND is maintained during an application limited period, which means the CWND is not increased merely by reception of ACKs, as long as during the previous RTT the flow did not fully use the available CWND. This seems to imply that the single flows experience severe application limitation on these high speed links. Application limitation happens when the application does not produce data fast enough [15] for two reasons. Either the application is transfering small amounts of data at a relatively constant rate to the TCP layer or the application is producing data in bursts separated from each other by idle periods. Based on this explanation we conclude that the iperf flow is application limited due to the former while the e-VLBI data flow is application limited due to the
latter. The focal interest being the e-VLBI data flow, eliminating or shortening the idle periods between the data bursts would rid the flow of application limitation. It seems then that single flow transport of e-VLBI data is limited by two distinct effects. Firstly, idle periods between data bursts, and secondly, network congestion not caused by this particular flow’s packet losses (since the single flow experienced zero packet loss throughout its lifetime). This network congestion must be hardware related (e.g. PCI bus conflicts, NIC performance) [16, 17] and can only be addressed by advances in hardware technology.

4.2. Parallel Flows

Our tests for parallel iperf flows are similar to those done by Hacker [18] but differ in that we list the CWND values in addition to the packet loss and throughput. We conducted two kinds of parallel flow tests; the first with two flows and the second with five flows. We did not test with all the possible number of parallel flows because our goal was to establish the general TCP interaction between parallel flows and the resulting throughput and thereafter compare these against the single flow test results. Having observed similar characteristics in the two and five flows tests, we extend our observation to apply to parallel flows in general considering some aspects [18] such as the anticipated increase of packet loss and RTT variation with more parallel flows. In addition the network over which we conducted our tests differs in terms of RTT and the supported MTU.

With parallel flows the aggregate throughput is much higher than what single flows achieve. The parallel flows are able to transfer data without application limitation surfacing because each flow does not need as much data to be fully utilized compared to a single flow, however the TCP congestion avoidance algorithm limits the further increase of the CWND since packet loss is experienced. It is under such circumstances that the use of modified congestion avoidance algorithms [8, 19, 20, 21, 22] may improve network performance.

4.3. Bottlenecks

For the single memory-to-memory data flow, the hardware in both the sending and receiving hosts as well as the application limited the throughput, because both CWND and RWND were stable throughout the connection’s lifetime. For the multiple memory-to-memory data flows the TCP congestion avoidance algorithm was the bottleneck because the CWND kept fluctuating implying the detection of packet loss. For the disk2net-net2disk e-VLBI data flow the receiver’s hardware limited the throughput because the RWND continuously fluctuated and was lower than the CWND, indicating that the receiving host was overwhelmed. Apart from the limitations discussed above, a number of other bottlenecks are possible:

• Undue retransmissions. TCP’s fast recovery mechanism may yield unnecessary packet retransmissions, thus reducing the effective throughput.
• Interface stalls. These happen when ever the network interface halts as it waits to receive data from upper layers. These stalls are interpreted as a sign of congestion, reducing the CWND and consequently the throughput.

• Vendor specific TCP implementations. Some TCP implementations have incorporated a mechanism of validating the CWND in cases where its value has been constant for a given period of time and this mechanism also contributes to reducing or maintaining a constant CWND depending on the situation.

5. Conclusions & Future Work

In this paper we show that application limitation can be a significant factor in single flows on high speed links. When parallel flows are used, the TCP congestion avoidance algorithm takes over as the limiting factor. We also note that packet losses other than those caused by network congestion can be quite important. Future work will include eliminating the idle time between data bursts during the lifetime of an e-VLBI data flow. We intend to do an analysis of the proposed aggressive congestion avoidance algorithms. We also intend to model TCP e-VLBI data flows and relate flow analysis with correlation to determine how correlation is impacted by packet loss and delay.

References


Multiagent Systems for Distributed Resource Allocation

Eric Ayienga, Elijah Opiyo, William Okello-Odongo, Kate Getao, Bernard Manderick and Anne Nowé, School of Computing and Informatics, University of Nairobi, P.O. Box 30197, Nairobi 00100

Wireless grids introduce additional challenges to the challenges already existing in resource allocation in the wired grid environment. Economic approaches and Multiagent Systems (MASs) are brought into one framework to solve problems of resource allocation and efficient QoS provisioning. Wireless grids are considered as complex systems with economies where multiple applications (consumers) compete for resources and services from network providers (suppliers). Resources are priced by suppliers based on demand, and users buy resources to satisfy their QoS needs. Optimization is done on network QoS parameters.

1. Introduction

A Grid can be considered as an automated resource allocation system. Its infrastructure is expected to support a wide variety of applications, with some having stringent resource requirements (inelastic applications) while others are flexible enough to adapt to whatever is available (elastic applications). Providing this service differentiation and resource assurance is often referred to as quality of service (QoS). Allocating the scarce network resources to satisfy the QoS requirements of multiple applications and users with widely different valuations for the underlying services presents a challenging resource allocation problem [MAU02]. This is due to; the existence of users who act in their own interest, resource constraints, rapidly changing and unpredictable demands, and many unreliable hosts that are physically and administratively distributed [LRA+04].

A wireless Grid complicates this further due to limited power of the mobile devices, the limited bandwidth, and the increased dynamic nature of the interactions involved [Ban05][ANA04]. A possible solution to this problem is overprovisioning through adding more resources. This is a costly solution but the only solution if the resources are already optimally allocated. Another solution relies on finding efficient mechanisms to optimally allocate the available resources.

Approaches to implement QoS in wireless Grid networks are specified in the IEEE 802.11e standard as cited in [MHB04]. These approaches are still in their formative stages and still have some limitations. This paper proposes a MAS solution that uses principles from economics to allocate scarce wireless local area network (wLAN) resources. Multiagent systems (MAS) offer strong models for representing complex and
dynamic real-world environments e.g. simulation of economies, societies and biological environments [LMSW05]. Economic techniques are considered because markets are effective mechanisms for allocating scarce resources in a decentralised fashion [NH05]. They achieve this based on the exchange of small amounts of information such as prices. They provide a natural way of viewing the resource allocation problem by ensuring that the individuals who value the resources most get them. In the proposed scheme, the pre-defined QoS requirements of applications are mapped to the actual resources using economic principles.

The resource allocation problem is considered as a market in which there are Seller/Resource agents and Buyer/QoS agents trading on items representing the resources [BACdK02]. The agents exhibit different acquisition capabilities which let them act differently depending on the current context or situation of the market. The scheme works by auctioning the resources and then updating the prices of the resources based on the laws of supply and demand. An applications preference for one QoS configuration over another is summarized by its utility function. For one particular resource allocation, there could be a number of different QoS configurations possible, and there has to be a trade-off between the various QoS measures.


2. Grid Computing

Grid computing allows the uniting of pools of servers, storage systems, and networks into a single large system so that the power of multiple-systems resources can be delivered to a single user point for a specific purpose [IBM03]. The main resources Grid computing is designed to give access to include, but are not limited to

- Computing/Processing power.
- Data storage/networked file systems
- Communications and bandwidth
- Application software

This pooling requires a significant hardware and software infrastructure. The hardware infrastructure is made up of supercomputers and their interconnection by high-speed networks. The software infrastructure monitors and controls the hardware to implement the user’s requirements. Because of its complexity the Grid is usually described in terms of its components and layered architecture.

2.1. Components of the Grid

The key components of a computational grid include the following [Min05]:

- Resource virtualization: The central concept of resource virtualization is the introduction of an indirection layer below the execution environment seen by applications and operating systems. This provides a basis for achieving flexible, secure sharing of resources.
• Security management: the Grid needs to take care that only authorized users can access and use the available resources
• Data management: data must be transported, cleansed, parcelled, and processed
• Services management: users and applications must be able to query the grid in an effective and efficient manner

2.2. Architecture of the Grid

Grids consist of a multi-layer architecture that ties physical devices such as servers, disk arrays, specialized equipment and network switches to applications. At a very high level, the layered architecture is composed of four layers

• The application layer: This is the layer that users/applications use to interact with the grid.
• The middleware layer: This is used to manage resources on the grid. It is a software function and is where most of the “intelligence” of the grid comes from
• The resource layer: Covers most of the resources found on the grid.
• The network layer: This physically links the resources together using guided and unguided media.

Most of the problems that a Grid solves involve data processing, network bandwidth or data storage. In grid computing, QoS is implemented at all the layers i.e. at the application, middleware, resources and network levels. To design a system that serves the dual purpose of QoS satisfaction and high resource utilization, the QoS requirements of the applications must be determined and the application data (characterized by application parameters) mapped to appropriate network parameters. This can be a challenging task given that application metrics do not map directly to network metrics [JS03]. Grid traffic is highly diverse and each traffic type has unique requirements. Network-level QoS mechanisms are required to enable the applications to use grid services efficiently.

3. Network support for QoS

Network QoS refers to a set of qualitative and quantitative traffic characteristics (e.g. throughput, service interval, packet size, delay, jitter, priority, type etc), which describes a traffic flow in support of a specific application. A flow is packet stream from a source to a destination (unicast or multicast) with an associated QoS and higher level demultiplexing information [CNRS98]. Traffic flows from different applications have different requirements and should be provided different QoS from the network. A QoS system is made up of several components that include QoS mapping, admission control and resource allocation [PaK03]. QoS mapping refers to the translation of the QoS representations from one layer to the next as seen in table 1.
Table 1: QoS Layers

<table>
<thead>
<tr>
<th>QoS Layer</th>
<th>QoS Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Frame rate, frame size &amp; resolution, response time throughput, security, price &amp; convenience</td>
</tr>
<tr>
<td>System</td>
<td>Buffer size, process priority, scheduling policy, caching policy, time quantum</td>
</tr>
<tr>
<td>Network</td>
<td>Bandwidth, throughput, bit error rate, end-to-end delay, delay jitter, peak duration</td>
</tr>
</tbody>
</table>

Admission Control is used to determine whether a network is able to support the requested traffic with the requested network level QoS parameters. QoS on a network is implemented by proper allocation of resources to provide better and more predictable network service by supporting bandwidth allocation, improving loss characteristics, avoiding and managing network congestion, metering network traffic, or setting traffic flow priorities across the network. The network resources can be classified into link bandwidth, transmission energy, processor time, and buffer space [FR00]. The QoS metrics used by a network to allocate its resources include bandwidth, delay, jitter, availability and loss probability shown in table 2.

Table 2: QoS Metrics

<table>
<thead>
<tr>
<th>QoS Parameter</th>
<th>Parameter Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>A measure of data transmission capacity and influences throughput (the amount of data successfully transmitted and received in unit time)</td>
</tr>
<tr>
<td>Delay</td>
<td>Length of time taken to transmit data from one point to another. It includes access Delay (the time between packet arrival and packet transmission by the sender) and End-to-end delay (the time between the arrival of a packet and its successful delivery to the receiver)</td>
</tr>
<tr>
<td>Jitter</td>
<td>Variation in delay. An important consideration in isochronous applications and has strict values for MPEG frames</td>
</tr>
</tbody>
</table>
Every transmission involves a negotiation between the application and the network. Negotiation is based on matching application requests to network resources based on the metrics. If the network resources are able to meet the requirements of the application, the connection is granted, otherwise it is refused. Sometimes, based on the negotiation between the application and the network, the application adjusts its request by making some tradeoffs in its parameters. Various techniques have been suggested for allocating network resources [Bha02]

• Overprovisioning: This is done through adding capacity e.g. bandwidth as seen in optical networks using wavelength division multiplexing (WDM).

• Efficient resource management. A cost effective solution for the provider and it ensures that applications get the specified QoS during the course of their execution. There are two approaches to this technique [Ban05]:
  
  Resource reservation: It is typically employed in centralized protocols. Network resources are distributed according to an applications traffic request and are subject to available bandwidth as seen in IETF’s Integrated Services (IntServ) model. This model has limitations of scalability [Ban05]. Prioritization of traffic streams. Employed in distributed protocols, where traffic is classified based on the application and resources are apportioned according to such classes of priority depending on availability and demand. This is as seen in IETF’s differentiated Services (DiffServ) model. The model has the limitation of providing fine grained QoS [Ban05].

For a wireless Grid, the network is typically implemented as a wireless local area network (WLAN). This uses the IEEE 802.11 specification which was updated to IEEE 802.11e to extend the WLAN medium access control (MAC) to support QoS in WLANs.

3.1 QoS in WLAN

A wireless local area network (WLAN) uses radio frequency (RF) technology to transmit and receive data over the air. The wireless medium has fundamentally different characteristics from a wired medium. When providing QoS, the MAC endeavors to provide QoS service guarantees within this inherently unpredictable medium. Therefore bandwidth and latency cannot be guaranteed like a wired system, especially in the unlicensed spectrum [CP05]. A WLAN network is susceptible to all the parameters listed in table 2. These parameters may cause variation in bandwidth, latency in data delivery, jitter and error rates. The adverse conditions prevailing in a WLAN network make QoS guarantees very important.
IEEE 802.11e is an extension of the basic IEEE 802.11 standard that extends the MAC layer to accommodate applications with QoS requirements. The new MAC allows two different types of stations to exist; QoS enhanced stations (QSTA) and non-QoS enhanced stations (STA). Channel access to an IEEE 802.11e QoS enhanced basic service set (QBSS) is controlled by the hybrid co-ordination function (HCF). The HCF allows two medium access schemes

- Enhanced distributed channel access (EDCA): is a contention-based channel access function that is designed for prioritized traffic and is similar to differentiated services (DiffServ) [CP05].
- HFC controlled channel access (HCCA): Uses polling and supports parameterized traffic similar to IntServ [CP05].

EDCA has the limitation of producing fine grained QoS while HCCA has limitations to do with scalability. In addition to these limitations, most of the QoS work in WLANs has concentrated on the throughput metric [PaK03]. More work needs to be done to consider the other QoS metrics as well [PaK03]. An intelligent, adaptive QoS solution could be the answer to the unpredictable nature of the medium leading to better utilization of the network resources [Ban05]. For these complex situations, agent-based approaches, which emphasize autonomous actions and flexible interactions, are natural models.

4. MAS based approach to Resource Allocation

Resource allocation deals with optimally allocating resource to the entity that values the resource most. But due to the issues raised in section 1, it still remains a difficult problem. The interaction, strategies and lack of perfect knowledge of Grid applications are analogous to issues faced by people in an economy. This makes it plausible to use MAS and principles from economics to solve the resource allocation problem [Hub96]. When considered as an economy, a wireless Grid has three main components; Resources, Agents, and Preferences.

4.1 Resources

A resource is an entity that is to be shared [Abb03]. Grid resource management is the process of identifying requirements, matching resources to applications, allocating those resources, scheduling and monitoring grid resources over time in order to run grid applications as efficiently as possible [NSW03]. Grid resources can be classified into both software and hardware.

4.2 Agents

In trying to solve the network resource allocation problem in grid computing, the option of agents representing applications and resources interacting in a MAS to reach an agreement on the usage of a resource is considered. Since there is competition for the resources, the agent interaction uses economic principles to achieve fairness and efficiency. This could take any of the forms shown in figure 1 below [BACdK02]:
Advances in Systems Modelling and ICT Applications

Figure 1: A Market Framework

- One application negotiating for one resource (Negotiation)
- Many applications competing for one resource (Auction)
- One application choosing from many resources (Reverse Auction)
- Many applications competing for many resources (Market)

Each agent is self-interested and is assumed to have no global information and its action is based on limited information. Since there is competition for resources the resource allocation system borrows from economic principles to achieve fairness and efficiency. The system has the following entities [KINS96]:

- Seller (Resource Agent): The objective of the seller is to maximize earnings from the sell of a resource. Their strategy is based on maximizing profit.
- Buyer (QoS Agent): The objective of the buyer is to minimize spending or buy the best resource at the cheapest price. Their strategy is to get resources to satisfy a QoS value cheaply.
- Trading Environment: The market model is the environment where the goods are exchanged. The resource at each location has its price determined by the residual amount and other environmental factors (e.g. supply and demand).

Sellers in this environment do not know the prices of resources at other locations when they determine their prices. Similarly, buyers do not know the current demand for resources when requesting for resources. A seller’s utility is affected by a buyer’s decision and similarly a buyer’s utility is affected by a seller’s decision.

Both the seller and the buyer rely on their strategies to maximize their utilities. The relative worth of a resource is determined using the principles of demand and supply (pricing). The auction model chosen is the Vickery Auction. In this type of auction participants submit sealed bids, having been told that the highest bidder wins the item but pays a price equal to the second highest bid. This type of auction is chosen because willingness to pay one’s true value is the dominant strategy (honesty is the best policy).

4.3. Preferences

In normal operation, applications on a Grid tend to compete for limited resources. In the proposed system, decisions about resources are made on an economic basis.
Grid resources are packaged as commodities with price being the deciding factor in resource use. Agents representing applications and agents representing resources use their preferences to make decisions. The preferences are the fundamental description useful for analysing choice. Utility through the use of utility functions is used to describe preferences. A utility function is a way of assigning a number to every possible consumption bundle such that more preferred bundles get assigned larger numbers than less preferred bundles.

For applications on the Grid, this involves modelling an application’s sensitivity to changes in QoS and the willingness to pay for a service level (i.e. how much they value the service). In this case, utility is expressed as a function of actual QoS parameters as shown in table 2. One possible approach of defining such a utility function might be to generate a single metric from multiple network metrics e.g. for an application with the following requirements; bandwidth B, delay D and loss probability L [WC96]. The formula would appear like

\[ f(p) = \frac{B(p)}{D(p) \times L(p)}. \]

The bigger the value of \( f(p) \), the more valuable the resource request. The idea is to mix various pieces of information into a single measure and give it a monetary value for bargaining decisions.

For the resource provider, utility is expressed as a function of the amount of resource available to the applications. In the case of wireless Grids this is the residual bandwidth, radio frequencies and energy. On the Grid, every service is associated with multiple levels of acceptable quality. Resources need to be regularly adapted to achieve optimised QoS in light of application demands.

At one extreme there are inelastic applications like real time, interactive voice employing a constant bit rate whose utility is modelled as a step function. At the other extreme there are elastic applications such as email and file transfer, that take advantage of whatever resources that are available. In between there are partially elastic applications. If bandwidth is considered as a resource, for these applications, utility functions are represented as the figures shown in figure 2 below:

**Figure 2 [DaS00]: Utility function: a) inelastic applications; b) perfectly elastic application; c) partially elastic application**

For elastic services, the sources can adjust their transmission rates in response to congestion levels within the network. Hence, by appropriately exploiting the elasticity using agents and economic principles, one can maintain high network efficiency.
An auction is a method of allocating scarce goods. Auctions are useful when selling a commodity of undetermined quality and also when the seller is unsure of the price of the he can get. In a wireless network the radio frequency channel resource is quite valuable and scarce. Based on demand, its valuation by different applications fluctuates.

In the Grid WLAN, the trading environment is assumed to be located at the AP. The resource agent is also located at the AP and allocates the AP’s resources. The resource agents make a decision on which QoS agent to grant the resource based on an applications QoS requirements and the availability of resources.

5. Economy in a grid network Environment

In HCCA, the polling mechanism is round robin. This is based on priorities of the data to be transferred. We propose to use MAS interaction in the polling process where instead of use of priorities, agents compete for the network channel using market-based techniques. The network can be considered both as a link between other resources and as a shared resource. When considered as a shared resource its components include bandwidth, transmission energy and RF channels.

5.1. Physical Architecture

Within the wireless network the IEEE 802.11 standard defines two kinds of services: the basic service set (BSS) and the extended service set (ESS). A BSS is made of stationary or mobile wireless stations and a central base station known as the access point (AP). An AP is a bridge that connects a wireless client device to the wired network. A BSS without an AP is called an ad hoc network. A configuration that has more than one BSS forming a single subnetwork is known as the ESS. In ad hoc networks negotiations can be considered. In the BSS, auctions are the most suitable while in the ESS, markets would be most suitable. Reverse auctions are rare but can work for the ESS. In this paper we only consider using MAS in the BSS through the use of Auctions. This topology uses the IEEE 802.11 protocols mentioned in section 4.

Figure 3: Wireless Grid Physical Architecture
5.2. Layered Architecture

In terms of protocols, the grid architecture is shown in figure 3. The MAS-based resource allocation will be at the middleware layer of the layered grid architecture. This can be as an additional component concerned with management of network resources within the Globus Toolkit as shown in figure 3.

Figure 4: Wireless Grid Layered Architecture

The APs are modelled as a hypothetical resource producers who “sell” their link resources to customers i.e. grid applications. Agents representing applications negotiate with agents representing APs for guaranteed transmission opportunities (TXOPs). The application agent’s utility function uses QoS parameters representing the application. AP agents price the link based on supply and demand. The negotiation involves both agents maximizing their utilities (the neediest case should win). There are two possibilities

- An ongoing transmission being terminated if a more needy application requests the channel.
- The Application adjusting its QoS parameters to lower values to access a resource

The incoming job requests come with a “budget” allocated (a measure of the application perceived value of the job). It is assumed that the relative worth of a resource is determined by the demand and the supply and will borrow from the principles of economics. Each agents are uses the output of the utility functions for negotiation and each takes actions independently of each other to maximize its utility. The system shall be evaluated with respect to the following criteria [WPBB01]:

- Market Equilibrium
- Resource Efficiency: This is measured in terms of throughput, packet loss, delay and jitter.
5.3 Agent Interaction

The agent interaction will be as shown in the figure below:

**Figure 5: Agent Interaction**

----

6. Conclusion

Because Grid resources are heterogeneous in nature, there are a lot of challenges in Grid resource management of resources that include processors, disks, data, networks, other services. Applications have to compete for resources to satisfy their QoS requirements. The proposed architecture is based on the premise that since the number of parameters used in decision-making is more there will be fine-grained application QoS improving on the DiffServ model. The next development stage is implementing the architecture and conducting experiments to test it against existing wireless network models. This is can initially be by simulation and later on, on actual networks

- Modelling wealth in terms of application /QoS agents and resource agents
- Coming up with pricing mechanisms for resources.
- Coming up with utility functions to model the QoS parameters of different applications.
- Using the above two in a MAS-based negotiation or auction mechanism to efficiently allocate resources.
- Protocol overhead in the additional agent layer introduced

**References**


[Fos02] Ian Foster. What is a grid? a three point checklist, 2002.


An Efficient Dynamic Admission Control for End-to-End Delay Guarantees in IP Networks

Kasigwa J., Baryamwineba V., and Williams D., Faculty of Computing and Information Technology, Department of Information Systems, Makerere University

In networks that support Quality of Service (QoS), an admission control algorithm determines whether or not a new traffic flow can be admitted to the network such that all users will receive their required performance. Such an algorithm is a key component of future IP networks as it determines the extent to which network resources are utilized and whether the promised QoS parameters are actually delivered. Our goals in this paper are threefold. First, we describe and classify a broad set of proposed admission control algorithms. Second, we evaluate the accuracy of these algorithms via experiments using both on-off sources and long traces of compressed video; we compare the admissible regions and QoS parameters predicted by our implementations of the algorithms with those obtained from trace-driven simulations. Finally, we identify the key aspects of an admission control algorithm necessary for achieving a high degree of accuracy and end-to-end QoS.

1. Introduction

Network applications such as streaming media, voice over IP (VoIP) and content distribution generate realtime flows that have tight delay requirements, but can tolerate certain level of missed deadlines or packet losses. Guaranteeing performance to such real-time flows involves provisioning resources during admission control and enforcing the usage of the assigned resources at run-time.

It is widely accepted that the today best effort Internet is not able to satisfactory support emerging services and market demands, such as Voice over IP (VoIP) and Videoconference. Differentiated services (DiffServ) are seen as the technology to support Quality of Service (QoS) even for real-time services in IP networks without the scalability problems of Integrated services (IntServ). Admission control is the set of actions taken by the network during the service establishment phase to check whether a service request is to be admitted or rejected. A new service request is admitted when the desired QoS for the new service can be satisfied, without causing any QoS violation to the already established services. An additional role of admission control is to optimize the use of network resources. The challenge is to design simple control functions that improve efficiency under any offered traffic conditions. In DiffServ, the lack of per-flow information and signaling in the core network imposes restrictions to the employed
admission control schemes, since checking whether sufficient capacity exists end-to-end might not even be feasible in the first place, e.g. when traffic crosses many domains or even intra-domain when an equal cost multi-path routing scheme is employed. The various admission control approaches differ in the methods they use to decide if there is enough capacity for the new service request and can be divided into three categories: (a) Admission control using apriori traffic descriptors, (b) Measurement based admission control (MBAC), and (c) Parameter-base admission control (PBAC).

In admission control using a priori traffic descriptors [1], it is assumed that there exists perfect knowledge of each traffic source type that will be used and, additionally, of the current number of established service instances. This information will enable admission control to compute the total amount of bandwidth required. However, since no traffic measurements are taken into consideration, if the provided traffic descriptors do not depict the actual behavior of the sources, or the appropriate traffic descriptors are not known a priori, the performance of this admission control scheme can be very poor.

MBAC shifts the task of traffic specification from the user to the network [2] in order to cope with the inherent problems of admission control using a priori traffic descriptors. Instead of users explicitly specifying their traffic descriptors, the network attempts to "learn" the characteristics of existing flows through real-time measurements. This approach has a number of advantages. First, the user-specified traffic descriptors can be very simple, e.g. peak rate, which can be easily policed. Second, an overly conservative specification does not result in over-allocation of resources for the entire duration of the service session. Third, when traffic from different flows is multiplexed, the QoS experienced depends often on their aggregate behavior, the statistics of which are easier to estimate than those of an individual flow. However, relying on measured only quantities for admission control raises a number of issues that need to be considered, such as the estimation errors [2], system dynamics and memory related issues.

Parameter-based admission control (PBAC) is based on some metric applied on probing packets sent by the end host/application along the transmission path [3]. A requirement is for the end-to-end route to be the same for probing packets and flows. Setup delays may be high and, furthermore, simultaneous probing by many sources can lead to a situation known as thrashing [3]. That is that even though the number of admitted flows is small, the cumulative level of probe packets prevents further admissions.

We propose a bandwidth management approach that overcomes the limitations imposed by this lack of per-flow state and signaling in the core network whilst allowing for efficient admission control schemes for real-time traffic. This approach takes into account statistical multiplexing gains due to the aggregation of sources, accounts for issues such as the increase or the decrease in the level of aggregation along the real-time traffic paths, and determines the granularity of the employed admission control schemes with respect to the candidate points of enforcement. Taking the implications of the bandwidth management approach into account, a novel framework for admission control is proposed and its performance is compared against existing approaches.
2. Bandwidth Broker Architecture

A Bandwidth Broker is an automated resource manager in the DiffServ architecture. The main purpose of a BB is to keep track of the current allocation of the network QoS resources [4]. Bandwidth brokers manage QoS resource allocation requests within a single or successive DiffServ domains based on the available resources and on the Service Level Agreements (SLAs) formerly negotiated between the customer and its service provider. If used for multi-domain resource management, bandwidth brokers negotiate among each other on behalf of the service initiator. Additionally, bandwidth brokers participate in transit domain communication by coordinating SLAs across domain boundaries. Traffic flows traversing multiple DiffServ network domains can only be served by enabling communication among successive ISPs. This can be achieved by entrusting the resources of each DiffServ domain to a centralized manager agent called a bandwidth broker.

The bandwidth broker architecture was first proposed for the Expedited Forwarding service using the DiffServ model [5]. The bandwidth broker architecture supports the proposed dynamic admission control framework in providing the end-to-end QoS guarantees with varying traffic flow requirements. With the bandwidth brokers in IP networks, core routers do not maintain any QoS reservation states, whether per-flow or aggregate. Instead, the end-to-end QoS reservation states are stored at and managed by a bandwidth broker. There are several advantages of deploying the proposed Dynamic Admission Control framework with the bandwidth brokers. Bandwidth brokers support the Dynamic Admission Control to solve the problem of inconsistent QoS states faced by the conventional hop-by-hop, static admission control approaches [6]. Furthermore, bandwidth brokers support the dynamic admission control framework in designing efficient dynamic admission control algorithms without incurring any overhead at core routers.

The bandwidth broker architecture supports the proposed Dynamic Admission Control to provide a unifying framework to characterize traffic flows, in terms of their abilities to support delay guarantees, both the per-hop behaviors of core routers and the end-to-end properties of their concatenation [7]. The proposed dynamic admission control takes into consideration of both per-flow end-to-end guaranteed delay services and class-based guaranteed delay services with traffic flow aggregation. Using the bandwidth broker architecture, dynamic admission control can be done on a per domain basis instead of on a "hop-by-hop" basis. The dynamic admission control approach coupled with the bandwidth brokers significantly reduces the complexity of the admission control algorithms in IP networks [8] [9].

The bandwidth broker architecture provides end-to-end QoS abstraction for scheduling mechanisms of core routers in the network domain to perform either per-flow or aggregate QoS control functions such as dynamic admission control and reservations setup with no or minimal assistance from core routers.
3. Dynamic Admission Control

In designing a dynamic admission control algorithm, one can conceivably have two goals. One is to provide a parameter that accurately estimates a priori the level of service failures that will result. The other is to achieve the highest possible utilization for a given level of service failures. The proposed algorithms, although embracing similar goals, differ in four important ways. First, some algorithms are principled, based on solid mathematical foundations such as Large Deviation theory, and others are ad hoc, in that they lack a theoretical underpinning. Second, the specific equations used in making admission decisions are quite different. Third, while all algorithms have a parameter that varies the level of achieved performance and utilization (by making the algorithm more or less aggressive), some algorithms attempt to calibrate this parameter and have it serve as an accurate estimate of the resulting performance, while others leave the parameter un-calibrated; in the latter case it is assumed the network operator will learn appropriate parameter settings over time. Fourth, the measurement processes used to produce an estimate of network load are very different; they range from a simple point sample estimate, to an exponentially weighted average, to estimates based on both the mean and variance of measured load. Thus, the space of measurement-based admission control algorithms is both heavily and broadly populated.

4. Experimental Design

We perform a number of simulation experiments to evaluate the accuracy and effectiveness of the proposed Dynamic Admission Control (DAC) algorithms and we make comparisons to Measurement-Based Admission Control (MBAC) and Parameter-Based Admission Control (PBAC) algorithms. We evaluate the accuracy of the proposed DAC algorithms via experiments using long traces of compressed video real-time traffic. We consider various scenarios with different traffic loads, QoS parameters and compare the DAC algorithms and QoS values obtained in simulations with those predicted by the MBAC and PBAC admission control tests. Throughout, we denote the link capacity by $C$ and the buffer size by $B$, and denote the subsequent traffic flow arrivals by $j$.

Throughout the experiments, we focus on the following performance metrics:

(i) The first is the link utilization of the traffic, which is the total average rate of all traffic flows divided by the link capacity. For the simulation, this average utilization is also the total number of bits transmitted by the sources in the simulation, divided by the total number of bits that the server can transmit over the duration of the simulation (the link capacity multiplied by the simulation time), i.e.

$$\text{Average utilization} = \frac{\text{total average rate of all traffic flows}}{\text{link capacity}}$$

(ii) Our second performance metric is the loss probability denoted by $P_l$, which refers to the fraction of bits dropped by a queue that has finite buffer space $B$. The blocking probability of each ingress edge node and packet loss probability
of the core node will be evaluated and compared with the existing admission control provisioning methods.

(iii) The third performance metric is the empirical fraction of packets that are dropped due to buffer overflow in a finite-buffer queue with maximum buffer space B. We denote this measure of loss probability by \( P_l \). In both cases, we consider a range of buffer sizes B which have a corresponding statistically-guaranteed delay bound \( d = C/B \), i.e.

\[
\text{Delay bound} = \text{buffer size} / \text{link capacity}
\]

(iv) The fourth performance metric is the Loss Curve. We refer to a loss curve as the relationship between loss probability and buffer size, which, for additive effective bandwidths is exponential.

(v) The last performance metric is the end-to-end QoS metrics for different traffic flow aggregates: blocking probability, bottleneck link delay, packet losses and end-to-end delays (throughput).

4.1. Simulation Experiments

We used NS (Network Simulator v2) developed by the VINT project at Lawrence Berkeley Laboratories [10] to obtain performance results for our proposed system. NS is a discrete event simulator designed for the simulation of Internet protocols.

The workload consists of a set of twenty 30-minute traces of MPEG and JPEG compressed video Expedited Forwarding (EF) traffic. With this collection of traces and an implementation of our dynamic admission control algorithms, we perform a set of trace-driven simulation and admission control experiments for a wide variety of traffic mixes and network capacities. To achieve this, we first implement a number of admission control algorithms from the aforementioned traffic classes and determine their respective admissible regions for various traffic mixes and QoS parameters.

The video is taken from an action movie, digitized to 384x288 pixels and compressed at 24 frames per second using the MPEG 1 compression algorithm with frame pattern IBBPBPBPBBP. The trace exhibits significant fast time scale rate variations which are primarily caused by the MPEG coder’s alternation between large intra-coded frames (I frames) and smaller inter-coded (P and B) frames pattern IBBPBPBPBB. Further details of these MPEG 1 frames can be obtained from Rose (Rose, 2002) [65]. For the simulations, we consider each frame to be transmitted at a constant rate over the frame time, 24 frame-per-second.

In each experiment, flows are randomly chosen from the 20 traces with randomly shifted initial phase. Each flow has exponential holding time of 600 seconds. Traffic flows arrive according to an exponential inter-arrival time distribution, with a mean time of 2 seconds. Each flow provides DAC algorithm with QoS requirements (delay, probability loss \( P_{loss} \)).

Our simulation scenario consists of a first-come-first-serve multiplex with buffer size B, MPEG and JPEG real-time traffic traces as the workload, and the link capacity of \( C = 45 \text{ Mbps} \), unless otherwise noted. In all experiments, we set the buffer size B
equal to $C$ times the required delay bound so that delay violation probability is the same as packet loss probability, that is to provide per-flow delay guarantees for the DAC algorithms to potentially improve the number of admitted flows and link utilization i.e.

Buffer size = Link capacity x delay bound i.e. $B = Cd$.

We therefore consider a realistic environment in our experiments, including a buffered multiplexer, heterogeneous flows, busy MPEG traces with complex temporal correlation with MBAC and PBAC mechanisms, multiple time scale and long range dependent traffic flow arrivals and departures, and QoS specified as delay bound and delay bound violation probability.

4.2. Scheme Validation and Traffic Characteristics

The dynamic admission control mechanism is verified through simulations for different classes of traffic and results are compared to MBAC and PBAC admission control methods. The above admission control methods are applied to different classes of traffic i.e. Expedited Forwarding (EF) and Assured Forwarding (AF) traffic classes to examine different behaviors using DAC algorithms.

Simulation experiments with different classes of traffic are performed and the following cases have been considered:
(i) Dynamic Admission Control (DAC): for EF and AF1 traffic.
(ii) Parameter-Based Admission Control (PBAC): for EF real time traffic, MBAC for AF1 real time traffic and AF2 non-real time traffic.
(iii) Measurement-Based Admission Control (MBAC): for EF real time traffic, AF1 real time traffic and AF2 non-real time traffic.

Connections are set up between the bandwidth broker and the edge routers. New traffic connections arrive at each edge router node with exponentially distributed inter-arrival times with a mean of 1.21 seconds, depending on the case. This will result in a total arrival intensity of 3.6 1/s. Holding times are also exponentially distributed with a mean of 100 seconds for Real Time EF and AF1 connections and 250 seconds for all other connections.

The simulation traffic consists of MPEG and JPEG video streaming [11]. Dynamic Admission Control nodes monitor and update the load on the links that are attached to its local router while the bandwidth broker applies exponential averaging on the loads received from all admission control agents. Dynamic admission control agents send their current link loads every 2 seconds to the bandwidth broker.

5. Simulation Results

To test the effectiveness and efficiency of our dynamic admission control algorithms in IP networks, we have carried out a number of experiments to evaluate the performance of the proposed dynamic admission control mechanism. The obtained results are compared to other admission control mechanisms.
5.1. Performance of Dynamic Admission Control

A perfect admission control algorithm would achieve the highest network resource utilization without violating the QoS requirements of any traffic. Consequently, the effectiveness of an admission control algorithm can be evaluated in two ways; one is by comparing the admission region achieved by the DAC algorithm with the true admission region, two is by comparing the actual services received with the QoS requirements of the traffic flows. Table 1 evaluates the effectiveness of our DAC algorithm in different scenarios in these two ways.

Expt I: Basic admission control methodology with Loss Probability $P_{loss} = 10^{-5}$

Results from three groups of experiments are shown in Tables 1 and 2. QoS is specified as (delay, $P_{loss}$), delay bound and delay bound violation probability:

Table 1: Basic admission control methodology

<table>
<thead>
<tr>
<th>delay (ms)</th>
<th>1</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{dac \times 10^{-5}}$</td>
<td>0.02</td>
<td>1.60</td>
<td>9.35</td>
<td>13.4</td>
<td>21.1</td>
<td>27.2</td>
<td>22.2</td>
<td>21.7</td>
<td>18.2</td>
</tr>
<tr>
<td>$U_{dac(%)}$</td>
<td>66.8</td>
<td>81.5</td>
<td>86.5</td>
<td>87.3</td>
<td>88.0</td>
<td>88.6</td>
<td>89.1</td>
<td>89.5</td>
<td>89.6</td>
</tr>
<tr>
<td>$U_{true(%)}$</td>
<td>71.4</td>
<td>81.6</td>
<td>84.2</td>
<td>84.7</td>
<td>84.8</td>
<td>85.4</td>
<td>86.6</td>
<td>86.8</td>
<td>86.9</td>
</tr>
</tbody>
</table>

$P_{dac}$ is the actual packet loss probability measured from simulations; $U_{dac}$ is the utilization achieved using our DAC algorithm; and $U_{true}$ is the highest possible utilization without violating any QoS parameters.

Expt II: Dynamic admission control with Loss Probability $P_{loss} = 10^{-5}$

Table 2: Dynamic admission control with schedulers

<table>
<thead>
<tr>
<th>delay (ms)</th>
<th>1</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{dac \times 10^{-5}}$</td>
<td>0</td>
<td>0.02</td>
<td>0.35</td>
<td>1.3</td>
<td>0.93</td>
<td>1.8</td>
<td>2.1</td>
<td>0.73</td>
<td>0.15</td>
</tr>
<tr>
<td>$U_{dac(%)}$</td>
<td>61.3</td>
<td>75.7</td>
<td>82.5</td>
<td>84.8</td>
<td>85.4</td>
<td>85.5</td>
<td>86.3</td>
<td>86.5</td>
<td>86.6</td>
</tr>
<tr>
<td>$U_{true(%)}$</td>
<td>70.3</td>
<td>80.7</td>
<td>84.3</td>
<td>85.3</td>
<td>86.1</td>
<td>85.6</td>
<td>86.8</td>
<td>87.7</td>
<td>88.2</td>
</tr>
</tbody>
</table>

To evaluate our DAC algorithm, we will compare $U_{dac}$ with $U_{true}$ and $P_{dac}$ with $P_{loss}$. If $U_{dac} < U_{true}$, the DAC predicted admission region smaller than the true admission region, and traffic flows are un-necessarily rejected; If $P_{dac} > P_{loss}$, the actual loss probability experienced is larger than promised value.
5.2. Processing of Admission Control Mechanisms

Table 3: Detailed processing of admission control mechanisms.

<table>
<thead>
<tr>
<th>End-to-End QoS Metrics</th>
<th>PBAC</th>
<th>MBAC</th>
<th>Proposed DAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of flow arrivals</td>
<td>24,003</td>
<td>24,003</td>
<td>24,082</td>
</tr>
<tr>
<td>Total number of accepted flows</td>
<td>23,725</td>
<td>23,725</td>
<td>23,633</td>
</tr>
<tr>
<td>Total number of rejected flows</td>
<td>278</td>
<td>278</td>
<td>449</td>
</tr>
<tr>
<td>Blocking probability (10−3)</td>
<td>1.2</td>
<td>1.2</td>
<td>1.1</td>
</tr>
<tr>
<td>Flows accepted in normal mode</td>
<td>4,826</td>
<td>10,347</td>
<td>23,406</td>
</tr>
<tr>
<td>Number of Flows accepted in critical mode</td>
<td>18,899</td>
<td>13,358</td>
<td>227</td>
</tr>
<tr>
<td>Acceptance probability in critical mode (10−3)</td>
<td>79.7</td>
<td>56.3</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 3 shows the statistical results for comparisons of DAC, MBAC and PBAC mechanisms as obtained from experiments when the normalized load is 1. The proposed DAC mechanism has a lower blocking probability than the existing PBAC and MBAC mechanisms. We can also see that DAC accepts most of the requested traffic flows (99%) at the path level in the normal mode without any link level bandwidth check. This means that the proposed DAC mechanism has an end-to-end scalability benefit for large IP networks.

5.3. Dynamic Admission Control Ratio vs. Bottleneck Link Load

We have used a simple topology, with one bottleneck link in the middle and two more links where the different multicast senders and receivers are attached (See Section 5.1.2). We simulated a single multicast session, with multiple sources, and receivers joining and leaving the group. The multicast session shared a bottleneck link with some background load with different intensities. The link capacity of core node is maintained at 45 mbps and the edge nodes is 18 mbps. Different simulations for DAC, MBAC and PBAC mechanisms were carried on the same bottleneck settings and arrival intensities.
We further conducted simulations experiments with EF traffic flows, and from the obtained results in Figure 1, the proposed DAC provided higher admission ratios and link utilization levels as compared to MBAC and PBAC. This is due to the dynamic admission control mechanism that provides high priority to real time traffic and consequent blocking of some Non-Real-Time - AF2 connections. We can also clearly see that the proposed DAC mechanism provides efficient network utilization as compared to MBAC and PBAC. This effect is best illustrated in Figure 1, where EF admission ratios and average link loads are substantially higher with DAC mechanism as compared to MBAC and PBAC admission control mechanisms. In all admission control mechanisms, the admission ratio decreases with the increasing arrival intensities, which is a result of some network congestion and increased processing of traffic flows being admitted into IP networks.

5.4. Bottleneck Link Delay, Packet Loss and Traffic Throughput

For EF and AF1 traffic classes, delay and packet loss stay very low with all the simulated admission control schemes, the selected load and reservation thresholds have a major effect on this. For EF packets, bottleneck link delay is always less than 1.6 ms (including the constant link delay of 1.0 ms). The real time EF video traces have packet loss of zero for DAC and MBAC admission algorithms.

Bottleneck link load which also refer to total link utilization has direct consequences on both delay and packet loss, this can be seen in Figure 2, where AF1 and AF2 traffic connections experience the highest delays and packet losses with PBAC as compared to MBAC and DAC. The obtained results for the proposed DAC closely approximate the end-to-end delays when using MBAC mechanism.
6. Conclusion

The simulation experiments have been carried using the NS-2 network simulator. From the dynamic admission control algorithms and simulation results, it has been proved that the proposed DAC mechanism guarantees a user end-to-end QoS requirements and provide better bandwidth utilization than the existing admission control algorithms. It has also been shown that the proposed DAC mechanism has scalability and more effective deployment ability to IP networks than existing MBAC and PBAC mechanisms. The proposed DAC mechanism provides better end-to-end QoS guarantees in terms of delay, jitter and packet loss as compared to the existing admission control mechanisms i.e. MBAC and PBAC.

References


Towards Web Design Frameworks (Wdfs)

Rehema Baguma, Faculty of Computing and IT, Makerere University.
rbaguma@cit.mak.ac.ug;
Ogao Patrick, Department of Information Systems, Faculty of Computing and IT, Makerere University, ogao@cit.mak.ac.ug, and
Tom Wanyama, Department of Networks, Faculty of Computing and IT, Makerere University, wanyama@cit.mak.ac.ug, P.O Box 7062, Kampala.

Web Information Systems has become an important area of research and practice, with the continued growth and use of the world wide web (web) and Internet technology. The use of the Internet and the web is now common in government, industry, health, commerce and education. The increasing rate of internet/web usage calls for efficiency and effectiveness in the development and deployment of web systems to provide high quality systems in the shortest time possible. One approach to this has been the web design frameworks proposed by Schwabe, Rossi, and colleagues over the last decade. This paper looks at what is entailed in the concept of web design frameworks, and considers their importance as a formal method of web application development.

Introduction

According to Schwabe and colleagues (Schwabe et al, 2001a), building web applications such as e-commerce is a time consuming task. It calls for careful design of the navigational architecture and user interface to make the applications usable. It also requires the designer to understand the user tasks while navigating the hyperspace to decide which navigation facilities to include, for example whether to consider defining indexes, guided tours, or landmarks according to the user needs. The designer has to bear in mind that the interface should help the user browse through the sea of web information by giving him cues and feedback on actions and by presenting the information in a clear and meaningful way.

Schwabe and colleagues noted that another cause of complexity with web applications is that web applications include complex behaviors. In addition to the primary functions, for example buying or selling for e-commerce applications, web applications are also integrated with the company’s internal business; often providing different views of corporate databases, and acting as integrators of other applications.

Another important dimension of web applications in comparison to conventional software is the need to reduce deployment and delivery times. Applications in the Web must be built quickly and, ideally with zero defects. Designers must improve not only development but also debugging and testing times. In addition, building applications in
the web involves using many different technologies.

Building reusable code and design structures has been recommended as a means to improve the process of building these applications (Schwabe et al, 2001a; Kong et al, 2005; Mittal et al, 2004). This paper reviews the concept of web design frameworks and emphasizes their importance as a formal method of web application development.

In this paper, a web application: is considered to be a structured set of software objects that may be navigated and processed in order to achieve one or more tasks (Wikipedia online encyclopedia).

A web application framework: is a generic definition of the possible web application objects, together with a generic definition of the application's navigational and processing architecture. A web application framework must define the set of possible objects to be navigated, how they can be structured in their navigational architecture and how they may behave (Schwabe et al, 2001b). In contrast, a web design framework: is the generic design of possible web application architectures, including conceptual, navigational and interface aspects, in a given domain. Web design frameworks must be environment and language independent. The generic designs are aimed at design reuse in web applications. A web design framework contains a reusable application model in a particular domain that can later be instantiated into specific applications in that domain (Schwabe et al, 2001a).

The difference between web application frameworks and web design frameworks is that the former deals with generic definition of objects in a particular domain whereas the latter deals with the generic design of the possible web architectures. Although both web application frameworks and web design frameworks must be integrated, the latter can be specified independently of the former (Schwabe et al, 2001a). Web design frameworks are environment and language independent whereas web application frameworks are programmed in a specific language.

The remainder of this paper is organized into three sections, which respectively motivate the need for web design frameworks, describe their main elements, and suggest future research.

**Why Web Design Frameworks are needed**

According to Schwabe and his colleagues (Schwabe et al, 2001b; see also Kong et al, 2005), building web applications is a complex and time-consuming process compared to conventional applications. Hence web practitioners still face considerable challenges as described in section 1. In addition, in web application development, the most important kind of reuse—that is design reuse, has been largely unexplored perhaps due to the non-object oriented nature of the web (Schwabe et al, 2001a). Although to-date object oriented technology has grown in use and popularity in software development, object orientation alone is still inadequate for developing applications for the web (ibid). Web applications require both an object (conceptual) model to specify usual behavior and a navigational model in which to define navigational components such as nodes, contexts, links and paths (Schwabe et al, 1998; Rossi et al, 1999). However, current object oriented approaches do not provide primitives for navigation design. Yet for web applications
to be successful, the navigational structure must be carefully defined (Schwabe, 2001a). The concern with web application design is not only in the behavior of domain classes but also the ways the user will navigate through them.

To help manage the web’s complexity, Schwabe et al (2001b) recommended use of web design frameworks as a novel concept to among other things, push design reuse in web applications. The major strength of web design frameworks is the ability to specify generic conceptual and navigational structures independent of each other.

Web design frameworks focus on reuse, combining the reuse of both design and code for efficient implementation of new applications. Web design frameworks decouple decisions at the object level from decisions at the navigation and interface level (Rossi, Shwabe and Garrido, 1998). Schwabe et al (2001b) refer to this behavior as separation of essential modeling concerns- that is, of application behaviour, navigational modeling and interface design. Separation of essential modeling concerns provides a sound ground for web software evolution and a basis for maximizing design and implementation reuse. In addition, classifying concerns as conceptual, navigational and interface models gives a rationale for design reuse. (Schwabe et al, 2001b).

Web design frameworks are a way to maximize reuse in the development process of web applications since a certain degree of commonality has been observed among navigational and interface solutions in similar application domains (Mittal et al, 2004; Schwabe et al, 2001a). For example, the activities triggered when users order an item online are usually similar regardless of site. Also the way users navigate to locate and purchase items online is also usually similar. According to Mitall et al (2004), about 85% of government operations are similar across different government solutions. Hence it saves development effort and cost to develop these processes once and then reuse them for many solutions. In addition, code and design reuse ensures consistency across solutions in the same domain.

Reuse in the development of web applications using web design frameworks can be achieved through reuse of interface templates in the form of HTML or XML descriptions; or reuse of information, accessing shared databases and components that exhibit some non-trivial behavior such as a code for implementing shopping baskets even though the supporting technology may not support reuse like HTML and XML; or the code may not be found in a single component.

Web design frameworks provide a bridge between software framework technology and web technology. They enable object relationships and behavior modeling in a specific domain and also allow specification of navigational architectures using non-object oriented tools. Web design frameworks help to specify the abstract architecture of a family of web applications. The framework specification includes both the common aspects of applications in the domain and the “hot spots” where the specifications of a particular application are accommodated (Schwabe et al, 2001a).

**Components of a Web Design Framework**

Separation of concerns provides a basis for characterizing a web design framework’s components (Schwabe et al, 2001b). A web design framework combines a generic
conceptual schema and generic navigational and context schemas to give a generic platform from which the underlying application model can be developed by adding or refining profiles or tasks and or defining different navigational topologies for specific user needs. In a web application, the user navigates over navigation objects that are views of conceptual objects. These views are defined opportunistically, according to particular user profiles and tasks. In addition, navigation objects must be organized into useful structured sets, called contexts. The structure of these sets defines the intra-set navigational architecture, whereas the set of conceptual relations, which are mapped onto navigation links, defines the inter-set navigational architecture. Then, interface objects map to navigational objects providing a perceptible appearance to the end user.

According to Schwabe et al (2001b), a web design framework is comprised of two main generic models, namely a conceptual model and a navigational model. The two models are mapped together by a view mechanism. The navigational model itself has two sub models, namely a generic navigational schema and a generic navigational context schema. Each model has points of flexibility (hot spots) whose purpose is to permit customization for design reuse. The structures of these components and their relationship is illustrated in figure 1 presented below:

**Figure:-1 Components of a web design framework, adopted from Schwabe, Esmeraldo, Rossi and Lyardet (2001).**

**Generic Conceptual Model:**

This component models what characterizes the application domain. It defines the abstract classes, possibly some concrete classes, and the relationships that make up the domain of application. The generic conceptual model reflects which objects the application deals with, their relationships and behaviours (Schwabe et al, 2001b).

Reuse in the conceptual model can be achieved through generic designs-- that is
the hot spots that are specified with object-oriented techniques. Hot spots are generic framework components that are customizable by replacing either class or subclass definitions or by code writing for a particular case. For example, for an online shop, the conceptual model can be extended by defining subclasses such as sub classing products into domestic, industrial or scholastic products. Designing for reuse at the conceptual model uses well-known object oriented techniques, an already well-established software design methodology. Therefore, more emphasis should be put on designing reusable navigational models according to Schwabe and colleagues (2001b).

Generic Navigational Model

The navigational architecture distinguishes web applications from conventional applications. The generic navigational model defines the navigational structure of a web application and how users will traverse the hyperspace via navigational objects such as nodes, links, indices and guided tours defined according to the intended user's profiles and tasks. In object oriented hypermedia design method (OOHDM), an approach for developing web design frameworks, an application is seen as a navigational view over the conceptual model. Nodes in OOHDM represent logical “windows” (or views) on conceptual classes defined during domain analysis. Different navigational models may be built for the same conceptual schema to express different views on the same domain. Links are derived from conceptual relationships defined in the requirements gathering stage. Therefore, node attributes are defined as object-oriented views of conceptual classes, allowing a node to be defined by providing access to attributes of different related classes in the conceptual schema. For example the orders class in an online shopping navigation schema imports the attributes product identification (unique identification of each product in the shop), payment mode (how the customer prefers to settle the payment), delivery mode (how the customer prefers the product delivered to her) and delivery date (when the customer wants the product).

Reuse in the navigational model can be achieved in different ways:

• By building completely new applications from the same conceptual model such as defining a new user profile. The application could be built to permit a shopkeeper to add new customer categories previously unknown to the system.

• By defining a generic navigational schema, which allows adding new nodes, links or classes and refining the definition of attributes. For example a product’s navigation options could be made in such a way that it allows further parameterization. In addition to textual and image descriptions, the shop owner could be allowed to add more navigation alternatives like audio clips.

• By achieving a generic navigational context schema- that is, defining abstract access structures and navigational contexts, For example alternatives could be established for grouping products into navigation contexts by manufacturer, manufacture date and country of origin.
Abstract Interface Design Model: It is important to recognize that there is a difference between navigation operations and interface operations. Not everything that happens in the interface is navigation related. Navigational design deals with internal functioning of the application while interface design deals with perceptible objects users interact with. Furthermore, interfaces need to be designed at an abstract level, to achieve independence of implementation environment. A given user profile might have different interface definitions according to a particular interface device to be used for example an internet browser or mobile phone. This brings in the abstract interface design model as the third vital component of web design frameworks in addition to the conceptual model and navigational model. The abstract interface design model specifies which interface objects the user will perceive. Interface behavior is declared by specifying how to handle external and user-generated events and how communication takes place between interface and navigational objects. OOHMDM uses a formal model of interface objects that considers the interface as a set of observers of navigational objects. User interface design is a critical activity in interactive applications, including web applications (Schwabe and Rossi, 1998). The abstract interface specification caters for the way different navigational objects look which interface objects will activate navigation, the way in which multimedia interface objects will be synchronized, and which interface transformations will take place.

Web Application design frameworks can also be built by systematically applying design patterns. According to Schwabe et al (2001b), design patterns capture known good solutions to recurring design problems. They represent design knowledge gathered from the solutions given by experienced designers that have encountered these same problems before possibly with slight variations. Applying design patterns is a form of design reuse. Rossi et al (1999) introduced navigational patterns as a way to record, convey and reuse design experience for example set based navigation, which recommends organizing objects into meaningful sets to help designers support tasks such as selecting items of interest in an electronic store. Schwabe et al (2001a) gave design patterns as the state of the art solution for collecting design experience such that design structures can be reused in different applications and contexts.

Conclusion

It is a common knowledge in the design community that designs rarely start from scratch but rather from already existing structures that are somewhat similar to what is needed. There is a need to find approaches to maximize reuse in the development process of web applications, since a certain degree of commonality has been observed among conceptual, navigational and interface solutions in application domains such as e-commerce, e-governance. Unlike in conventional applications, a web application’s navigational topology to a great extent determines an application’s success. Web development processes must capture and represent effectively such aspects for families of applications in each domain. Web design frameworks allow this practice to become a systematic documented activity that is independent of the implementation environment. The systematic reuse of behavioral, navigational and interface design
structures as described by web design frameworks is a key approach for maximizing reuse in web application development.

This paper has reviewed recent work in the field of web design frameworks in order to raise awareness of these frameworks and their importance to building more efficient and effective web applications.

Future work could investigate the potential of encapsulating solutions to web design problem domains into web design frameworks. These solutions could be generic designs for reuse across domains, such as a web design framework for e-commerce solutions, or they could address specific problems in the domain, such as security in e-commerce with a web design framework for secure e-commerce.

References


The Future of Intelligent Networks in Developing Countries

Narcis T. Rwangoga, ntrwangoga@cit.mak.ac.ug, and Martin Ngobye, mngobye@cit.mak.ac.ug, Faculty of Computing and IT, Makerere University

Telecommunication networks have become important not only for the academic community but also for users in business and industry. As globalization takes shape, this trend applies to both developed and developing countries. Unfortunately for many developing countries, the telecommunication networks are still dominated by the public switched telephone networks (PSTN), which are not able to support rising demand for services through networked systems. Innovative solutions have to be devised to plan, design and implement telecommunication networks that are able to seamlessly provide services demanded through the network today. This applies to both wired and wireless networks. The demand for network based services has risen in developing countries, considering how wireless technologies for cellular and personal communications have been extended to most areas in developing countries. Intelligent Networks (IN) is a promise that countries which are implementing telecommunication networks can use to deliver network based services. In this paper, we explore and discuss why IN provides the promise and requirements for their implementation. The objective is to provide a basis a tool for telecommunication network planers and implementers can use for IN structure implementation in developing countries where resources are highly constrained.

Introduction

Infrastructure for basic services has been expanding rapidly in developing countries. However this is not keeping pace with the needs, especially in telecommunications. Communication costs in developing countries are far higher than that in the rich countries. A number of countries have launched major plans for network expansion. Present telecommunication networks of the developing countries provide only basic service what is generally known as Plain Old Telephone Service (POTS). Manual and electromechanical switching is still extensively used.

All over the world, the term intelligent network (IN) is used to describe a network architecture that applies to all telecommunication networks according to J.M. Patel (1998) [1]. At the heart of this concept is an individual, software-defined communication profile for customers of telecommunication services. The IN combines important functions and data in a central location and provides them in only one or just a few nodes. The intelligent network provides an intelligent, distributed database access
capability from a plurality of service switching points (SSPs) to data and functions stored in one or just a few service control points (SCPs) for the purpose of controlling the service according to D. P Satapathy (1998) [2]. In an intelligent network, the logic for controlling telecommunications services migrates from traditional switching points to computer-based, service independent platforms. This provides network operators an open platform provisioned with generic service components that can interoperate with elements from different vendors, based on published, open interface standards. This platform can be used to develop new and different services.

In developing countries, the traditional telecommunications environments still exist, where companies act both as network operators and service providers. D.P. Satapathy and J. M. Peha (1996) [3] define the network operator as the entity that owns and operates the network infrastructure. The same authors define the service provider as an entity that offers services to the subscribers. The service provider uses the network infrastructure of a network operator to deliver the service to the subscriber but is responsible for the management and development of the service. Service offerings are still likely to be driven by technological availability rather than customer need since with the traditional telecommunication architecture, most of the network infrastructure is based on proprietary interfaces with bounded capabilities. This environment results in long development times and large investments to deploy services. New technological capabilities, privatization and deregulation, and changes in market and customer demand have driven the need for new approaches to install networks. Operators and service providers need to add new features rapidly to attract and retain customers. Intelligent networks can play an important role in providing such new features and services.

The Intelligent network architecture

What we call the intelligent networks today has had its roots in the frustrations of network service providers with limited, highly specialized capabilities of the switching systems that were available when advanced network services first begun to be deployed in long distance network. Within traditional public switched telephone networks, the hierarchy of switching equipment and software must be upgraded each time a new service is added to the network. This is a complex and costly process. Further, network switches could not provide new number translation, routing and charging capabilities. As telecommunications services have evolved, the need to reduce the overhead for service use has increased along with the need to simplify maintenance and service upgrades or additions. The Intelligent network essentially separates these services from switching equipment and organizes a centralized system so that providers need not perform major modifications on multiple switches when they introduce new services. According to J. M. Peha (1998) [4], the first step in intelligent network development was to create separate service data in a centralized database outside the switching nodes. The second step was to separate the service programs, or service logic, and define a protocol that would permit the interaction between switching systems and intelligent nodes containing the service logic and data.

For service switching points and service control points (intelligent nodes) to work,
common channel signaling, or out-of-band signaling is required as opposed to the traditional in-band signaling. Relying on out-of-band signaling, or signaling system 7 (SS7) protocols, provides the mechanism to place service logic and service data into dedicated network elements that can remotely handle call control and connection. SS7 also enables intelligent applications to communicate with other applications and to access databases located in various parts of the network. Certain network elements can be distinguished in every IN as shown in figure 1. Service switching points (SSPs) are stored program control switches that interface to the SS7 signaling network. The SSP embodies the call control function (CCF) and service switching function (SSF) entities.

**Figure 1. Intelligent Network Functions and Functional Relationships**

The SSF recognizes Intelligent Network service calls and routes the appropriate queries to the service control function (SCF) that resides in a service control point (SCP) via the SS7 network through signaling transfer points (STPs). STPs are high-capacity, high-reliability packet switches that transport signaling messages, using large routing databases, between the IN nodes. SCP commands are used by the SSP to process calls. The SCP is a fault-tolerant high capacity, transaction-processing entity that provides call handling information in response to SSP queries. The Service Management Point (SMP) provides operation, administration and maintenance functions for the IN. The intelligent peripheral (IP) provides enhanced services or functions under the control of an SCP, possibly relayed by an SSP. As seen in Figure 2, the IN architecture is fundamentally based on SS7 and its protocol architecture. A common signaling transport capability known as the message transfer part (MTP) handles the corresponding open systems interconnection (OSI) physical, data-link, and network layers. The next level, signaling
connection control part (SCCP), augments the MTP by providing both connectionless and connection-oriented message transport, as well as enabling addressing capabilities for message routing. The transaction capabilities application part (TCAP) provides procedures for real time transaction control. The final layer, IN application protocol (INAP) defines the operations required between IN network elements, such as SSPs and SCPs.

**Figure 2: IN Protocol Stack**

All of these basic elements form the infrastructure in Intelligent Network, which supports the notion of separating service-control functions from service switching functions, to realize more rapid services development and deployment. Another equally important concept in Intelligent Network has been the notion of service independence. Here, the primary goal is to identify and create generic sets of reusable service components that could be used to build new services and loaded into SCPs to generate new services rapidly. To provide a framework that would lead toward Intelligent Network engineering standardization, the Intelligent Network conceptual model (INCM) was developed (Figure 3). The INCM, which is solely a tool for describing IN capabilities and characteristics, is composed of four “planes” that represent different aspects of implementing Intelligent Network services. This model depicts the relationship among services and service features, global service logic, distributed service logic, and the physical network entities such as SCP and SSP. These planes include the service plane, the global functional plane, the distributed functional plane and the physical plane as shown in figure 3. The service plane describes services from a user’s perspective, where a service consists of generic blocks or service features that make up part or all of a service. The global functional plane deals with service creation and is comprised of the service independent blocks that will be used to create service features. Global service logic defines how service independent blocks are linked together to form features and how these service independent blocks interact with another basic service independent block known as the basic call process (BCP).
The BCP is the process that optimally supports services that do not require special features and is basic to the processing of all services. The distributed functional plane defines a set of functional entities that perform specific actions. Service independent blocks are implemented through a specific sequence of functional-entity actions performed by those functional entities. Table 1 describes functional-entity components as well as their relationship to the IN physical entities.

**Table 1: In Physical and Functional Entities**

<table>
<thead>
<tr>
<th>Physical Component</th>
<th>Distributed Functional Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service Switching Point (SSP)</td>
<td>Call Control Function (CCF)</td>
<td>Controls call processing and provides network connection services</td>
</tr>
<tr>
<td></td>
<td>Service Switching Function (SSF)</td>
<td>Supports IN triggering during call processing and access to IN functionality</td>
</tr>
<tr>
<td></td>
<td>Specialized Resource Function (SRF)</td>
<td>Supports the interaction between the call processing software on the switch and the service control function</td>
</tr>
<tr>
<td></td>
<td>Call Control Agent Function (CCAF)</td>
<td>Supports specialized network resources generally associated with caller interaction; provides user access to the network</td>
</tr>
<tr>
<td>Service Control Point (SCF)</td>
<td>Service Control Function (SCF)</td>
<td>Executes IN service logic and influences call processing on the switch via its interface to the SSP</td>
</tr>
<tr>
<td></td>
<td>Service Data Function (SDF)</td>
<td>Manages customer and network data for real-time access by the SCF in the execution of an IN service</td>
</tr>
<tr>
<td>Intelligent Peripheral (IP)</td>
<td>Specialized Resource Function (SRF)</td>
<td>Supports specialized network resources generally associated with caller interaction</td>
</tr>
</tbody>
</table>
The need for Intelligent Networks

According to a study conducted by Stanford University, USA, (1999) [5], it is reported that many scientists have predicted that a global network of affordable multimedia computers, online libraries, student-centered “learning ware” and enhanced human communications in general will improve access to high-quality education on a scale that simply cannot be accomplished today. But how are developing countries, where even traditionally laid network infrastructures are not yet fully in place, prepared to participate in the global network? Ever since the explosion of information technology took center stage in developing countries, more advanced research have been dedicated to provide network based services according to resources available. How will such networks match the rising demand for network based resources? Is it the time to consider pursuing more of intelligent network services and less on legacy network infrastructures? The network is the one element of the infrastructure that touches all others, from the middleware and applications to the servers and end users. It is, therefore, a logical place to implement changes that can cost-effectively scale to impact the entire networked system. When capabilities are administered at the endpoints, changes must be made at every distributed node or server, causing management complexity and operational costs to rise exponentially. If, however, these capabilities can reside in the network, where it is easier to centrally manage changes, they can scale more efficiently and simplify operations. Also, in this era of increased industry regulations and scrutiny of corporate governance, centralized management allows for better oversight and enforcement of business policy. For networks to truly become a reliable and cost-effective foundation from which to optimize business transparency and agility, they must not only hide complexity from the user, but they must also actively participate in the delivery of applications and services and be adaptive to the changing requirements of the business. The three elements of an intelligent networking strategy are a systems-level approach to network design, including how the network integrates with the rest of the information technology infrastructure; active participation by the network in the delivery of applications and services; and policies for linking business objectives and processes to network rules.
Implementation Challenges


The first challenge is at Systems-level. To date, the networking industry has tended to focus on solving customer problems one at a time, adding features, capabilities and intelligence only at the individual product level. With each individual product having its own management and feature design, the operation, management and maintenance of networks have become more complex and expensive as infrastructures have grown. This lack of a system-level approach has created technology islands and the need for additional components that can further increase complexity, add to operational costs and duplicate functionality. We now require a network that is more closely tied to, aware of and responsive to the needs of the applications, resources and devices connected to it. The increasingly complex tasks of intelligent IP networks will require more sophisticated functions, such as better content awareness, seamless encryption and filtering, greater quality of service and more adroit traffic shaping. These features can’t be deployed as pockets of technology, but must be integrated throughout the infrastructure. System wide security, end-to-end performance controls, service-level resiliency and system wide management visibility can be achieved only by looking at the IT infrastructure as a highly integrated and open system, not as a series of interconnected boxes selected by product-level features and benefits. Some networking vendors are already taking steps in this direction by creating networking subsystems that tightly integrate the functionality of several previously separate components—router, switch, firewall, intrusion-detection/prevention system, wireless access point, IP telephony, etc. as one way of reducing systems integration costs and simplifying operations and management.

The second challenge concerns applications and Services. For networks to become more intelligent, they must be able to make better informed decisions regarding the handling of particular applications. An intelligent network can also enable applications to make better decisions. The network must not only look deeper into the payload of individual packets, to understand what type of application it is, and what it is trying to do; it must be capable of examining streams of packets as well. Already, with the adoption of Web services, traditional firewall technology, which focuses on packet-level network traffic alone, can’t adequately protect the infrastructure. Web services using XML or SOAP, for example, require application-centric, message-level inspection to protect against this new level of vulnerability and should be an integrated component of any next-generation security solution. In the future, network components will be designed with intelligence that interacts with business applications to enhance their performance. For instance, an ASIC on a router could look into packet payloads from an order entry system during an order crunch. In this scenario, the network would understand what’s happening in Layers 4 through 7. If it is the end of the quarter and the transaction server has orders piling up, the network could inspect packets to look for big orders from important customers with short delivery times, route the transactions to
the appropriate server, flag the management system and create a follow-up transaction to ensure that the order was fulfilled on time. Of course, this scenario must be dictated by a business policy, which is why policy control is an imperative component of the intelligent networking strategy.

The third challenge is how to enforce policy. As mentioned earlier, the network touches every element of the infrastructure; therefore, it is in a unique position not only to monitor the transfer of information, but also to enforce policies in a very coherent way. Policy controls enable the network to be managed through policies and policy domains, creating a higher-level systems management capability that reduces risk of change as well as management and administrative costs. Linking network policy to business policy provides a framework to adapt the intelligent infrastructure to the specific needs of a particular business. It also has the potential to substantially reduce management complexity and therefore operational costs. For instance, an organization may have a security policy that states that all users must have the latest antivirus signatures and patches installed. But with thousands, or even hundreds of thousands, of endpoints to simultaneously control and monitor, this policy is largely unenforceable. However, the network, designed with a higher degree of interoperability between PCs, servers and policy appliances can automatically check when a new client wants to be connected to the network to see if the new device adheres to the organization’s security policy. This example of policy-based intelligent networking already exists today through the joint efforts of leading networking and antivirus software vendors.

**Services supported by intelligent networks**

As developing countries are racing against their counterparts in the west, in closing the gap of unequal opportunities in technological advancements, the need for Intelligent Network services today are paramount. Most mobile networks have deployed services and yet many are yet to be accessed by customers. An intelligent network can be used to implement the following IN services, among other things:

(i) Roaming services: Mobility dictates a need for technology or standards that make it possible for different networks to talk to each other. Subscribers want to be able to use the same voice-activated services that they use in their own cities when they travel across boundaries.

(ii) Data-service Capabilities: Handset displays allow customers to use various message services. One, called short message service (SMS), works very much a pager. It allows phones to send and receive messages in addition to making or taking telephone calls. SMSs require many SS7 messages just to set up the signaling and the mechanism to get the data through the wireless network. It requires a significant amount of checks and balances, finding the database, pulling up the message, encapsulating it with the right header information to route it to the correct user, and finally sending it out like a phone call.

(iii) Multi-media: Video applications are new to enterprises. Until recently, the infrastructure needed to capture and transmit video reliably was prohibitively
expensive. Advances in network infrastructure, capture devices, and playback platforms now make enterprise video a feasible alternative to meetings and company events. Innovative companies are finding new ways to use video technology for business.

Recommendations

There is need to analyze weak points in existing intelligent networks. There is also a need to test unimplemented network configurations as early as the planning stage so as to determine the configuration that has the best performance under given basic conditions. For example, the question arises as to how powerful the SCP or its individual processors must be or what capacities must be provided, and possibly in what form, for the individual services. The question also arises as to whether and how better performance can be achieved, possibly by changing the process management functions within the IN processors. A further critical element in operating an intelligent network is the overload protection mechanism which is supposed to stabilize the system as closely as possible to its performance limit. In practice, corresponding overload protection parameters can be determined only with great difficulty, which means there is a need to determine the optimum overload protection settings as a function of other network configurations and independently of network operation and to transfer them to the real network.

Conclusion

For a global network of affordable multimedia computers, online libraries, student-centered “learning ware” and enhanced human communications to become a reality in the developing countries, some decisions have to be made and a supportive environment created, including regulatory issues, choice of technology, cooperation with developed countries and support from international bodies. Developing countries have special opportunities and concerns regarding computer networking technology. They have a unique opportunity to leap frog into implementing advanced systems bypassing the older network. The IN provides a good opportunity of achieving this objective in a short time with very modest investments. However this needs a careful evaluation of the available technology options versus the needs. There is no doubt that IN would contribute to modernization of network based services in developing countries.

References


Part Four

ICT and Legal Applications
Identifying Sensitive Knowledge for Law Enforcement Agencies

Habibu Atib, habibu. atib@research.vu.edu.au and John Zeleznikow, john.zeleznikow@vu.edu.au, School of Information Systems and International Corporate Governance Research Institute, Victoria University, Australia

Law Enforcement and Intelligence agencies need to effectively manage knowledge to further their goals of crime detection and prevention. Techniques such as computer profiling, link analysis and Knowledge Discovery from Databases are being used to assist these goals. Since the terrorist attacks in the United States of America, on 11 September 2001, there has been an increased concern about security. This has led to a fundamental change in the balance between the security needs of the community and the civil liberties of the individual. Whilst Knowledge Discovery from Databases can indicate patterns of deviant behaviour and indicate potential criminals, these results indicate patterns and not evidence. In this paper we examine the ethical issues caused through the use of crime data mining. We shall introduce the notion of Knowledge Discovery from Databases in relation to Law Enforcement agencies and consider previous research such as Link Analysis and crime detection tools such as COPLINK CONNECT and FLINTS. We introduce several approaches for managing security knowledge including the concept of sensitivity management.

Introduction

Law enforcement investigators are faced with the challenge of determining the relevant facts using their judgement, intuition, and deductive power to help preserve the security of citizens and the country in general.

(Chen et al. 2004b) claim that security concerns since the terrorist attacks on 11 September 2001 has increased significantly amongst many nations. Law enforcement and intelligence agencies are actively collecting domestic and foreign information to prevent future attacks. They are drowning in a sea of data, but operating with little intelligence or knowledge.

According to (LeBeuf 2001) and (Reed 2005), to be able to respond immediately, law enforcement agencies must be properly equipped, trained and organized to fight crime.

Among the skills required are knowledge experts who are able to discover previously unknown and interesting facts about the data (Chen et al. 2004a). Further information can be found in the in ((Han & Kamber 2001), (Hand et al. 2001), (Selfridge 1995) and (Bartlmae & Riemenschneider 2000)).
According to (Frawley et al. 1991), Knowledge Discovery from Databases (KDD) is the ‘non trivial extraction of implicit, previously unknown and potentially useful information from data’. Data mining is a problem-solving methodology that finds a logical or mathematical description, eventually of a complex nature, of patterns and regularities in a set of data (Fayyad et al. 1996a).

KDD techniques have been widely used in financial fields such as banking and insurance to analyse customer databases for marketing and sales purpose or credit risks (Chen et al. 2004a) and (Bartlmae & Riemenschneider 2000). Until recently, they have received little attention in the policing and legal sectors. Currently, much information about terrorists is being captured for national security purposes.

A good analysis of the use of KDD in law can be found in (Stranieri & Zeleznikow 2005). The recent acts of terrorism, following the activities in the United States of America, on 11 September 2001, have led to an increased analysis of the use of KDD to conduct profiling of potential terrorists.

**KDD, Link Analysis and Crime Detection**

(Chen et al. 2004a) state that by increasing efficiency and reducing errors, crime data mining techniques can facilitate police work and enable investigators to allocate their time to other valuable tasks. They present a general framework for crime data mining that draws on experience gained with the Coplink project. They developed a crime data-mining framework, with which they describe three examples of its use in the Coplink project: named-entity extraction, deceptive-identity detection, and criminal-network analysis.

The use of KDD for crime detection poses many problems. “A major challenge facing all law-enforcement and intelligence-gathering organisations is accurately and efficiently analysing the growing volumes of crime data (Chen et al. 2004b). If KDD is properly applied, it will enhance the work of intelligence agencies in the process of identify useful links and patterns, which can be used for the investigation and detection of crimes.

Various techniques have been used to construct criminal investigation decision support systems. Statistics has been used to analyse evidence (see for example (Aitken 1995) and (Schum 1994)). Areas investigated include DNA testing, fingerprints, footwear and ballistics.

In law enforcement, intelligence analysts often refer to nodes and links in a criminal network as entities and relationships (Sparrow 1991). The entities may be criminals, organizations, vehicles, weapons, bank accounts, etc. The relationships between the entities specify how these entities are associated together.

Law enforcement agencies and intelligence analysts frequently face the problems of identifying possible relationships among a specific group of entities in a criminal network. Such tasks can be fairly time-consuming and labour-intensive without the help of link analysis tools.

Previous work in the area has included the creation of links between transactional database records of individuals who have related financial transactions in order to
identify money-laundering networks (Goldberg & Wong 1998); the Link Discovery Tool, which uses shortest path algorithms to discover association paths between two individuals, that on the surface appears to be unrelated (Horn et al. 1997); the COPLINK system which is based on finding links between database elements of person’s organizations, vehicles and locations (Hauck et al. 2002a).

Since the creation of the United States Homeland Defence Agency, as a consequence of the Acts of Terrorism on September 11 2001, there has been a rapid increase in commercially available products that claim to perform link analysis.

Before data can be visualised, the preceding step is to gather the relevant data. Most police data is scattered over distributed information sources. To find relevant information, a police officer needs to know which data sources offer specific sets of data and how to access them, as well as understanding each individual source’s query language and user interface. He or she must then manually integrate retrieved data. There are several examples of important software for data integration, based on pooling the various diverse information sources available, that can be brought to bear on a problem.

An example tool is COPLINK (Chen et al. 2003a), (Chen et al. 2003b) and (Chen et al. 2004a), first created in 1998 at the Artificial Intelligence Laboratory at University of Arizona at Tucson. COPLINK is marketed as a ‘one-stop access to point for data to alleviate police officers information and cognitive overload.’ The COPLINK system consists of two major components: COPLINK CONNECT is designed to allow diverse police departments to share data seamlessly through an easy-to-use interface that integrates different data sources available, that can be brought to bear on a problem.

FLINTS - ‘Forensic Led Intelligence System’ (see (Leary 2004)) integrates diverse data sources, including both ‘hard’ (DNA, finger-prints, shoe-prints) and ‘soft’ (behavioural) forensic data. The first version of FLINTS gave officers the ability to build a graphical pattern of links between crimes and criminals, previously thought to have no connection. Discovering these links has resulted in thousands of hours saved, hundreds of crimes solved and many criminals convicted. The FLINTS II version automatically trawled through various computer systems and pulled out the appropriate information, which linked criminals to other criminals and crime, working in real-time.

FLINTS III has the capability to deal with offender network analysis and assist in identifying groups of offenders. The new element of geographical profiling means officers can locate crime ‘hotspots’ by either incident or crime type, displaying the information by area. Comparative and seasonal analysis maps show emerging trends and developing hotspots, which are presented in maps and animated formats for the user. FLINTS was developed to support the detection of high volume crimes within the West Midlands in the United Kingdom, through a judicious choice of queries to evidential databases of DNA, fingerprints, footwear and tool-marks. Analysis of the data reveals patterns, associations and links which would not have been detected had each evidence type been managed in separate systems.
FLINTS is a new approach to knowledge management in that it releases the inherent power in large data collections used by law enforcement agencies. Through a judicious choice of questions, knowledge about scenarios, links, stories and connections between many types of data and many types of events as well as many people and many locations can be inferred. Results are visualized to aid analysts understand the chains of links and then contemplate new searches for new links. FLINTS is considered as decision support system aiding the analysts and investigators while identifying relevant information amongst a mass of data. The strength of the system is to be able to identify what should have been considered as ‘obvious’ links between people and crimes but are hidden in mixed masses of data.

Law enforcement agencies deal with sensitive information, some of which is of national security concern. This concern needs to be addressed when applying KDD to this domain.

The challenge here is to determine how the community can benefit from using KDD to detect and prevent criminal activities while at the same time ensuring that professional ethics within law enforcement agencies are not violated.

In particular, we must consider how we can:
1) Separate sensitive information from the knowledge generated during the statistical analysis and data mining.

2) Identify all sensitive patterns/links from every available pattern.

In this paper, we will look at the identification of sensitive knowledge from the point of view of law enforcement investigators. This will help us build an enhanced understanding of dealing with information or data in a security environment.

Managing Security Knowledge

According to (Hauck et al. 2002b), the area of knowledge management (KM) has attracted immeasurable amount of attention in recent years, and it is evident that knowledge management exists at the enterprise level. Traditionally, knowledge was stored manually on paper and in individuals’ minds. Access to and utilisation of knowledge is a prime challenge when managing security knowledge. (Hauck et al. 2002b) claims these problems make KM acquisition and interpretation a complex and scary process in the intelligence area. (Zeleznikow et al. 2005) claim that Knowledge discovery from databases process within law enforcement environment can be described using the following stages:

a) Start by examining all the collected police information or data.

b) Secondly try to analyse and understand the reasons for storing this kind of information; the developer should be able to understand the suitable task for the intended users and which of tasks could ideally be automated.

They further state that several systems have been developed to support the investigation processes. These include:

- Geographical Information Systems (Pins in maps)
- Clustering and linkage analysis algorithms.
• Data mining for profiling.

These systems address performance effective management but do not consider issues of sensitivity in a security environment. This originates from two angles:
a) These technologies were originally developed for less sensitive domains such as marketing, insurance and sales.
b) The historical background of assuming sensitivity is predefined.

We have previously considered sensitive information in terms of human data such as name, sex, age etc. Recently, there has been a growing awareness of the ethical issues of misusing such data, especially in medical domains.

Sensitivity is time dependent and also it depends on the person or organisational being considered. What person X or organisational M considers being sensitive may not considered sensitive to Person Y or Organisational N (Sherman 1998) claims that the outcomes of police performance in implementing guidelines and evaluating agencies, units, and officer should depend on evidence-based research. Hence we cannot assume sensitivity for information to be predefined. It must depend on research within a given domain and environment.

**Sensitivity Management and Future Research**

(Han and Kamber 2001) state that a data mining system has the potential to generate thousands of patterns or rules. Not all of the patterns are useful or interesting. Hence we need to define what is an interesting pattern and how can we generate all the interesting patterns and only the interesting patterns.

A pattern is interesting if:
a) the pattern is easily understood by humans;
b) the pattern is valid (with some degree of certainty) on new or test data;
c) the pattern is potentially useful; or,
d) the pattern is novel.

A pattern is also interesting if it validates a hypothesis that the user wished to validate, or resemble a user’s hunch. An interesting pattern represents knowledge.

Several objective measures of pattern interestingness exist, based on the structure of discovered patterns and of the statistics underlying them. The concepts of support and confidence are examples of objective measures of pattern interestingness. In general, each interestingness measure is associated with a threshold, which may be controlled by the user.

Although objective measures help identify interesting patterns, they are insufficient unless combined with subjective measures that reject the needs and interests of a particular measure. Subjective interestingness measures are based on user beliefs in the data. These measures find patterns interesting if they are unexpected (contradicting a user’s belief) or other strategic information on which the user can act.

It is often unrealistic and inefficient for data mining systems to generate all of the possible patterns. Instead, user-provided constraints and interestingness measures
should be used to focus the search. Association rule mining is an example where the use of constraints and interestingness measures can ensure the completeness of mining.

[Oatley et al 2006] have used a variety of data mining techniques (classification and association rules, neural network clustering, survival analysis and Bayesian belief nets, case-based reasoning, ontologies and logic programming) to support police in detecting the perpetrators of burglary from dwelling houses, a volume crime with a low detection rate.

Patterns of interest when using data mining for crime detection often involve detecting outliers rather than conventional patterns. Noise is typically assumed to consist of random discrepancies in data that are due to measurement anomalies and are not a feature of the distribution of data. Data objects that are grossly different from or inconsistent with the remaining set of data are called outliers. [Han and Kamber 2001] note that many data-mining algorithms attempt to minimise the influence of outliers or totally eliminate them.

The outlier-mining problem can be viewed as two subproblems:
a) define what data can be considered as inconsistent in a given data set; and
b) find an efficient method to mine the outliers so defined.

The problem of defining outliers is nontrivial. If a regression model is used for data modelling, than an analysis of the residuals can provide a good estimate for ‘extremeness’. The task is more difficult when finding outliers in time-series data, as they may be hidden in trend, seasonal or other cyclic changes. Data visualisation methods are weak in detecting outliers in data with many categorical attributes or in data of high dimensionality, since human eyes are good at visualising numeric data of only two to three dimensions.

The minimization or elimination of outliers can lead to the loss of important hidden information – such as in the case of fraud detection, where outliers may indicate fraudulent activity. For example, surrounding the acts of terrorism of September 11 2001, there were reports about the existence of a number of foreign students who were interested in learning to fly planes, but not land them. Intelligence forces ignored these outliers, since their emphasis focused upon traditional knowledge discovery techniques.

In traditional KDD, the goal of the KDD system is to learn how standard decisions are made. Hence, data selection, pre-processing and transformation have been used to eliminate outliers. But in crime detection KDD it is essential to develop techniques for searching for outliers.

Our approach to handling sensitive crime knowledge

A sensitivity knowledge base (SKB) will be constructed to enable us to address the ethical issues related to the use of data mining techniques in the criminal domain. We will be working with domain experts (primarily law enforcement officers) to profile possible sensitivity data/information, which will be grouped into two categories, classified and non-classified sensitivity information. Rules will be developed to perform this classification.
Once a prototype has been designed, it will be used to demonstrate how the proposed sensitivity knowledge base can be used to address the sensitivity concern while managing criminal records. Law enforcement agencies will evaluate the system.

An officer on duty captures criminal information into the system; then system should be able to automatically assist in identifying whether sensitive data/information exists in the record. Data/information from the SKB will be of importance while the system tries to detect sensitivity information.

Below is a proposed flowchart of the expected prototype.

**Figure 1: Sensitivity Knowledge Based Generator**

```
Crime Sensitivity Information

Non-Classified Sensitivity
Develop rules

Classified Sensitivity
Develop rules

Sensitivity Knowledge Base
```

**Figure 2: Sensitivity Checker flowchart**
SKB – Sensitivity Knowledge Base
RC – Registered Crime
SC – Sensitivity Checker
CC – Category Checker
RCI: Record as Classified Information.
RNCI: Record as Non Classified Information

**Future Research**

A further issue of great concern is the validity of using results derived from data mining. After all data mining merely indicates likely patterns of fraudulent activity, not proof of illegal actions. This topic will be the focus of our future research.

We are developing ideas on how sensitivity needs to be addressed, but we need further assessment with the help of domain experts from law enforcement agencies. These experts will provide us with advice for building a crime data model to manage sensitivity within law enforcement agencies.

**References**


Using JAD to Bridge the Design-Reality Gaps; a Major Cause of IS Projects’ Failures in the Developing Countries

Euphraith Muthoni Masinde, School of Computing and Informatics, University of Nairobi – Kenya, muthoni@uionbi.ac.ke/muthonimasinde@yahoo.com, P.O. Box 15697 Nairobi, 00100 GPO Kenya

Information Systems (IS) projects failure is ‘a gap between what the users expect from an IS and how well these expectations are met by the perceived performance of the delivered system’. IS projects fail more than they succeed. IS failure rates in the Developing Countries (DCs) are much higher than those in the Industrialised Countries (ICs) because among other reasons, the gaps tend to be exaggerated by the huge difference between the ideas/IS projects and the political /behavioural realities in the DCs. These chronic failure rates have continued to place the DCs on the wrong side of the digital divide, turning IS projects and ICTs in general into a technology of inequality. Solution: employment of Joint Application Development (JAD); a software development methodology that will involve the stakeholders in the entire process of IS implementation. This paper explains how JAD can be used to eradicate most of the causes of IS projects’ failures in the DCs using the University of Nairobi case study. The CHAOS Ten Success factors have been employed to analyze data for nine IS projects.

Introduction

According to research by the Standish Group, on average, only 16% percent of all IS projects in the world ‘succeed’. Despite this lame-duck status, IS permeate just about every aspect of life in the ICs and their failure cause havoc everywhere (Donaldson and Jenkins, 2000). Incomplete requirements specifications and lack of user involvement are the two most common factors that cause these projects to fail or be cancelled. Not involving the users will mean that the final product is what the developer(s) thought the users needed rather than what the users actually needed. Several authors have referred to this situation as ‘gaps’. Heeks(2002) called them design-reality gaps while Linda (2000) called them expectation-perception gaps. Lyytinen and Hirschheim (1987) referred to this situation as Expectation Failure. One way of bridging or at least reducing these gaps is to employ a software development methodology that actively involves users in the entire process; this is where JAD comes in.
According to Heeks (2002), most Information Systems projects in DCs fail totally or partially. IS failure of course is not a DCs’ malaise, for instance, in the USA, in 1995, the cost effort ploughed into computer projects that were subsequently cancelled plus the cost of project overruns were estimated by the Standish Group at a spectacular $140 billions! No one can accurately give the figures that measure the software failure rates in the DCs; this is not even possible in the ICs. Some approximates in the latter are that ¼ projects totally fail, 1/3 to 3/5 partially fail while the rest succeed (The Economist 2000). There is no known proof that the figures in the DCs are higher or lower than these but practical reasons such as lack of technical and human infrastructure may push the failure rates in the DCs upwards. The reality of the failure rates is more mundane and lack of literature in this area and lack of IS projects’ evaluations (the little found is on individual projects’ case studies) makes the study of failure/success rates of IS projects in the DCs more complicated. Sahay and Walsham (1995) summarises the situation of IT in the developing countries as follows: ‘The process of IT use in the Developing Countries is a complex phenomenon and it typically involves actors at various levels. It is important to study the interaction of these actors on the process of IT implementation and use” (p. 118)

JAD has been in use for close to 30 years now and it is a methodology aimed at involving all key stakeholders [1] in the entire development process. JAD has been refined over the years incorporating features such as electronic meeting systems by Carmel et al (1992); that differentiates Electronic-JAD (E-JAD) from Traditional-JAD(T-JAD). Various software tools available today can be used in automating most of the tasks [2] that are carried out during JAD sessions. The studies of the uses of JAD have also been carried out, such as in the ‘an exploratory study of JAD in information systems delivery’ by Davidson (1992)

Measuring success of a IS project is a very difficult task and whatever parameters one uses; no absolutely satisfactory results may be attained. Authors of various pieces of literature in this area have proposed different ways of measuring success. One way is by looking at the technical properties of the system, the fit between organization needs and system capabilities and also customers’ satisfaction proposed by Christine and Paul (1999). Whatever the criterion used, in this paper, systems will be classified into three success scales: Successful, Challenged/Partial Failure or Failed (Standish Group).

The University of Nairobi adopted the JAD as a methodology in 1999 and has since then reversed the trend of massive IS projects’ failures that had existed since early 70s. The success is quantified by the very high successful rate of IS projects that have been put in place so far.

**JAD – An Overview**

**Definition**

Joint Application Design (JAD) is a structured process in which users, managers, and analysts work together for several days in a series of intensive meetings to specify or review systems requirements. The systems development personnel at IBM developed it in the late 1970s. JAD has evolved over time to include other phases (design, coding,
etc) of software development, hence acquiring the name Join Application Development. It is the latter definition that is used in this paper. There is a close correlation between JAD and Rapid Application Development (RAD); to some extent, JAD is a tool for RAD success (Hoffer et al). Despite the different definitions and forms that JAD has acquired, the key characteristic of JAD is the facilitated sessions. Each JAD session has well-defined objectives, detailed agenda and guidelines, visual aids and final documents containing all the decisions made by the group. JAD should be used after the high-level requirements have been developed and should consist of 3 major phases as shown in Figure 1 proposed by Allan and Mary, 2000. Each session is supported by a session leader, a facilitator, a scribe, technical specialists (e.g. in Database Design and user interface design) and domain expert.

![Figure 1: High-level process diagram of the JAD process](image)

**Advantages/Disadvantages of JAD**

Numerous articles, case studies, and other related studies have shown a number of benefits of using JAD. These have been summarised by Alan on the website ([http://www.carolla.com/wp-jad.htm](http://www.carolla.com/wp-jad.htm)) as: saves time, eliminates process delays and misunderstandings and improves system quality; It is one of the best ways to reduce function creep, most of which results from poor initial requirements. By properly using transition managers, and the appropriate users, the typical cultural risk is mitigated while cutting implementation time; It also avoids bloated functionality, gold-plating, and helps designer’s delay their typical “solution fixation” until they understand the requirements better; Lays the foundation for a framework of mutual education, separate brainstorming, binding negotiation, and progress tracking; Finally, JAD helps avoid the requirements from being too specific and too vague, both of which cause trouble during implementation and acceptance.
JAD has its own share of disadvantages; top on the list being cost. JAD can push the cost of the entire project upwards in terms of people’s time and money. Bringing many people from different levels together in a room may mean that some may not give their ideas; cannot challenge their bosses and the same people may later turn around and reject the system (Hoffer et al).

**JAD Principals**

Authors suggest general principles of JAD such as to involve all the major stakeholders or stakeholders’ representatives, ensuring that JAD teams have support from upper management, involving a technical facilitator with skills in both systems analysis and group dynamics and ensure that each stakeholder has a representative empowered with decision-making. Further, each session should be short (2 – 4 hours), each session must produce JAD minutes, which contains attendees’ resolutions, action items, and open issues. The facilitator sends copies to all team members and their managers. Defined in most of the literature also are the JAD tasks such as identifying all stakeholders and clarifying executive goal. (Hoffer et al).

**Is Success/Failure**

**Categories of IS Projects’ Success/Failure**

Success/failure of software projects has been classified by authors Standish Group, Puri et al, 2000, Kitiyadisai, 2000, Benjamin, 2001 into three categories: 1) Successful – in which most stakeholder groups attain their major goals and do not experience undesired outcomes, 2) Challenged/partial failure – where major goals are unattained or significant undesirable outcomes such as over-budgets and over the time estimates are experienced. This category of projects is difficult to be assessed because the failure/success may subjective. 3) Failed/total failure – is where the project is never implemented, cancelled or implemented and immediately abandoned.

**Failure Factors**

May (2000) gave some of the major specific causes of software failures as: Poor user input leading to systems that do not meet their needs, Stakeholder conflicts; either the ‘stake holders’ of the system are not well defined or they are not willing to work together, Vague requirements, Hidden cost of going “lean and mean” which means over-reducing the number of employees while maintaining the same deadlines, Failure to plan – ‘failure to plan is planning for failure’, Communication breakdown between people in the various levels in the software project. Architecture that may not be flexible, late failure warning signals. Lientz and Rea, 1999 analyses the cause so IS projects failures in form a list with 25 reasons (p. 12-14) which all seem to fit in the above categories.
Success Factors

The factors that determine whether a particular project gets into either of the above categories are many (addressing all the factors that lead to software failures). Research as shown that the nature of the project matters too. For example, the size of the project, (smaller projects tend to have a higher success rate than larger ones) and company size (bigger companies have better chances to succeed). All authors in this area agree that the factors that make IS projects successful are not especially technical, e.g. McConnell, 1998. McConnell argues that user involvement is a critical survival skill because it ensures that the users will use/like the final product. He summarises it in the statement; “Ask users what they want, show them what you intend to build and ask them how the like it.” CHAOS study identified ten weighted successful factors for IS project (Standish Group):

Table 1 – The CHAOS Ten Success factors

<table>
<thead>
<tr>
<th>Success Criteria Points</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. User Involvement</td>
<td>20</td>
</tr>
<tr>
<td>2. Executive Support</td>
<td>15</td>
</tr>
<tr>
<td>3. Clear Business Objectives</td>
<td>15</td>
</tr>
<tr>
<td>4. Experienced Project Manager</td>
<td>15</td>
</tr>
<tr>
<td>5. Small Milestones</td>
<td>10</td>
</tr>
<tr>
<td>6. Firm Basic Requirements</td>
<td>5</td>
</tr>
<tr>
<td>7. Competent Staff</td>
<td>5</td>
</tr>
<tr>
<td>8. Proper Planning</td>
<td>5</td>
</tr>
<tr>
<td>9. Ownership</td>
<td>5</td>
</tr>
<tr>
<td>10. Others</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

Information Systems Failure - A Crisis

The Oxford English Dictionary defines crisis as ‘a decisive moment, a time of danger or great difficulty, a turning point’. In this paper, IS crisis will be used to refer to all the problems that are encountered in the IS projects. For software industry, IS crisis has been with us since the birth of the industry (software) itself. Pressman, 1994 proposes the use of the term ‘chronic affliction’ instead of ‘IS crisis’ because of its longevity and reoccurrence. IS crisis has led to devastating losses that are documented in literature.

There is a large body of literature addressing IS projects failures in the ICs. Some of the case studies include: The London Ambulance Service Computer-Aided Despatch (LASCAD) System Project which is one of the most frequently quoted UK examples of IS failures in recent times (Dalcher and Tully, 2002, Christian Lundestad, 2003, Paul, 1999,Paul, 1995). “On the 27th October 1992, an IS made the lead story on the BBC’s nine o’clock news; the new computerised system established at the headquarters of the London Ambulance Service (LAS); LASCAD, failed and that as a result, the lives of 20-30 people may have been lost.” (Paul, 1991), (p. 1). The system was introduced in an atmosphere of mistrusts by staff and there was incomplete ‘ownership’ of the system by the majority of its users. There was disorganisation, low staff morale, friction between
management and the workforce and an atmosphere of hostility towards computing systems (Paul, 1995). Other cases documented by Flowers, 1996 and Glass, 1998 are: the FBI Virtual Case File that was delivered one year late with only $170 million wasted; Licence Registration System for Washington State Department that was delivered late, $40 million wasted and was never used. Closer to our lives, the Microsoft’s new Sender ID technology quarantines unwanted email but it cannot tell the healthy from the sick (IEEE Spectrum January 2006).

Contrary to this, there is no adequate literature (known to the author) on IS projects failures in the DCs. Some of the projects documented include: 1) the Accounts and Personnel Computerisation Project of Ghana’s Volta River; the project partially failed because of the lack of involvement of some lower-level staff (Tetty 2000). 2) The Touch Screen Kiosks for remote rural communities in South Africa’s North-West province failed soon after implementation. The kiosks were not useful to the local users because the content therein was either not updated or did not make sense to them (Benjamin 2001). 3) Workflow System for a South African tyre manufacturing firm that was never used (Calitz 2000).

Walsham and Sahay (2006) analysed most recent (2000 onwards) literature on IS in the DCs under the topic ‘Research on Information systems in Developing Countries: Current Landscape and Future Prospects.’ They classified the literature into 3 categories: 1) Those that addressed how to deal with the challenges facing IS practitioners 2) The ones dealing with the role of technology and 3) Those that proposed suitable theories and methodologies. In the first category, one of the challenges is ‘local adaptation and cultivation’. Bringing IS to a new local context involves some implicit elements of cultural transfers and mutual learning. Proposals by authors in this area (Bada, 2000, Ehikhamenor 2003, Makome 2003, and D’Mello 2003) converge to the point that understanding the local context is crucial in IS implementation. In the second category, the issue of ‘standardisation versus localization’ was analysed. The authors (Braa and Hedberg 2002 and Thompson 2002) propose localization rather than standardisation. Recommended also is the issue of evaluating the applicability of each technology (e.g. GIS) to make sure that it does not conflict with the local way of perceiving and knowing. In the third category of literature evaluated (Bada 2002, Heeks, 2002, Braa and Hedberg 2002 and Madon and Sahay 2002), suggested that new IS theories for the DCs were required.

**Solution to IS Gaps**

In all the above, lack of cohesion between the IS project developers and the stakeholders was the major contributor to the failures. Consequently, what was delivered was not what the stakeholders were all along expecting.

**Expectation-Perception Gaps**

This expectation-perception gap can be understood and analysed by a gap model that was proposed by Linda, 2000 as shown in figure 2.
Linda analysed 6 gaps that arise as a result of developers working in isolation from the users/stakeholders: Cognition Gap (Gap 1) – the difference between ‘what the users needs’ and ‘what they think they needs’. It is as a result of users’ inability to cognate upon their information needs; Comprehension Gap (Gap 2) – the difference between ‘what the user need’ and ‘what the developers think the user need’; it is the developers’ inability to comprehend users’ information needs; Expression Gap (Gap 3) – the difference between ‘the developers’ understanding of users’ needs’ and ‘the translation of developers’ understanding into requirements specifications’. This is affected by the developers’ mental constructs (e.g. perceptive process, values, ethics, motives, prejudices, intellectual ability, experience, etc); it is caused by developers’ inability to translate the perceived information needs of users into requirements specifications; Delivery Gap (gap 4) – difference between ‘a system as specified’ and ‘a system as delivered’; this is the developers’ inability to transform specified needs for information provision into systems deliverables; Utility Gap (Gap 5) – difference between ‘a system delivered’ and ‘a system in use’; it is usually as a result of users’ inability to utilize the delivered systems to satisfy their information needs; Expectation-Perception gap (Gap 6) – a function of gaps 1, 2, 3, 4 and 5. Solution: keep Gap 6 closed by closing gaps 1 through 5 by having IS professionals becoming service providers to users rather than the becoming proprietors of the IS.
A similar model was proposed by Lyytinen and Hirschheim (1987). They categorised IS failures into: 1) Correspondence Failure – a management perspective of IS failure where there is lack of correspondence between the objectives set out and the results of evaluating the resultant IS; 2) Process Failure – is the unsatisfactory development performance; the development process cannot produce a workable system or the system is produced but the project runs over budgets in terms cost, time and other resources; 3) Interaction Failure – is the mismatch of the requirements and the resulting IS, a situation that leads to a system that is hardly used; 4) Expectation Failure is a superset of Correspondence, Process and Interaction Failures. Lyytinen, 1998 broadened the notion of Expectation Failure to generalise failures into two: Development Failure and User Failure.

**Design-Reality Gaps**

Heeks, 2002 presents these gaps (in the case of DCs) using seven dimensions: Information, Technology, Process, Objectives and values, Staffing and skills, Management and structures and Others; yielding the ITPOSMO mnemonic (figure 3). These factors lead to very huge gaps essentially between the rationality of the software projects design (hard) and the political/behavioural realities (soft) of the DCs’ organisations.

**Figure 3. Design-Reality Gaps**

![Design-Reality Gaps Diagram]

**Termination Failure**

Sauer, 1998 ‘lightens’ the definition of IS failure to only when the development or operation ceases; called Termination Failure He proposed a triangular model with three interacting components: Information System, Supporters and Project Organisation. Beynon-Paul, 1999] suggested the replacement of Supporters with Stakeholders yielding the model in figure 4
Sauer introduced the concept of flaws that may be corrected within any innovation process at a cost. E.g. program bugs, hardware performance, organisational changes etc. a build up of uncorrected flaws may lead to termination failure.

It is clear from above that most of the IS projects failures are not caused by the technical shortcomings but rather by the social issues around the IS projects. Huy et al, 2006 argues that, to some extent, there might be need to reform IS education to make the learning process effective for IS professionals. The IS professionals may be lacking on the area of ‘System Thinking’ hence being unable to deal with organizational and social issues of IS projects that they handle.

Putting Linda’s and Heeks models together results in the situation depicted in figure 5. In the model, it is clear that the ‘expectation gaps’ for IS projects in the DCs are so huge because of the ‘design-reality gaps’
Using Jad To Improve Is Projects’ Success Rates In The Dcs

Not much literature exists on the assessment of IS projects success/failure in the DCs. The little that exists brings out the fact that failure rates are higher than those in the industrialized countries. Even rarer, is literature on how to deal with the high failure rates. The few approaches to addressing this include theoretical ones (Baruah, 2000 and Barrett, 2001) as well as those based on soft systems ideas. The latter recognizes that social and organizational factors are more likely to determine the success/failure of a software project in the DCs more than elsewhere in the world. An IS project that tries to match values, perceptions and assumptions of the key stakeholders has higher chance of succeeding (Heeks 2002). A balance however need to be reached to avoid a situation where the projects do not change the environment at all in which case it would beat the logic of having the system in the first place.

Local Improvisations

Just like everywhere else in the world, IS projects in the DCs fail more than they succeed because of the gaps between what is developed and what dominant stakeholders wanted. What is unique about the DCs is the fact that the ideas/software systems come from the industrialized countries (totally different environment) making the gaps even wider (figure 5) hence, increasing the failure rates. The solution to averting the trend of software failures in the DCs therefore lies in bridging the above gaps. One proposal by Heeks, 2002 is the use of local improvisations using bi-directional approach; changing local realities to make them closer to the IS projects design; and/or change the IS projects’ design to make them closer to the DCs organisational realities. In doing so, the success will still depend on how well the initial requirements are acquired and that the direction of gap-bridging will vary from one project to another. Some proposal of supporting this gap-bridging include: exposing organisational realities, improving local IS capabilities, educating the ‘carriers’ of the industrialised innovations (such as donors, consultants, ICT vendors, DC personnel trained according to traditional industrialised curricula) on the realities of the DCs and analysing both the ‘how’ and the ‘what’ in relation to the software project.

Why JAD?

There are still some loose ends here; how do we formalise the whole process of bringing the stakeholders together? JAD solves this via JAD sessions because they provide both qualitative and quantitative consultation with all the stakeholders of the project in a formal set up.

Based on the CHAOS Ten Success Factors, JAD can be used to achieve a very big percentage of success as follows: In JAD, users (or user representatives) are involved in the software development (20%). JAD sessions involve managers, hence assured of executive support (15%). The representation of most stakeholders during sessions ensures that the business objectives are clearly stated (15%) and that all the stakeholders will feel that they own the software (5%). Important too is the fact that once the requirements are stipulated, they are bound to be firm (not changing always) since most
of the important decision makers are represented (5%). This is already 60% success chance by just adopting JAD!

**UoN – CASE STUDY**

**Background Information**

The University of Nairobi (UoN) is the oldest, largest and best-established public university in Kenya and the larger Eastern Africa region. Its origin can be traced back to 1951 when its precursor, the Royal Technical College of East Africa was established (www.uonbi.ac.ke). The University of Nairobi is located at the heart of the capital city, Nairobi, with an enrollment of over 30,000 students pursuing diploma, and degree courses in most areas of study. The University’s Information and Communications (ICT) Center is in charge of Management Information Systems’ (MIS) support among its responsibilities. The University has always operated on some kind of ‘semicomputerized’ systems until early 90s when these ‘systems’ could no longer support the University’s operations. This was triggered by a revolution in the University, aimed at expanding the University in terms of students’ population, increasing the number of programs and relaxing most of the programs regulations to accommodate Module II students (Kiamba, 2003). Most of the software that had been put in place (both in-house developed and off-the-shelf) had just remained unutilized and later discarded. The factors that contributed to this are: - resistance by the end-users, lack of support by the University’s top management, lack of support by vendors of these systems and use of obsolete hardware and operating systems. With the pressure from the dynamism in the new way the University was being run, there was urgency to acquire the essential information systems and ensure very high level of success in the implementation of these systems. Hence, the choice of the JAD methodology by the Director of the University’s ICT Center. Today, this decision has turned out to be a success story for the University.

**IS Failure at the Uon**

Among the many causes of software failures mentioned in this paper, user resistance contributed to 60% of reasons as to why software projects at the UoN failed. Other reasons that contributed to the failures are: incompetent staff, because most of the users were not computer literate, lack of support from the University’s top managers who did not see the need to fund such systems, use of inappropriate tools such as obsolete operating systems, computer hardware and programming tools and lack of experienced project managers. From document reviews and informal interviews conducted, the failure causes were distributed as follows:
Advances in Systems Modelling and ICT Applications

Part 4: ICT and Legal Applications

Chart 1: Causes of IS

- Users Resistance
- Obsolete Tools
- Lack of Support by Vendors
- Lack of Support by UoN Management
- Others

New Software Development Approach At The UoN

In the early 2000, having realized the critical need of a functioning MIS department at the UoN, the top management put in place a well structured computing unit (the ICT Center) for the University. MIS was then re-established under this center, staffed with very competent project managers and supplied with excellent software development tools. Modular approach to software development was used with Students Management Information System (SMIS) and Human Resources Management Information System (HRMIS) pioneering. Halls Management Information System (HMIS) followed two years later. For each of the above projects, the end users (at all levels) were identified and together with the technical people (systems analysts/programmers, network engineers, database administrators among others), formed a team that worked together on the projects from the projects’ initialization to commissioning. To ensure that the top management of the University supported the projects, the relevant managers/representatives were on several occasions invited to attend some of the teams’ meetings. Also, the minutes of the teams’ sessions were always copied to these managers to update them on the progress of the projects. Further, out-station seminars were held at intervals of 6 months to update stakeholders on the progress of the projects.

However, developing these systems has not been a smooth ride all through. Most of the systems were completed long after the deadlines and went way beyond their budgets. Quite a bit of user resistance was experienced especially because most of the University employees were still computer illiterate at the time of commissioning the systems. The overwhelming support from the top management especially the Vice Chancellor helped address most of the hiccups. In summary, the systems are a big success and have made a significant mark in the history of the University of Nairobi.

Data collection and Analysis

In determining the success/failure of the IS projects, Sauer’s definition of IS project success/failure was used. A sample of 30 users of all the new systems was requested to
fill questionnaires regarding their view of the success/failure of the systems. Informal interviews were conducted for 10 stakeholders of the systems; these were mainly middle level managers (such as the Examination Officers, Finance Officers, etc) and the systems analysts. The questionnaires/interviews used a framework of the CHAOS 10 success factors of measuring success. On average, the following were the results.

**Chart 2: UoN IS Projects Success Levels**

![Chart 2: UoN IS Projects Success Levels](image)

**Chart 3: UoN IS Projects**

![Chart 3: UoN IS Projects](image)

**Conclusion**

From this paper, it has been shown that more than 60% success of IS projects can be achieved by employing the use of JAD methodology. During JAD sessions, stakeholders are actively involved and their ideas/views incorporated. This is a critical success factor in the DCs. Potential IS ‘gaps’ are bridged continuously.
Having worked for the UoN, JAD will work for most organisations in the DCs and this may reverse the current trend in the IS industry in these countries. Further quantitative as well qualitative research is required into actual situation of IS failure/success in the DCs. Also, more ‘experiments’ to prove that the model proposed in this paper need to be carried out using data from other institutions within the DCs.

**Footnotes**

[1] Key stakeholders in relation to an information system in a given organization will include the top management, the sponsor, the end-users and IT technical people.

[2] Some of the JAD tasks that are automated includes brainstorming, outlining, matrix analysis, voting and prioritizing, strategic plan development, business process re-engineering, requirements definition, prototype evaluation, implementation plan development and system migration assessments.

**References**


Heeks Rechard 2002 “Failure, Success and Improvisation of Information Systems Projects in Developing Countries”. The Development Informatics working paper series; Published by: Institute for Development Policy and Management, University of Manchester, Precinct Centre, Manchester, http://www.man.ac.uk/idpm/idpm_dp.htm#devinf_wp

Hoffer Jeffrey A., Joey F. George and Joseph S. Valacich; Modern Systems Analysis and Design; 3r Ed. Pearson Education


Kiamba Crispus., 2003; The Experience of The Privately Sponsored Studentship and Other Income Generating Activities at The University of Nairobi A Case Study Prepared For A Regional Training Conference On Improving Tertiary Education In Sub-Saharan Africa: Things That Work!: Accra, September 23-25, 2003


Sauer C 1993; Why Information Systems Fail: A Case Study Approach; Alfred Waller, Henley-On-Thames, 1993


The Economist, 2000. No gain without pain, Government and the Internet Survey, 24

Survey of Data Mining Methods for Crime Analysis and Visualization

Okwangale Fredrick R., Department of Networks, Faculty of Computing and Information technology, Makerere University and Dr Ogao Patrick, Head of Department, Department of Information Systems, Faculty of Computing and Information technology, Makerere University

Crime prevention is a primary concern of police as they perform their central role of protecting the lives and property of citizens. But the police force is usually relatively very small compared to the crime prone population they have to protect making them more of a reactive rather than preventive force. Police often have at their disposal vast amounts of least utilised crime data (such as crime incident reports) which if analysed could reveal some hidden information such as crime committing trends useful in crime prevention. Use of Information Systems techniques such as data mining and Geographic Information Systems for analysing these data is promising in boosting the police efforts. This paper reviews the applicability of various data mining methods and Geographic Information Systems in crime analysis and visualization in mainly poor planned settings characterised by missing electronic data a common phenomena in the developing countries like Uganda. The focus is on criminality of places rather than the tracing of individual criminals. The review tends to reveal that a combination of Geographic Information Systems and data mining techniques that can work under unclean data are best suited for use in the poorly planned settings.

1. Introduction

Security of citizens is the major concern of the police. Rather than focusing on enforcement and incarceration police can deter crime through the knowledge benefits that derive from information and its associated technologies The police force can employ information technology to turn police officers into effective problem solvers and to leverage their intellectual capital to pre-empt crime (Brown et al., 2003)[4]. However one challenge to law enforcement and intelligence agencies is the difficulty in analyzing large volumes of data involved in criminal and terrorist activities (Chen et al., 2003)[14]. A variety of techniques in data mining and Geographic Information Systems (GIS) are surveyed in their applicability in use in environments of less organized data.
2. Overview of crime, data mining and Geographic information system

2.1 Crime

Crime is referred to as a comprehensive concept that is defined in both legal and non-legal sense (Akpinar et al, 2005)[2]. From the legal point of view crime is the breaking or breaching of the criminal law (penal code) that governs a particular geographical area (jurisdiction) aimed at protecting the lives, property and the rights of citizens belonging to that jurisdiction. Crime is an offence against a person (for example murder, and sexual assault), or his/her property (for example, theft and property damage) or the State regulation (for example traffic violations) (Akpinar et al, 2005)[2] (Oxford Dictionary of Current English). In non-legal terms crime is a set of acts that violate socially accepted rules of human ethical or moral behavior (Akpinar et al, 2005)[2]; for example acting against a ritual in some society.

Crime occurs in a variety of forms which police informally categorizes as being either major or volume. Major crime consists of the high profile crimes such as murder, armed robbery and non-date rape. These crimes can either be one-offs or serial. In the case of serial crimes it is relatively easy to link crimes together due to clear similarities in terms of modus operandi or descriptions of offenders. This linking is possible due to the comparatively low volume of such crimes. Major crimes usually have a team of detectives allocated to conduct the investigation. In contrast volume crimes such as burglary and shoplifting are far more prevalent. They are usually serial in nature as offenders go on to commit many such crimes. Property crimes, such as domestic burglary offences, committed by different individuals are highly similar and it is rare to have a description of the offenders (Adderley and Musgrove, 2001) [1]. Table 1 shows the classification of crime (Chen et al., 2003)[14].

Table 1. Crime types at different levels. Source: (Chen et al., 2003)[14]

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Violations</td>
<td>Driving under the influence of alcohol, fatal/personal injury/property damage traffic accident, road rage</td>
</tr>
<tr>
<td>Sex crime</td>
<td>Sexual offences</td>
</tr>
<tr>
<td>Fraud</td>
<td>Forgery and counterfeiting, frauds, embezzlement, identity deception</td>
</tr>
<tr>
<td>Arson</td>
<td>Arson on buildings</td>
</tr>
<tr>
<td>Gang / drug offences</td>
<td>Narcotic drug offences (sales or possession)</td>
</tr>
<tr>
<td>Violent crime</td>
<td>Criminal Homicide, armed robbery, aggravated assault, other assaults</td>
</tr>
<tr>
<td>Cyber crime</td>
<td>Internet frauds, illegal trading, network intrusion /hacking, virus spreading, hate crimes, cyber piracy, cyber pornography, cyber-terrorism, theft of confidential information.</td>
</tr>
</tbody>
</table>
Also a single detective officer may have a large number of different volume crimes to investigate at any point in time. With a view to satisfying this demand, police forces around the world employ specialist crime analysts, people who have specialist training in a variety of disciplines including investigation techniques, criminal psychology and information technology. It is their task to assist investigating officers by analyzing crime trends and patterns, identifying links between crimes and producing packages which target an individual or group of offenders linking them to a series of crimes (Adderley and Musgrove, 2001)[1].

Most, if not all, of current systems both manual and computerized revolve around the investigation of crimes already committed. They are, therefore, reactive. In the developed countries like the UK, a majority of crime prevention forces use different types of relational database management systems (RDBMS) for recording and subsequent analysis of crime. Standard or interactive queries are written to produce patterns of crime, offending and various statistics (Adderley and Musgrove, 2001)[1] but its is a common phenomena in the developing countries to find mainly manual criminal record books used alongside the pin-up maps for crime incidence location.

2.1 Data Mining

Data mining deals with the discovery of unexpected patterns and new rules that are “hidden” in large databases. It serves as an automated tool that uses multiple advanced computational techniques, including artificial intelligence (the use of computers to perform logical functions), to fully explore and characterize large data sets involving one or more data sources, identifying significant, recognizable patterns, trends, and relationships not easily detected through traditional analytical techniques alone. This information then may help with various purposes, such as the prediction of future events or behaviors. (Reza et al, 2001)[14]

The development of new intelligent tools for automated data mining and knowledge discovery has led to the design and construction of successful systems that show early promise in their ability to scale up to the handling of voluminous data sets.

Theories of crime and delinquency tend to be discipline-specific and are dominated by psychological, sociological, and economic approaches (Reza et al, 2001)[14]

Data Visualization

Visual methods are powerful tools in data exploration because they utilize the power of the human eye/brain to detect structures. A number of data mining tools for visualization exist, a histogram and Kernel plots being the most basic used for displaying single variables. Scatter plots for the display two variables at a time and reveal correlation, if any, between them. And for more than two variables, scatter plot matrices are often used (David et al, 2001)[7]. GIS also provides a powerful visualization tool through display of maps that allow the exploration of spatial patterns in an interactive fashion (Pfeiffer, 1996) [12].
2.2 Geographic Information Systems

Over the past few years Geographic Information Systems (GIS) has become a standard tool for crime analysts in many police departments, regardless of their size (see for example McEwen and Taxman 1994; Rossmo, 1995). One of the inherent advantages of GIS is its ability to integrate information from a variety of sources into one user interface. In turn, this allows for spatial analyses that would either not have been possible or at a minimum far more difficult prior to the advent of GIS (Olligschlaeger, 1997)[11]

2.3 Crime Analysis

Crime being inherently spatial phenomena (Ratcliffe, 2004)[13], taking place within a given location and at a specific time, it is logical to analyse crime in terms of spatial crime analysis. The advent of geographic information systems (GIS) has given behavioural scientists the ability to map and track a wide range of social phenomena such as crime incidences. Much more information about what is happening “on the ground” is now available. This enhanced analytic capacity has presented police departments with the opportunity to better serve and protect the people and places in their care. Furthermore, as GIS assists one in seeing and understanding behaviour patterns, it also provides stakeholders with opportunities to join together in partnerships for the common good (Holzman et al, 2003)) [8]. A data pattern is an expression in some language describing a subset of the data or a model applicable to that subset (Fayyad et al, 1996)[6]. There are at least three types of police crime information (primary, secondary, and tertiary), intelligence (prospective, retrospective, and applied), and operational strategies (preventive, prospective, and reactive), each of which interacts in a complex fashion with technology (Manning, 1992) [9]. Crime data analysis aids police in transforming this information from one type to another.

Crime Spatial Data Analysis

Spatial data analysis utilizes statistical analysis methods that address specific issues relating to spatial data, including spatial dependence (autocorrelation) and spatial heterogeneity. These issues run counter to traditional statistical assumptions of heterogeneity and independence of sample data. If these issues are ignored, then analysis results might not be valid. In combining the powerful tools of GIS to integrate and manipulate spatial data with rigorous statistical methods, spatial data analysis shows great promise for criminology, criminal justice, and law enforcement research and practice. (NIJ, 2006)[10]

3. State Of Crime Data At The Uganda Police

Most of the crime data for the Uganda police is manually recorded in criminal record books (CRB) in form of statements made by arrested suspected criminals, reports by informers and information gathered by the police themselves. Pin-up maps are used to capture the locations of crime incidents (Uganda Police online, 2005)[15] The deployment areas tend to be specific.
4. **The Data Mining Techniques for Crime Spatial Data Analysis and Visualisation**

There are a number of data mining techniques for crime analysis and visualisation but the choice of which to use depends on the features of the problem, the data and the objectives (David et al, 2001) [7]. This paper is limited to the Exploratory Data Analysis (EDA) and specifically crime spatial data analysis. Typically, EDA techniques are interactive and visual, and there are many effective graphical display methods for relatively small low-dimensional data sets. The difficulty in visualisation of points increases with the number of variables involved (David et al, 2001) [7]. The methods used in crime spatial data analysis can be classified into those concerned with visualisation of data, those for exploratory data analysis and methods for the development of statistical models (Pfeiffer, 1996) [12]. The methods are surveyed categories of the data based on the categories of crime data to be analysed - point patterns.

**Point Patterns**

Spatial point patterns (SPP) are based on coordinates of events such as locations of crime incidences and may also include the time of occurrence. All or a sample of point pattern may be plotted on the map. The aim of SPP analysis is to detect whether the point pattern is distributed at random, clustered or regular. SPP is typically interpreted as analysis of clustering. A dot map is commonly used to represent SPP. The tool effectively used for analysis of clustering effects is the K function. This method assesses clustering of crime incidences in detection of hot spots (Kingham et al, 1995) where time and space relationship analysis is required, the methods used are Knox’s method, Mantel’s Method and K-nearest neighbour method. All the three methods require the production of distance matrices of the spatial as well as temporal relationship between crime incidences. Knox’s method requires critical distance in time as well as space defining closeness has to be set but the determination of these critical distances requires subjective decision. Mantel approach does not however require use of critical distances but uses both time and space matrices. It is however insensitive to non-linear associations. The K-nearest neighbour is based on the approximate randomisation of the Mantel product statistic (Pfeiffer, 1996) [12].

5. **Conclusion**

Exploratory data analysis makes few assumptions about data and it is robust to extreme data values. It is possible to use simple analytical models with EDA. The methods that are robust to missing data are useful in the data mining of crime data where data is not so precisely collected. The distances between crime locations are normally not easily available to the police in the areas that are not well planned. The poorly planned areas are best represented by dividing them into area clusters and the analysis is done based on the clusters. The methods that support clustering are therefore best suited for the crime analysis of the poorly planned settings. The manual pin maps are best replaced by the use of GIS.
References


Informing Regional ICT Policy: Case Study of Trends in ICT Indicators of OECD, EU, COMESA and EAC

Joseph Muliaro Wafula, muliaro@jkuat.ac.ke/muliaro@yahoo.com, Institute of Computer Science and Information Technology, Jomo Kenyatta University of Agriculture and Technology, PO Box 62000, City Square 00200, Nairobi, Kenya,
Anthony J Rodrigues, tonyr@uniobi.ac.ke, School of Computing and Informatics, University of Nairobi, PO Box 30197, Nairobi, Kenya and
Nick G Wanjohi, Institute of Human Resource Development, Jomo Kenyatta University of Agriculture and Technology, PO Box 62000, City Square 00200, Nairobi, Kenya

Economic and social transformation of nations has been brought about by technology. Both policy analysts and academic researchers have been interested in understanding and articulating these phenomena especially, when governments are constantly needing information about the performance of their own country relative to their partners and competitors. Several methods are currently in use including use of e-readiness indices. However, the criteria of generating the e-readiness indices are not uniform. For one to produce an e-readiness index, at least two indicators have to be summed, subtracted, multiplied or divided yet research has shown that of now there are no meaningful theories that justify such algorithms. To make it worse, the selection of the ingredients of the e-readiness index depends heavily on the value judgment of the scholars as well as on the availability of the data. As it stands now, one of the best options for informing policy makers and governments is to use indicators. This paper discusses trends and analysis of ICT indicators such as mobile phone subscribers, mobile communications revenue, annual telecommunication investment, international bandwidth and mass media usage in relation to ICT policy, e-strategy and enabling environment for OECD, EU, COMESA and EAC based on the 2005 ITU and DHS databases. Correlation analysis among indicators was done in attempt to establish emerging commonalities and differences among these regions and the accompanying lessons noted. The future of wireless telecommunications that are converging with mobile computing devices in offering Internet connectivity and access to most COMESA countries including EAC are discussed.

Introduction

The successful use of ICT applications requires improved awareness in the public and business sectors, better education and improved literacy rates, user involvement in designing and implementing new services and applications, policies for improved public access to networks, and a readiness on the part of governments and other
stakeholders to assume responsibility for selecting and giving priority to a wide range of policy and practical initiatives (Mansell, 2002). More so, the government needs to be the first to incubate new ICT applications if it expects faster and sustainable adoption of the technology. Frieden (2005) argues that a successful ICT incubation appears to require government involvement, albeit with a light hand that stimulates and rewards investment, reduces unneeded regulatory scrutiny, and promotes global marketplace attractiveness without “tilting the competitive playing field” to favour a specific technology or company. It has been observed that nations as diverse as Canada, Japan and Korea provide insights on how to achieve maximum success in ICT development and the roles governments can effectively assume. These and other nations offer insights on how government-led integration of technology incubation and development can generate ample dividends. These governments readily encourage private enterprise and direct foreign investment in technology ventures. The best practice ICT development observed in most of these nations demonstrates the benefits from long-term involvement by honest, technologically sophisticated government officials who understand the stakes involved and work conscientiously to establish a transparent, efficient, flexible and positive business environment for the long run. Frieden (2005) puts it that governments can enhance ICT development by articulating from the top a broad vision of what ICT can do for a nation and its citizens, while leaving to community champions the flexibility to propose specific, “bottom-up,” projects that aggregate the supply of services needed to support the build out of a telecommunications infrastructure. In line with the “bottom-up” approach, Gillett et al (2004) through research have observed that local governments have become involved in promoting the development of advanced communication services in various ways such as acting as stewards of local economic development through improvement of efficiency and quality of government service delivery through e-government.

In Canada, Korea, Netherlands, and Sweden, government institutions for telecommunications, broadcasting, and Internet policy and rules are organized in a variety of ways. In all cases, there is some kind of distinction between the authorities for telecommunications and the authorities for broadcasting. Responsibility for the Internet tends to fall to the telecommunications authority. All the four countries have some rules and designated public authorities with mandates to promote national culture and identity in the media. This is enforced through a variety of quotas for domestic programs and limits on foreign programs that are used in the media (Wu, 2004).

The future of the Internet has been perceived to be broadband and hence the industry groups and analysts have stressed the importance of broadband access for continuing the evolution of advanced communication services and overall economic growth (Gillett et al, 2004). It has been noted that broadband development thrives when it becomes a national priority (Frieden, 2005). The question is, under what circumstances does broadband qualify to become a national priority? Different views have been raised in response to this question. Firth and Mellor (2005) argue that broadband has the potential to offer nations improved quality of education and health services, improved connectedness of government with society, and provision of jobs and prosperity.
Most nations showing interests in broadband are doing so on the understanding that broadband will bring social and economic benefits (Firth and Mellor, 2005).

Convergent trends in telecommunications and broadcasting technologies and markets have called forth a re-examination of universal service provision in the communications sector and presents chances for major reforms (Simpson, 2004). Currently, the Information and Communication Technologies (ICT) sub-sectors are in a process of technological convergence and the key determining factors in this process are the liberalization of the telecommunications markets and technological change (Bore’s, Saurina and Torres, 2003).

The research conducted by Gillwald (2005) reports that the international reform agenda has been very unevenly and expediently applied in many developing countries, usually focusing on privatisation at the expense of other reform drivers. The research observed that South Africa in practice, just like many other developing countries, its reform agenda prioritised privatisation as the mechanism that would most rapidly redress the imbalances experienced in provision of telecommunications services. In theory, privatisation was supposed to reform the incumbent, and to make it operate more effectively, so that connectivity would be improved through low-cost access to an expanding network. But the focus on this one reform lever happened at the expense of the other levers such as competition, which was needed for the reform machinery to work. The efficiencies brought by competition into the market were therefore not realized forcing prices to remain high for telecommunication users and consumers whose choices remained limited.

Therefore, there is need for ICT policy now to balance the reform agenda by ensuring that the telecommunications market is structured in a manner that minimizes regulatory complexity, allowing the regulator to focus on measures to induce investment in network roll-out, encourage services innovation, improve consumers’ choice and service quality, develop market efficiencies, and effectively target subsidies to those who most need them. Policy must enable fair competition that will drive down costs, so that services become more widely affordable. The increased demand will give operators the economic incentives to expand the coverage of their networks and services.

Fan (2005) reported in the study on OECD countries that where competition was most advanced, there was highly developed Internet access. The research also observed that the extent to which ISP and telecommunication markets were open was closely connected to allowing foreign investment in infrastructure. Taking a particular case of one of the OECD countries, in this case Australia suggests strongly that foreign involvement in the telecommunications and Internet sectors is very positive for Internet access expansion (Fan, 2005). This calls for strong policy considerations for enhancing competition through adequate opening up of telecommunications and Internet sectors to spur information infrastructure development. It has been suggested that without appropriate development of information infrastructure, the disparities already experienced by rural and remote communities will be further exacerbated as the reliance of goods and services over computer-mediated networks increases. (Bandias and Vemuri, 2005)
Policy-makers around the globe are putting in place statutes designed to foster knowledge-based economic activity. The two broad global strategies witnessed across both G7 and developing economies are as follows: Investment in telecommunications systems and bandwidth, where information is viewed as a resource for improving commercial and industrial competitiveness and productivity (Grantham and Tsekouras, 2004). In the case of Africa, it has been demonstrated that wireless ICTs have the unique ability to circumvent the limitations posed by under-investment in fibre optic networks in less prosperous localities and national territories due to lower infrastructure costs. A number of mobile businesses such as Paybox (http://www.paybox.co.uk) have enabled owners of any networked mobile phone—who also has a bank account—to make secure electronic payments to vendors or individuals. The same is not true of wired electronic payments. Developing countries, particularly the Sub-Saharan Africa, should develop policies that create incentives to financial institutions and mobile service providers to collaborate and offer such services e.g. payment of water bills, electricity bills, parking bills. This could encourage the development of more mobile phone applications including government service applications. This is because, despite the high cost of mobile services compared to the fix services, the convenience and flexibility of pre-paid mobile services have clearly spurred their adoption on a massive scale, especially in Africa (Gillwald, 2005). In addition, Oyelaran-Oyeyinka and Lal (2005) argue that, first; countries need to adjust their telecommunication and economic policies to promote public as well as private investments in ICT that in turn might further boost economic growth and, secondly; governments in developing countries need to encourage the use of PCs through policies that enable soft loans on purchase of computers for individuals and academic institutions. Such policies should not only address PCs, but also mobile phones so as to further broaden and deepen penetration.

In Japan, user needs have driven developments of the mobile Internet such as “i-mode” that is accessed via mobile phones (Kenichi, 2004). Mobile Internet systems allow for a short message service, email, web browsing, and additional advanced services such as picture mail. One of the technological reasons for the extraordinary success of i-mode is that NTT DoCoMo adopted Compact HTML (C-HTML) as the language for i-mode websites, instead of the more standard Wireless Application Protocol (WAP). C-HTML is a compatible subset of HTML for terminals. NTT DoCoMo’s i-mode enables content providers to more easily enter the market than WAP, because it is easier to create i-mode websites in C-HTML than in WAP. Kenichi (2004) explained that the history of the mobile Internet in Japan showed that user needs had promoted the mobile Internet in Japan more than anything else. The Japanese government policy had emphasized technological development of mobile phone systems such as PHS and IMT-2000 without any regard or attention to user needs, hence missed out i-mode that focused on services rather than technology. Investigations in Japan found that mobile media users were more active in personal communications. The experiences in Japan revealed that user needs deserves top attention. In fact the Japanese experience after 1995 demonstrates that user needs have brought about the high penetration rate and unique usage patterns. In general, ICT policies should strive to articulate the needs of the people and call for appropriate applications to be developed to address the needs.
The need for preserving digital information and making it accessible for the present and future is expected to be captured by all nations, however, research has revealed that Sub-Saharan Africa (SSA) countries are ill equipped in this direction, if any (Ngulube, 2004). In a study carried out in Botswana, Ghana, Kenya, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, Tanzania, Uganda and Zimbabwe by Ngulube (2004) discovered that the countries did not have the capacity to manage electronic records. Therefore it is evident that most countries in SSA seem not to be seriously addressing the issues relating to the preservation of digital records and archives. This makes the need for inclusion of policies statements on the management of electronic records in national ICT policies worthy of attention and consideration.

The Economic and social transformation of nations has been brought about by technology. Both policy analysts and academic researchers have been interested in understanding and articulating these phenomena especially, when governments are constantly needing information about the performance of their own country relative to their partners and competitors. Several methods are currently in use including e-readiness indices. However, the criteria of generating the e-readiness indices are not uniform. For one to produce an e-readiness index, at least two indicators have to be summed, subtracted, multiplied or divided yet research has shown that of now there are no meaningful theories that justify such algorithms. To make it worse, the selection of the ingredients of the e-readiness index depends heavily on the value judgment of the scholars as well as on the availability of the data.

As it stands now, one of the best options for informing policy makers and governments is to use indicators. This paper discusses trends and analysis of ICT indicators such as mobile phone subscribers, mobile communications revenue, annual telecommunication investment, international bandwidth and mass media usage in relation to ICT policy, e-strategy and enabling environment for COMESA, EAC, EU and OECD based on the 2005 ITU and DHS databases. Correlation analysis among indicators was done in attempt to establish emerging commonalities and differences among these regions and the accompanying lessons noted.

In general this paper discusses trends of ICT indicators with a view of informing sub-regional and regional ICT policy. A period of ten years starting from 1993 to 2002 was studied for COMESA, EAC, EU and OECD. In some cases the full range could not be covered due to lack of data for some years, however, it was sufficient in helping propose meaningful and significant regional policy statements.

**Methodology**

This research is based on a positivist paradigm where it is believed that reality exists objectively and independently from human experiences (WenShin and Hirschheim, 2004). Hence quantifiable and measured ICT indicators such as GDP, mobile phone subscribers, total telephone subscribers, mobile communications revenue, annual telecommunication investment, international bandwidth and the usage of newspapers, radio and television here referred to as mass media usage (MMU), were studied and possible inferences made empirically based on secondary data obtained from ITU and DHS.
Using the ITU data, the annual average per indicator (AAI) for all the regions was calculated by summing the absolute values of all member countries and dividing by their total number. The normalized annual indicators (NAI) were obtained by dividing AAI by the mean population per year for the region. NAI was then multiplied with 100 to obtain the annual average of ICT indicator per 100 inhabitants whose details are given in appendix A. The correlation analysis was done using normalized ICT indicators per 100 inhabitants. For purposes of studying general trends and comparison of ICT indicators, the absolute values of the indicators were plotted, interpreted and validated using relevant literature review on ICT policy, strategy and regulation. DHS data was used to study MMU trends in EAC sub-region using percentage values of selected indicators.

**ICT Indicators Trend Informing Policy, Strategy And Regulation**

The trend in bandwidth with time for OECD countries indicates that bandwidth was growing with time as shown in Fig.1. This confirms the remarks made by Grandtham and Tsekouras (2004) that most developed countries are strategizing for knowledge-based economy through investing in telecommunications systems and bandwidth. In the OECD countries the broadband Internet service is defined as 256 kbps downstream and 64 kbps upstream (Wu, 2004). Therefore all the OECD countries have sufficient broadband that enable innovation and evolution of advanced communication services and overall economic growth in accordance with the school of thought of Gillet et al (2004). For such a trend to be observed, it means OECD countries have made broadband a priority (Frieden, 2005) so as to guarantee the future of Internet in those countries. Broadband has been known to have great potential of improving quality of education and health services; connectedness of government with society; and provision of jobs and prosperity (Firth and Mellor, 2005).

The future of the Internet has been perceived to be strongly dependant on broadband and hence the industry groups and analysts have stressed the importance of broadband access for continuing the evolution of advanced communication services and overall economic growth (Gillet et al, 2004). Though on average, COMESA countries have broadband as shown in Fig.1, there is need for it to be given national priority in order for it to develop enough to support national services and economic growth. It has been noted that broadband development thrives when it becomes a national priority (Frieden, 2005).
It is worthy noting that OECD countries have allowed competition to advance, which could be responsible for the high development of Internet access (Fan, 2005). For instance, Australia has a policy that allows direct foreign investment and involvement in telecommunications and Internet. This has had a very positive effect on Internet access development.

Therefore, one way of realizing sufficient broadband in COMESA region that can support e-applications including those identified and recommended by COMESA such as ASYCUDA, CPIS and REPSS, is to pursue a strong policy considerations for enhancing competition through adequate opening up of telecommunications and Internet sectors to spur information infrastructure development. Without appropriate development of information infrastructure, the disparities already experienced by rural and remote communities will be further exacerbated as the reliance of goods and services over computer-mediated networks increases (Bandias and Vemuri, 2005). In fact, policy-makers around the globe are encouraging investing in telecommunications systems and bandwidth as a major strategy towards building a knowledge-based economy (Grantham and Tsekouras, 2004). Firth and Mellor (2005) do support the idea of COMESA going for broadband since it has the potential to offer nations improved quality of education and health services, improved connectedness of government with society, and provision of jobs and prosperity.

How do we explain for some similarity in the trends of some indicators of COMESA and OECD when we know that most COMESA countries do not have even a national
ICT policy and/or the regulators lack capacity to perform as expected? It is true that the OECD countries are reaping the benefits of good ICT policies and strategies combined with a suitable enabling environment. Nevertheless, the case of most COMESA countries can be explained using the Japanese experience with the i-mode for mobile Internet (Kenichi, 2004). This experience revealed that neither technological advantages nor ICT policy is strong enough to initiate a new type of telecommunication service but the user needs can do. Among the COMESA countries, mobile services are expensive as compared to fix services yet because of need, convenience and flexibility it has been adopted on a massive scale (Gillwald, 2005). Which means, the need comes first then policy and technology can follow for effective and sustasinable adoption and application of ICT for development. Just as is argued in the case for OECD countries, the mobile phone adoption in COMESA countries is not dependant on GDP. Better results could be achieved if COMESA countries articulated to the public and private sector ICT policies that are founded on the needs of their people. Especially coherent national and sub-regional ICT policies and strategies would support the introduction of new regulatory frameworks that promote production and use of ICT including the marginalized people. The idea of adoption of a common ICT policy in COMESA is necessary in order to support the creation of an enabling environment through harmonization and integration of regulatory regimes and adoption of a holistic e-strategy. Critically speaking, it is not necessarily the harmonization of regional ICT policy framework that is limiting the ability of the region to attract investment in regional and national ICT operations as stated in the ICT policies for COMESA document, but also the legal and regulatory framework could be contributing.

Introduction of wireless telecommunications that are converging with mobile computing devices would offer Internet Connectivity and access to most COMESA countries than through the traditional PC/modem, hence promoting Internet usage in COMESA countries. This argument is supported by the trend observed of increase in mobile phone subscribers (see Fig. 2). Though the number of personal Computers shown in Fig. 2 below also suggests an increasing trend, it is not rapid enough as that of mobile subscribers. In 1998, most countries had on average 20,000 mobile subscribers and the number of PCs on average was about 160,000. By the year 2002, the number of mobile subscribers was about 420,000 while that of PCs was about 140,000 which, meant a ratio of mobile subscribers to PC of 6:1. This further confirms that in Africa wireless ICTs have the unique ability to circumvent the limitations posed by under-investment in wired networks.
Levi (1998) identified the telephone as the future channel of communication in Africa on the basis of the fact that the central element of the African mode of communication was oral. In Africa, telephone can be used to promote oral communication as well as sustain kinship relationships. In view of this observation and the onset of new technologies, EAC sub-region in Africa has been selected to study the trends in mass media usage in order to suggest how it can inform ICT policy. Kuyvenhoven (2004) argues that new and low-cost communication technologies (ICT) can play a major role in speeding-up the process of information diffusion and improving market efficiency, especially in the fields of input provision and marketing outlets. It has also been observed that rapid changes in information and communication technologies have lowered costs and opened up new opportunities in print media, Internet, radio and television. The impact of these rapid changes need to be harnessed in relation to its ability to shape political, economic, and social landscapes (Andriantsoa, 2005).

**EAC case of MMU**

According to data obtained from ITU (2005), the percentage households with televisions in EAC countries by the year 2002 was on average approximately 10%, whereas that of radios over the same period was on average approximately 60%.

Considering a ten year period (1993-2003), the trend observed using DHS data (2005) show that the urban MMU and the rural MMU in Kenya by the year 2003, more than 80% of both urban and rural Kenyans were reading newspapers and watching...
television. Listening to radio seemed to have been on decline so that by the year 2003, less than 50% of the people were listening to the radio weekly. A declining trend in use of radio noted could be attributed to the increased affordability of TVs, which are a better channel of communication since they appeal to both visual and audio.

Basing on DHS data (2005) MMU in Tanzania was observed to be generally poor. The urban area recorded on average of about 30% of its people using mass media. This case was worse off in the rural where approximate MMU fell to 15% of the population.

In the case of Uganda, the DHS data (2005) revealed that less than 50% of Ugandans were reading newspapers and watching TV over the period ranging 1995 to 2000/01. Only the radio had more than 70% of the population using it. The situation is worse in the rural where approximately 10% of the population were using mass media.

Andriantsoa (2005) argue that in order to have a vibrant and competitive media establishment several important public investments are needed such as:

- A legislative framework favoring and protecting free speech
- A system for allocating broadcast rights and transmission bands for radio and television in order to ensure both order and quality of transmission over the airwaves and;
- Investments in training and professional development for journalists.

Underdevelopment in the above listed areas could be responsible for the low usage of mass media observed in EAC countries.

Though there is pressure and need to embrace new technologies, it is important to realize that even the old technologies such as radio and TV have not been exhaustively used as seen in the case of EAC. For instance, AM radio is one medium that can be a carrier of information and knowledge to the people over a wider region and can help in the preservation and exchange of local knowledge (Rodrigues and Wafula, 2004). Such technology has not been used exhaustively in COMESA and EAC. More so in the area of passing relevant information to communities in the form and languages they understand. By choosing to limit such technologies, communities have been denied the opportunity to coordinate their production, distribution and consumption of what they produce. Entrepreneurs have always been there but due to political reasons and fear, they have not been fully allowed to provide these services. Particularly in Africa, people live in communities with already established mass cultures. Therefore ICT policies and strategies should enable radio and TV or any technology targeting masses to take advantage of this and effectively in help in alleviating poverty through provision of timely and relevant information.

Fears have been registered in relation to difficulties experienced in controlling mass media and preservation of culture with the advent of new technologies. For that matter, a culture runs the risk of losing its footing in the surging flood of information sweeping the world if an effort is not made to place its products in the global market. Georgette (1997) defines culture as a continuously evolving entity that closely interacts with its
environment. Therefore there is need for ICT policies to support development of culture-based products as a way of preserving and sharing it.

**GDP and ICT Indicators**

Correlation Analysis Results. Correlation analysis was done between averaged values of selected indicators in COMESA, EAC and EU blocs to establish their level of interdependence. The selected results applied in this paper are as shown in Table 1 below. The details of how the correlations were arrived at are given in appendix A.

**Table 1: Correlation between averaged values of selected ICT Indicators normalized per 100 Inhabitants**

<table>
<thead>
<tr>
<th></th>
<th>Gdp Us$</th>
<th>Internet Hosts</th>
<th>Total Telephone Subscribers</th>
<th>Mobile Subscribers</th>
<th>Personal Computers</th>
<th>Internet Users</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EU-GDP (US$)</strong></td>
<td>1</td>
<td>0.367</td>
<td>0.281</td>
<td>0.245</td>
<td>0.410</td>
<td>0.288</td>
</tr>
<tr>
<td><strong>COMESA-GDP (US$)</strong></td>
<td>1</td>
<td>-0.544</td>
<td>-0.497</td>
<td>-0.330</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>EAC-GDP (US$)</strong></td>
<td>1</td>
<td>0.501</td>
<td>0.423</td>
<td>0.380</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Correlation Results Interpretation and Validation. According to the correlation analysis done using SAS Software v8 and tabulated in Tables 1 above, GDP generally appear to have a weak relation with ICT Indicators such as Internet Users, Mobile Subscribers, Total Internet Hosts, Total Telephone Subscribers and Personal Computers. In all these cases, a maximum and minimum value of correlation of 0.501 and –0.544 respectively were obtained. Therefore the observed changes in these indicators need to be attributed to other factors other than GDP. This paper is particularly interested in identifying the role of policy, strategy and regulations factors.

The trends in the annual mean GDP across COMESA, EAC, EU and OECD shown in Fig.3 suggest negligible changes in GDP with time over a ten year period. The sudden change in mean GDP in COMESA between 1994 and 1995 was due to enormous drop in GDP of Angola. It can be observed from Fig.3 that the mean GDP of EU and OECD are close and way above that of COMESA and EAC by magnitudes of the order $10^3$. The trends are parallel with little hope of COMESA and EAC approaching GDP of EU and OECD. Given the EU and OECD are advanced in ICT adoption and usage, it remains to be seen if the strong GDP position is a key factor. Norris (2000) observes for instance that, once a country rises above the US$ 9000 per capita GDP
mark, the online population expands exponentially. According to this rule, Fig. 4 clearly shows that COMESA and EAC fall below this threshold while EU and OECD are way above it. However, this may not necessarily be the case for all region, therefore other contributing factors need to be identified.

**Figure 3.** Trends in annual mean GDP for COMESA, EAC, EU and OECD regions. Data points Source: ITU. (2005). World Telecommunications Indicators Database. 8th Edition.

**Figure 4.** Average GDP per capita trend in COMESA, EAC, EU and OECD
In the case of the mean total telephone subscribers (TTS) per 100 inhabitants, an increasing trend is observed (see Fig.5). This trend suggests that the TTS per 100 inhabitants was not dependent on GDP. According to Norris (2000), this observation cannot be explained on the basis of GDP per capita for COMESA and EAC. This is confirmed by the correlation value of -0.497 and 0.423 observed between GDP and TTS (see Table 1) for COMESA and EAC respectively.

Figure 5. Mean Total telephone Subscribers per 100 Inhabitants trends in COMESA EAC, EU and OECD. Data points Source: ITU. (2005). World Telecommunications Indicator database. 8th Edition.

Despite the prediction by Norris (2003) that countries with GDP per capita of less than US$ 9000 are not expected to adopt ICT at high rate, Fig.5 shows that COMESA and EAC are doing so and at a higher rate than EU and OECD. This observation can be largely attributed to communication needs that for too long had not been met due to poor infrastructure development. With the advent of new technology, remote and rural access has been made possible. This further gives room for ICT policy, strategies and regulations to enhance and fasten towards meeting such needs. This is something COMESA and EAC have the potential to implement within a short time, which can in-turn cause the GDP per capita to grow against the hypothesis of Norris (2000). Figure 6 shows a significant growth rate in the average total telecommunication service revenue in COMESA and EAC as compared to EU and OECD, which is a good indicator in support of this view.
Proposed Regional Policy Statements

The observed trends in ICT indicators of COMESA, EAC, EU and OECD articulated in this paper shade new insights as well as inform about the kind of generic regional ICT policy statements that ought to be in place. The generic regional ICT policy statements proposed can be grouped in eleven categories as follows:

Strategic ICT leadership

i) Encourage participation in regional and global governance of ICT

ii) Encourage use of ICT to mainstream gender and other empowerment issues

iii) Encourage collaboration with other countries on regional projects

iv) Support enhancement of capacity for research and development in ICT

v) Encourage government leadership in incubation of new ICT applications.

vi) Encourage government leadership in articulation of ICT vision to the nation and its citizen

vii) Support government encouragement of communities to propose specific bottom up projects that aggregates the supply of services needed through local government in order to create a demand for information infrastructure development.
viii) Encourage CBOs and local authorities to own ICT initiatives so as to attain the necessary critical mass needed to spur socio-economic development.

ix) Encourage discussion on the nature of power relations between and among all stakeholders in order to stimulate and enable participatory policy making, which is regarded as a best practice ingredient for better policies.

**Human Capital**

i) Promote Mainstreaming of ICT in the education system.

ii) Encourage development of enumeration and incentive packages for skilled ICT personnel in order to minimize high turnover.

iii) Support enhancement of capacity for research and development in ICT.

iv) Encourage institutions of higher learning to undertake R&D activities in collaboration with telecommunications service providers and manufacturers.

v) Encourage ICT policy stakeholders to base their contribution in the policy making process on research findings so that they understand their roles, limits and responsibilities.

**Institutional Framework**

i) Stimulate ICT innovation.

ii) Support the development of appropriate mechanism for coordination and implementation of ICT policy.

iii) Promote institutional rationalization for purposes of co-coordinating ICT policy implementation.

**Legal and Regulatory Framework**

i) Promote balanced telecommunication reforms in privatization and liberalization through creation of simple and explicit regulatory systems.

ii) Encourage open competition in order to attract Foreign Direct Investment (FDI).

iii) Promote establishment of technologically neutral framework for ICT licensing.

iv) Encourage national organizations that have rights of way to contribute in the development of the national information infrastructure.

v) Promote duty free zones to attract ICT investment.

vi) Promote publication of the rights and obligations of ICT consumers.

vii) Encourage development of a national standard interconnection model that specifically address the issue of Reference Interconnection Offer (RIO) for major operators as well as issues of universal services obligations.
viii) Encourage publication of information and mechanisms of administration and management of domain names.

Information

i) Support establishment of the office of Information Commissioner and its functions.

ii) Support the public to obtain access to the greatest extent possible consistent with the public interest and the right to privacy, to information in the possession of government, public bodies and specified private bodies.

iii) Encourage promotion of national culture and identity in media.

iv) Encourage stakeholders and development partners to support creation of local content in order to preserve the knowledge and culture of traditional communities.

v) Promote sharing of information and data through provision of timely and quality statistics.

Information Technology Services

i) Encourage adoption of E-applications.

ii) Support preservation of digital records and archives for future referencing and posterity.

iii) Encourage public and private sectors to develop and deploy Open Source software.

iv) Offer incentives for individuals to own PCs and mobile phones.

v) Encourage development of e-applications that are a shared vision among all stakeholders.

vi) Develop e-applications that address social, economic, political and cultural needs of the time.

Telecommunications Services

i) Encourage investing in telecommunication systems and broad bandwidth.

ii) Promote utilization of all installed ICT infrastructure to be optimally utilized and synchronized.

iii) Promote direct interconnectivity between mobile cellular service providers and other service providers including sharing of infrastructure.

iv) Encourage development of National Geospatial Data Infrastructure.

v) Promote development of uniform numbering schemes in order to simplify interfacing and Interconnection.

vi) Broadcasting
vii) Promote development of broadcasting legislation that cover policy advisory and dispute resolution for the sector as well as take into consideration the overlaps due technological convergence and integration.

viii) Support the protection of the public, especially minors, from unregulated pornographic and violent programming in the broadcast media.

ix) Promote broadcasting which is for, by and about specific geographical communities or communities of interest, whose ownership and management are representative of those communities that pursue a social development agenda and which is not-for-profit.

x) Support establishment of commercial broadcasters that contribute to the promotion of culture and empowerment of the poor and vulnerable groups in society while remaining commercially viable.

xi) Encourage broadcasters to play a crucial role in providing a level playing field in the electronic media for all political actors so as to promote diversity, good governance, human rights and democracy.

xii) Encourage gradual transformation from analogue to digital broadcasting.

xiii) Support the conversion of studio production and communication technologies from analogue to digital and the development of necessary capacity to operate as a digital broadcaster.

xiv) Support streamlining of the operations of the film and music industries and promote local production and talent.

xv) Stimulate the growth of advertising industry as a major source of income for the broadcasting media.

**Radio Frequency Spectrum**

i) Promote frequency allocation based on international standards while keeping in mind public interest objectives.

**Postal Services**

i) Encourage postal corporations to have a Universal Service Obligation in the provision of equitable access to quality and efficient postal services.

**Universal Access**

i) Encourage establishment of a Universal Service Fund.

ii) Encourage annual allocation of funds equivalent to a reasonable proportion of nation’s GDP for ICT deployment, diffusion and universal access in partnership with the private sectors and development partners.

iii) Encourage ICT operators to have social obligations.
Conclusion

How do we explain for some similarity in the trends of some indicators of COMESA and OECD when we know that most COMESA countries do not have even a national ICT policy and/or the regulators lack capacity to perform as expected? It is true that the OECD countries are reaping the benefits of good ICT policies and strategies combined with a suitable enabling environment. Nevertheless, the case of most COMESA countries can be explained using the Japanese experience with the i-mode for mobile Internet (Kenichi, 2004). This experience revealed that neither technological advantages nor ICT policy is strong enough to initiate a new type of telecommunication service but the user needs can do. Among the COMESA countries, mobile services are expensive as compared to fix services yet because of need, convenience and flexibility it has been adopted on a massive scale (Gillwald, 2005). Which means, the need comes first then policy and technology can follow for effective and sustasinable adoption and application of ICT for development. Just as is argued in the case for OECD countries, the mobile phone adoption in COMESA countries is not dependant on GDP. Better results could be achieved if COMESA countries articulated to the public and private sector ICT policies that are founded on the needs of their people. Especially coherent national and sub-regional ICT policies and strategies would support the introduction of new regulatory frameworks that promote production and use of ICT including the marginalized people.

Introduction of wireless telecommunications that are converging with mobile computing devices would offer Internet Connectivity and access to most COMESA countries than through the traditional PC/modem, hence promoting Internet usage in COMESA countries. This argument is supported by the trend observed of increase in mobile phone subscribers (see Fig. 2).

Levi (1998) identified the telephone as the future channel of communication in Africa on the basis of the fact that the central element of the African mode of communication was oral. In Africa, telephone can be used to promote oral communication as well as sustain kinship relationships.

Though there is pressure and need to embrace new technologies, it is important to realize that even the old technologies such as radio and TV have not been exhaustively used as seen in the case of EAC. For instance, AM radio is one medium that can be a carrier of information and knowledge to the people over a wider region and can help in the preservation and exchange of local knowledge (Rodrigues and Wafula, 2004). Such technology has not been used exhaustively in COMESA and EAC. More so in the area of passing relevant information to communities in the form and languages they understand.

Fears have been registered in relation to difficulties experienced in controlling mass media and preservation of culture with the advent of new technologies. For that matter, a culture runs the risk of losing its footing in the surging flood of information sweeping the world if an effort is not made to place its products in the global market. Therefore there is need for ICT policies to support development of culture-based products as a way of preserving and sharing it.
The trend in bandwidth with time for OECD countries indicates that bandwidth was growing with time as shown in Fig.1. This confirms the remarks made by Grandtham and Tsekouras (2004) that most developed countries are strategizing for knowledge-based economy through investing in telecommunications systems and bandwidth.

Though on average, COMESA countries have broadband as shown in Fig.1, there is need for it to be given national priority in order for it to develop enough to support national services and economic growth.

Despite the prediction by Norris (2003) that countries with GDP per capita of less than US$ 9000 are not expected to adopt ICT at high rate, Fig.5 shows that COMESA and EAC are doing so and at a higher rate than EU and OECD. This observation can be largely attributed to communication needs that for too long had not been met due to poor infrastructure development. With the advent of new technology, remote and rural access has been made possible. This further gives room for ICT policy, strategies and regulations to enhance and fasten towards meeting such needs.

According to the correlation analysis done using SAS Software v8 and tabulated in Tables 1, GDP generally appear to have a weak relation with ICT Indicators such as Internet Users, Mobile Subscribers, Total Internet Hosts, Total Telephone Subscribers and Personal Computers. In all these cases, a maximum and minimum value of correlation of 0.501 and –0.544 respectively were obtained. Therefore the observed changes in these indicators was attributed to other factors other than GDP.

The trends in the annual mean GDP across COMESA, EAC, EU and OECD shown in Fig.3 suggested negligible changes in GDP with time over a ten year period. It can be observed from Fig.3 that the mean GDP of EU and OECD are close and way above that of COMESA and EAC by magnitudes of the order $10^3$. The trends are parallel with little hope of COMESA and EAC approaching GDP of EU and OECD. Given that the EU and OECD are advanced in ICT adoption and usage, it emerged that the strong GDP position was not a key factor. Norris (2000) observes for instance that, once a country rises above the US$ 9000 per capita GDP mark, the online population expands exponentially. According to this rule, Fig.4 clearly shows that COMESA and EAC fall below this threshold while EU and OECD were far above it. However, it was noted that other factors mainly need and others such as policy, strategy and regulation contribute.

References


PART FIVE

E-Commerce
Bridging Africa’s Digital Divide: Building Sustainable ICT Infrastructures

Joseph M. Kizza, Department of Computer Science, The University of Tennessee-Chattanooga, TN 37403

Introduction

Computers in particular and information communication technology in general have, in the last fifty years, driven development in every country. There is documented evidence that computers and information technology indeed speed up and enhance development. However, the rate of growth has been uneven. In industrialized rich countries where resources and capacity are abundant, such growth and the rate of growth has been phenomenal. Developing countries, however, have not been so lucky to benefit substantially. There are observed and documented complex problems that need to be overcome before any meaningful development can be registered.

The papers in this proceedings are all focusing on the theme of ICT and development. “The Second International Conference on Sustainable ICT Capacity in Developing Countries – SREC06” out of which this proceedings grew was a platform and a forum where ICT-based development was discussed over a course of four days. Papers were presented in four tracks: Computer Networks, Information Technology, Information Systems, and Computer Science and Software Engineering. There were seven keynote papers to headline the conference. Joseph M. Kizza in “Bridging Africa’s Digital Divide: Building Sustainable ICT Infrastructures” takes note of the changing fortunes of Africa with the acquisition of computing technologies, especially wireless, resulting from the miniaturization and the plummeting prices and how these will help Africa to quickly leapfrog into the 21st Century with rapid development not seen in generations. Janet Aisbett in “Information Quality and Information Systems Success” reviews key models of IS success in the context of the prevailing IS. She identifies information quality as a constant factor in the models, and poor content quality as a continuing contributor to perceptions of failure. She concludes that fundamental research into information quality is still needed, and provides an example of such investigations concerning new forms of representation. Dilip Patel in “An Organisational Climate Awareness Toolkit For Nurturing the Effectiveness of Team/Group Interactions” discusses the findings of a study of organisational climate issues in several commercial organizations and outlines an organisational awareness toolkit. Anthony J. Rodrigues in “ICTs In Developing Countries: Contexts, Challenges...
And Interventions” discusses the barriers that Africa must overcome including lack of: information coordination, coordination of physical connections and technical personnel cooperation. He notes that this is exacerbated by telecommunications monopolies and obsolete regulatory frameworks, limited involvement of research institutions in network building and diffusion, and language barriers. There is a need to lower the overall costs of creating a competitive workforce and to do this continuously on a 24/7 basis for people specially those in rural areas. In “A Fundamental View on the Act of Modeling”, H.A. (Erik) Proper and P. van Bommel, discuss the role of models and modeling in the information system development life-cycle as part of their ongoing research effort to better understand the act of modeling. They describe a formal framework by which the process of modeling can be regarded as involving the selection of more and more refined interpretations in terms of the underlying meta-model of the modeling language used. The resulting framework will be used to create a laboratory setup in which to consequently and closely study (and support) modeling processes. Lastly Andrew Vince in “Indexing a Discrete Global Grid” discusses a method for computer representation and manipulation of global data based on a multi-resolution subdivision of a regular polyhedra. In particular he considers the problem of efficiently indexing the cells of such a discrete global grid.

 Presenters in each of the four tracks discussed in depth the latest research and developments in the area protocols and best practices that are suitable for development and the how-to in the implementation of some of these notable solutions and best practices.

 In Computer Networks, five papers were presented. In “Multiagent Systems for Distributed Resource Allocation”, Eric Ayienda et al discuss the additional challenges introduced by wireless grids that are above and beyond those already existing in resources allocation in the wired grid environment. Since wireless grids are considered as complex systems with economies where multiple applications (consumers) compete for resources and services from network providers (suppliers), optimization techniques can be done on network QoS parameters to solve problems in wireless networks. In “On Network Measurement and Monitoring of end-to-end paths used for e-VLBI”, Julianne Sansa et al present and discuss the bottlenecks currently limiting the transfer speeds of astronomical data from radio telescopes across the world over high speed links to the central processing centre in the Netherlands. The processed data is used to produce images which are used by radio astronomers. They image sensitivity of continuum observations scales with the data rate, so the higher the data rate, the more sensitive these observations will be. They observed that while the required high data rates should be attainable on these network links, however it is not the case. So they discuss how to uncover the prevailing bottlenecks and propose solutions to address them. Narcis T. Rwangoga and Martin Ngoye in “The Future of Intelligent Networks in Developing Countries” discuss the growing demand for network based services in developing countries and how wireless technologies for cellular and personal communications have extended these network services to most areas in developing countries. They explore how Intelligent Networks (IN) is a promise that countries
which are implementing telecommunication networks can use to deliver network based services. Habibu Atib and John Zeleznikow in “Identifying Sensitive Knowledge for Law Enforcement Agencies” take a had look at how Law Enforcement and Intelligence agencies can effectively manage knowledge to further their goals of crime detection and prevention. Techniques such as computer profiling, link analysis and Knowledge Discovery from Databases are discussed. Finally Kasigwa, Baryamureeba and Williams in “An Efficient Dynamic Admission Control for End-to-End Delay Guarantees in IP Networks” describe and classify a broad set of proposed admission control algorithms, evaluate the accuracy of these algorithms via experiments using both on-off sources and long traces of compressed video, compare the admissible regions and QoS parameters predicted by they implementations of the algorithms with those obtained from trace-driven simulations, and identify the key aspects of an admission control algorithm necessary for achieving a high degree of accuracy and end-to-end QoS of a computer network.

In Information Technology, five papers were presented. Aramanzan Madanda in “ICT Liberalisation and Changing Gender Relations in Contemporary Uganda: A Research Agenda”, identifies research needs in the area of gender and ICT in Uganda, especially in relation to adoption/non adoption and changing gender relations at individual and community levels. Madanda takes a gender perspective in this discussion of the current Ugandan ICT policy framework and ICT diffusion and adoption, drawing on arguments advanced by liberal economic theory of technology diffusion and adoption and feminist theory to illuminate key gender and ICT research issues in Uganda. Elijah Omwenga in “ICT for Educational Use: Cases of Implementing a Partnership Approach Model for Reaching Target Groups” discusses how institutions of higher learning can use ICT to further their mission. Two phases to archive this are advanced: phase one of awareness and advocacy through seminars and workshops, sensitisation on the relationship between ICTs and wealth creation and training of stakeholders and trainers; phase two of taking the programmes to the people – the students and target groups. Omwenga also proposes a partnership model that can be applied to help realise the goals of effectively reaching the target groups. The model recognises a three-stage pyramidal communication model involving target groups and partners, Networking and Collaboration and finally policy strategising on thematic areas. Joseph Wafula et al in “Informing Regional ICT Policy: Case Study of Trends in ICT Indicators of OECD, EU, COMESA and EAC”, discuss trends and analysis of ICT indicators such as mobile phone subscribers, mobile communications revenue, annual telecommunication investment, international bandwidth and mass media usage in relation to ICT policy, e-strategy and enabling environment for OECD, EU, COMESA and EAC based on the 2005 ITU and DHS databases. Correlation analysis among indicators is also done in an attempt to establish emerging commonalities and differences among these regions and the accompanying lessons. Asifwé Rubanju in “The Impact of ICT on Universities: Classroom/Lecture Theatre Design and Curriculum Delivery”, describes a research done to determine the impact of Information and communication technologies on class/lecture room design and curriculum delivery. Finally P. J. Ogao in “Beyond the
Computer Graphics Course: A Case for African Universities”, proposes a knowledge continuity strategy to the computer graphics course by outlining a visualization course content that supports science and society.

In Information Systems, ten papers were presented. In “Information and Communication Technologies (ICTs) in Small and Medium Enterprises (SMEs): Findings From Uganda” reports findings from a research carried out on 351 SMEs in Uganda with a varying degree of formality to help understand the potential of ICTs on SMEs, by understanding the entrepreneurs behind the SMEs, their information practices, the needs that dictate the way they operate as well as the environments in which they operate. Benedict Oyo and Ddembe Williams in “An Exploration of the Factors that Affect Quality Assurance Decision Making in Higher Education”, argue that system dynamics modelling approach is a valuable alternative for quality assurance research in higher education. This approach is emphasized owing to the fact that, a greater portion of university quality problems and their solutions do not have quantitative foundations mostly because such systems involve qualitative elements that are difficult to quantify and model using other approaches. A framework is used to demonstrate the theoretical value of system dynamics modelling in combining diverse factors responsible for quality assurance in higher education and arriving at a consistent decision about the quality of awards. Euphraith Masinde in “Using JAD to Bridge the Design-Reality Gaps: A Major Cause of IS Projects’ Failures in the Developing Countries” explains how Joint Application Development (JAD), a software development methodology that will involve the stakeholders in the entire process of IS implementation, can be used to eradicate most of the causes of IS projects’ failures in developing countries using the University of Nairobi case study. The CHAOS Ten Success factors have been employed to analyze data for nine IS projects. In “Using Developmental Research to Design a Web Instructional Design Subsystem”, J. K. Njenga and A.J. Bytheway report on a process and findings of using development research design and methodology to design an instructional design subsystem. The development research used involves a review of the literature with the aim of exploring, analyzing, integrating and synthesizing the broad field of learning and instructional theories, paradigms and best practices with the design of the instructional design subsystem. They also discuss how the subsystem can be improved by rigorous attention to the formulation of theory. Rehema Baguma et al in “Towards Web Design Frameworks (WDFs)” discuss Web Information Systems, the increasing rate of internet/web usage calls for efficiency and effectiveness in the development and deployment of web systems to provide high quality systems in the shortest time possible, the concept of web design frameworks, and their importance as a formal method of web application development. Ddembe Williams in “Revisiting Dynamic Synthesis Methodology: A reference theoretical framework and tool for business process modelling and analysis”, provides a useful and systematic reference point for researchers who wish to work in the business process modelling and generally to encourage careful work on the conceptualisation and execution of Dynamic Synthesis methodology in business process and to a wider process-modelling field. Ddembe also suggests that the
potential usefulness of the Dynamic Synthesis Methodology is in aiding researchers and managers in improving both building and testing theories in understanding business processes and strategic modelling and analysis. Gilbert Maiga and Ddembe Williams in “Towards a Reusable Ontology Framework for Biological Information Integration”, provide the research background and approach to an ongoing study that aims to develop a reusable framework for the integration of biological and clinical research data. A theoretical basis for the reusable framework is given and the proposed approach for developing and validation of the framework using the Protégé ontology development environment is also outlined. In “Survey of Data Mining Methods for Crime Analysis and Visualization”, Okwangale Fredrick R and Ogao Patrick review the applicability of various data mining methods and Geographic Information Systems in crime analysis and visualization in mainly poor planned settings characterised by missing electronic data, a common phenomena in the developing countries like Uganda focusing on criminality of places rather than the tracing of individual criminals. And finally Ssemaluulu Paul and Williams Ddembe in “A System Dynamics Tool for Evaluating IT Investment Projects”, investigate five different methodologies taking into account the suitability or goodness of the framework, bias, focus and complexity. The simulation tool was used to analyze how different variables interact to affect the total benefits of an information system. It was observed that only a strong interaction of people, information, and technology can improve business performance, and consequently lead to Information Systems success.

In Computer Science and Software Engineering, seven papers were presented. Venansius Baryamureeba in “On Solving Large Scale Linear Systems Arising From Interior Point Methods for Linear Programming” proposes a strategy for solving the linear systems arising from interior point methods for linear programming. Further, Baryamureeba proposes how to construct a preconditioner for the iterative approach for solving linear systems. In “Architectural Building Blocks For Reusable Service Oriented Architecture With Integrated Transaction Processing”, Benjamin Kanagwa proposes a generic high level architecture that specifies the structural elements for transaction centered service oriented architecture. Elisha T. O. Opiyo et al in “Multi-Agent systems Scheduler and Open Grid Computing Standards”, propose a scheduling model based on multi-agent systems and review market based approaches to the resource allocation problem for decentralized contexts. They interestingly note that increased reliance on agents and Web-services may one day lead to a situation where the grid computer is the Internet and the Internet is the grid computer. Peter Waiganjo Wagacha and Dennis Chege in “Adaptive and Optimisation Predictive Text Entry for Short Message Service (SMS)” describe the development of a generic mobile phone predictive text application they experimented with using local Kenyan languages. They also present an adaptive learning model for improving text entry speed that incorporates a specific user’s word usage habits. In “Kinyarwanda Speech Recognition for Automatic Voice Dialing Systems”, Jackson Muhirwe describes the design and training of a Kinyarwanda language speech recognition system that could be used by developers to create applications that will help those who want to use and speak Kinyarwanda. Katherine W.
Getao and Evans K. Miriti in “Computational Modelling in Bantu Language” propose a derivational approach to discovering the structure of compound Bantu words that does not depend on large amounts of prior knowledge. This technique is amenable to machine learning methods such as reinforcement learning. Through this they demonstrate a new approach to natural language processing that may be of specific application to Bantu languages and to other language groups with a strong inflectional strategy. Finally in “The Influence of Job Physical Characteristics on Their Schedulability in Multi-cluster Systems”, John Ngubiri and Mario van Vliet investigate the influence of jobs’ physical characteristics on their schedulability in a multi cluster system using the Fit Processor First Served (FPFS) scheduler and their relative strength in determining schedulability as well as their sensitivity to scheduler parameters.

This proceedings will be a valuable teaching and research tool for scholars, students and practitioners of ICT-based development as well as an enriching experience for all interested in ICT-related issues.
An Organisational Climate Awareness Toolkit For Nurturing The Effectiveness of Team/Group Interactions

Shushma Patel and Professor Dilip Patel, shushma, dilip}@lsbu.ac.uk, BCIM, London South Bank University, 103 Borough Road, London SE1 0AA, UK

The development of innovative systems solutions for complex problems remains a challenge. To be able to develop modern information systems, we must include the use of multi-disciplinary teams, such as technical and non-technical specialists. Team members must work together and IS professionals can no longer work independently to design their own information system, but instead need to carefully include the users in the process. One area that needs to be considered carefully is how the IS project teams are created. A major factor in successful IS solutions is awareness of the organisation climate. In this paper, we investigate organisation climate, and propose three levels of organisation climate and demonstrate that nurturing the organisational climate awareness helps increase and improve the effectiveness of team/group interactions. We discuss our findings of studying organisational climate issues in several commercial organisations. We also outline an organisational awareness toolkit.

1.0. Introduction

It is recognised that group situations confront the individual with considerable role problems and identity problems, which in turn cause problems for organisations. Handy (1979) identified the following problems: - The problem for organisations – individuals must be organised into groups in order to make most effective use of their mix of skills and abilities, but too much emphasis on the group may blunt the individual contribution; whilst too much emphasis on the autonomous individual may hinder the development of group identities. The problem for individuals – groups or families are desirable psychological homes. Without them individuals can become too egocentric to be effective in organisational situations and can deprive themselves of the richness of interpersonal relationships. To submerge oneself in a group too deep may involve the sacrifice of someone’s individuality.

By studying these two problem areas found in group interactions it is evident that there are three generic elements involved in organisational theory research: -

• the organisation as a whole,
• the teams/groups within the organisation,
• and the employees that work in the teams/groups for the organisation
In this paper, organisational climates are described and their impact on team performance is investigated. An overview of the organisational toolkit is presented and a summary of the results of the application of the toolkit on a range of companies are presented.

2.0. Organisational Climates

The interaction between the employees in the teams/groups affects the organisation’s overall aims and goals. There is clear and strong evidence that the climate of organisations predicts their subsequent performance and affects the performance of the individuals, groups, teams and departments within them. Michie and West (2002) suggest that innovative and effective organisations are places where members have a shared belief of an appealing vision of what the organisation is trying to achieve. They view organisations as a high level of interaction, discussion, constructive debate, and influence among the members as they go about their work. In turn this creates high levels of trust, cooperative orientations, and a sense of interpersonal safety characterize interpersonal and inter-group relationships. Within the context of organisational climate there are two well-established and defined levels of climate; psychological climate that focuses on the day-to-day feelings, attitudes and behaviours of the employees within the organisation (Hellreigel & Slocum, 1974; Howe, 1977), and organisational climate, which also focuses on the day-to-day feelings, attitudes and behaviours, but at the organisational level (Ekvall, 1991; Schneider, 1990 & 1994; James & Jones, 1989). However, by reviewing research and literature on the levels of climate, there is evidence of a third climate level (Powell & Butterfield, 1978), which, lies somewhere between the other two levels of climate, and refers to the day-to-day feelings, attitudes, and behaviours of the teams/groups found within an organisation. This climate level has been categorised or referred to as the following; team climate, unit climate or group climate (Arvidsson, Johansson & Akselsson, 2003; Salas, Staigt & Burke, 2003; Lad Burgin, 2001). In our research we define three climate levels and their boundaries within an organisation in terms of nurturing organisational climate awareness to improve the effectiveness of team/group interactions:

- Personal climate (refers and replaces psychological climate),
- Sub-climate (refers and replaces team climate, unit climate or group climate),
- Organisational climate.

In order to study organisational climates within the context of organisational theory, metaphors can be used to illustrate one concept with the attributes normally associated with another (Bruton-Simmonds, 2003). If we think of organisations metaphorically, we can dislodge our normal way of thinking about organisations and search for new ones. This research proposes a new organic organisational metaphor to help nurture organisational climate awareness. Table 1 shows the mapping between the proposed tomato plant organisational metaphor and proposed three levels of organisational climate.
People are not always conscious of the way they see things and the guiding metaphors (Belfer, 2002) can become bridges to support the exchange of ideas. Every metaphor is best suited to a particular purpose. Each metaphor provides us with a different perspective on information and communication.

3.0. Organisational Climate Awareness Toolkit

To assess the current levels of climate, we need to have a mechanism for extracting information. The toolkit proposed in this research is made up of a combination of existing organisational study tools that are normally used separately to analyse a particular level of climate in an organisation. Table 2 shows the reasons for selecting specific tools to make up the organisational climate awareness toolkit.
Table 2 Chosen tools for the organisational climate awareness toolkit

<table>
<thead>
<tr>
<th>Tools / Level of feelings, behaviours &amp; attitudes</th>
<th>Climate level</th>
<th>Relationship to the Tomato Plant Organisational Metaphor</th>
<th>Purpose &amp; Pros</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personality Test 1</td>
<td>Personal Climate Level = Behaviours, Feelings &amp; Attitudes</td>
<td>Tomato seeds = organisational employees</td>
<td>Measure the personality levels in a fun, light hearted &amp; entertaining way. Easy to understand &amp; fill in; Fun process; Charts individual strengths &amp; vulnerabilities; Reliable measurement method</td>
</tr>
<tr>
<td>Organisational Climate Index (OCI)</td>
<td>Sub-Climatic Level = Behaviours, Feelings &amp; Attitudes</td>
<td>Tomato segment = groups; Tomatoes = teams</td>
<td>OCI offers a language to help you understand the organisational strengths &amp; needs. Easy to understand &amp; fill in; Charts organisational strengths &amp; vulnerabilities; Extensively researched &amp; tested; Reliable measurement method</td>
</tr>
<tr>
<td>Personality Test 2</td>
<td>Personal Climate Level = Feelings</td>
<td>Tomato seeds = organisational employees</td>
<td>Measure the emotional levels in a fun, light hearted &amp; entertaining way. Easy to understand &amp; fill in; Fun process; Charts individual strengths &amp; vulnerabilities; Reliable measurement method</td>
</tr>
<tr>
<td>Team Skill Management</td>
<td>Personal Climate Level = Behaviours, Feelings &amp; Attitudes</td>
<td>Tomato segments = groups; Tomatoes = teams</td>
<td>Designed to help manage the self in a team, create working partnerships, encourage interdependence between members of a team &amp; understand how a team should fit into the organisation. Charts individual strengths &amp; vulnerabilities; Easy to understand &amp; fill in; Extensively researched &amp; tested; Reliable measurement method</td>
</tr>
<tr>
<td>Personal Test 3</td>
<td>Personal Climate Level = Behaviours, Feelings &amp; Attitudes</td>
<td>Tomato seeds = organisational employees</td>
<td>Measure the level of contentment with life in a fun, light hearted &amp; entertaining way. Easy to understand &amp; fill in; Fun process; Charts individual strengths &amp; vulnerabilities; Reliable measurement method</td>
</tr>
</tbody>
</table>

Corporate Character Questionnaire

Sub-Climatic Level = Behaviours, Feelings & Attitudes
Organisational Climate Level = Behaviours, Feelings & Attitudes
Tomato segments = groups; Tomatoes = teams
Tomato plant = organisation
Introducing a new way of thinking about corporate culture. The authors have identified four distinct types of corporate culture: Networked, Mercenary, Fragmented & Communal, & provide advice on how to change the culture of their corporations. The ethical issues of manipulating the ways in which people relate to one another are also addressed. Charts organisational strengths & vulnerabilities; Easy to understand & fill in; Extensively researched & tested; Reliable measurement method

Are You Tactful or Undiplomatic?

Personal Climate Level = Behaviours & Attitudes
Tomato seeds = organisational employees
Measure level of tact in a fun, light hearted & entertaining way. Easy to understand & fill in; Fun process; Charts individual strengths & vulnerabilities; Reliable measurement method

Are You in the Right Job?

Personal Climate Level = Feelings & Attitudes
Tomato seeds = organisational employees
Measure satisfaction with your current job in a fun, light hearted & entertaining way. Easy to understand & fill in; Fun process; Charts individual strengths & vulnerabilities; Reliable measurement method
Climate has an influence on behaviour and has been found to determine the successful achievement of a variety of educational products such as cognitive achievements, satisfaction, motivation and personal development (Belfer, 2002). This influence encourages climate to be a fundamental element on its own. Metaphorical models can facilitate the information exchange and develop the communication framework that promotes a learning climate. This paper proposes a framework, toolkit and theories to investigate and evaluate the following hypothesis:

“An organisation can be nurtured through organisational climate awareness in terms of understanding the feelings, attitudes and behaviours that can be found at the different climate levels within the organisation in order to increase the effectiveness of team/group interactions.”

Multi-level models are designed to bridge micro and macro perspectives, specifying relationships between phenomena at higher and at lower levels of analysis (for example, individuals and groups, groups and organisations, and so on). Accordingly, a multi-level theoretical model must specify how phenomena at different levels are linked. Links between phenomena at different levels may be top-down or bottom-up. Many theories include both top-down and bottom-up processes (Klein & Kozlowski, 2000). This research investigates and explores multi-disciplinary areas within the context of organisational theory looking at both top-down and bottom-up point of view.

4.0. Experimental Design

In order to explore and evaluate the hypothesis above, the proposed toolkit was used in a three part case study investigation:

- Part 1 – toolkit pilot on small teams/groups = aim at testing the usability of the proposed toolkit
- Part 2 – toolkit used on a range of organisations in different industries = aim at testing the generic use of the toolkit
- Part 3 – toolkit tested on a larger sample team/group = aim at testing the usability of different sizes of teams/groups

The selected organisations were from a judgement sample of twelve organisations ranging from large international organisations such as investment banks to small local supermarket and universities. Each part of the case study investigation had a post case study investigation questionnaire, which was given to the participants a week after they were given the results of the organisational climate awareness toolkit. The results of the post case study questionnaires were used to measure whether the toolkit had nurtured the organisational climate awareness levels and if there were any effects on the team/group interactions.

5.0. Results

The organisational climate awareness toolkit results showed that the current personal climate level across the 12 organisations varied in terms of feelings, attitudes and behaviours, which is normally expected because as human beings we are all different.
Therefore, we have individual personality types that are affected by external factors that in turn affect our feelings, attitudes and behaviours. This means no one can have the exact same personal climate as someone else. Whereas the results for the current group/teams sub-climate showed that the majority of organisations’ groups/teams are communal and therefore, much space is shared either formally or informally. Formal social facilities are supported by extensive informal socialising, food and drink spread into ‘work’ space. There are communication flows easily inside between levels, departments, and across national cultures, but outsiders may feel excluded. People use work time to socialise and they are not penalised for doing so. People get to know each other quickly, and many have known each other for a long time. Allegiance will be professional rather than organisational. Also the majority of the organisations’ group/teams have a number of individuals with the required mixed range of skills needed to contribute to the overall effectiveness of group/team sub-climate. By looking at the current personal climate and sub-climate it is obvious that the groups/teams throughout all the 12 organisations possessed individuals with the right mix and range of skills required to make the group/team effective. However, it was the personal climate that affected the groups/teams level of effectiveness because everyone has different feelings, attitudes and behaviours. Therefore it is important to understand and be aware of this in order to be able to work and interact effectively as a group/team. The results from the current organisational climate offered a small insight into the complicated holistic view of the feelings, attitudes and behaviours at the higher abstract organisational level. Therefore in order to get a clearer overall view of the organisational climate, the sample size needs to cover the whole organisation, not just the small sample size used in this study.

6.0. Conclusion

By comparing the breakdown analysis of each post case study investigation questionnaire results for all organisations used it is clear that the organisational climate awareness toolkit has had a positive effect on all the three levels of climate within each organisation. Participants have indicated that the organisational climate awareness toolkit has had an affect on their personal climate in terms of increasing the awareness of not only of their personal feelings. Attitudes and behaviours, but also of their group/team members too, which in turn helped to improve their interactions among the group/team. This has also caused changes in the sub-climate by creating an atmosphere of awareness towards others and their roles among the group/team. The overall effect on the organisational climate is that the participants understand and are more aware of the feelings, attitudes and behaviours that surrounds them and how they fit in, but most importantly how they should react to them in order to be more effective as a group/team.

In summary, this research provides an important insight into the composition of teams for success. It is important for Information Systems project managers to be aware of personality-type differences, as well as awareness of appropriate skills-mix. Therefore performance is to some extent related to personality type in determining team composition.
7.0. References


ICTs In Developing Countries: Contexts, Challenges And Interventions

Anthony J. Rodrigues, School of Informatics and Computing, University of Nairobi, P.O.Box 30197, Nairobi 00100, Kenya

Internet connectivity requires more than just simply installing phone cables or wireless networks. In the case of Africa it also requires overcoming specific barriers. Some of these are a lack of: information coordination, coordination of physical connections and technical personnel cooperation. Furthermore this is exacerbated by telecommunications monopolies and obsolete regulatory frameworks, limited involvement of research institutions in network building and diffusion, and language barriers. The world is undergoing a major change in learning. Internet technologies have fundamentally altered the technological and economic landscapes. Successful e-learning depends on strategies that optimize use of the technology within the cultural context of a nation. There is need to lower the overall costs of creating a competitive workforce and to do this continuously on a 24/7 basis for people specially those in rural areas. We also examine the current problems in the establishment and use of information networks and explore ways in which these valuable networks can be sustained so as to exploit more fully their potential.

Introduction

The Internet In Africa

Africa is the world’s second largest continent and the world’s least computerized continent. It has been affected by post-colonial difficulties, civil wars, and extensive health, educational, demographic, and economic problems. These have been exacerbated by the influence of cartel mentalities residual and emerging involving both foreigners and locals to various degrees. Thirteen percent of the world’s population resides in Africa (United Nations Population Division, 1998) but in some countries teledensity is as low as one telephone per 1,000 people (World Bank, 2000). According to a May 2000 study the number of African Internet users was somewhere around two and a half million. This worked out at about one Internet user for every 700 people, compared to a world average of about one user for every 35 people, and a North American and European average of about one in every three people (Jensen, 2000). Internet connectivity requires more than just simply installing phone cables or wireless networks. In the case of Africa it also requires overcoming specific barriers such as:

- state monopolies;
• poor local economies and shortage of capital especially foreign currency;
• scarce computing resources and lack of secondary equipment;
• unreliable sources of electricity;
• inadequate technical training;
• varied political climates and lack of regional cooperation;
• cultural considerations; and
• African social practices.

At the Inet 96 workshop, sponsored by the Internet Society (ISOC), participants reviewed current “Internetworking” barriers and discussed partial solutions. These brainstorming sessions highlighted the difficulties and frustrations encountered by many African countries as they grappled with the information age (Adam, 1996). The barriers were outlined into six main points and included:

Lack of information coordination. Lack of coordination at the national, sub-regional, regional, and international level resulting in redundancy and duplication of effort. Everyone wants to coordinate, but no one wants to be coordinated. Simply put in another way, there were too many managers, and not enough people willing to act as staff. Further intensifying this problem was the lack of information about who was doing what in terms of connectivity, what the costs were, and what the overall plan was.

Lack of coordination of physical connections. At the time of the Inet 96 meeting there were no plans for developing a backbone to link countries to each other. Networking solutions tended to be country specific. The resulting mix of independent connections demanded people with expertise and technical knowledge to develop gateways between these links. In addition many countries still used analog links that were difficult to integrate to newer communication technologies. Together with competition for resources and donor requirements these problems made active collaboration very difficult. Only recently the Eastern African Submarine Cable System (EASSy) was established to develop and implement a submarine cable system to provide fibre optic telecommunication facilities to the East coast of Africa, linking Northern and Southern Africa international gateways to the system for onward worldwide connectivity. The project was initially expected to come into service in the second quarter of 2007.

Lack of technical personnel cooperation. Cooperation and coordination between system managers and networking advocates had been difficult. There had not been enough opportunity for interaction between technical personnel.

Telecommunications monopolies and obsolete regulatory frameworks. Prohibitive government regulations made operating commercial services difficult. In many countries telecommunications were controlled totally by the government. There was little incentive on the part of public companies to build an efficient communication system.

Minimum involvement of research institutions in network building and diffusion. The need for the involvement of the academic community was stressed.
Language barriers. Language and illiteracy are central to the problem of integrating communities into the Internet. Most countries use different languages than English or French. Network user interfaces to these languages were unavailable and developments in transliteration were in the very early stages of development.

The World Bank has conducted several studies on the Internet in Africa. A study (World Bank, 1997) concluded that the most vital issues were reducing the cost of connectivity and overcoming government and regulatory barriers. The typical structure for telecommunications in many African countries follows:

- The provision of all services is entrusted to a state-owned operator (SOO) on a monopoly basis;
- The sector Ministry is supposed to supervise the SOO and assume the function of setting sector policy;
- Sector regulation is shared between the SOO, the sector Ministry, and the Cabinet. The SOO often takes over the responsibilities of the government on policy and regulation matters, resulting in a very dominant, inefficient SOO (World Bank, 1997).

The report went on to say that authorizing an Internet service and subsequent regulation of the network would depend on several issues:

- The current Telecommunications Act in the country;
- The nature of the service (commercial vs. non-profit) Who is/are the sponsor(s);
- Whether the SOO can be a part of a joint venture;
- Whether the sponsors are requesting to lease lines from the SOO or a license to build their own network;
- Whether the SOO can provide the required technical requirements (for example speed and quality); and
- Whether the planned Internet network would be used as a private network (by one institution or in a single building) or a “shared private network”.

Two years later in a further study the World Bank reiterated that in order to expand access and use of the Internet in Africa it was necessary to provide the following (World Bank, 2000):

- Low cost and reliable access to international bandwidth;
- Low cost and reliable local bandwidth connectivity;
- Countrywide reliable local cost access to ISPs;
- Low cost access to network equipment;
- Widespread public access to networked computers;
- An educated and trained user and provider base; and
- Support for the development of national and African Internet content.

The report concluded that this could only be achieved by:
• Liberalization of the telecommunications network;
• Liberalization of Internet service provision;
• Lowering of tariffs on computer and telecommunications equipment;
• General tariff balancing with possible support for local cost ISP access;
• Support for community access to the Internet; and
• Support for training in the use of the Internet.

The Internet Service Providers

The Internet is largely dependent on the telephone network (made up of the cost of the line and the cost of local and long-distance charges), the availability and affordability of access equipment, and at an even more fundamental level the pervasiveness of telematics (the mixing of hardware and software with human skills, organizational skills, and knowledge transfer). It is not enough however, to just have a telephone network. The telephone network must also be able to handle an Internet backbone. This is a network cable with very large throughput capacity, dedicated to Internet traffic, and usually leased from the public network.

There are two main problems. Some countries have connectivity within the country and just need to connect to the international Internet backbones; others also need to improve access to connectivity within the country. Country specific backbones are established and maintained by Public Telecommunication Operators (PTOs).

The task of improving network infrastructures can often be very difficult. For example the average delay in obtaining a phone in Ethiopia and the Sudan is 10 years and in Kenya 5.6 years (World Bank, 2000). The cost of connecting to the international Internet backbones can also be high. This is because the developing country Internet Service Providers (ISP’s) have to pay the full cost of connection to backbones in other countries (typically the United States), including costs for local interconnections (called peering) and transit traffic. Ironically it usually costs less to connect to the United States than to connect with other countries within a region. Of course once a connection is established it can be used by all users from anywhere in the world. Thus United States ISPs and users obtain free connectivity to overseas sites.

Of the multinational ISPs, AfricaOnline, founded by three Kenyan students studying at MIT and Harvard, , is one of the largest operations and has set up servers in the Cote D’Ivoire, Ghana, Kenya, South Africa, Swaziland, Tanzania, Uganda, and Zimbabwe amongst others , and plans to continue its expansion. UUNET is in Botswana, Namibia, South Africa, Swaziland, and Zimbabwe, while Swift Global is in Kenya, Tanzania and Uganda. Mweb (a South African based ISP) has moved into Namibia, Uganda, and Zimbabwe. The major international Internet suppliers are, inter alia, AT&T, BT, Global One/Sprint, UUNET/AlterNet, MCI, NSN, BBN, Teleglobe, Verio, and France Telecom/FCR.
Connection Alternatives

For countries without high-speed networks, undersea cables and satellites can help with the connectivity problem. For coastal states that already have an internal network in place, connecting to a cable landing point is likely to involve only small additional costs. For land-locked countries, or for countries with too small a demand to justify a cable landing point, the costs of connecting to a cable may offset entirely the initial cost advantage. For example, no one single African country generates enough traffic to justify the construction of a cable linking it to any other country or group of countries. For these countries, satellite systems may be the preferred transmission option. Satellites have an advantage in that they can potentially offer a means of connecting to the network via mobile handsets. This can be an important consideration for areas with a dispersed population. Satellites can also price capacity asymmetrically. It may be that the ideal connectivity is a hybrid of landlines, undersea cable and satellite.

African countries have the possibility of accessing: INTELSAT and RASCOM satellites. RASCOM (The Regional African Satellite Communications Organization) was established in May 1992. The plan is to set up a telecommunications infrastructure based on space technology which will provide telecommunications services at very low costs not to exceed US$0.10 per minute. RASCOM would also put an end to telephone calls within Africa being routed via Europe or America. Unfortunately due to funding issues the implementation of RASCOM has been slow (Rascom’s Space Technology, 1999). Initial projections seemed to indicate that for Rascom 1 & 2 satellites the operational phase would be between 2005 and 2006 according to the ITU website (www.itu.int/ITU-D/afr/).

The Eastern African Submarine Cable System (EASSy) established to develop and implement a submarine cable system to provide fibre optic telecommunication facilities to the East coast of Africa, linking Northern and Southern Africa international gateways to the system for onward worldwide connectivity. The project was initially expected to come into service in the second quarter of 2007.

Cost

While the poor telecommunications infrastructure acts as a barrier to developing the Internet, the high costs associated with connect time is the primary barrier in the path of ordinary citizens. ISP charges vary greatly between $10 (typically for e-mail only access) and $100 a month, depending on the market, the varying tariff policies of the PTOs, and the different national policies on access to international telecommunications bandwidth. For example, in 2000 Ghanaian ISPs paid approximately $2500 for a half-circuit but Kenyan ISPs paid $8000 for the same use of a half circuit due to extra charges levied by the PTO. In 2005 the cost of the same has gone up to US$9546 for a 2Mbps half circuit. These charges are passed on to the users. According to the Organization for Economic Cooperation and Development forty hours of Internet access in the United States costs US$20 a month including telephone. Although European costs are higher (on average US$45), these figures are for four times the amount of access and all of these countries have per capita incomes which are at least ten times greater than the African average (Jensen, 2000)
Large Users

Examples of the power of the Internet are already in place in the form of the African Virtual University, WorLD, the UN International Trade Center’s Virtual Handcraft Exhibition Center, and SatelLife.

The African Virtual University (AVU) is an Internet based program initiated by the World Bank. Its primary focus is to serve the educational needs of the countries of Sub-Saharan Africa by building world-class degree programs that support economic development by educating and training world-class scientists, technicians, engineers, business managers, health care providers, and other professionals. It links sixteen English speaking and nine French speaking universities in Sub-Saharan Africa with instructors from universities in developed countries. It offers instruction in computer science, computer engineering, and electrical engineering. AVU has transitioned from a World Bank funded project to an independent non-profit organization headquartered in Nairobi, Kenya.

The USAID Leland Project is an interesting example of aid to Africa with a view to opening up markets for the developed world. The Leland Initiative is a five-year $15 million United States government funded project to connect African countries to the Internet (USAID Leland Initiative, 1999). John L. Mack, director of African and Middle East trade and development policy at the US State Department summarized the goal of the project as “trade not aid”. He also added that interconnectivity is a powerful force for democracy though in fact there is no empirical evidence linking already developed electronic media such as radio and television to the spread of democracy in Africa.

Currently twenty one countries have signed up including Kenya, Benin, Botswana, Cote d'Ivoire, Eritrea, Ethiopia, Ghana, Guinea and Guinea Bissau.

Global Development Network: Need For Communication Between Academics And Politicians.

It is the task of academics and research institutes worldwide to analyze the complex interrelationships of our times, to describe them and to translate them into a language that will not just be understood by political decision-makers but also by the players of civil society. A great many of the insights that have already been gained and that are important for a more viable future have not been translated into practice so far because there is too little common language between academics and politicians (Wiezorek-Zeul 1999).

The idea of the Global Development Network can be an important step towards facilitating equitable access to information and overcoming the existing Babel of languages:

- the strengthening of research institutions in the South,
- global networking between development “think tanks,” and
- mutually enriching exchange between the academic and the political arena are important steps towards putting existing knowledge to use on a global scale.
This is not about a knowledge network for knowledge’s sake. Sight should not be lost of the central purpose: improving the living conditions of the people in developing countries and creating a basis for a common future that is worth living. It is the welfare of the people that must be the focus. This is why, from a development perspective, special importance must be accorded to applied research for the benefit of developing countries, for instance in the agricultural or health sectors.

In creating a Global Development Network, an awareness should be created that the transfer and exchange of knowledge is not just a technical challenge as:

- Knowledge equals power,
- Knowledge is politically explosive,
- Knowledge must be embedded in its socio-cultural context; and
- Local knowledge is at risk of being lost.

Knowledge equals power. The current dispute in the WTO over the protection of intellectual property shows the amount of power linked to knowledge and the political and economic interests at stake. Industry in the rich countries is demanding better protection in marketing the results of its research and inventions. Only if there are prospects of high profits, it is argued, will there be sufficient incentives for private enterprises to invest in research and development. From a business point of view, that makes sense. However, it is also understandable that the developing countries fear being excluded from important technical developments, and often even being denied the benefit that others are deriving from local knowledge and genetic material from their own countries, for instance, in the field of medicine.

Knowledge is politically explosive. There are governments that still insist on:

- monopolizing knowledge,
- deny students the opportunity to participate in international exchange programs, or
- muzzle their national press.

However, academic freedom and freedom of speech are important elements of good governance and indicators of the development prospects of a given country. As bridges are built between the academic and the political arena in the Global Development Network, it is necessary to secure the political commitment of all those involved to facilitate dialogue at all levels between the academic arena, policymakers, and civil society on the options for development and, even more decisive, their implementation. The media, too, has an important role to play in this context, as it ensures transparency and monitoring, and the dissemination of knowledge at the level of the people.

Knowledge must be embedded in its socio-cultural context. Information is now capable of being transmitted across the globe at an incredible speed. The communications revolution witnessed over the last few years is considered one of the main driving forces behind globalization. However, the fast and global transfer of knowledge may also imply some risks. Knowledge is no mass-produced commodity that can be applied in the same way everywhere and immediately. Global knowledge must be adapted to local conditions.
Local knowledge is at risk of being lost. About 80 per cent of all Internet sites are in English, even though a mere ten per cent of all people in the world understand the language. The industrialized countries hold 97 per cent of all patents worldwide. These two figures give an indication of the dominance of one part of the North in the knowledge sector. If the misconceived path for the global networking of knowledge is followed, this dominance may increase even further. We need to preserve local, indigenous knowledge and to be able to use it globally.

The North also has a lot to learn from the South. Local knowledge in the developing countries needs to be taken more seriously, and exchanged. The North is not the only source of knowledge. Accordingly, South-South exchange, or exchange within one country or one region, needs to be reinforced so as to make better use of knowledge there. In developing countries local knowledge - for instance that of rural people - is increasingly becoming accepted and used as a resource for innovative solutions to problems, rather than being discarded as too traditional and, therefore, obsolete. However, exchange still continues to be difficult in many cases, because just as in the developed world there is distrust, deep-seated habits, and prejudice.

Academia And The Information Society

The need to revitalise Kenyan Universities and to renew the sense of urgency in acknowledging their role in solving societal and developmental problems cannot be understated. The World Summit on Information Society, held in Geneva in 2003, referring to research and development recommended the following:

- Promote affordable and reliable high-speed Internet connection for all universities and research institutions to support their critical role in information and knowledge production, education and training, and to support the establishment of partnerships, cooperation and networking between these institutions;
- Promote electronic publishing, differential pricing and open access initiatives to make scientific information affordable and accessible in all countries on an equitable basis;
- Promote the use of peer-to-peer technology to share scientific knowledge and pre-prints and reprints written by scientific authors who have waived their right to payment;
- Promote the long-term systematic and efficient collection, dissemination and preservation of essential scientific digital data e.g., population and meteorological data in all countries;
- Promote principles and metadata standards to facilitate cooperation and effective use of collected scientific information and data as appropriate to conduct scientific research;
- Governments, through public/private partnerships, should promote technologies and R&D programmes in such areas as translation,
iconographies, voice-assisted services and the development of necessary hardware and a variety of software models, including proprietary, open source software and free software, such as standard character sets, language codes, electronic dictionaries, terminology and thesauri, multilingual search engines, machine translation tools, internationalised domain names, content referencing as well as general and application software; and

- Invite relevant stakeholders, especially the academia, to continue research on ethical dimensions of ICTs.

E-learning In Developing Countries: Challenges And Opportunities

The world is undergoing a major change in learning. Internet technologies have fundamentally altered the technological and economic landscapes. Successful web-enabled learning depends on strategies that optimize use of the technology within the cultural context of a nation. There is need to lower the overall costs of creating a competitive workforce and to do this continuously on a 24/7 basis for people specially those in rural areas, see El-Sobhy 2003.

E-learning Challenges: The challenges facing e-learning implementation in developing countries include, interalia:

- ICT Infrastructure;
- Culture, Leadership, and e-Learning Strategy;
- Local Content;
- Copyright Issues; and
- Instructors & Learners;

The lack of telecom infrastructure: Telecom infrastructure is a must for e-learning. Most developing countries have limited Internet Bandwidth and Internet penetration is still low.

The role of leadership in building a durable e-Learning strategy: Without a clear strategy for e-learning, no progress will be achieved. Leaders in both public and private sector should play an important role in building this strategy. Several efforts towards using the technology for learning have not been sustainable because many leaders have underestimated the complexities of the interactions between e-Learning and the working environment and how difficult it is to change people’s perception. An appropriate e-Learning strategy not only addresses issues of technology and learning effectiveness, but also the issues of culture, leadership, justification, organization, talent, and change.

Cultural challenges: The e-learning tenet is to incorporate learning into the daily environment, to make learning accessible anytime anywhere, to make learning come to the beneficiary at work, at home, and even while traveling. To sustain e-learning, a strong culture is required that embraces e-learning as an important activity in any organization to the extent of ingraining it into the day-to-day culture.
**Limited local content:** Developing countries facing a problem of offering local learning content. Except for a very few countries, the learning content is provided mainly from European or American content providers.

**Copyrights issue:** The copyright issue is extremely important since we have several stakeholders in the e-learning process that will have copyrights:

- the content provider;
- the copyrights for the conversion of the content from its raw version to a web-enabled version;
- in some instances there is a copyright for the localization of the material; and
- the party that built the platform through which the final material is offered.

Role of instructor and of the learner: The instructors and learners are important parts of e-learning system and some roles and functions of both are changing. The role of the instructor can be coaching, observing, offering hints, reminding and/or giving feedback, whereas the learner will learn to be more individually responsible for the pace of study becoming more and become increasingly dependant on on-line material than text books.

The quality of e-Learning relies heavily on the preparation and the talent of the instructor. Instructors who deal with e-Learning should have special skills. They should be able to deal with remote learners and realise their needs through their typing mode and sketchy words. In addition, they should be capable of using the new technological means of instructing, like web cams, e-mail, chatting facilities, slides, sketches, and video and audio clips. This implies having good typing skills, ability to setup and deal with multiple programs at the same time, and deal with many learners simultaneously. In addition they orchestrate so they can manage and encourage the learners to communicate with each other through the discussion forum.

E-Learning learners should possess the following characteristics:

- Ability to learn independently and view learning positively;
- Ability to make the best use of their time, be self-disciplined, and enjoy working alone most of the time;
- Ability to express themselves clearly in writing;
- Skilful in computer and Internet use;
- In need of knowledge, but incapable of attending traditional training or education; and
- Problem solving abilities.

**E-Learning Opportunities:** Despite the challenges facing e-Learning in the developing countries, it may be considered a gold mine of business opportunities. The e-Learning services are value-added technical services and can help in the creation of several jobs. The opportunities are categorized below:
Enhancing the ICT Infrastructure: Since citizens in most developing countries lack sufficient access to the net, they need to use community centers or private cyber cafes. Investors or SMEs in developing countries can establish such a network of centers across the country. They can franchise the establishment of such centers. Those centers can be learning centers which will give the citizens in rural areas access to educational and learning materials in different topics. It can also provide some other community services for the society.

Regional projects for Internet Exchange Points “IXP” will be needed. Such projects will help route the Internet traffic within the groups of developing countries instead of going outside to Europe, USA, and Japan and come back again. This will help in better using the limited bandwidth and will lead to reduce the cost for the ISPs.

Local Content Development: Many developing countries are rich in cultural heritage. Unfortunately the production, distribution and access of local content are still rare. Digitizing the countries’ cultural heritage and making it available for people from around the world is a good business opportunity. e-Learning can be used to teach people around the world the famous local languages and cultures of many developing countries. It can be used to help local people learn how to preserve their culture by making handy crafts (which could be sold by other e-commerce sites). Developing local content will help in raising the level of education and in alleviating the illiteracy problem in developing countries. It can help in raising the level of awareness between citizens about the challenges facing them such as, the environmental problems of deforestation and erosion, sanitation and health, and with the HIV/AIDS pandemics.

Support Services Providers: To provide local e-Learning, organizations are required which will provide supporting services. These supporting services can be translation services to develop the content not only in English but also in Swahili and the ethnic languages of Kenya. Content localization is a critical supporting service to e-Learning, since some of the content will be acquired from foreign content providers. Such courses need not only to be translated, but also some sort of localization to adapt the content, the examples, and the cases to the local community.

Such projects can be a good example for regional cooperation and across borders joint business. Both the translation and localization can be done through the use of the Internet between different parties in different countries. This will lead to the presence of specialized centres for translation and localization across the developing countries.

Technical Support Providers: Providing e-Learning services, will require support from some other technical support providers. This may include Web developers, software developers, multimedia developers, graphics designers, database administrators, WAN and LAN administrators, data and networks security systems specialists, data entry persons, maintenance specialists & technicians, networks installation specialists & technicians, and so on.
Those providers will be small and medium companies, and some of them will provide their services across borders. This will result in further collaboration between business owners and regional entrepreneurs.

Marketing Support Providers: Providing e-Learning services also requires support from marketing and publicity agents. Although e-Learning will be done through the net, it also needs some sort of market surveying, data analysis, marketing strategies and plans, marketing campaigns, designing publicity brochures, designing marketing and sales kits, designing the advertising campaigns, preparing mailing lists, designing the promotional materials and kits, arranging for conferences and exhibitions etc.

After successfully implementing e-Learning projects in different developing countries, a second phase of collaboration between the different business parties will be needed. In this phase a joint organization can be established to be able to export those services to other countries or regions. e-Learning can be a first phase or a cornerstone in establishing a joint services market.

Sustainability Of National And Sectoral Information Networks

We now examine the current problems in the establishment and use of information networks and explore ways in which these valuable networks can be sustained so as to exploit more fully their potential. The issues range from infrastructural to policy ones. The focus of these being the involvement of national governments to spearhead initiatives (Nshimbi 2001). Sustainability of information networks in Africa, whether at a national or sectoral level is a very difficult task considering the:

- disparities in telecommunication infrastructure;
- restrictive government policies and regulatory frameworks; as well as a
- lack of the necessary logistics to identify and develop quality content.

At a more operational level, the sustainability of these networks includes:

- ongoing maintenance;
- security;
- efficiency in the delivery of content and the long-term use; and
- affordability, accessibility and stability of the networks.

Infrastructure (National) At a national level, many issues have to be dealt with to sustain the telecommunications infrastructure which forms the backbone of information networks. The infrastructure in many African countries does not adequately meet the demands of growing populations and is, in most case, outdated. It is also mostly confined to the urban and semi-urban areas, with very scanty coverage in rural areas. With the involvement of private sector companies, collaborating with the public sector companies, most African countries are seeing an improvement in the services being provided and new investments in the development of telecommunications infrastructure.

Content (National) Content, at a national level, has been mainly confined to Central Statistical Offices, (CSO’s) which, in most cases, do not have a systematic data collection
standards to facilitate integration of data sets over a region. There is also a scarcity in the human resource required to manage the content. For many African countries most of the useful content is not in electronic format and this poses another challenge.

Local content development has to be encouraged to help share indigenous knowledge and build local capacity. The development of content can increase traffic flow to the sources of information thereby making the information networks sustainable.

Both aspects require deliberate national policies like those in Egypt where there is increased political commitment and a ministry has been set up specifically for ICT issues. Mauritius gave incentives to Internet users in its year 2000 budget by reducing the cost of Internet access for dial-up and for international links by satellite, leading to a rapid increase in connections to the Internet. The Government also liberalised access to the international gateway and all primary and secondary schools were provided with one free access to the Internet. Incentives on corporate tax were also be extended to service providers and IT training schools. Most governments are also relaxing their regulations, reducing license fees and allowing for greater competition in the telecommunications industry, thereby empowering local enterprises. Governments could also eliminate duty on ICT equipment being imported into the country as is the case in Malawi, Mauritius and Kenya and also allow for local call charges for Internet access from anywhere in the country. This local call service is currently available in 16 African countries, 5 of which are COMESA countries i.e. Ethiopia, Malawi, Mauritius, Kenya, Uganda and Zimbabwe.

Sectoral (Infrastructure) Regionally, the issue of infrastructure as being a hindrance to economic development has long been recognised and many regional initiatives, such as RASCOM and COMTEL have been started which aim to make more effective use of infrastructure and keep the costs to the end-user as low as possible.

A further setback in terms of infrastructure is that most of the ICT equipment is not manufactured locally. It has to be imported and incurs extra import and duty expenses making it out of reach for the average citizen and for most of the local learning institutions.

COMTEL was set up with the main objective of routing telephone traffic within the region and reduce on transit fees and also set up local infrastructure using the most appropriate technology. This is being done with the involvement of the National Telecommunications Operators (NTO’s) in each of the COMESA Member States. A reduction in telecommunications tariffs would facilitate the increase in Internet access.

Sustainability. The sustainability of information networks, from the infrastructural and content point of view can therefore be summarised as follows:

- Private/public sector partnerships – to involve a greater part of the population;
- Less restrictive government policies and regulations;
- Awareness for the general public to increase appreciation and use of the information networks;
- Identify and develop quality content (indigenous African content for an
African information network);

- Adopt a System for acquiring the content – generation, collection and translation;
- Market maturity to allow competition and exploitation of new technologies such as VOIP;
- Use of more widespread and efficient technologies e.g. wireless;
- Sharing of national experiences to encourage best practices;
- Centres of Excellence to build human expertise;
- Local dial national access to reduce on the costs of accessing the Internet access;
- Improved ICT Training Programmes;
- Collaborative efforts within regional organisations and funding bodies;
- Affordability and accessibility of networks up to the grassroots;
- Involvement of other utility providers, such as electricity companies, to offer services even in remote areas; and
- Continuous evaluation of the information to ensure its relevance.

Even with the current problems of inadequate funding, poor infrastructure and state-run monopolies in Africa, the sustainability of the networks remains a priority in order that the continent may close or reduce the gap on the digital divide and secure more opportunities for trade and development through these networks. The resultant increase in intra and external trade will develop the national and regional capital markets and allow for more innovations within the information systems.

Therefore, as a way forward in ensuring their sustainability, all key players should promote national ICT policies and agitate for the active involvement and commitment of governments. Professional associations serve as a good platform to be able to contribute to the development and sustainability of ICTs and convince decision makers of the importance of the technology.

**Policy Recommendations**

It is however often difficult to persuade governments to spend money on computers when faced with drought and malnutrition. It is hard to persuade people that the connectivity that can promote and handle e-commerce can also provide telemedicine, distance education, and agricultural help with pest control, farm practices and drought management. Electronic access can minimize the rural-urban distinction. Information can cure illness and bring food. The challenge is to show the correlation between poverty and lack of access to information while many health problems can be cured with access to clean water.

Prior to 1990 the West was known as the 1st World, the Communist bloc the 2nd World, and the developing counties of Africa, Asia and Latin America as the 3rd World. The concept was derived from the three Estates of French society before the Revolution
in 1789. After 1990 the term North is used to refer to the industrialized, developed countries; and the term South to the underdeveloped areas of the world, concurs and points to the elitist nature of Internet society (United Nations Development Program Human Development Report, 1999). The report is concerned that there will be a two-tier technology society. The first society (the North) has access to plentiful information at low cost and high speed; the second society (the South) has its quality of access impeded by time, cost, uncertainty of connection, and outdated information. Time alone will tell whether the argument that information in itself does not feed, clothe, or house the world, can be ameliorated by the fact that it has the capacity to create wealth that can be converted into food, clothing, and shelter and recognized as such in Policy documents.

With the aforesaid in mind we recommend the following:

- Liberalization of the telecommunications network;
- Liberalization of Internet service provision;
- Lowering of tariffs on computer and telecommunications equipment;
- General tariff balancing with possible support for local cost ISP access;
- Support for community access to the Internet;
- Support for training in the use of the Internet;
- Support for the development of national and African Internet content;
- Promote local and global development networks with private and public sector partners augmented by academia;
- Promote Research and Development Initiatives in the private and public sector using Universities for contract research;
- Encourage and Revitalise traditional and other local knowledge to solve developmental problems in situ;
- Develop a coherent and feasible e-Government Strategy prioritizing sector-based applications such as e-health and e-education and training within the context of an integrated e-government system;
- Design e-applications using Architectures at various layers namely: Process, Information, Data (including spatial data) and ICT Infrastructure so as to reduce the risk of system failures;
- Implement these e-applications in concert with appropriate local government infrastructures and communities to enhance sustainability;
- Create an enabling environment including a legal framework; and
- Review regularly the information emanating from the system to ensure its relevance.

References


El-Sobky, Ahmed. (2003) e-Learning in Developing Countries: Challenges and Opportunities

RITSEC Jensen, Mike 1999 The Internet in Africa: an overview, Banjul, Gambia


Architectural Building Blocks For Reusable Service Oriented Architecture With Integrated Transaction Processing

Benjamin Kanagwa, bkanagwa@cit.mak.ac.ug and Ezra K Mugisa, emugisa@cit.mak.ac.ug, Faculty of Computing and Information Technology, Makerere University, P.O.BOX 7062, Kampala, Uganda

With increase in the number of complex services, their interaction, performance, quality of service, security and composition have become critical issues in Service Oriented Architecture (SOA). At the same time, the structure of services, their exposure and management reflect key architecture and design decisions that enable Service Oriented Computing (SOC). The layout of architectural elements plays a significant role in mitigating these critical issues. Services typically encapsulate some business activity, and their interactions directly imply business interactions. We note that most SOA issues are native to a specific architecture and are influenced by the transactional interactions within the architecture. In this paper, we therefore propose a novel top-level architecture that knightly integrates transaction-processing capabilities into SOA. The contribution of this paper is a generic high level architecture that specifies the structural elements for transaction centered service oriented architecture.

Introduction

SOA has gained more popularity of recent largely due to web services (Hashimi, 2003) and partially due natural evolution in software engineering. There is a systematic progression from Object oriented, to distributed objects, components and then Service Oriented Systems. SOA is defined as an architectural concept, an approach to building systems that focus on loosely coupled set of components (services) that can be dynamically composed (Graham et el). It is also defined as a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains (Oasis, 2003).

Service Oriented Architecture (SOA) defines the services of which the system is composed, describes the interactions that occur among the services to realize certain behavior, and maps the services into one or more implementations in specific technologies (Nimal et al, 2003). SOA is implemented by a collection of supporting technologies that compliment each other to provide a uniform view to services by abstracting the potential discrepancies in service consumers. For instance, SOA instances that target network accessible services will define technologies that enable
access mechanisms possibly through API (Application Programming Interface), remote procedure or some messaging system. Those intended for dynamically discoverable services would provide technologies that enable service discovery possibly through registries or some form of advertising.

A fully encapsulated SOA would support a collection of network-resident software services accessible via standardized protocols (guarantee for interoperability), whose functionality can be automatically discovered and integrated into applications or composed to form more complex services. Complex services will increase as more organizations realize the power of SOA as a cost effective integration solution to invaluable legacy systems and business transactions.

Services are the logical manifestations of some physical resources (like databases, applications, devices or humans), grouped as a process (set of actions that an organization is prepared to execute) that an organization exposes to the network (for example the web, hence web services) (Webber et al, 2003). Services exposed by different organizations are built to extend their services for business purposes.

The increase in the number of complex service compositions, service interaction, security, performance and quality of service (QoS) have becomes critical issues in architecting services. Since services typically encapsulate some business activity, their interactions directly imply business interaction. However each business has a set of business rules that govern its internal and external behavior as it interacts with other businesses. To avoid potential violation of any business rules during service interactions, SOA architecture should be built with a strong Service Transaction Processing Architecture (STPA) based on trusted architectural patterns. As mentioned earlier, STPA requirements are different from traditional transactional systems. Traditional systems typically employ ACID (Atomics, Consistency, Isolation, Durable) with two-phase commit protocols (2PC) (Bernstein, Newcomer, 1997).

The service composition transactional requirements and the corresponding architecture derive uniqueness from the SOA paradigm principles. Services are by nature autonomous, loosely coupled, distributed, do not share data and service transactions are generally long-lived. The transactional implications of these properties is not in the scope of this paper, but we rather focus on the architecture structure that inherently supports transactional processing without compromising the SOA paradigm. We also show how STPA knits with other SOA architectural structures.

The rest of the paper describes transactions in SOA, the proposed architecture and its realization.

Transactions In Soa
The need for transactions in SOA was realized beginning with 2001 as the need and paradigm shift to service oriented architectures started to be evident. Web services vendors have since developed several approaches leading to competing specifications.

The specifications try to provide a guaranteed outcome between loosely coupled interacting systems. There are two major transactions specifications, notably Business Transaction Protocol (BTP) (Ceponkus et al., 2002) and Web Services Coordination
(WS-C) (W3C, 2004) /Web Services Transactions (WS-T) (IBM, 2004). BTP is a protocol that enables orchestration of business processes between loosely coupled software services to achieve consistent outcomes from the participating business parties (Mark little, 200). BTP defines two transaction types, Atoms and cohesions. The two transaction types are relaxation in the classic ACID transaction management and referred as extended transactions based on 2-phase commit protocol. The foundation of both protocols is traditional ACID transactions with relaxation in atomicity, consistence, isolation and/or durability. Such transaction specifications that relax some of or all of the ACID properties are referred as extended ACID transactions (Mark little, 2003).

Transactions in SOA pose new challenges due the nature service compositions. Services are typically autonomous, distributed, loosely coupled, and may encapsulate long running transactions (Linthanmaphon and Zhang, 2004). For this reason, traditional transactional approaches do not properly work in SOA for mainly two reasons; (i) Traditional ACID models require that resources be locked during the transaction until it either committed or aborted. Composite service transactions may involve long running transactions and locking resources for that duration is very expensive and hinders scalability. (2) ACID transactions often do not provide suitable recovery techniques (apart from the transaction abort). However Composite services require robust recovery techniques since composite services involve autonomous systems with unique business implications. At the same time composite services usually have more complex failures. Whereas some service transaction specifications have been developed, no effort as of now has emphasized how to apply the low-level protocols to build a high-level transaction based architecture in SOA, leave alone how the service transaction architecture would knit with other SOA architectural structures.

Justification

The above transactional protocols do not address architectural issues in terms of placement and interaction with other service oriented computing artifacts. They address low-level transaction protocols. They provide options for implementing the transaction-processing component in the proposed generic architecture.

Service performance, security models and QoS heavily rely on the actual placement of services in a specific service oriented system. For example, performance relies on the location of files, size of data transfers and so on. However, performance is partly a function of the transactional requirements of the system that is being built. The level of transactional support in services plays significant role in mitigating other issues associated with services. We thus propose a high level-generic architecture with inherent transactional support. By identifying the subcomponents of the top-level structures, this architecture can be used to implement any SOA instance.

The Proposed Soa Architecture

The proposed architecture identifies five top-level architecture building blocks as in figure 1
STPA defines specifications and protocols that provide transactional support for service compositions. It ensures that services participating in business transactions through service compositions execute gracefully even in catastrophic circumstances. STPA uses the Service Interaction architecture to manage service interactions.

**Messaging Architecture**

The messaging architectures define specifications for message structures and semantics. Messaging is considered an important part of SOA since it inspires loose coupling and interoperability. The messages can contain data, or instructions that facilitates interaction in the composite service. The messaging architecture is used by all components except the communication architecture as indicated in the figure 1.

**Service Interaction Architecture**

This is a collection of structures that enable composite services to meaningfully interact. It includes all workflow and business processing structures. Services can only interact after they know about each other and thus service discovery is an important sub component of the Service Interaction architecture. It uses the Messaging architecture to relay message during service interactions.

**Communication Architecture**

This describes the low-level communication architecture that is responsible for the actual transfer of messages during service composition and execution of transactions. It is used by the messaging architecture for the actual message transfers.

**Service Architecture**

This defines the architectural primitives used in the construction of and subsequent exposure of a service. It specifies the structure and semantics of service interface contracts a well as its realization.
Mapping Architecture To Implementation

In this section we describe how to map the proposed architecture onto the web service set of technology. Since our focus is high-level structures, we do not provide code-level implementation in this paper. The actual implementation of the above architectural components within the SOA paradigm is a function of the specific SOA instance and its target environment. As earlier noted, a SOA is implemented by a collection of supporting technologies that complement each other. In this section we map the above architecture to web services, which is an implementation of SOA.

In Web services, the STPA structure can either be built on top of WS-C/T or BTP as the underlying foundation for transaction management framework. The Messaging Architecture structure can be built on top SOAP as the messaging protocol. SOAP relies on Hyper Text Transfer Protocol (HTTP) and other protocols for communication, which are part of the communication architecture. Service Interaction Architecture can be based on any of the various standards for workflow or business processing protocols.

The wider adoption of BPEL4WS (Business Process Execution language for Web services) (Weerawarana, 2002) by web services community makes it a better choice as the underlying protocol. The Service Architecture is realized using Web Services Description Language (WSDL) (W3C, 2003) as the language for service contracts. The actual service functionality is realized by defining an interface and its implementation. The actual implementation (coding) can be in any programming language using any programming model.

Future Work

Future work will be mainly centered on STPA since its the major component in service compositions. Its important to formally specify and reason about the architecture using some form of Architecture Definition Language (ADL). But before the ADL, we intend to identify more fine-grained architectural primitives from which the entire architecture can be drawn. Architecture Patterns that relate to transactions management are the current prime concern. We shall also show how to map the architecture to domain specific SOA instances.

Conclusion

The architecture realizes the importance and uniqueness of the transactional requirements of service compositions. It is structured with transactional capability as an integral part of any SOA instance. The architecture neatly maps onto web services implementation except for the STPA whose protocols are rather in infancy stage. With its generic nature, it can be used to generate any SOA instance.
References


Erik Christensen, Francisco Curbera, Greg Meredith, and Sanjiva Weerawarana, Web services description language (wsdl) 1.1-w3c Note.


Mark Little, (2005) Transactions and web services, COMMUNICATIONS OF THE ACM vol. 10 no. 46.


Information and Communication Technologies (ICTs) in Small and Medium Enterprises (SMEs): Findings from Uganda

Ali Ndiwalana, ally@dicts.mak.ac.ug, Directorate for ICT Support, Makerere University, Kampala, Uganda,
Steve Esselaar, F.F., Esselaar.S@pdm.wits.ac.za, 2 LINK Centre,
Wits University, Johannesburg, South Africa,
Tusubira, tusu@dicts.mak.ac.ug, Directorate for ICT Support, Makerere University,
Kampala, Uganda, and
Christoph Stork, Christoph_stork@web.de, 2 LINK Centre,
Wits University, Johannesburg, South Africa

The potential contribution of information and communication technologies (ICTs) to the development of small and medium enterprises (SMEs) can only be assessed by first understanding current information practices and needs in such enterprises. This paper reports findings from a questionnaire and interview survey of 351 SMEs in Uganda with a varying degree of formality. The SMEs were randomly selected from Kampala (central), Fort Portal (west), Lira (north), and Mbale (east). We find that informal businesses have a higher profitability in terms of fixed assets employed than semi-formal ones, which in turn have a higher profitability than formal businesses. The mobile phone has overtaken the computer as a tool in supporting the running of a business in spite of the lack of well-designed phone or SMS based business information systems. While we have data on ICT possession and usage within SMEs, we still don't have a good understanding about their potential impact to help make more useful interventions and better justify their costly investments. We would like to avoid the usual techno-centric approach where technology is given a pedestal and instead leverage the technology in a user-centric manner by first understanding the entrepreneurs behind the SMEs, their information practices, the needs that dictate the way they operate as well as the environments in which they operate. This survey is a first attempt in this direction.

Introduction

This paper reports findings from a questionnaire and interview survey of 351 small and medium businesses (SMEs) in Uganda. The SMEs were randomly selected from Kampala (central), Fort Portal (west), Lira (north), and Mbale (east). The survey was done as part of the SME E-access and E-usage index (2005) carried out in 13 African countries by the ResearchICTAfrica network (Esselaar et al., 2006). The results discussed
in this paper relate only to the Uganda survey. The aim of the survey is to understand the impact of ICTs on private sector development with particular emphasis on the SME sector within the context of developing countries.

The SME sector has an important role to play in the present and future economic development, poverty reduction and employment creation in developing economies (Hallberg, 2000). Stern (2002) stresses that the SME sector is the sector in which most of the world's poor people are working. The sector largely exceeds the average economic growth of national economies in many countries and contributes significantly to employment creation. Accordingly, governments and donors alike have recognized the important role of the SME sector for overall development. As a result, many government policies are geared towards supporting their growth through a variety of programmes that range from tax incentives to technical assistance, from regulatory provisions to policy interventions, training and other types of business development services (O'Shea and Stevens, 1998).

SMEs in general face a number of constraints in relation to survival and growth. These tend to vary with business or entrepreneur-specific factors, SME sector, their location and other factors, many of which are cross-cutting (Blili and Raymond, 1993, Liedholm, 2002). In many SME surveys that have been done, usually in developed countries, the most conclusive have focused on issues such as access to information and to capital and financial intermediation. The scarcity of information is in big part due to SMEs’ tendency to rely more on informal sources of information (Moyi, 2003, Duncombe, 2004, Sawyerr et al., 2003) while access to finance is hampered by a lack of or poor record keeping, that would otherwise provide a basis for financial institutions to extend them credit facilities. ICTs have the potential to help SMEs better leverage their informal information resources while minimizing their limitations on one hand. On the other, they can help SMEs to improve their business processes and keep better records, providing a basis on which financial institutions can determine their viability and hence extend them credit for enterprise development. ICTs can play a catalytic role within the SME sector by reducing transaction costs (thereby increasing efficiency) and increasing access to markets.

The SME E-access & E-usage Index (2005) aims to look from the demand side at ICT access and usage. The main objective is to identify the information practices and needs of SMEs as well as obstacles that they face in their daily business activities. A good understanding of these issues will be useful in two ways; from the regulatory perspective, we can provide guidance to the relevant authorities and various stakeholders in the development arena in creating relevant policy initiatives and interventions that stimulate economic growth and employment. From the private sector perspective, we shall identify potential niches for applications and services that do address needs relevant to the SMEs.

While the intent is to understand the potential of ICTs on SMEs, we avoid the usual techno-centric driven approach where technology is given a pedestal and instead provide a solid foundation from which to leverage the technology in a user-centric manner by understanding the entrepreneurs behind the SMEs, their information
practices, the needs that dictate the way they operate as well as the environments in which they operate. This paper presents key results from this survey.

**Methodology**

The SMEs were sampled based on a target list identifying SMEs within different sectors. While the bulk of these are mainly in Kampala, we did include other parts of the country to get a better picture of business activity within the SME sector in the country as a whole. The survey included FortPortal in the west, Lira in the North as well as Mbale in the East. Access to SMEs that would provide us with accurate data was an issue. No random sampling procedure was used but the businesses were selected based on their profile. We used a combination of sources that include business directories, worked with a number of small business organisations i.e. Uganda Gatsby Trust and personal contacts.

A tremendous effort was invested in training enumerators to understand how the business makes its money. Gathering financial information from SMEs is not an easy task. It required enumerators to build up trust with the entrepreneurs and gather an understanding of how businesses operate. Several check backs were built into the questionnaire to check for consistency of responses while the interviews were conducted. This was meant to enable enumerators to uncover wrong or inaccurate information given to them while they were still in the field.

A typical session involved guiding an SME owner or manager through a series of questions divided into modules that relate to information about the business, ICTs deployed by the business, business financials as well as the business climate. The session usually lasted under 45 minutes.

The most common reason that led entrepreneurs to provide inaccurate information included the fear for higher taxation since they imagined that our enumerators were from Uganda Revenue Authority. Others included fear of competition as well as poor or absence of record keeping. Enumerators were trained to assist SMEs without record keeping in estimating the values of fixed assets and other financial figures.

Our sample is not statistically “representative” and as such our findings cannot be extrapolated to account for the whole country. Our methodology has been geared towards getting collecting more authentic data given the lack a national survey that would have enabled us to achieve generalisable results.

**SME Definition**

SME definitions vary from country to country and are sometimes sector specific. They may be defined by the number of employees they have, by their rate of turnover, by their total borrowing from the financial system, by who or which entity owns majority shareholding or various combinations of all the above. For this survey, the definition is based on a recommendation from the African Development Bank, which defines SMEs as having less than 50 employees.

This paper distinguishes between survivalists and businesses. Survivalists engage in business activities with the aim of generating enough income for day-to-day
consumption, rather than growing a business that generates a sustainable stream of income. Survivalists usually do not distinguish between business and personal finances, do not keep records, do not pay taxes and are not registered with any authority. Money from sales or services rendered is predominantly consumed immediately for private purposes (same day, week or month). A micro and small business in contrast has some degree of formality. This might include a business bank account, formal work contracts for employees, a physical address with contact details, registration with receiver of revenues and other authorities and so forth. The line between survivalists on the one hand and informal businesses on the other is ambiguous and varies from country to country. The distinction between formal and informal businesses equally varies from country to country.

Only businesses that complied with the following characteristics were included in the survey:

- Physical presence (shop, workshop, house were the business is operating from) with contact details (a minimum of two out of these three: Street number, post box, telephone/cell-phone number);
- Business must operate with the aim of generating sustainable income streams;
- Business should be independent, i.e. not be a branch of a larger business;
- Business must have less than 50 employees.

The analysis of this paper is based on the responses of 351 SMEs.

**Formality Index**

The data collected from SMEs was used to classify responding SMEs into informal, semi-formal and formal businesses. Table 3 below shows the various variables that contributed to the formality index.

**Table 3: Computing the Formality Index.**

<table>
<thead>
<tr>
<th>Question</th>
<th>Value</th>
<th>Index Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.1 Form of ownership?</td>
<td>Sole proprietor, Partnership</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CC, Pty</td>
<td>0.5</td>
</tr>
<tr>
<td>D.3 Is your business Registered with the Receiver of Revenues? (pay tax?)</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.5</td>
</tr>
<tr>
<td>D.4 Is your business Registered for VAT?</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>D.11 How many of your employees have a WRITTEN EMPLOYMENT contract?</td>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>One ore more</td>
<td>1</td>
</tr>
<tr>
<td>D12 Does your business strictly separate business finances from personal finances?</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table 4: Classification of Formality Index

<table>
<thead>
<tr>
<th>Formality Index</th>
<th>Index points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal</td>
<td>1.5 points or below</td>
</tr>
<tr>
<td>Semi-Formal</td>
<td>2 points or more and less than 3.5 points</td>
</tr>
<tr>
<td>Formal</td>
<td>3.5 or more points</td>
</tr>
</tbody>
</table>

PTYs and CCs usually require registration with various ministries such as finance and/or trade and Industry, whereas sole proprietors and partnership do not necessarily need to be registered with either in most countries surveyed. A SME registered for VAT is also more likely to be formal than one that is only registered for income tax since VAT handling requires sophisticated record keeping. Having written employment contracts for employees also contributed toward the formality index. Having a written contract allows employees to enforce the rights and minimum wages as stipulated in labour laws while those without contracts can often be hired and fired at will. Whether a business is strictly separating personal from business finances and the sophistication of record keeping and accounting was also included in the formality index.

The maximum value a business could achieve in terms of the formality index is 4.5. Businesses were then categorised into informal, semi-formal and formal. The breakdown of this classification is shown in Table 4. Table 5 displays the breakdown of the sample in terms of the International Standard Industrial Classification (UN, 2002) and formality categories.

Table 5: Sample breakdown by ISIC and formality

<table>
<thead>
<tr>
<th>ISIC Description</th>
<th>Formality Index Category</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Informal</td>
<td>Semi-Formal</td>
</tr>
<tr>
<td>D: Manufacturing</td>
<td>42</td>
<td>22</td>
</tr>
<tr>
<td>F: Construction</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>G: Wholesale and retail trade; repair of motor vehicles, mot</td>
<td>34</td>
<td>51</td>
</tr>
<tr>
<td>H: Hotels and restaurants</td>
<td>15</td>
<td>13</td>
</tr>
</tbody>
</table>
Ict Possession And Usage

SMEs in the survey have invested in a range of ICTs as shown in Figure 8 below. It is clearly apparent that the mobile phone is becoming ubiquitous and has over taken the other forms of communication. Interestingly, the degree of access to ICTs generally follows the degree of formality of the business in all cases except that of the mobile phone where the trend is reversed. While this could be due to a combination of reasons, we presume it primarily relates to the easy of getting a mobile phone. In Uganda, a customer just walks into a shop and pays for a SIM card and a mobile phone (or just SIM card if they already own a mobile phone) and is instantly connected onto the network. In addition, the mobility and flexibility that the mobile phone affords might also play a role. A counter argument could be that people in more formal businesses have more access to a wider variety of other ICTs and hence exhibit a lesser need for the mobile phone compared to their less formal counterparts.

Figure 8. ICT Possession within surveyed SMEs
In terms of usage, the mobile phone is still more prominent for business communications. Mobile phones are used more often for keeping in contact with customers and clients compared to any other form of communication. 87% of SMEs in the sample used the mobile phone for this purpose compared to 37% using fixed line telephones. When it comes to ordering supplies, there is still a difference, albeit less significant. 60% of SMEs are using mobile phones compared to 28% using fixed lines.

The crossover between business and personal is also more pronounced amongst mobile phone users compared to fixed line phone users. 18% of SMEs use the fixed line phone for personal use, compared to 80% of mobile phone users. This could be a reflection of the mobility that allows for more flexibility when using the mobile phone compared to a fixed phone. It could also mean that the mobile phone is seen as a shared resource compared to the fixed line phone. 66% of owners of SME’s use the mobile phone the most compared to 8% of owners using a fixed line phone.

A much higher percentage, 60% of the businesses do own a mobile phone but no fixed phone compared to the opposite (i.e. those that do own a fixed phone but no mobile phone) 7%. Adding to the perception that mobile phones play a more important role within SME’s is the feeling that fixed line phones are too expensive with 26% of SMEs stating that fixed line phones are too expensive compared to 19% stating that mobile phones are too expensive. Interestingly, of those SME’s that do not have a fixed line phone, over 11% state that it is a service that is not available in their area, compared to just over 3% who say the same thing regarding mobile phones. This reflects better service coverage for the mobile phone, another factor that may explain its ubiquity.
94% of SMEs rate a mobile phone as important or very important compared to 78% that rate a fixed line phone or 77% a computer as important to very important for their business activities.

While 19% of SMEs believe that there is no need to own a computer, 32% also believe that they are too expensive. This jumps to a higher percentage on both counts when taking computers with Internet connections into consideration. 43% of SMEs believe that there is no need for a computer with an Internet connection and 35% believe that they are too expensive. The role that cyber cafés play in supporting SME’s is underscored by the fact that 18% of SMEs that do not have an Internet connection do use cyber cafés.

77% of SMEs rate a computer as important to very important, but this drops dramatically when rating the importance of an Internet connection. 41% believe that an Internet connection is important to very important. Compare this to the 94% that rate the mobile phone as important to very important and its clear that the mobile phone has overtaken the computer as the most practical method of running a business. However, the mobile phone clearly has its limitations because 39% of SMEs that do not have access to computers anticipate using them in the future and a further 29% anticipate using the Internet in the future. This speaks to the possibility that computer and Internet usage compared to mobile phone usage is a question of cost and accessibility rather than usefulness.

ICT Usage Intensity among businesses is calculated using the ICT Usage Index and the ICT Possession Index (usage/possession). The ICT intensity shows the extent to which businesses employ ICT for business purposes in terms of what they have in ICT equipment and facilities.

The ICT possession index reflects what businesses have in terms of ICT equipment and facilities. A point of 1 is given for each the ICT devices owned by a business and a maximum of 6 points is obtained should a business have all 6 items. Table 6 below shows this index.

**Table 6: ICT Possession Index**

<table>
<thead>
<tr>
<th>Index value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, the business has one or more working telephones!</td>
</tr>
<tr>
<td>Yes, the business has one or more working mobile!</td>
</tr>
<tr>
<td>Yes, the business has one or more working fax machines!</td>
</tr>
<tr>
<td>Yes, the business has one or more post boxes!</td>
</tr>
<tr>
<td>Yes, the business has one or more working computers!</td>
</tr>
<tr>
<td>Yes, the business has an Internet connection!</td>
</tr>
<tr>
<td>Maximum Value</td>
</tr>
</tbody>
</table>
The ICT Usage Index was developed by awarding one point for any employment of ICT facilities and equipment in carrying out business transactions. This gave a maximum total of 15 points should a business have and use all the ICT facilities and equipment mentioned in this study for business purposes. This is shown in Table 7 below.

### Table 7: ICT Usage Index

<table>
<thead>
<tr>
<th>Description</th>
<th>Index value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, the business uses the telephone to communicate with clients and customers!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the telephone to order supplies!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the mobile to communicate with clients and customers!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the mobile to order supplies!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the fax to communicate with clients and customers!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the fax to order supplies!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the post box to communicate with clients and customers!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the post box to order supplies!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the computer to communicate with clients and customers!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the computer to order supplies!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the Internet to communicate with clients and customers!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the Internet to order supplies!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business sends SMS or Text Messages for Business Purposes!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business receives SMS or Text Messages for Business Purposes!</td>
<td>1</td>
</tr>
<tr>
<td>Yes, the business uses the internet for business purposes!</td>
<td>1</td>
</tr>
<tr>
<td>Maximum Value</td>
<td>15</td>
</tr>
</tbody>
</table>

The results of the Kruskal-Wallis Test for the mean comparisons for the ICT Possession, ICT Usage and ICT Usage Intensity (usage/possession) shown in Table 8 below indicate that there is a significant difference in the mean ranks for all three indices across formality. Formal SMEs possess more ICT equipment and make more use of it than semi-formal ones; and semi-formal SMEs more than informal ones. However, informal SMEs have the highest ICT Usage Intensity, followed by the semi-formal ones, with the formal SMEs having the least. The high ICT usage in the formal sector can be attributed to their higher employment of such facilities in carrying out their business activities.
Table 8: Mean Rank comparison using Kruskal Wallis Test for ICT Possession, Usage and Usage Intensity (Usage/Possession) Indices with grouping variable = formality

<table>
<thead>
<tr>
<th>Formality Categories</th>
<th>N</th>
<th>Mean Rank</th>
<th>Chi-Square</th>
<th>df</th>
<th>Asymp. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ICT Possession Index</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal</td>
<td>151</td>
<td>119.82</td>
<td>95.276</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>Semi-formal</td>
<td>139</td>
<td>203.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>61</td>
<td>251.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>351</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ICT Usage Index</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal</td>
<td>151</td>
<td>126.01</td>
<td>68.127</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>Semi-formal</td>
<td>139</td>
<td>205.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>61</td>
<td>231.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>351</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ICT Usage Intensity Index</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal</td>
<td>148</td>
<td>185.21</td>
<td>7.121</td>
<td>2</td>
<td>0.028</td>
</tr>
<tr>
<td>Semi-formal</td>
<td>138</td>
<td>174.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>61</td>
<td>144.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>347</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ICT Impact on Profitability

Informal businesses have a higher profitability in terms of fixed assets employed than semi-formal ones, and semi-formal ones a higher profitability than formal businesses as highlighted in Table 9 below. This confirms intuition. Informal businesses usually operate at a far higher gross profit than formal ones. They are not bound to minimum wages, can hire casual labour whenever needed, pay mostly no tax and operate on less infrastructure than formal businesses.

Table 9: Mean Rank comparison of profitability (after tax profit divided by total fixed assets) by formality categories

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal</td>
<td>151</td>
<td>178.93</td>
</tr>
<tr>
<td>Semi-formal</td>
<td>139</td>
<td>177.87</td>
</tr>
<tr>
<td>Formal</td>
<td>61</td>
<td>164.48</td>
</tr>
<tr>
<td>Total</td>
<td>351</td>
<td></td>
</tr>
</tbody>
</table>

Kruskal Wallis Test

<table>
<thead>
<tr>
<th>Grouping Variable: formality index categories</th>
<th>Chi-Square</th>
<th>df</th>
<th>Asymp. Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.959</td>
<td>2</td>
<td>.619</td>
</tr>
</tbody>
</table>
Profitability here is defined as after tax profits (which is the same as pre tax profits for most informal businesses since they do not pay taxes) divided by the total value of fixed assets. The Kruskal Wallis test shows that the result obtained is not significant.

Chowdhury & Wolf (2003) used modified Cobb-Douglas production functions to investigate labour productivity and returns for a survey conducted in Tanzania and Kenya. Their main findings indicate that there is no significant impact of ICT investments on return performance of SMEs and that indeed investments in ICT negatively impact labour productivity. Their results however ignore the fact that informal businesses are likely to have less ICT equipment than formal ones, yet their profitability is higher as shown in Table 9 above.

Also using production functions might not be an appropriate approach for dealing with SMEs. Production functions assume a certain relationship between inputs and outputs (Hill et al., 1997). Most SMEs do not produce one product or service but generally deal in a variety of things, concentrating on whatever generates most money at any given time. An Internet Café might offer secretarial services and sell digital music at the same time. A lock and key smith can also fix cars on the side or do spray painting. It is often difficult to judge whether an SME is a manufacturing, a service or a retail/wholesale business since borders lines are fuzzy.

Wolf (2001) mentions that the focus on production processes might be too narrow and that ICTs might exert their influence through product-quality improvements and improved services. ICTS might additionally help SMEs in the administration of their businesses and enhance procurement and marketing processes.

A better approach might be to focus on SMEs as a business rather than on single production processes. Care should be taken to account for the direction of causality since businesses might purchase more ICTs as a consequence of good profits or conversely, their profitability might increase due to leveraging ICTs.

OBSTACLES TO ICT ADOPTION

The biggest obstacle cited as a hindrance to wider ICT usage within SMEs is the high cost of ICTs. There is need for government to implement policy and regulatory changes aimed at reducing the cost of ICTs and encouraging more of their usage within SMEs in order to foster economic growth and employment.

The second obstacle cited was the lack of need for ICTs. While we don’t necessarily agree with this, given the high utilisation of ICTs exhibited within the sample through for example the use of mobile telephony, we postulate that this is meant to reflect the lack of adequate applications and services that leverage this infrastructure to provided added value to the SMEs.

Lack of awareness and knowledge about what ICTs can do to help businesses become more efficient in their operation as well as the lack of financial resources to invest in ICTs were the other major obstacles highlighted in the survey. Limited access to financial resources was particularly relevant for informal SMEs where about 10% of the businesses indicated it as the main obstacle to ICT adoption compared to about 4% of the semi-formal and 5% of the formal businesses as shown in Figure 10.
Interestingly, the lack of awareness and knowledge about ICTs was highlighted in a similar manner across formality with informal businesses at 8%, semi-formal at 7% and formal at 7% despite the variance in education levels of SME owners, where owners of informal businesses are on average less educated compared to owners of semi-formal and formal SMEs, as indicated in Figure 12.

**Figure 12. SME owners’ highest education level attained**
At the time of this survey, power failures and outages was one of the least mentioned obstacles, but things have since changed significantly with more businesses experiencing daily power outages as a consequence of inadequate power supplies within the country to cope with the growing demand. Informal and semi-formal businesses were the least concerned about power failures at 1%, with formal businesses citing it as an obstacle 4% of the time.

Discussion

While the role of SMEs in promoting economic growth is gaining increasing support, their precise contribution to providing employment and alleviating poverty remains unclear. Part of the problem is that there is a paucity of data around SMEs. There is no centralised database on SMEs making representative surveys very difficult. In addition, there is a general lack of basic statistics such as how much employment SMEs provide, their average turnover, their profitability, life-span and product range.

Despite Uganda’s tremendous growth in terms of access to ICTs, it is not commensurate with the growth in terms of economic development. While this is generally attributed to the high ICT service costs vis-à-vis our income levels, we would like to posit that it is mainly due to the lack of appropriate applications and services that SMEs can use to leverage this infrastructure for more economic benefit. The ubiquitous mobile phone is used more for conversation than say access to related business information and relevant data sources.

In this paper, we have reported on a small and medium enterprise (SME) Survey carried out in Uganda as part of the Research ICT Africa network (RIA). The paper demonstrates that the lack of significant impact on company performance can be attributed to the failure to distinguish between the formal and informal sector. Using a formality index, we classify respondent SMEs into informal, semi-formal and formal businesses. While it is clear ICTs are input factors for both formal and informal SMEs with a positive correlation to profitability, use of ICTs might be the cause or a result. ICTs could be the result in the sense that SMEs purchase more ICTs as a consequence of good profits or a cause in the sense that SME profitability increases due to ICT usage.

We do find that informal businesses have a higher profitability in terms of fixed assets employed than semi-formal ones, which in turn have a higher profitability than formal businesses. This is understandable, given that increasing formality encumbers a business, ensuring that it must follow certain laws in its operation.

The impact of these findings from a policy point of view require further study which is outside the scope of this paper. Despite this, we would like to make a few observations that may guide related future work. Firstly, the mobile phone has overtaken the computer as a tool in supporting the running of a business. Given their prevalence and accessibility, well-designed phone or SMS based business information systems may have a significant impact on the profitability of SMEs.

Secondly, the traditional focus on formal businesses particularly in terms of financial support undermines the role that the informal sector plays in the economy. The fact
that it is more difficult to measure does not lessen its impact. In this light, the fact that informal businesses are more profitable than formal ones raises the issue that some businesses might prefer to be out of the formal environment. The move towards mobile telephony given its (current) limited functionality might indicate that the benefits of moving into the formal economy are out-weighted by its costs. A focus on the reduction of input costs of informal businesses (such as ICT costs) could encourage a faster migration from informal to formal sector.

Thirdly, focusing on mobile technology has both advantages and disadvantages. Advantages may include; providing an SME with a low cost base coupled with the ability to easily communicate with suppliers and customers given the financial constraints, leveraging people’s familiarity with mobile phones as opposed to having to teach them how to use computers, utilising an already existing and extensive network infrastructure that is largely used for entertainment or other social aspects. A latent disadvantage may be the limited functionality mobile technology currently offers. Potential mobile phone information systems could be used to track inventory, provide cash flow and income statements or even more basically help keep better business records.

Fourthly, we should start to move beyond collecting data on ICT possession (infrastructure, penetration, etc) and usage (purpose, barriers, etc.) within SMEs and start to collect and analyse data on the impact of various ICTs. That way, we shall be in position to make more useful interventions and better justify the costly investments that SMEs relate with ICTs.

Acknowledgements

We would like to acknowledge all the SME owners who cooperated with us during the survey, the enumerators for data collection, Mariama Deen-Swarray for data analysis and ResearchICTAfrica for supporting the survey.

References


PART SIX

ICT and Education
An Exploration of the factors that affect Quality Assurance Decision Making in Higher Education

Benedict Oyo and Ddembe Williams, Department of Information Systems, Makerere University, benoyo@gmail.com, d.williams@cit.mak.ac.ug, P.O. Box 7062 Kampala, Uganda

Increased use of quality assurance in the higher education context has made a profound impact in the areas of research, teaching and administration. In response to the necessity for reforming higher education system in line with the needs and expectations of the community and business sector, new approaches and quality assurance practices have been sought in higher education. This paper argues that system dynamics modelling approach is a valuable alternative for quality assurance research in higher education. This approach is emphasized owing to the fact that, a greater portion of university quality problems and their solutions do not have quantitative foundations mostly because such systems involve qualitative elements that are difficult to quantify and model using other approaches. A framework is used to demonstrate the theoretical value of system dynamics modelling in combining diverse factors responsible for quality assurance in higher education and arriving at a consistent decision about the quality of awards.

Introduction

Debate over the extent to which quality assurance techniques can be applied in higher education has been ongoing for over a decade in many institutions around the world (Cheng, 2003). The concept of internationalism and effectiveness in higher education have brought about new dimensions as governments in most parts of the world have considered clear agenda for higher education, acknowledging that more explicit assurance about quality is needed (Mizikaci, 2006).

With increased demands for higher education (HE) all over the world, HE institutions are faced with an opportunity and a challenge to maintain quality of education. To this end, national approaches to extra-educational scrutiny have been established. Although national bodies concerned with licensing of HE institutions exist, quality assurance (QA) in higher education entails above and beyond, government policies and national bodies’ regulation, but requires commitment to quality undertakings independently by these institutions. Education, and in particular higher education itself, is also being driven towards commercial competition imposed by economic forces Seymour (1992). According to Freeman (1993) this competition is the result of the development of
global education markets on the one hand, and the reduction of governmental funds that force public organizations to seek other financial sources, on the other.

Universities now actively compete against each other for students and resources. In the developing countries and Uganda in particular, in the last decade alone, the number of public universities has grown from one to four and nationally recognised private universities from none to sixteen. For the case of UK, Telford and Masson (2005) reported that UK, national newspapers publish annual league tables of all UK Universities, the score for each university being based on a wide range of criteria and these are seen as having an impact on student recruitment.

Feigenbaum (1994) believes that “quality of education” is the key factor in “invisible” competition between countries since the quality of products and services is determined by the way that “managers, teachers, workers, engineers, and economists think, act, and make decisions about quality”. In a related analysis, Jackson (1997) noted that, at a time of diminishing unit resources to support teaching and learning, the totality of the steady-state environment in which quality assurance activities were undertaken less than a decade ago has been transformed into one of continuous change, adaptation, experimentation and innovation.

There is an increasing tension as institutions are caught up in a dilemma assuring quality of HE in an atmosphere of large student population. The solution to this dilemma is not to reduce the student population but to explore new QA methods and frameworks that allow for coexistent of quality and large numbers. The framework proposed in this paper, incorporates the use of system dynamics (SD) methodology in studying QA challenges and problems in HE. This method is considered ideal for studying the dynamics of soft variables such as; student quality, teaching quality, and research quality that underpin quality of higher education, as well as predicting effects of institutional quality initiatives.

Quality Assurance Problems In Higher Education

Administrators of contemporary universities face the challenge of maintaining the quality of the fundamental functions of a university, namely teaching, research, academic and professional service, while trying to serve the students, under the pressure of limited resources in terms of faculty, facilities and income (Gürüz et al. 1994; Ulusoy, 1994). These problems and possible solutions have been studied both on macro and micro levels by many researchers. Some of these studies were based on certain quantitative research, for instance, (Mahmoud and Genta, 1993; Saeed, 1993) yet, a great portion of literature on university problems and their solutions do not have quantitative foundations, mostly because such systems involve qualitative (human) elements that are difficult to quantify and model.

The link between the qualitative and the quantitative aspects of the problem is important and therefore deserves more research (Mahmoud and Genta, 1993). A possible approach, in order to model and explore this link, is system dynamics methodology that employs a set of techniques that allows quantitative and realistic representation of variables that are typically non numerical in nature but qualitative (Luna-Reyes and Anderson, 2002). Further more, Jehazkova and Westerheijden (2001),
in their analysis of a next generation of quality assurance in higher education found that problems of quality assurance in higher education relate to the following scenarios:
(i) Serious doubts about educational standards and outcome awards
(ii) Doubts about efficiency of high education systems and institutions
(iii) Doubts about innovation capacity and quality assurance capacity of the institutions
(iv) Need to stimulate sustainable quality culture in institutions.

They did not however, determine the challenges of evaluating these scenarios to foster decision-making. Telford and Masson (2005) call for a shift from the traditional peer-review-based assurance system to new models that can link quality in higher education with cost effectiveness. In another call, Mizikaci (2006) challenges governments at different stages of development of their higher education sectors, to adopt more explicit assurances about quality. Faced with quality demands, higher education needs to explore new integrated approaches to QA with focus on quality of awards that should be reflected by confidence in the system. The framework proposed in this paper, demonstrates the theoretical value of system dynamics modelling in combining diverse factors responsible for quality assurance in higher education and arriving at a consistent decision about the quality of awards. More specifically, the focus of SD in this case is in its capability to; model different possible scenarios, handling non linear relationships and incorporating feedback loops as well as coping with high levels of variability within/between the modelled variables.

Quality Assurance Approaches And Systems In Higher Education

A distinctiveness in approach to quality assurance in literature such as continuous quality improvement (Roffe, 1998), benchmarking (Camp, 1989; Kempner, 1993), accreditation (Harman & Meek, 2000; ENQA, 2004), total quality management (Ewel, 1983; Sherr and Lozier, 1996), satisfaction/dissatisfaction paradigm (Tolman, 1932, p.428 as cited by Telfor and Masson, 2005), peer review (Green, 1993), internal, interface and future quality assurance (Cheng, 2003) and system dynamics (Owlia & Aspinwal, 1996; Salhieh and Singh, 2003; Barnabe, 2004) has been applied in higher education. In practice, some universities have developed computer applications for QA, these include;

i) Quality Measurement System (QMS2000) that was developed by the University of Louisville in partnership with Dey Systems, a Louisville technology company that specializes in quality and measurement solutions. QMS2000 is a relational, interactive information system that includes data from students, alumni, faculty, staff, and employer satisfaction surveys that are linked to corresponding databases at the university. QMS2000 is on-line, operating in a networked; client-server environment that permits licensed users access to designated components of the system at any time from designated desktops at the university. QMS2000 users generate reports and perform advanced statistical analyses drawing from the databases Welsh and Dey (2002). Appendix A shows the QMS2000 architecture.
ii) UNIGAME, a dynamic simulation game for strategic University management was developed by Barlas and Diker in 2000. Their model focused specifically on long-term, dynamic, strategic management problems, such as growing student-faculty ratios, poor teaching quality and low research productivity. It yielded numerous performance measures about the fundamental activities in a university: teaching, research and professional project activities. The simulation model was built using system dynamics methodology and was validated/verified by standard tests, using data from Bogazici University, Istanbul, Turkey. Necessary changes have been made to turn the model into an interactive one and the gaming interface was built using VENSIM software. They claimed that the research results obtained suggested that UNIGAME promises not only a useful technology to support strategic decision making, but also a laboratory for theoretical research on how to best deal with strategic university management problem. Appendix B shows the global sector used to develop UNIGAME.

In order to meet the current QA demands in HE and prepare for future challenges, there is need for integrating QA approaches in HE, a possible tool in achieving this is SD. Cheng (2003) found that, most quality assurance mechanisms depend on one or a combination of a limited number of methodologies, the most important of which are self-studies or self-evaluation; peer review by panels of experts; use of relevant statistical information and performance indicators; and surveys of key groups, such as students, graduates and employers. He further added that at national level, most common forms of assessment are ‘horizontal’ reviews of disciplines within the institutions and ‘vertical’ evaluations of institutions by external bodies against defined minimum standards, which directly affect the outcome award.

In this paper we have proposed a framework that integrates conceptual analysis of some of the different QA approaches with factors that contribute to HE quality in terms of teaching quality, research quality, and student quality. The framework assumes that a licensed HE institution starts with minimum strategic quality planning and in the course of its operation, new quality assurance strategies are developed in research, teaching and students’ sections, which as a whole contribute to quality of awards.

Factors that affect QA decision making in higher education

In this section we explore factors that affect QA in terms of quality of outcome awards, quality assurance capacity in higher education, and quality culture in higher education.

Quality of outcome awards

This is affected by;

• Availability of learning resources (infrastructure)
• Quality of teaching
• Students support and guidance
Advances in Systems Modelling and ICT Applications

- Curriculum design, content and organization
- Student evaluation of course content and delivery
- Student projects and nature of supervision

**Quality assurance capacity of a faculty**

This is based on analysis of research and teaching quality in HE by the following parameters;
- Funding priorities
- Infrastructure
- Volume and quality of research
- Patents and awards
- Technological transfer

**Quality culture of a faculty**

This is determined by analysis of QA policies and techniques in the following categories;
- Entry qualification
- Internal self evaluation
- National and international accreditation
- Peer review
- Benchmarking to facilitate transfer of good practice
- Community service, such as consultations and training

In order to establish factors that augment QA in higher education and their levels of effect, system dynamics modelling approach has been proposed in this paper. Through this approach, a dynamic hypothesis for QA has been demonstrated.

**The Value Of System Dynamics Modelling**

**Definition**

SD is the study of dynamic behaviour of a variety of complex systems, generally in the domain of human activity systems such as business, organisational management and other social systems (Coyle, 1999; Goodman, 1989). It has been used to address practically every sort of feedback system (Wilson and Pettigrove, 2002). As a quantitative method, SD uses mathematical formulas with numerical values for parameters under study. The focus of a SD study is not a system but a problem (Forrester, 1985 as cited by Owlia and Aspinwal, 1996). The problems that one addresses from the perspective of SD have at least two features in common. First, they are dynamic (involve quantities which change over time). Secondly they involve the notion of feedback where, item x affects another item y and y in turn affects x perhaps through a chain of causes and effects. Studying a link between x and y and independently the link between y and x
cannot predict how the system will behave. Only the study of the whole system as a feedback system can lead to correct results.

Illustration

Galbraith (1999) noted that, a university in its operating environment is an example of a complex social system, since it is characterized by the interaction of closed chains of causality (feedback loops) that together define the system structure and hence how the system behaves over time. Feedback loops are of two kinds, positive (reinforcing) loops, and negative (balancing or stabilizing) loops. Reinforcing loop (R) is a representation of growing or declining actions. In this case, a positive loop response reinforces the original perturbation. Balancing loop (B), also referred to as counteracting, is a goal seeking feedback process that seeks stability or return to control.

When a feedback loop responds to a variable change opposing the original perturbation, the loop is negative or goal seeking (Goodman, 1989).

In a university, an example of a positive loop is the process by which an increase in infrastructure improves quality of teaching, which in turn improves on courses and content. Similarly an example of a negative loop is the process by which an increase in volume and quality of research decreases the average student fees, and an increase in average student fees in turn increases institutional revenue. An illustration of these two feedback loops is shown in Figure 1. As indicated, the polarity of the arrow pointers represents an increase (+) or a decrease (−) on the influenced factor as a result of an increase in the influencing factor.

Dynamic Hypothesis Of Quality Assurance

The interactions as in Figure 1 show the effect of student fees, research quality, staff quality, and courses and content on the quality of outcome award as a major determinant for QA. Williams (2003) emphasises that a test of a good theory lies in its ability to predict, shape or change the surrounding world. Figure 1 offers a useful basis for research on quality assurance since it provides a theoretical analysis of the factors that determine the quality of outcome award. The feedback structure in Figure 1 contains four dominant feedback loops, of which three are reinforcing loops (R), and the other one is a balancing loop (B). Crossed influences with a “||” indicate a time delay in the same direction.
The interactions among these key variables throughout the QA process determine the quality of outcome award of a HE institution. The dynamics as illustrated in Figure 1 can be categorized in two major loops namely:

- **Award-quality loops (R1 and R2)**
- **Cost-quality loops (B1, R3 and R4)**

**Award-quality loops (R1 and R2)**

The reinforcing loop R1 is a major loop that can be used to evaluate the quality of outcome award by verifying the effect infrastructure on the design of courses, and students’ progress. The loop R2 is a major loop contributes to outcome award by increasing quality of teaching, which in turn increases the reputation of the institution overtime, which creates confidence in the outcome award.

**Cost-quality loops (B1, R3 and R4)**

Balancing loop B1 is a dominant loop that combines cost with quality, i.e., this loop demonstrates that increase in quality of research creates new avenues for income and would lead to decrease in student fees while maintaining quality since an institution is defined by its research output. On the other hand, reinforcing loops R3 and R4 demonstrate that increase in institutional revenue increases infrastructure. Further increase in staff quality increases quality of teaching which in turn increases an institution’s reputation. The issues of concern in this dynamic hypothesis have been further illustrated in the conceptual framework as shown in Figure 2.
Modelling qualitative data

Solving soft problems using system dynamics modelling as for the case of quality assurance in HE, raises concern since SD uses mathematical formulas with numerical values for parameters under study. However, the question for system dynamics appears not to be whether to use qualitative data but a question of when and how to use it. What method should be used to gather data? From whom should data be gathered? At what stage in the modelling process might qualitative data be an appropriate, perhaps even a preferable information source? How should we analyse and use qualitative information, expert judgment, and group consensus? How is qualitative data linked to model structure? (Luna-Reyes and Anderson, 2002).

The lack of an integrated set of procedures to obtain and analyse qualitative information creates, among several possible problems, a gap between the problem modelled and the model of the problem. That is to say, it is not always easy to understand the links between reality and the assumptions or formulations in the model. This gap is more noticeable especially when the model involves the use of soft variables, such as “student satisfaction”, “teaching quality”, “pressure to decrease cost of education”, and “perceived faculty productivity”.

The problems associated with the quantification and formulation of qualitative variables has led some experts in the field to the conceptualisation of a qualitative system dynamics practice (Wolstenholme, 1990). In some cases, the uncertainty associated with the quantification of qualitative variables has caused experts to believe that the results from ensuing simulations could be misleading, or at least, very fragile (Coyle, 2000). He further stated that qualitative data dominates the pre-modelling stage, stopping short of the actual formulation stage at the point of “system description”. Wolstenholme (1990) had in an earlier publication shared this point of view by considering that the phase of diagram construction and analysis could be considered itself as a qualitative branch of system dynamics. The use of qualitative data has, at the very least, been cause for debate about whether or not to simulate based upon qualitative materials. In QA study, qualitative data can be collected and analysed through case study, followed by formulation of a simulation model to provide prescriptive insights. This approach is underpinned by Dynamic Synthesis Methodology coined by Williams (2004), which combines case study research and simulation modelling methods.

Relevance Of System Dynamics Modelling To Quality Assurance In Higher Education

The increasing demand for HE globally has created a dilemma on how to assure quality as well as to provide means of measuring quality of awards given by these institutions. This paper proposes a SD modelling approach for studying QA and quality of award in HE and uses a framework (Figure 2) to demonstrate the interaction between key variables in assuring quality of outcome awards. SD has been considered most relevant than other approaches in this study due to the following reasons:

• Deals with non-linear relationships and incorporates feedback loops as illustrated in Figure 1.
• The study of QA may start as partially structured and becomes more refined, structured in later stages as more understanding is gained from the process.

• Offers capabilities to model different possible scenarios. i.e HE quality outcome award can be modelled sequentially, concurrently or nestedly through research or teaching or student output.

• Copes with high levels of variability within/between the modelled variables, including human behaviour.

• Can be used to identify key variables to avoid unnecessary data collection and in some cases, expert opinions for fitting available data can be sought, as illustrated in figure 3.

Conceptual Framework

A major goal of development of quality assurance techniques in higher education is to create an increasing level of quality learning initiatives, research and students. The main concern in this paper is the identification and analysis of the dynamic factors that affect; teaching quality, research quality and student quality vis-à-vis a higher education institutional quality assurance policy and techniques.

The proposed framework in Figure 2 has been derived from dynamic hypothesis in Figure 1. The framework assumes that a licensed higher education institution starts with minimum strategic quality planning underpinned by: economic, political, environmental and societal structures where the institution is located. In the course of an institution’s operation, new quality assurance strategies are developed in research, teaching and students’ sections respectively. Quality assurance techniques are employed at stages marked by arrows 1 to 3 and a summative study of the whole quality assurance process is done using system dynamics modelling approach. Analysis of results using this approach defines the institution’s quality output that is then fed back on the strategic quality planning for continuous improvement.
Quality of higher education is perceived as a composition of planned supply of higher education by institutions based on evidence of quality assurance indicators. A quality assurance indicator in this paper is referred to as the outcome of institutional education relative to the minimum standards against which education is provided by the institution. Institutional policy makers and staff need to understand the dynamic interactions between quality assurance techniques and policies, research and student quality as illustrated in the conceptual framework. The usefulness of this framework is in its ability to provide a conceptual analysis of QA factors that determine HE quality output through integration of research quality, teaching quality, students’ quality and techniques that can be modelled using SD. Variables necessary for data collection have been identified in Figure 2 will offer capabilities to model quality in HE under defined priorities in the areas of research, teaching or student domains.

**Future Work And Conclusion**

Quality assurance has emerged as the single most critical factor for the survival and growth of an institution. In most higher education institutions, a number of quality assurance approaches have been used and yet quality assurance problems have remained persistent. In many circumstances, HE institutions have implemented centralised QA mechanisms with primary focus on management satisfaction rather than academic development, improvement and innovativeness.

In this paper, we have proposed a framework that provides an environment for undertaking QA activities in HE with focus on: research quality, teaching quality, and student output quality, alongside an institution’s quality planning. The framework
has been used to demonstrate the theoretical value of system dynamics modelling in combining diverse factors responsible for quality assurance in higher education and arriving at a consistent decision about the quality output of a higher education institution.

To further operationalise the proposed framework, we intend to validate the dynamic hypothesis for QA in Figure 1 by quantifying the identified factors and subsequently carrying out a simulation. Analysis of the simulation results will be made in order to gain insights of the current state of quality of awards for an institution as well as to predict future quality by varying quantified values of the identified factors.

References


Barnabe, F. (2004), From Ivory Towers to Learning Organizations: the Role of System Dynamics in the ”Managerialization” of Academic Institutions, Department of Social and Business Studies University of Siena, Italy


Ishida


Mizikaci, F. (2005), A systems approach to program evaluation model for quality in higher education, Quality Assurance in Education Vol. 14 No. 1, pp. 37-53


Technology and specifically the web-based, is a moving target – web-based learning and teaching technology is being used and developed at rapidly. Diverse methods that are adaptable to changing features of technology as well as a guiding methodologies to select which technology features should be fixed or flexible in the design are. This paper reports on a process and findings of using development research design and methodology to design an instructional design subsystem. The development research used involves a review of the literature with the aim of exploring, analyzing, integrating and synthesizing the broad field of learning and instructional theories, paradigms and best practices with the design of the instructional design subsystem. The complacent cases are abstracted and used in the design of the subsystem. The research design perspective enables making of principled design approaches informed by knowledge of instructors’ particular generalizing processes and the research in instructional design. This paper also discusses how the subsystem can be improved by rigorous attention to the formulation of theory, if more iterations are used, each leading to revision, refinement and modification of the instructional design subsystem and perhaps increasing the appeal of the theory in-use.

Introduction

The need to offer research that breaks from the traditional educational research that does not often lead directly to practical advances or build strong linkages between research-based insights and improved practices motivated the study reported in this paper. The break was meant to realign educational systems by improving the coordination between research, design, development and practice (Burkhardt & Schoenfeld, 2003). Developmental research reported in this paper is a process, or research approaches whose intent is to produce knowledge with the main aim of improving the processes of instructional design and development. Developmental researches in general are “those studies that involve the production of knowledge with the ultimate aim of improving the process of instructional design, development, and evaluation” (Richey, Klein & Nelson, 2004: 1099).
The developmental research process starts with a researcher choosing methods based on the questions to be addressed. However, the questions, issues or topics of the study themselves may change as the researcher’s “conception of the reality of the “world” being studied changes” leading to research methods being “adjusted, expanded, modified or restricted on the basis of the information acquired” (Savenye & Robinson, 2004: 1050). The goals of developmental research as developing creative approaches to solving human teaching, learning, and performance problems while at the same time constructing a body of design principles that can guide future development efforts. The research methods used in developmental research are observational, correlational, experiments, quasi-experiments, grounded theory, and case studies (Reeves, 2000).

The notion of “developmental research” by name is still unclear and more often being confused with research concerned with the study of human growth and development. Richey et al (2004), in a bid to clear the confusion, described the simplest form of developmental research as used in instructional design as either: -the study of the process and impact of specific instructional design and development efforts; or a situation in which someone is performing instructional design, development, or evaluation activities and studying the process at the same time; or the study of the instructional design, development, and evaluation process as a whole or of particular process components. (Richey, Klein & Nelson, 2004: 1099)

Developmental Research For This Paper

The main characteristics of developmental research can be understood from the purpose, focus and techniques of the research.

The purpose of the study used here identified as optimizing and gaining sound basis for instructional development activities. This is achieved through the design of a prototypical instructional subsystems and designing methodological directions of the design, development, evaluation and refinement. This approach addresses the issues of the validity and/or effectiveness of a particular technique or model, conditions and procedures that facilitate its successful use, explanations of successes or failures encountered in its use, a synthesis of events and/or opinions related to its use. The effectiveness and validity issues are used progressively in the enhancements of the technique or model or in building a new one.

The focus of the developmental research is to reduce human problems or complexities in their daily life or work processes through the design of a product or an activity. The focus of this study is to design a web-based instructional system that can ease or reduce the complexities of the instructional design activities.

The techniques used are mainly extensive search and refinement of the existing descriptive research in instructional design that would lead to streams of knowledge that can be used in the design, development and production of the subsystem. These techniques have three main phases. In phase one, the problem and its inconveniences are derived, assessed and documented. The second phase is the definition of the characteristics of the final states. This, when put in congruent with the model developed usually forms a plan of what need to be done and when. The final phase is the actual
development which is also incorporates testing and production of the developed product. The final phase deals with the use and decision making regarding the use of the product together with its future improvements.

**Developmental Research In This Study**

The developmental research approach is a highly cyclical and iterative process, each iterative sequence forming a basis for refinement. Developmental research used take the form of design experiments. A “Design experiment” as first described by Brown (1992) is an eclectic approach to educational research. The choice of design experiments was necessitated by the pragmatic approach of design experiments that deals with complexity of contexts, for example, the identification of the theories, assumptions, and methods underlying instructional design whose outcome is a practical visible action or tool. It is open-ended and multi-perspective with its meaning being usefully realized and considered within the contexts it is being applied.

The research design applied in this study had four iterative phases viz, analysis phase, model design phase, model implementation phase and model try-out and evaluation/validation phase.

**Research Design, Results And Discussions**

Due to the nature of development research, the research design results and discussion are presented together in this paper

**Defining the problem**

The research topics focused on the main area of instructional design in web-based learning management systems: analysis, design, development, and implementation and to some extent evaluation of model or process, program, or tool; and identification of the general development principles of situation-specific recommendations. Three major items are discussed under the problem definition and these are focus of the problem, framing of the problem; and identifying limitations.

**Focusing on the problem**

Focusing on the problem involves concentrating on a particular aspect of the design, development, implementation and evaluation process, without focusing on the variables, or the type of media to be used (Reeves, 2000). This establishes the research parameters, determines how the research was to be conducted, and how much of the design process is to be addressed.

The problem focus is on solving problems associated with the user of web learning management systems. These problems are lack of:

- Web-based instructional systems that are based on well grounded research (Hardre, 2003; Roschell, Kaput, Stroup & Kahn, 1998).
- An in-built expert instructional design module to guide instructors during the creation of the learning materials. (Roschell et al, 1998).
3.1.2 Framing the problem

Framing the problem in this case involves looking for means to counter the limitation identified when focusing on the problem. In this study, it took a lot of discussion with peers and mentors. After the discussion, there were positions that were made clear and identified that should be included in the problem statement. While the approach to the solutions was the implementation of a web based instructional design subsystem that implements and uses the existing theory and models on instructional design, issues about reporting on a system and the evaluation of the completed system and learning materials developed through its processes were contentious. However, with the agreement that the aim of the research was to come up with a designed product that uses research on instructional design that could be used to help instructors during the process of creating learning materials for the web.

The research question was iteratively refined and revised. The revision came up with the following question which was adopted as the main research question for this study:

**How can instructional design for web-based learning be optimized through the use of existing research?**

**Identifying limitations of the study**

The aims of identifying limitations in developmental research are to recognize the unique conditions or confines that arise due to the context specific nature of this kind of research. Identifying the limitations of the study was as contentious as identifying the problem itself. Of particular interest in developmental based researches are the unique conditions or limitations that arise due to the context specific nature of this kind of research. These limitations affect the extent to which the generalization of the conclusions of the study may be done. With this in mind, the following assumptions were identified as:

- The system designed will improve the modularity, quality, speed, and ease of developing reusable learning materials for web based courses.
- Though there might be no time to test the quality of the learning materials developed, it is projected that the resultant learning materials will be of high quality, and will improve the efficiency and effectiveness of learning.
- An instructional design model to be designed will be identified from literature. The identification of the model will be through critical evaluation and examination of the models available.
• The model identified will be ideal or adaptable and generalizable for use in the design of instructional materials for all the disciplines and subjects which have the potential of being offered online.

• The instructional design model identified can be abstracted and reduced to simple computable modules.

During the progression of the study these limitations are taken into consideration at all stages.

**Review of related literature**

Review of the literature dwelt on: the instructional and learning theories – the different paradigms used in the design of learning and instructional materials; the instructional models - the various ways researchers and practitioners have tried to give a structure or systems or an approach to the process of instructional design; the widely used instructional methods; the various learning styles and learner preferences for web learning; the theories and the models of instructional design together taking into considerations the instructional methods and the various issues of the learners’ learning styles; and the implication the whole literature that has been identify has for the design of web-based instructional design system.

### 3.3 Research procedure

Even though developmental researches occurs in natural environments and enhance credibility of the research, it creates some methodological dilemmas to researchers. Therefore, the following three areas are considered: a) the participants, b) the research design and c) the data collection, analysis and reporting procedures and mechanisms (Reeves, 2000).

#### 3.3.1 Participants

Participants in developmental research are the sources of data. The participants in were the research who test theories and develops new findings, the solution providers who facilitate the design the subsystem and the educators/instructional designers who are the beneficiaries of the designed subsystem.

#### 3.3.2 Research design

**3.3.2.1 Creation of a developmental context**

The initial instructional design subsystem was derived from the models that will were identified during the literature review. In particular, the researcher was in favour of a modification of the generic ADDIE model. A graphical layout of all the processes and procedures that were envisaged was made on paper and then submitted to the instructional designers/educator for their review and comments. The design was made on paper to avert the feeling from the participants that a lot has been put into the design of the subsystem, and as such, their comments would be limited so as not to give the designers a lot more work to do. After every review with the experts, the results were communicated and discussed with the solution developers.
The model was first presented to the instructional designers and later to the application developers who do the actual implementation of the system. This was aimed at formalizing the design.

A number of issues arose during the process and they include:

- Educators and instructional designers were expecting to see a ‘product’ they can use while on the other hand the software developers wanted to get the full specifications at once so that they can start working on the system.
- Even the expert instructional designers do not subscribe to one approach or model of instructional design.
- There was a great deal of iteration, mainly on instances where the participants had to validate/compromise what they thought.
- A number of the instructional designers were reluctant to use any of the models but instead wanted to adopt “what works best for them”. Most of the instructional designers consulted their existing content to form the instructional strategy and base their approaches on the content.
- A number of instructional designers and educators did not know what they want until they were presented with the model. This prompted their thoughts and they contributed towards what they saw as a desired model.
- The time taken during this phase was considerable and the management in charge of the software development thought a lot of resources were being tied up for quite sometimes.
- The development team was concerned that most of the proposed features for the instructional design subsystem could have been dealt with without necessary writing a completely new subsystem.

3.3.2.2 Design of initial system

Initially, the design was adopted from the literature review. This design was printed on paper and then taken to expert instructional designers for criticism, comment and review. After a consensus model was arrived at, the design of the system was made using design patterns.

The initial inquiry into the subsystem was done during the formalization of the design and alignment with the literature when it was subjected to the expert instructional designers in the process explained in section above. The aim of this initial inquiry was to derive a design that can be used for the instructional design subsystem; compare the design with the original design from the existing literature; and provide detailed functional specifications for the resultant instructional design subsystem.

3.3.3 Data collection, analysis and reporting

Data collection

In the first iteration, interviews with educators and instructional designers were
conducted to provide information on how the instructional designers view their work process in relation to the model designed from the literature review. The instructional designers were given a design sample in order to determine their reaction to a typical design scenario.

**Data analysis**

Initial analysis was done from literature and then corroborated by educators and the instructional designers. This lead to the validation of the model in the design of the subsystem. The analysis explored details on the specific educator/instructional designer knowledge and skills and the design task that could be supported and improved in the system.

The data collected during the iterations was analyzed to come up with a generalized view of how the instructional designers go about their work, and determine their initial reactions to the designed model.

The design agreed upon did not show much difference from what was from the literature. The only major difference that was recorded is that even though a model may be specified, instructors and educators do not always subscribe to it – neither do they follow all the distinct steps as set out in the model.

**Reporting: revision, refinement and modification of the subsystem**

The initial design was refined, through modification and revision following agreements was reached with the expert instructional designers. The iterations will be continuous. Based on the main observations of the preceding iteration and data analysis, the subsequent subsystem will be refined and modified to accommodate the computable views and improvements suggested by the instructional designers (subjects). Before the actual modifications are made, the results of the existing state of the subsystem will be subjected to the expert instructional designers who were used in the first iteration for review and also weighed against the existing literature for identification of generalized improvements or modifications that can be made.

**3.4. Results and conclusion**

Developmental research contributes to instructional design as field of knowledge based on understanding of the new procedural models, generalizable principles or the lesson learnt in a particular project. This research can identify two results, the research design and the instructional patterns.

Managing the scope of the work was a challenging task because of a number of issues:

- The research approach used has both the research and design intertwined. This meant that there was a great deal of iterations and revisions before a final design could be reached.
- Identification of patterns for the instructional design process features arrived at, giving the ‘problems’ identified as core to the process a ‘solution’ that could be generic. This was a challenging task because the problem-solution pair had to be unique and at the same time computable.
• The complexity of the instructional design process, and the relationship between the instructional theory and instructional practice proved a challenge to getting all the requirements for the design of the subsystem easily.

• The limited research in the area of web instructional systems – partly due to the fact that the web is a relatively new technology and the research is taking shape in the area, led the researcher in adopting the existing theories to come up with an ‘eclectic’ approach to the design of the web instructional design subsystem.

Conclusion And Recommendations

In developmental research, design and research efforts are intertwined, with research efforts being typically iterative and successive efforts focusing on different aspects of learning, instructional design, development and evaluation, and the whole instructional system or product change.

Design experiment approaches utilized qualitative research methodologies, by creating and testing new ways to collect data and analyze – through interaction with the literature, experts and practitioners in the field of instructional design and educational material development at large. The iterative nature advocated in these approaches allowed questions of various scope and complexity – in the instructional design field to be studied.

The findings of successive implementations can form a rich base of information to refine theories about instructional design. The procedures of a design-experiment methodology used entailed creation of a developmental context for the web-based instructional design subsystem, creation of the initial model based on the existing knowledge (instructional design models, theories and paradigms) and technologies (in this case web-based technologies), testing and implementation of the model and finally an iterative inquiry into the effectiveness of the model. The results of the inquiry phase are used to revise, refine and modify the model iteratively over several rounds.

References

Burkhardt, H, and A Schoenfeld (2003). “Improving educational research: toward a more useful, more influential, and better funded enterprise”, Educational Researcher, Vol 32, No. 9, pp3-14


Survey of Data Mining Methods for Crime Analysis and Visualization

Okwangale Fredrick R. Department of Networks, Faculty of Computing and Information Technology, Makerere University, Kampala – Uganda and
Ogao Patrick, Head of Department, Department of Information Systems, Faculty of Computing and Information Technology, Makerere University, P. O. Box 7062, Kampala – Uganda

Crime prevention is a primary concern of police as they perform their central role of protecting the lives and property of citizens. But the police force is usually relatively very small compared to the crime prone population they have to protect making them more of a reactive rather than preventive force. Police often have at their disposal vast amounts of least utilised crime data (such as crime incident reports) which if analysed could reveal some hidden information such as crime committing trends useful in crime prevention. Use of Information Systems techniques such as data mining and Geographic Information Systems for analysing these data is promising in boosting the police efforts. This paper reviews the applicability of various data mining methods and Geographic Information Systems in crime analysis and visualization in mainly poor planned settings characterised by missing electronic data a common phenomena in the developing countries like Uganda. The focus is on criminality of places rather than the tracing of individual criminals. he review tends to reveal that a combination of Geographic Information Systems and data mining techniques that can work under unclean data are best suited for use in the poorly planned settings.

1. Introduction

Security of citizens is the major concern of the police. Rather than focusing on enforcement and incarceration police can deter crime through the knowledge benefits that derive from information and its associated technologies The police force can employ information technology to turn police officers into effective problem solvers and to leverage their intellectual capital to pre-empt crime (Brown et al., 2003)[4]. However one challenge to law enforcement and intelligence agencies is the difficulty in analyzing large volumes of data involved in criminal and terrorist activities (Chen et al., 2003)[14]. A variety of techniques in data mining and Geographic Information Systems (GIS) are surveyed in their applicability in use in environments of less organized data.
2. Overview of crime, data mining and Geographic information system

2.1 Crime

Crime is referred to as a comprehensive concept that is defined in both legal and non-legal sense (Akpinar et al, 2005)[2]. From the legal point of view crime is the breaking or breaching of the criminal law (penal code) that governs a particular geographical area (jurisdiction) aimed at protecting the lives, property and the rights of citizens of belonging to that jurisdiction. Crime is an offence against a person (for example murder, and sexual assault), or his/her property (for example, theft and property damage) or the State regulation (for example traffic violations) (Akpinar et al, 2005)[2] (Oxford Dictionary of Current English). In non-legal terms crime is a set of acts that violate socially accepted rules of human ethical or moral behavior (Akpinar et al, 2005)[2]; for example acting against a ritual in some society.

Crime occurs in a variety of forms which police informally categorizes as being either major or volume. Major crime consists of the high profile crimes such as murder, armed robbery and non-date rape. These crimes can either be one-offs or serial. In the case of serial crimes it is relatively easy to link crimes together due to clear similarities in terms of modus operandi or descriptions of offenders. This linking is possible due to the comparatively low volume of such crimes. Major crimes usually have a team of detectives allocated to conduct the investigation. In contrast volume crimes such as burglary and shoplifting are far more prevalent. They are usually serial in nature as offenders go on to commit many such crimes. Property crimes, such as domestic burglary offences, committed by different individuals are highly similar and it is rare to have a description of the offenders (Adderley and Musgrove, 2001) [1]. Table 1 shows the classification of crime (Chen et al., 2003)[14].

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Violations</td>
<td>Driving under the influence of alcohol, fatal/personal injury/property damage traffic accident, road rage</td>
</tr>
<tr>
<td>Sex crime</td>
<td>Sexual offences</td>
</tr>
<tr>
<td>Fraud</td>
<td>Forgery and counterfeiting, frauds, embezzlement, identity deception</td>
</tr>
<tr>
<td>Arson</td>
<td>Arson on buildings</td>
</tr>
<tr>
<td>Gang / drug offences</td>
<td>Narcotic drug offences (sales or possession)</td>
</tr>
<tr>
<td>Violent crime</td>
<td>Criminal Homicide, armed robbery, aggravated assault, other assaults</td>
</tr>
<tr>
<td>Cyber crime</td>
<td>Internet frauds, illegal trading, network intrusion /hacking, virus spreading, hate crimes, cyber piracy, cyber pornography, cyber-terrorism, theft of confidential information</td>
</tr>
</tbody>
</table>
Also a single detective officer may have a large number of different volume crimes to investigate at any point in time. With a view to satisfying this demand, police forces around the world employ specialist crime analysts, people who have specialist training in a variety of disciplines including investigation techniques, criminal psychology and information technology. It is their task to assist investigating officers by analyzing crime trends and patterns, identifying links between crimes and producing packages which target an individual or group of offenders linking them to a series of crimes (Adderley and Musgrove, 2001)[1].

Most, if not all, of current systems both manual and computerized revolve around the investigation of crimes already committed. They are, therefore, reactive. In the developed countries like the UK, a majority of crime prevention forces use different types of relational database management systems (RDBMS) for recording and subsequent analysis of crime. Standard or interactive queries are written to produce patterns of crime, offending and various statistics (Adderley and Musgrove, 2001)[1] but its is a common phenomena in the developing countries to find mainly manual criminal record books used alongside the pin-up maps for crime incidence location.

2.1 Data Mining

Data mining deals with the discovery of unexpected patterns and new rules that are “hidden” in large databases. It serves as an automated tool that uses multiple advanced computational techniques, including artificial intelligence (the use of computers to perform logical functions), to fully explore and characterize large data sets involving one or more data sources, identifying significant, recognizable patterns, trends, and relationships not easily detected through traditional analytical techniques alone. This information then may help with various purposes, such as the prediction of future events or behaviors. (Reza et al, 2001)[14]

The development of new intelligent tools for automated data mining and knowledge discovery has led to the design and construction of successful systems that show early promise in their ability to scale up to the handling of voluminous data sets.

Theories of crime and delinquency tend to be discipline-specific and are dominated by psychological, sociological, and economic approaches (Reza et al, 2001)[14]

Data Visualization

Visual methods are powerful tools in data exploration because they utilize the power of the human eye/brain to detect structures. A number of data mining tools for visualization exist, a histogram and Kernel plots being the most basic used for displaying single variables. Scatter plots for the display two variables at a time and reveal correlation, if any, between them. And for more than two variables, scatter plot matrices are often used (David et al, 2001)[7]. GIS also provides a powerful visualization tool through display of maps that allow the exploration of spatial patterns in an interactive fashion (Pfeiffer, 1996) [12].
2.2 Geographic Information Systems

Over the past few years Geographic Information Systems (GIS) has become a standard tool for crime analysts in many police departments, regardless of their size (see for example McEwen and Taxman 1994; Rossmo, 1995). One of the inherent advantages of GIS is its ability to integrate information from a variety of sources into one user interface. In turn, this allows for spatial analyses that would either not have been possible or at a minimum far more difficult prior to the advent of GIS (Olligschlaeger, 1997)[11]

2.3 Crime Analysis

Crime being inherently spatial phenomena (Ratcliffe, 2004)[13], taking place within a given location and at a specific time, it is logical to analyse crime in terms of spatial crime analysis. The advent of geographic information systems (GIS) has given behavioural scientists the ability to map and track a wide range of social phenomena such as crime incidences. Much more information about what is happening “on the ground” is now available. This enhanced analytic capacity has presented police departments with the opportunity to better serve and protect the people and places in their care. Furthermore, as GIS assists one in seeing and understanding behaviour patterns, it also provides stakeholders with opportunities to join together in partnerships for the common good (Holzman et al, 2003)) [8]. A data pattern is an expression in some language describing a subset of the data or a model applicable to that subset (Fayyad et al, 1996)[6] There are at least three types of police crime information (primary, secondary, and tertiary), intelligence (prospective, retrospective, and applied), and operational strategies (preventive, prospective, and reactive), each of which interacts in a complex fashion with technology (Manning, 1992) [9]. Crime data analysis aids police in transforming this information from one type to another.

Crime Spatial Data Analysis

Spatial data analysis utilizes statistical analysis methods that address specific issues relating to spatial data, including spatial dependence (autocorrelation) and spatial heterogeneity. These issues run counter to traditional statistical assumptions of heterogeneity and independence of sample data. If these issues are ignored, then analysis results might not be valid. In combining the powerful tools of GIS to integrate and manipulate spatial data with rigorous statistical methods, spatial data analysis shows great promise for criminology, criminal justice, and law enforcement research and practice. (NIJ, 2006)[10]

3. State Of Crime Data At The Uganda Police

Most of the crime data for the Uganda police is manually recorded in criminal record books (CRB) in form of statements made by arrested suspected criminals, reports by informers and information gathered by the police themselves. Pin-up maps are used to capture the locations of crime incidents (Uganda Police online, 2005)[15] The deployment areas tend to be specific.
4. The Data Mining Techniques For Crime Spatial Data Analysis And Visualisation

There are a number of data mining techniques for crime analysis and visualisation but the choice of which to use depends on the features of the problem, the data and the objectives (David et al, 2001)[7]. This paper is limited to the Exploratory Data Analysis (EDA) and specifically crime spatial data analysis. Typically, EDA techniques are interactive and visual, and there are many effective graphical display methods for relatively small low-dimensional data sets. The difficulty in visualisation of points increases with the number of variables involved (David et al, 2001)[7]. The methods used in crime spatial data analysis can be classified into those concerned with visualisation of data, those for exploratory data analysis and methods for the development of statistical models (Pfeiffer, 1996)[12]. The methods are surveyed categories of the data based on the categories of crime data to be analysed - point patterns.

Point Patterns

Spatial point patterns (SPP) are based on coordinates of events such as locations of crime incidences and may also include the time of occurrence. All or a sample of point pattern may be plotted on the map. The aim of SPP analysis is to detect whether the point pattern is distributed at random, clustered or regular. SPP is typically interpreted as analysis of clustering. A dot map is commonly used to represent SPP. The tool effectively used for analysis of clustering effects is the K function. This method assesses clustering of crime incidences in detection of hot spots (Kingham et al, 1995) where time and space relationship analysis is required, the methods used are Knox’s method, Mantel’s Method and K-nearest neighbour method. All the three methods require the production of distance matrices of the spatial as well as temporal relationship between crime incidences. Knox’s method requires critical distance in time as well as space defining closeness has to be set but the determination of these critical distances requires subjective decision. Mantel approach does not however require use of critical distances but uses both time and space matrices. It is however insensitive to non-linear associations. The K-nearest neighbour is based on the approximate randomisation of the Mantel product statistic (Pfeiffer, 1996) [12].

5. Conclusion

Exploratory data analysis makes few assumptions about data and it is robust to extreme data values. It is possible to use simple analytical models with EDA The methods that are robust to missing data are useful in the data mining of crime data where data is not so precisely collected. The distances between crime locations are normally not easily available to the police in the areas that are not well planned. The poorly planned areas are best represented by dividing them into area clusters and the analysis is done based on the clusters. The methods that support clustering are therefore best suited for the crime analysis of the poorly planned settings. The manual pin maps are best replaced by the use of GIS.
6. References


5. Colleen MCCue, Emily S. Stone, M.S.W., And Teresa P. Gooch, M.S., 2003, “Data Mining and Value-Added Analysis” FBI Law Enforcement Bulletin


ICT for Educational use: Cases of implementing a partnership approach model for reaching Target Groups

Elijah I. Omwenga, emwenga@uonbi.ac.ke, School of Computing and Informatics, University of Nairobi, P. O. Box 30197, Nairobi 00100, Kenya

The question that comes into the minds of most of the universities in Africa today is on how to use ICTs to strengthen their resolve to reach out to potential client students who did not make it to the higher learning institutions through the normal school progression programmes. One can identify two major phases to dealing with this issue. First, there is need for awareness and advocacy, seminars and workshops, sensitisation on the relationship between ICTs and wealth creation and training of stakeholders and trainers. Phase two will largely involve taking the programmes to the people – the students and target groups. In this paper we describe a proposed partnership model that can be applied to help realise the goals of effectively reaching the target groups. The model recognises a three-stage pyramidal communication model involving target groups and partners, Networking and Collaboration and finally policy strategising on thematic areas. The application of this model and implications in the use of ICTs for improving teaching and learning especially among women have been analysed and suggestions for improvements made.

Introduction and Background

One of the reasons for the dismal economic performance of African Countries despite their large natural resources endowment is the inability to harness their resources for the benefit of their population. African countries lack the scientific and technological capacity to develop, process and utilise the resources with which they are endowed. There is need for scientific and technical manpower to maintain and operate industries, build infrastructure and provide services to agriculture, industry and other economic sectors. Governments and donors are encouraged to look at the development of science and technology capacity in Africa as an important route to sustainable development (ANSTI, July 2005). Information and Communication Technology (ICT) has been recognised as a key player in the implementation of activities and services that enhance sustainable development (Omwenga, 2004). Several non-governmental organisations have embarked in projects that use ICTs to empower the African people through partnership and collaborations approaches. This study looks at one such programme.

The objective of this study was to investigate intervention mechanisms for supporting the use of ICTs for national development. The study considered a numbers
of key success cases which have implemented various ICT programmes in the African region.

The approach taken recognised that there are many organisations that are promoting the use of ICTs for national development and as such there was need to sample these organisations with a view of getting a general picture of what goes on. Projects in three countries namely Kenya, Nigeria and Ghana were considered. The reason for selecting these counties was mainly due to the fact that the sponsor of this study, Abantu for Development, a non-governmental organisation (NGO), established in the south with a northern presence, had several activities going on in these countries (Ngunyi, et al. 2000). The organisation has regional offices in Kenya, Ghana and the UK with an operational secretariat in Kaduna – Nigeria. In the strategic plan of 2001-2004 (Abantu, 2001), the main objective of this NGO is to enhance the capacity of African people, especially women, to participate in development and increase their participation in political and economic structures of their countries. Using the partnership approach, the organisation recognises ICTs as a driving force towards achieving these objectives at all levels.

Study Methodology

The Approach. This study used the questionnaire approach complemented with physical verification of the information through filed visits. This approach was economical and more effective. It enabled the project team to collect information from several institutions at low cost because there was limited travel involved. However, the researcher travelled to the regional offices in Ghana, Nigeria and of course Nairobi where he was based. Over 40 questionnaires were distributed. Several documents, publications, and websites were accessed and useful information was collected.

The Survey tool. The main issues that were being investigated focused on finding out how the partner institutions were dealing with challenges posed to the implementation of their mandates in the use of ICTs for economic development and education in general. It was imperative that the study take into account specific aspects that were of paramount importance to the sponsors: empowerment of the women organisations in strengthening their electronic capacity. In this connection, the data collection instrument explicitly requested for information regarding this aspect.

Stakeholders in the Implementation of the ICT Programmes

To implement the Strengthening of ICT Capacity Programmes requires a multisectoral approach, which involves many players. At least six different types of respondents participated in this evaluation study. Of these six categories, 45.9% were staff directly involved with the activities of the various organisations leaving approximately 54% of the respondents whom we could categorise as stakeholders. These included organisations that were classified as either Partners, Beneficiaries or Collaborators (Omwenga, 2002). Some of these institutions played double roles of Partners and Beneficiaries depending on how one looks at the relationship and role of the organisation. Partners are mainly institutions which deal with ICT issues and whom parent organisations have selected
to work with. On the other hand, collaborators are organisations, which parent organisations contract to carry out, on their behalf, essential services such as radio programme production and media linkages. Collaborators could also be other peer organisations working for the common goal of wealth creation among the African communities. Policy makers include members of parliament, local council personnel, government agencies and others. Resource persons are the trainers, researchers and scholars who participate in workshops and seminars.

Each of these stakeholders had a part to play towards a successful implementation of the ICT programmes and in fact one measure of the success of this programme was the number of policy makers that have been brought on board. Other stakeholders included the donors and members of the boards of trustees of the sponsor of this study.

**The Partnership Model approach**

The technique of partnership to reach out to the target groups was found to be a cost effective means of quickly disseminating information (Omwenga, 2002). We elaborate further the concept of partnership by identifying into this category tertiary institutions, schools, or NGOs that are involved in the use of or application of ICTs in furtherance of educational needs as provided by or supported by parent organisations such as a University or a commercial educational enterprise. The pyramid below gives a pictorial view of this approach.

**Fig. 1: The Partnership Model - reference (Omwenga, 2002)**
Figure 1 above shows the 3-level model for implementing cooperation and partnership in ICT use and application. The responsibilities of either party (the partner and the Organisation) are clearly spelt out and the partners are mandated by the organisation to carry out specific functions that will help in the realisation of objectives including leeway to engage in franchising. During the study it was realised that most partners did not have specific programmes that targeted such issues as ICT for poverty eradication. Organisations encouraged them to introduce programs that would advance the objectives of poverty alleviation or wealth creation.

The model operates at three levels: policy formulation at the top, networking and establishment of partnerships in the middle level and programme implementation at the lower level; mainly through partners, collaborators and beneficiaries.

On behalf of individual organisations, the partners would mobilize the target groups who are largely NGOs, commercial entities, tertiary Institutions, colleges or universities. Thereafter, tailored ICT training programmes, workshops and seminars are mounted for selected participants.

**Implementation Strategies**

All the partners surveyed had adopted a number of strategies for implementing their
ICT programmes on the ground. Below is a summary of the main methods employed. The list is not exhaustible.

Table 1: Implementation Strategies

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Objectives</th>
<th>Characteristics</th>
<th>Target participants</th>
<th>Assessment by respondents</th>
<th>Challenges</th>
<th>Observations/ Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workshops and training on ICT</td>
<td>Intermediary means of reaching to lower levels. Strengthen organisations. Training of Trainers. Unique, Targeted training</td>
<td>At least 2 workshops within 2 years. Application of affirmative action. TOTs – cost sharing</td>
<td>Women, Middle level people</td>
<td>Important</td>
<td>Resources, Need Infrastructure, Timeframes, Sustainability</td>
<td>Invest more on IT, Establish training centres, increase frequency</td>
</tr>
<tr>
<td>Seminars</td>
<td>Raising awareness</td>
<td>Information dissemination, Documentation</td>
<td>Women, relevant NGO representatives</td>
<td>Important</td>
<td>Cost, Sustainability</td>
<td></td>
</tr>
<tr>
<td>Media Type</td>
<td>Purpose</td>
<td>Recommendations</td>
<td>Caveats</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>----------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Publications</strong></td>
<td>Information dissemination on ICT, Awareness, Advocacy, Education</td>
<td>Newsletters/magazine – a regular publication that contains columns on ICT, Brief training manuals on ICT</td>
<td>All – but make targeted, gender-sensitive programs</td>
<td>Magazine on ICT popular in Ghana but short or late supplies.</td>
<td>Explore repackaging for different groups, Need to expedite delivery to regional centres in Nigeria.</td>
<td></td>
</tr>
<tr>
<td><strong>Radio</strong></td>
<td>To sensitise, inform, educate</td>
<td>Not implemented in Kenya. Up and running in Accra. Temporal stoppage due to cost at Kaduna</td>
<td>All – but make targeted, gender-sensitive programs</td>
<td>Relevant especially to the poor rural women who are the main target</td>
<td>Expensive. Can be time consuming to produce quality programs. Introduce radio programmes in Kenya, Negotiate for reinstatement of the programme in Kaduna. Introduce programs in local languages.</td>
<td></td>
</tr>
<tr>
<td><strong>T/V</strong></td>
<td>To discuss topical issues on ICT – usefulness, Awareness campaigns, educate</td>
<td>On rare occasions staff from individual organisations participate</td>
<td>All – but make targeted, gender-sensitive programs</td>
<td>Relevant</td>
<td>Expensive. Need very competent staff who can articulate issues well. Encourage staff to participate in such programs whenever.</td>
<td></td>
</tr>
<tr>
<td><strong>Print media</strong></td>
<td>To educate, sensitise</td>
<td>Organisations have warm relationship with most media houses in all regions. Some produce their own publications</td>
<td>All – but make targeted, gender-sensitive programs</td>
<td>Required</td>
<td>To cultivate and sustain good relationship. Need to put records straight in case of some negative criticism. Need to take advantage of the warm relationship.</td>
<td></td>
</tr>
<tr>
<td>Provision of Equipment</td>
<td>To strengthen capacity and delivery of service</td>
<td>Can be a necessary action if such equipment will help the partners advance organisations’ goals even better</td>
<td>Partners</td>
<td>As far as possible</td>
<td>Cost, sustainability</td>
<td>Encourage partners to budget for such equipment</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>---------</td>
<td>------------------</td>
<td>---------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Technical advise</td>
<td>To strengthen capacity and delivery of service</td>
<td>Troubleshooting computer systems, Information on effective means of running the programs</td>
<td>Partners</td>
<td>Useful</td>
<td>Such expertise may not be readily found in most organisations</td>
<td></td>
</tr>
<tr>
<td>Facilitation during programs</td>
<td>To support the implementation of ICT programs</td>
<td>Organisations absorb some administrative cost whenever possible</td>
<td>Partners</td>
<td>Useful</td>
<td>Can be costly</td>
<td>Partners tend to view this as a right rather than as a privilege</td>
</tr>
<tr>
<td>Email and Internet</td>
<td>More and more people have started using email as a means of communication; thanks to training given by the organisations. This is probably one of the main objectives of these organisations that were surveyed. Whereas we recognize that these are not the only forms of ICTs, their role is crucial to the empowerment of the people.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It is interesting to note that different forms of ICTs are being used as a vehicle to the attainment of objectives of reaching the target groups. Each strategy has a different impact-cost ratio and organisations employ multiple strategies simultaneously. This is likely to have a bigger impact over a short period of time.

Strengths of the partnership strategy. Almost all the partner representatives interviewed appeared quite committed to do their best with the limited resources. They fairly well understood the objectives of the programme and what was expected of them. The technique enables Organisations to free themselves of some of the concerns of implementation in order to concentrate on core activities such as policy formulation and networking.

Limitations of the Strategy. This strategy has worked fairly well but a number of limitations were observed. For instance, some partners innocently tend to believe that it is their right to receive whatever assistance Organisations were capable of giving. They seem to hold the notion that they are doing the organisation’s work. In some cases, some of the partners believe that parent organisations need to assist them even more, so that they are able to realise the mandated duty of strengthening the organisations ICT capacity.

It was not clear from the information gathered if all the partners involved in this programme were actually doing anything specifically related to the ICT objectives as set out by individual organisations. Some of them who are doing commendable work in other areas appreciated the tenets of the ICT program but lacked the conceptual and logistical prowess to see the implementation of the programmes on the ground.

**Analysis of the model**

Strategy versus Consistency with objectives. Objectives and strategies can sometimes overlap and one objective can be subsumed in another. Some objectives and strategies that have not been explicitly discussed in this section are either implied or overlap.

The objective of Access to and Use of ICTs. The organisations surveyed wished to achieve this objective through the production of gender-sensitive guides on ICTs, provision of information and advice, development of human resources for ICT use and conducting skills workshops for partners.

The data on the four areas from the respondents is summarised in the table below.
Table 2: Respondents Views on Access and Use of ICTs

<table>
<thead>
<tr>
<th>Activity</th>
<th>n (number of valid responses)</th>
<th>Mean (1-Nothing, 2-Not much, 3-Fairly Active, 4-Done very well, 5-Excellent)</th>
<th>Floor of Mean (1-Nothing, 2-Not much, 3-Fairly Active, 4-Done very well, 5-Excellent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Producing Gender Sensitive guides</td>
<td>18</td>
<td>3.5</td>
<td>3</td>
</tr>
<tr>
<td>Providing Information and Advise</td>
<td>26</td>
<td>3.92</td>
<td>3</td>
</tr>
<tr>
<td>Development of Human resources for ICT use</td>
<td>21</td>
<td>3.3</td>
<td>3</td>
</tr>
<tr>
<td>Conducting skills workshops for partners</td>
<td>22</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Holding workshops and seminars</td>
<td>26</td>
<td>4.1</td>
<td>4</td>
</tr>
<tr>
<td>Strengthening African women’s presence in cyberspace</td>
<td>22</td>
<td>3.3</td>
<td>3</td>
</tr>
</tbody>
</table>

We notice from Table 2 that out of a possible total of 40 responses for every question, there were a substantial number of respondents who did not give their opinions (missing data items). Failure for a respondent to give an opinion on a questionnaire item could mean that the respondent does not have sufficient information for him/her to give an opinion or conversely it would imply that according to the respondent, there are no such activities taking place. The researcher chose not to take either of these positions and instead excluded such data items.

Using column 4 of the table we can conclude that most organisations have done very well (4) in conducting skills workshops and seminars for women organisations but is only fairly active (3) in all the other areas.

The strategies used to achieve this objective are sound. The results of this analysis indicate that within the short period that the ICT programs have been in existence, substantial progress has been made in enabling access to and use of ICTs.

The strategy of Demystifying ICTs. Table 3 below shows that approximately 91% of the respondents said that the Strengthening of the Electronic Capacities of African Organisations Programme has helped to demystify the concept of Information and Communication Technologies among the target groups. This percentage represented 33 respondents of whom 21 said the programme had done at least very well. Fig. 2 on the right is a representation of the same information graphically.
Table 3: Respondents Opinion of Extent to which ICT has been Demystified

<table>
<thead>
<tr>
<th>Extent of demystifying ICTs</th>
<th>Valid</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>Low</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Very Good</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Excellent</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
<td>95.0</td>
</tr>
<tr>
<td>Missing System</td>
<td>2</td>
<td>5.0</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Fig. 2: Graphical Representation of Opinion on Demystification

If we extrapolate this result by making reasonable conclusions that the respondents were representatives of a wider society that had had a chance to participate in one or more of the programmes, then we can conclude that the strategies employed to achieve this level of acceptance are valid. However, from the data available we can only guess that training may have contributed a greater percentage towards influencing respondents to rank the level of demystification thus. But it would as well have been through the print media, the Gap Matters magazine special column on technology or indirectly through a friend who was a beneficiary of one of the training programmes.

It was observed that, in particular, more women are now able to contribute positively towards issues that affect them using the Internet. In Nigeria (Kaduna State) and Ghana, there has been an influx of women preparing to contest state government seats than ever before. This has been directly attributed to Gender and Conceptualisation seminars held prior to the start of the campaigns.

Given this result, we can conclude that this strategy is consistent with the objectives of raising awareness and removal of the fear of handling computer-based ICTs.
Recommendations from the study

We can draw a number of recommendations from this study. We highlight and discuss them below.

**Process ownership.** There is need to re-emphasise that the partner bodies should own the process and there is need for them to re-examine the reasons for their existence in the first place. This will make them be accountable for their actions including failures.

**Implementation Influence.** Each individual organisation, for example a University accrediting an institution such as a college, should establish a mechanism of having a bigger say in the running of the partners in order to influence the implementation of its programs through them. We can draw a parallel with a University – an organisation – accrediting an institution or college – a partner – to implement the programmes of the University. In this case the University shall require to monitor and ensure that quality assurance structures are adhered to.

**Memorandum of Understanding.** Organisations need to draw verifiable memoranda of understanding (MOU) between the partners and themselves so that either party shall hold the other responsible in event of failure to deliver.

**Demarcation of responsibilities.** For the partnership relationship to work out well, clear demarcation of responsibilities should be spelt out in the MOU between the two parties involved.

**Measuring Progress.** A number of organisations surveyed have developed criteria for measuring progress. This criteria needs to be adopted to suit the ICT programmes and be used to evaluate the partners’ progress.

**Monitoring.** There is need for close monitoring of activities. Organisations should consider making it their mandate to carry out reviews of the work of the partners with a view of advising and using the feedback to bring the partners back on course in event to major deviations.

**Research.** Organisations need to carry out research on various issues on ICTs within localised environments since the local situations could be a factor on the best way of introducing ICTs and the impact of the same on poverty alleviation within the localised area.

**Conclusion**

This study focused on a model of inclusive participation of stakeholders in the implementation of programmes that promote the use of ICTs for economic development at all levels. We have looked at a three-tier model that recognises the need for providing information through various channels during the initial stages of decision making and implementation of the programmes. This model identifies this first stage with the target groups or partners who are the ultimate beneficiaries or implementers of the programmes. These target groups, in turn, need to partner with other organisations on whose behalf they are carrying out the activities. This sets the stage for networking and close collaboration for effective interpretation of the action plans and mandates. Higher up in the pyramid are the corporate bodies and organisations such the United Nations, Non-governmental, or Universities. There are variations of the model in each case which is beyond the scope of this paper.
Clearly this model is only effective if the recommendations identified above are implemented; particularly the need for partners to vouch to own the process and be accountable for the outcome of their activities. The false belief that parent organisations (the highest in the pyramid) should provide everything for building the ICT infrastructure even during the implementation phase is rather short-sighted and has contributed to the apathy that was observed during this study. This is a major problem though, that cuts across other sectors of government in most developing countries. It is always believed that the government should provide. We must reverse the trend to that of asking “what shall one do for the organisation or government”.

The study established that for effective delivery of ICT services, there is need for a clear demarcation of responsibilities between the stakeholders (Organisation, partners, beneficiaries, collaborators) right through the pyramid. Even in the education sector, the situation is not any different. For instance the University (organisation) provides the ICT infrastructure and environment for learning. It partners with the parents who bring their children to university to learn and in turn pay taxes and other fees to meet the expenses. The students are the beneficiaries. The roles are clear for each of these groups: the university builds the infrastructure while the parents meet recurrent and maintenance expenses and the students utilise the facilities to learn.

References


The impact of ICT on Universities: Classroom/Lecture theatre Design and Curriculum Delivery.

Asifiwe Collins Gyavira Rubanjju, School of Education, Department of Higher Education, Makerere University, P.O. Box 7062, Kampala, casifiwe@educ.mak.ac.ug

Empowering learners to engage in meaningful, challenging and enlightening tasks is the aim of all educators and Information and Communication Technology has the potential to play a powerful role in every university - both inside and outside lecture room/classroom. Institutional responses to ICT influences have inevitably brought about changes in the following areas: how teachers redesign and present the curriculum, in the curriculum itself, in the uses made of resources and the way current lecture rooms/classrooms and buildings are being reconfigured. This study utilized a qualitative research methodology and was carried out in selected universities to assess the impact of Information and communication technologies on class/lecture room design and curriculum delivery. There are a variety of philosophical approaches to the integration of ICT. In addition to that, a range of practical solutions in overcoming the universal problems of staff and student access to computers, professional development for educators, and internet access was revealed. It is also recognized that rapid advances in technology will inevitably “age” traditional observations in educational delivery of curriculum and classroom design. Overall, there were far sights by lecturers, and administrators creating learning opportunities for students using ICT.

1. How is the availability and use of ICT changing the use of existing lecture theatre/classroom spaces?

Introduction:

Some of the universities in Uganda are just over ten years old while others such as Makerere were founded seventy years ago. It is worth notable that some universities started as colleges but were elevated to university status. The classroom/lecture theatre spaces we operate in were designed to reflect the traditional and transmissive style of curriculum delivery. In addition, they were designed with little, if any thought being given to investigation-based, group learning, let alone fibre-optic cabling. While new buildings are constructed and others are renovated, the reality for most universities is that existing spaces must, in the short term at any rate, be adapted to accommodate new learning technologies.
Different thoughts about teaching spaces

The fixture of even one computer in the classroom/lecture theatre can have a profound effect on the way students learn and the way the classroom operates. Teachers integrating computer use into the curriculum soon modify their classrooms to reflect the changes in student learning behavior that inevitably emerge. Creating space in the classroom for computers and peripherals such as a printer, network connection and large monitor initiates a rethinking process by the teacher, leading to re-evaluating how classroom activities and learning experiences work best.

Early responses to ICT often involved creating a technology centre, or a dedicated building to house computers and peripherals, with students being taken to these facilities when work with computers was needed. Computer laboratories were developed with time being booked for a whole class to use the technologies; this arrangement still exists in all universities visited. Alongside laboratories, however, a variety of solutions have been developed. These enable in-class access to computer facilities in several departments in the three Universities visited and it was these arrangements that were of interest in this research.

The anytime, anywhere access to information sources, and “ubiquitous” computing enables students to engage directly with expert sources when they are needed; and the sight, sound, touch experience becomes a powerful motivator in learning. Bernard Holkner says, “The Cast of Players in a student’s learning experiences has increased dramatically.

Convergent technologies now allow experts, peers and collaborators to join the student’s world, enriching learning experiences.”

Professor Hedley Beare in 1998 recognized the role of the teacher becoming more fluid and covered by “a range of professional educators – tutors, instructors, mentors, learning theorists, curriculum planners and experts, assessors, curriculum writers, assignment markers, editors and student counselors”. A mix of on-site and distance programmes seems to be a possibility for the student of the future.

What is the ideal lecture theatre?

Brett Hunter believes today’s classrooms need Internet access for research, distributed multimedia curriculum on line, access to digital libraries, distance education courses and remote collaborative tools. Information on demand for students also includes video, live video broadcast, desktop and videoconferencing. “The use of voice (for activities such as interviews, speeches, background music, explanations) and Video (for live conferences within and between schools) will change the way schools operate”.

In terms of configuration, today’s lecture theatre needs changed design involving creating two main areas for student use: grouped and networked computer facilities, and an area designated for group-working space.

Tomorrow’s lecture theatre

Paul Butler believes that future classrooms will be characterized by:
i) Access to on-line resources which use a powerful combination of video, multimedia, text and graphics. These are prepared by specialists in a centralized resource development facility and delivered to individuals or groups by technology.

ii) Provision for the teacher to teach the whole class or part of the class, assisted by technology as appropriate

iii) Provision for all students to learn the same way or to choose ways which suit their own individual learning styles, assisted by technology as appropriate.

iv) Access to individualized curriculum pathways, managed by technology

v) Access to individualized diagnostic testing and assessment of progress, managed by technology

vi) Students moving independently between learning areas as necessary, managed by technology

vii) Flexible room layout and furnishing, large-screen video display

viii) Individualized access to network resources – wireless networking; cheap, light-weight notebook computers; continuity of access to network resources away from school

What is really happening in classrooms/lecture theatres?

My research revealed that, most of the lecture rooms remain apparently untouched by technology. A few lecture room configurations have changed to incorporate easy access to the computer(s) and to facilitate the discussions and problem solving that inevitably follow their use. However there are no installations in place at the moment in most of the lecture theatres in the universities in the study. Universities in my research group made a number of differing responses to ICT. Through grouping these responses, I have created the following list of observed trends in the use of lecture room spaces with ICT:

i) Rearranging the lecture theatre design.

ii) Creating new spaces

iii) Creating dedicated, flexible lecture room Space

iv) Lecture room redesign for maximum flexibility

Re-arranging lecture theatre design:

My most frequently-viewed, changed lecture room design involved creating two main areas for student use: grouped and networked computer facilities and an area designated for group-working space. However, most of the lecture rooms did not have any evidence of new configurations for easy ICT installations. In some lecture theatres of Makerere, there were adjustable chairs and/or desks that enabled safe use of computers (lap tops) for lecturers during instruction. In other theatres there were network terminals (Computing and IT lecture rooms, Women and Gender studies complex, and
department of food science and technology), public address systems and patch panel. In one of the lectures (institute of languages), there was also a television set. This however was not the case in the Kyambogo and Kampala International Universities.

Easily-moved desks: Movable desks in classrooms are individual and can be reconfigured for the many different functions in the room. In some lecture theatres in Makerere University, it was discovered that there are movable chairs and desks while in most other lecture theatres; there has not been any configuration. This was also observed in other Universities in the study.

Creating new spaces:
There is no established evidence of reconfiguration of lecture theatres in the three universities under the study. However, the study recognized that the new buildings set up are spacious. The new spaces were designed to allow maximum student space both inside the “lecture room” and in the outer corridors.

Creating a dedicated, flexible lecture room space:
With increased number of student population at the three universities, there has been need to construct more buildings. These buildings are flexible with enough lecture theatre space. The rooms can be configured to suit individual students’ needs with a number of network outlets where PC’s and lap tops can be placed to ease student and lecturers’ access to computers.

Lecture theatre redesign for maximum flexibility:

   The advent of ICT has allowed – many would say, demanded – the teacher to develop more individualized approaches to the learning needs of students. The immediacy brought to lecture room experiences by contact with on-line experts, the world’s best libraries and encyclopedias via the Internet means that curious students can learn far beyond the planned lesson.

Using laptop and desktop computers:

   In the search for flexible teaching styles, a few faculties in more affluent circumstances have engaged parents indirectly in the purchase of, mobile learning technologies (laptop computers) as well as desktop computers for students. This has helped students—especially resident post graduate- (twenty-four hours) access to ICT and the Internet.

2. How is ICT use changing the way lecturers/teachers and administrators approach curriculum delivery?

The impact of ICT on student learning:

Much has been written elsewhere regarding ways ICT can be used in the school curriculum and these are not included in this paper, except by way of illustration. My research focused on curriculum practices reflecting fundamental changes in understanding about how the new technologies could support and enhance learning.
The Need For ICT Integration In Institutions Recognized: In each of the universities visited, they recognized the need for change to integrate ICT into curriculum delivery. In Makerere University, the ICT policy was produced and is being implemented in phases. ICT standards have been developed and published on the university intranet and in the University’s strategic plan booklet. Funding comes from government: about 1% funding by University income which is faculty generated; and bilateral funds from donors such as Sida/SAREC of Sweden, Carnegie Corporation of New York. Respondents from Kyambogo and Kampala International Universities showed no knowledge of funding for the ICT programs known to staff who participated in the study.

Research studies have identified the key changes in student learning behavior attributed to the use of computers, the Internet and other learning technologies. Researchers Robert Bracewell of McGill University, Therese Laferriere of Laval University and the consulting firm of Reginald Gregoire, Inc collated summary observations and reference points from a wide range of research papers pertaining to the integration in the classroom of information and communication technologies. Their results grouped primarily, around the specific learning achieved, student motivation and the relationship of students to knowledge. Secondly, they grouped observations related to the consequences of appropriate use of technologies on the teaching function of teachers, including planning of teaching, intervention with a group of students and the assessment of learning. They identified “participants having knowledge and skill in computer use” as a prerequisite to effective classroom use and noted that many studies researched did not deal with this issue for students or teachers.

Changes in student behavior attributed to ICT: Their fourteen observations relating to the changes brought about by effective classroom integration of ICT were as follows:

i) The development of various intellectual skills is noted (e.g. reasoning and problem solving, learning how to learn and creativity)

ii) Specific content that is learned using the new technologies is broadened and deepened

iii) Students demonstrate a greater spontaneous interest in a learning activity

iv) The time and attention devoted to learning activities increases when students use ICT

v) The ease of access to information sources develops the research spirit

vi) Broader co-operation among individuals within and beyond school is enabled through technologies

vii) The availability of simulation, virtual manipulation, graphic representation and rapid merging of data contributes to linkage in knowledge and leads to more integrated and better-assimilated learning

viii) Teachers gain information on new instructional resources and availability of support for their use much more readily with ICT

ix) Teacher co-operation with others both within and beyond the school when planning activities
x) The orientation of planning is more towards students performing real work in cooperation with other students
xi) Relationships between teachers and students more interactive and guiding, rather than transferring information from teacher to student
xii) A different vision of teaching and learning; learning seen more as continuous research than a body of facts
xiii) Assessment of learning uses more demanding methods
xiv) More effective diagnosing of specific difficulties

It is of critical importance that changes such as the above be accommodated when planning suitable programmes of learning for 21st Century students in universities.

Curriculum Changes need to be planned: Marie Jasinski states: “There are eight defining principles education will have to meet in order to satisfy market demand in the knowledge economy with its convergent technology infrastructure.”

These are:
1) Lifelong learning
2) Learner-directed learning, with the teacher becoming the facilitator, diagnostician and therapist
3) Learning to learn so that individuals can plan and realize their own learning
4) Contextualized learning
5) Customized learning, designed to meet different needs, preferences and cultural practices
6) Transformative learning, enabling the changing of belief systems to overcome disability and disadvantage
7) Collaborative/co-operative learning
8) Just-in-time learning, as individuals choose from the global supermarket of opportunities.

These principles combine to give a radically changed model of education from last century’s teacher-centered schools delivering set curricula, chosen and sequenced by the teacher. Students respond intuitively to technologies that have surrounded them all their lives. For them, “the Internet is instantly interactive and the user can control what happens. It responds to the individual and is an empowering medium that allows them to do things that their teachers don’t understand or can’t do.”

Reasons for Delay in Integration of ICT in Universities: In this research, a number of respondents (lecturers and administrators) cited the following reasons for the delay (in no particular order).

• Lecturers’ lack of ICT qualifications
• Too little time to plan and learn the skills effectively
• Lack of money leading to limited access to computers
• Expensive software
• Timetable restrictions
• Lack of creativity
• Limited availability of equipment such as data projectors
• Unwillingness to change
• Difficulty in linking ICT to the curriculum
• Needing IT facilities in lecture theatres rather than laboratories

_The use of ICTs is changing the ways universities operate:_ Changes are occurring, however and these will be discussed under the following headings:

1. Growth of on-line Learning
2. Rethinking what is being taught
3. Real-life learning experiences
4. Teachers collaborating to share and develop expertise

**Growth of on-line learning**

_Web-Based Courses:_ A key area of rapid growth has been within the discipline of on-line learning, both in web hosted environments and in packaged form on CD. The use of web-based courses is proliferating in Makerere University. Some faculties are providing wider curriculum choices and more individualized programmes through the use of web courses. However, there is still a gap in creating such online content due to lack of sufficient knowledge by the lecturers in most of the departments and faculties.

_Consortia of Schools:_ There are no documented co-operative ventures between universities that are in place. Unlike in other countries where there has been a consortium of schools and other institutions combining to create virtual school, that has not been the practice in Uganda and specifically in the Universities of interest in this research.

_Powerful School Intranets:_ Makerere University has worked hard and produces some online courses for students on its intranet. Some of these courses have been prepared, supplemented by courseware on CD. Resources developed by respective lecturers are now being uploaded on the web using the Learning Management Systems such as Blackboard and KEWL.

**Rethinking what is being taught**

_Authentic contexts:_ In this study, findings show that lecturers in the universities of interest using ICT effectively soon find that much of the repetitive “busy work” and unreal, contrived exercises are no longer relevant in an environment that allows students direct access to engaging information sources. This realization has led to new approaches being made to curriculum content and the encouragement of information literacy skills development. Students are being encouraged to use higher order thinking and given strategies for powerful learning in this new environment.

_Using higher Order Thinking:_ Makerere University other than any university in the study, has policies regarding the integration of ICT in the curriculum have seen
a greater emphasis on digital presentations- which has lifted senior scores and oral communication considerably. The “hunting and gathering” side of research is being closed down resulting in much more analysis and synthesis of information. Various lecturers in different departments administer and receive assignments from students online. This is however common with the post graduate students due to their smaller population compared to the undergraduate courses where the population is bigger.

**Teachers collaborating to share and develop expertise:** In all the visited universities, lecturers had no opportunity to meet and collaborate virtually through shared networks and web-spaces. These cooperative sharing and learning communities if availed could help transform the way lecturers work and their attitudes towards sharing and learning from each other.

**Conclusion:** Through the course of this study, there visionaries- whether they were lecturers or administrators- that created learning opportunities for students using ICT. Strategies for encouraging (or requiring) whole staff take-up of integrated curriculum use of ICT exists in the strategic plans of almost all universities visited. Once ICT becomes an integral part of student learning, teaching styles and classroom organization cannot remain unchanged.

**Tomorrow’s university:** From the arguments presented above, the following points were derived:

- Collaborative endeavors will become more common among universities.
- There will be a greater emphasis on communication, community and creativity for high, value-added organizations
- “The right connections and the right tools” become more important, emphasizing the need for effective technologies
- Interactive, asynchronous discourse becomes an important learning method
- Universities must recognize the need for authentic learning and that it cannot be assessed in old ways or against “old” criteria;
- The learning process becomes dynamic, exciting and fun using technology to learn in ways never before possible.

Information and communication technologies are beginning to have an impact on curriculum and classroom/lecture room design in each of the three universities visited. However there is a lot still needed to transform the traditional lecture room design and curriculum delivery methods to those where ICT will be a necessity. The use of the information gained during this research will drive organizations and individuals towards differing solutions in response to the needs of their students and learning communities. It is important that educators and administrators collaborate and learn from the mistakes, discoveries and best practice from other universities and researchers. There is much to learn from each other and much to gain for students.

It has also been noted that where rigorous examination systems and prescribed learning outcomes control the curriculum, it is much harder for innovative use of technology in the curriculum to occur.
References


Department of Education and Employment and Qualifications and Curriculum Authority Information and Communication Technology: The National Curriculum for England Key Stages 1 – 4, 1999


The over-capacity of telecommunications bandwidth in most of the world, coupled with the emergence of sophisticated workflow software has enabled business to move operations to where the talent is creating a 24 hour global enterprise. Sub-Saharan Africa (excluding South Africa), especially Science and Technology (S&T) research institutions are yet to take advantage of the immense collaborative synergies the information age has brought. This paper describes the African Virtual Environment Collaborative (AFRO-VELAB) that aims to remedy this shortfall by providing a medium where African S&T researchers can readily form research teams – based on areas of expertise – and take on sophisticated collaborative research projects that each could never attempt by going it alone.

Introduction

The twenty-first century, dubbed the information age, has seen remarkable increases in productivity and creativity, using low-cost, high-bandwidth information networks to facilitate collaboration between business units spread out across the world. No longer must business bring talent to a few central hubs located in the host country. Businesses can now tap talent at the source, using the abundance of band-width, coupled with the emergence of sophisticated work flow software to facilitate electronic and ‘face-to-face’ (via video-conferencing) communication and collaboration amongst employees in regional centres spread out across the world. The abundance of band-width, primarily in the form of fiber-optic cables, is as a result in the over-investment in capacity during the dot.com boom of the late nineties. “[Work flow software platforms], enable you to create virtual global offices – not limited by either the boundaries of your office or your country – and to access talent sitting in different parts of the world and have complete tasks that you need completed in real time. And so 24/7/365 we are all working.” (Friedman, 2006).

The main beneficiaries of this information-age spurned global collaboration have been the Americas, Europe, India, Japan, India and South-East Asia. Conspicuously absent from this new geographic boundariless world is Africa, especially sub-Saharan Africa, excluding South Africa (henceforth referred to as the region). In the region, Internet traffic, where available, primarily travels over relatively slow and expensive
Advances in Systems Modelling and ICT Applications

Part 6: ICT and Education

In the next few years however, the situation is expected to improve dramatically.

Several projects are underway that should significantly reduce the reach and cost of connectivity. For example, the 9,900 km East African Submarine Cable System (EASSy) will lay an undersea fibre optic cable from Durban South Africa to Port Sudan in Sudan, with landing points in Djibouti (Djibouti), Mogadishu (Somalia), Mombasa (Kenya), Dar-es-Salaam (Tanzania), Toliary (Madagascar), Maputo (Mozambique) and Mtunzini (S. Africa). Through land-based fibre optic cables the EASSy is to be connected to land-locked countries, Ethiopia, Uganda, Rwanda, Burundi, Malawi, Zambia, Zimbabwe, Botswana, Swaziland and Lesotho. Leading Telcom operator MTN Uganda’s strategic planning general manager, estimates that as a result of the cable project, “The cost of transmitting telcom traffic [in the region] will come down by 80 per cent” (Barigaba, 2006).

The business community in the region will undoubtedly take advantage of the potential improved connectivity brings. What of the regions science and technology (S&T) research institutions? Will they be left on the sidelines? This paper argues that improved connectivity, coupled with appropriate work flow software with embedded information systems will give researchers in African Science and Technology institutions a level playing field and new opportunities for excellence, similar to the lifeline these innovations gave software programmers in Bangalore, India.

The Africa Virtual Environment Collaborative (AFRO-VELAB) hopes to capitalize on the inevitable improved connectivity in the region to significantly foster collaboration amongst African S&T researchers, primarily, and between them and researchers in the rest of the world. It will foster the creation of synergies between the regions researchers, allowing them to attract research funds, and to make significant contributions to the knowledge-base relevant to Africa’s problems, in a way that none of them could if they went it alone.

**Contextual Example: Research On Attractant-bait Mosquito Traps Relevant To Africa**

**Introduction.**

Malaria continues to be a major public health problem in much of sub-Saharan Africa. Ogot (2004) discussed potential research areas, relevant to Africa, for the further development of attractant-baited traps as a viable avenue for the reduction of adult mosquito populations, and made a presentation on the challenges to forming viable research teams in this regard. Attractant-baited mosquito traps, first introduced in the United States for widespread consumer use in the mid-1990s, have the potential to provide an environmentally friendly alternative to insecticide use for the control of adult mosquito populations. Although several variations exist, all work around the same basic principles. Mosquitoes are attracted to their mammalian hosts by detection of carbon dioxide (CO\(_2\)), water vapour (H\(_2\)O) and other odors that mammals exhale, in addition to body heat. With reference to Figure 1, mosquito traps produce these mammalian attractants to lure the mosquito towards them.
Representative Products and their Applicability to Africa. Consider the Mosquito Magnet® (illustrated in Figure 2) produced in the US that runs off propane and electricity. With reference to Figure 3, three of the mosquito attractants (carbon dioxide, water vapor and heat) are produced from the controlled combustion of propane gas that is readily available as liquid propane gas (LPG) in tanks. The trap also has an electric fan, placed near the exit area of the attractants from the trap, that sucks in the unsuspecting mosquitoes into a net where they die from dehydration. Some models use a thermal electric generator to convert the excess heat produced from propane combustion into electricity, thereby eliminating the need for an external electricity source. Other models also release 1-Octen-3-ol, a natural flying insect attractant (Ramoni, 2001) that has been shown to increase the attractiveness of mosquito traps by several orders of magnitude (Kline, 2002).
With reference to the functional breakdown diagram in Figure 3, the manner in which two functional areas (attract production and mosquito trapping mechanism) of this trap is manifested would prevent its wide use in sub-Saharan Africa, especially rural areas. Specifically

1. The traps require the purchase of a 9 Kg propane tank after every 20 days of continuous operation. This cost, not to mention the initial investment of these traps, would be beyond the financial means of most of the target population.

2. Most rural areas do not have a regular supply of electricity. Some traps use a sticky tape as an alternative to the electric fans. This may present a suitable option for the African-context.

Potential Research Areas. Research continues in the developed world to design the next generations of mosquito traps that are more effective in significantly reducing adult mosquito populations, than those on the market today. These efforts, however, are all within the context of the developed world especially as relates to the purchasing ability of the end-user and their available resources. In order for these devices to be applicable for use in Africa, Africans must initiate research in areas that make the traps relevant for use in their context. Example areas include (Figure 4):
1. Development of attractant production methods that use locally available resources. For example, the use of biogas – produced within the trap itself – may be a viable replacement for propane. Biogas can be produced by anaerobic digestion of animal waste and consists primarily of methane (50%-80%), carbon dioxide (20%-50%). Two of the desired attractants, carbon dioxide and heat, are produced by the anaerobic digestion process. Novel methods need to be developed to effectively use these attractants as well as those produced from methane combustion.

2. Develop trapping mechanisms that do not require the use of electricity. Creation of sticky surfaces from local materials has potential, as does design of systems that can develop suction without the use of fans – for example the use of the convection currents from the combustion process.

3. As much of the trap itself as possible should be constructed from locally available materials. Costs can further be reduced if the end-user is able to construct/assemble the traps from a combination of a very small number of inexpensive purchased parts and other components that can be found around a typical homestead.

If one wanted to put together a team of researchers from the region to tackle the above problem and develop a viable device, how would one do it? Information on the research capabilities and interests of the region’s S&T researchers is virtual non-existent on the Internet in a manner that is readily searchable. For example, if one was searching for the region’s leading researchers on bio-gas, one would have to follow a circuitous route: Search relevant journals for bio-gas research articles; screen articles found for...
relevant bio-gas research areas of interest to you; screen sub-set of articles for regional scientists; contact scientist. Not a simple task. Yet, as this example shows, the need for Africa-based scientists, as part of the research team, would be imperative due to their extensive knowledge of local materials and conditions. Finally, even if a team could be formed, how would it effectively work together given the limited budgets of most sub-Saharan institutions, not to mention the significant time-wasted travelling?

AFRO-VELAB aims to harness the power afforded by modern information-embedded work flow software coupled with improving connectivity in the region to address these problems.

**AFRO-VELAB**

**Genesis:** The Africa Virtual Environment Collaborative (AFRO-VELAB) project originated during discussions at the “Collaboration of Researchers at African S&T Institutions and African Scientists in the Diaspora” workshop held at the 1st ANSTI Conference of Vice-Chancellors, Provosts, and Deans of Science in Engineering and Technology held in Accra, Ghana from 15-17 November 2005. At the workshop several impediments to effective collaboration were identified, that it was hoped AFRO-VELAB would address. These include (Ogot, 2005):

1. Lack of access to academic journals.
2. A serious lack of information about sister institutions on the continent. Need for better leverage of resources that exist on the continent.
3. Most current assistance/collaborations from overseas does not target engineering. Health, gender issues, and social sciences have thriving collaborations. An opportunity currently exists to establish collaborations for engineering.
4. All collaborations must be mutually beneficial and conceived by both parties to ensure success.
5. One should not assume that all researchers in the region or the Diaspora are willing to or are seeking collaboration. Effective collaborations must be of the willing, not the coerced.
6. Collaborations should not be restricted to S&T universities only. Significant research is carried out at government and donor-sponsored research centres within African countries, e.g. ICIPE, ILRAD, etc.
7. Scientists seeking collaborations must be proactive and not wait for things to come to them.
8. Collaborative projects should be directed at ‘Research for Development’ to enhance the impact of African universities on the development of African rural societies.

The AFRO-VELAB, therefore, is being developed in response to the above, and the current difficulty in identifying researchers and research capabilities at Africa’s S&T institutions. Such a system would greatly increase the visibility of African researchers and significantly increase intra- and inter-continent collaboration. Specifically, the secure virtual collaborative would be used to:
1. Rapidly form research teams through its database of African/Diaspora researchers (categorized by area of expertise) seeking collaboration. The database would include both university and government laboratory researchers.

2. Generate multi-institutional cross-disciplinary research proposals in response to Request for Proposals.

3. Provide an online collaborative environment to carry out research across geographic boundaries. The environment would also provide an efficient means to share information and to collectively work on documents (more efficient and better tracking than email attachments).

4. Keep projects on schedule by incorporating project management tools. The system would include major languages used by S&T institutions in Africa: English, French, Arabic and Portuguese.

**Overview Of Afro-velab**

The following sections will present mock-up screens of Phase I of the proposed system to give an indication of its capabilities.

**Secure Environment.** All users would need to be registered and be required to login (Figure 5). Registration and user accounts will be through member institutions or directly with AFRO-VELAB. At the time of registration, contact information, research interests, etc will be collected and categorized.

**Figure 5. Login Page**

![Login Page](image)
Research Groups and Projects. Users can belong to as many research groups as they like. Once logged in, all active research groups that the user is a member of are visible on the left, categorized by research projects (Figure 6). Also available is a list of tools available. These include email, a list of tasks (to do list), a calendar, etc.

Creating a New Research Group. As earlier alluded to, the largest perceived benefit of AFRO-VELAB lies in using the embedded information systems to rapidly put together research groups in response to a call for proposals on a particular research area. The system addresses one of the greatest challenges for collaborative research in Africa: access to information on research interests of different scientists and engineers at the multitude of science and technology research institutions on the continent. With reference to Figure 7, one can search for a researcher via a series of pull down menus that categorize scientists first by general researcher areas, then more specific later on.

Once the desired research area is selected, all scientists in the system under that area and their current affiliation are listed (Figure 8). Selection of one of them, David Otieno in this example, brings up the researcher's contact information, research areas and current Curriculum Vitae. One then has the option to contact the researcher to see if they are interested in participating, and if so, add them to your group. All group information and documents (similar to Figure 6) immediately become visible to the new group member.
Figure 7. Pull-down menus used to search for a scientist with a desired research interest

Figure 8. List of researchers in desired area and detailed information available if one selected

Figure 9. Clicking on a group lists all available documents as well as tools available to the project
Workflow Tools. Once groups have been formed and research initiated, AFRO-VELAB provides workflow tools that allows the sharing of documents, discussion groups, team calendar and tasks lists, as well as project management tools (see Figures 9 and 10).

**Charting The WAY Forward**

Dr. Ogot is currently leading the initiative to develop AFRO-VELAB as a vehicle to facilitate active collaboration amongst researchers in Africa’s S&T institutions, and between continent researchers and those in the Diaspora in research institutions outside the continent.

Development and Beta-Testing of the System (2 years). Working closely with the working group, partners – including the ANSTI, Association of African Universities (AAU), Western Hemisphere Diaspora Network (WAHDN) – and based at the Pennsylvania State University, efforts are underway in the development of the system. Beta-testing will follow shortly thereafter at the 10 working partner institutions listed in Table 1. Other institutions are welcome to serve as beta-testers as well as lend their expertise in the development of the system.

In parallel, development of training programs and training materials will proceed. It is envisioned that we will train trainers at each of the partner institutions. The trainers will then be responsible for all training of new users at their institutions. During the beta-testing work will begin in populating the system’s database with researchers seeking collaboration, their research interests and current contact information. This data will be gathered during the each user’s registration process. The system will also be configured to automatically issue a reminder to users on a semi-annual basis, to update their contact information. Initial seed funding for the development phase of the system has been sought from several developmental and science agencies (e.g., US NSF, EU, Canadian IDRC) and from commercial companies. We view the long term sustainability for the system may use a combination of subscriptions from participating universities and research institutions, combined with development grants from governmental agencies, private foundations and commercial companies.
Table 1. Initial Beta-Testing Institutions

<table>
<thead>
<tr>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty of Engineering and Technology, University of Botswana</td>
</tr>
<tr>
<td>University of Port Harcourt, Nigeria</td>
</tr>
<tr>
<td>Ecole Nationale Supe’rieure des Travaux Publics, Cameroon</td>
</tr>
<tr>
<td>University of Benin, Nigeria</td>
</tr>
<tr>
<td>Faculty of Engineering and the Built Environment, University of Johannesburg, SA</td>
</tr>
<tr>
<td>Obafemi Awolo University – Main campus, Nigeria</td>
</tr>
<tr>
<td>University of Buea, Cameroon</td>
</tr>
<tr>
<td>Howard College Campus, University of Kwazulu-Natal, S. Africa</td>
</tr>
<tr>
<td>University of Nairobi, Kenya</td>
</tr>
<tr>
<td>Moi University, Kenya</td>
</tr>
<tr>
<td>University of Kwazulu-Natal, S. Africa</td>
</tr>
</tbody>
</table>

Full Deployment (Year 3+). Once beta-testing is complete, there will be active recruitment of participating institutions on the continent and internationally. As with the beta-testers, a series of training of trainer sessions will be held at each newly added institution. The system will also be ported from Penn State University computer systems to a commercial hosting company in anticipation of the need for significant increase in capacity and bandwidth due to higher use. In addition, we aim to shift base of operations from Penn State University to Nairobi, Kenya where ANSTI is based.

The addition of features will follow soon thereafter including other languages used in Africa’s S&T institutions, mainly French, Portuguese, and Arabic; global research and notification of relevant research opportunities; and in close collaboration with African universities, development, hosting and deployment of remotely-controlled over the Internet educational laboratories that can be used by all institutions to mitigate their lack of adequate undergraduate educational infrastructure.

Other additions to the system could include provision of web-based science and mathematics software, digitization and hosting of all Masters Thesis and Dissertations, in addition to proceedings from regional technical conferences.

Concluding Remarks

Sub-Saharan African S&T institutions have lagged behind their western counter-parts for decades mainly due to lack of adequate funding and access to current information. The information age has given the region’s researchers the opportunity to leverage the capabilities of information-based work flow software and improved connectivity, to create synergies amongst each other that can foster significant collaboration allowing them to play a larger role on the world S&T stage, especially as pertains to solutions to problems relevant to the continent. The AFRO-VELAB, a work-in-progress, is
hopefully a step in that direction. By creating an enabling environment, researchers can spend more time working on solving Africa’s problems, and less on ‘re-inventing the wheel’ due to lack of information about work on-going at sister institutions or experience frustration at their inability to effectively collaborate with each other, forming teams to compete for large international research grants. We welcome any institutions, organizations or individuals willing to participate in the development and deployment of the system.

References


Part Seven

ICT and Gender
ICT Liberalisation and Changing Gender Relations in Contemporary Uganda: A Research Agenda

Aramanzan Madanda, Department of Women and Gender Studies, Makerere University, P.O. Box 7062, Kampala Uganda

Growth and use of modern Information and Communication Technology (ICT) in Uganda is comparatively recent. Rapid growth has been made over the last five years. Uganda’s ICT initiatives are implemented under the so-called “private sector led” policy framework. The objective of this paper is to identify research needs in the area of gender and ICT in Uganda, especially in relation to adoption/non adoption and changing gender relations at individual and community levels. The paper takes a gender perspective in discussing the current Ugandan ICT policy framework and ICT diffusion and adoption. It draws on arguments advanced by liberal economic theory, the theory of technology diffusion and adoption and feminist theory to illuminate key gender and ICT research issues in Uganda.

Introduction

The growth and spread of Information and Communication Technology globally has been mainly after the collapse of command economies led by the Soviet Union and the concomitant growth and spread of market economies led by the West (United States of America, Western Europe and to an extent Japan). The new technology, as Huws (1995) describes it, has enjoyed significant recognition. The United Nations Economic Commission for Africa claims the utilization of ICT is a way to help Africa reach major development goals such as: improvement in the quality of life; enabling economic integration; improved trade and improved linkages with the global community (UNECA, 2003)

Uganda’s policy framework recognizes ICT as a development catalyst. The country has in place institutions, including a unit in the office of the president, charged with ensuring that ICT benefits are harnessed. The country is proposing to set up a Ministry for ICT in the near future. The commencement of ICT liberalization in Uganda from the mid 1990s has been within the broad framework of liberal economic thinking. As such Uganda’s ICT policy framework provides for what is known as “private sector led” development (ROU, 2003).

Uganda’s ICT policy defines ICT as encompassing a wide range of communication and information technologies (ROU, 2003). The Association for Progressive Communications (APC, 2003:9) defines “modern” ICTs to include three categories:
The concept ICT has also been defined to refer to “communicative use of information technologies to bridge geographical distances when people interact with other people, activities and organizations” (Thulin and Vilhelmson, 2004:477). However, one of the important contestations for this theoretical discussion is that the general understanding and perception of ICT is frequently narrow. ICT is often widely conceptualized as a technical area where social and gender concerns should be peripheral.

The objective of this paper is to identify research needs in the area of gender and ICT in Uganda. The term gender refers to the economic, social, political and cultural attributes and opportunities associated with being male or female. In most societies, men and women differ in the activities they undertake, in access to and control over resources, and in participation in decision-making. Women as a group often have less access than men to resources, opportunities and decision-making. These inequalities are a constraint to development because they limit the ability of women to develop and exercise their full capabilities, for their own benefit and for that of society. The nature of gender definitions (what it means to be female or male) and patterns of inequality vary among cultures and change over time. A recognition of this variability assists in the analysis of specific situations and development of appropriate strategies for change. (Madanda, Ahikire, Kwesiga & Tanzarn, 2004:15)

Gender is often distinguished from sex. Sex is the biological difference between females and males (women and men or boys and girls). Sexual differences are universal. They are the same throughout the human race and involve women’s as well as men’s bodies. Men produce sperm, women bear and breastfeed children. These biological differences between women and men do not change. Although sex switchovers are possible with transmutation, transformed men and women have more limited sexual functionalities (ibid).

A useful cultural linkage between gender and technology has been made by Li and Kirkup (2005). Citing Silva (2000) they assert that from a “social constructionist perspective, the relationship between gender and technology is one where: ‘both gender and technology are processes; they are shaped, or acted out, in interaction’” (Li & Kirkup, 2005:11). “This argument suggests that both gender and technology change along with the societies they are part of, and the whole is both culturally and historically shaped. Gender identities in different national cultural contexts embody different expectations of the people performing them” (ibid).

**Research Problem, Purpose and Scope**

As earlier intimated, a wide array of literature places a lot of hope in ICT. ICT is presented as a cross-cutting tool and platform for remedying a cocktail of developing country problems: poverty, poor governance, disease, market shortage, gender inequality, female disempowerment, low quality education and constrained access to information. It is presented as an instrument with potential to leapfrog poor nations through development stages, making them reach their goals sooner than expected, or as a tool for destroying powerlessness through mobilising communities for empowerment
and social change (Isaacs, 2002; Hawkins, 2002; Aloo, 1995).

There is evidence that ICT sector liberalisation and opening up telecommunications to competition, though often opposed by monopoly operators, can bring in more investment and force down end user prices, making access more affordable, notably to poor women (Hafkin, 2002). For example, Uganda has experienced a tremendous growth in ICT under a liberal policy framework in terms of the spread and improvement in quality of services, infrastructure, distribution of hardware and usage. While the country has made modest strides in the use of computer technology, the use of mobile telephony has been especially phenomenal (ROU, 2005).

The pace and factors affecting computer technology and mobile telephone technology diffusion, adoption and usage in Uganda are yet to be fully documented. Further, the influence of computer and mobile telephone technology diffusion, adoption and use on gender relations in the household and the broader community, as well as the contribution towards women’s as compared to men’s socio economic empowerment, is not adequately understood. Conversely, the influence of gender relations on the distribution of benefits of computer technology and mobile telephony is yet to be clearly illustrated.

The purpose of this paper is to present a theoretical discussion of the basis for assessing the benefits and opportunities apparently brought about by the growth and spread in computer technology and mobile telephony within Uganda’s liberalised ICT policy framework to various categories of rural, urban, literate, illiterate, rich or poor women and men. We will make a theoretical assessment of factors that affect diffusion and adoption of computer and mobile telephone technologies by gender, and how the resulting usage affects living conditions and gender relations at household and community levels. The paper intends to argue theoretically that gendered power relations, gender ideology and cultural values influence how, when and what women relatively gain from the “booming” liberalised ICTs with mobile telephony and computer technology as examples. Ideology is used to refer to “the shaping of the way in which people [women and men] think the begetting of a … ‘consciousness’ and the imposition of a corresponding ideology or system of beliefs or values” (Crotty, 1998: 121). We will end this paper by identifying some key research issues and gaps that need to be addressed in the gender and ICT area.

Theoretical Debates

The complexity of the topic for discussion requires a number of theoretical perspectives. We will use three theoretical stances to make our case, namely: liberal/neo-liberal economic theory; theory of technology diffusion and adoption; and Feminist Theory.

Liberal/Neo-liberal Economic Theory. The development and application of ICT has led to the emergence of the so called information or knowledge society. The view of ICT as a tool that will allow developing countries to reach their development earlier than expected brings up a number of issues. First, it places the analysis of ICT in the realm of development theory. ICT development in Uganda has specifically been “private sector led.” This strategy places Uganda’s ICT development path into the
liberal/neo-liberal theoretical predisposition. This theoretical stance has roots traceable from Adam Smith’s theory of the spontaneous order of the market place. Adam Smith (1723-1790) was a key figure in the Scottish enlightenment who offered a scheme of the gradual evolution of society in which the key ideas were the division of labour driven by technical innovations via the regulation of the invisible hand of the market place that generated material and moral advance (Preston 1996: 48). In the post second world war period, the neo classical liberal free market approach has been influential. This classical position that touts the power of the free market was developed further by the neo-liberal development theorists. It is one on which the neo-liberal theoretical school of thought is firmly anchored.

The key institutional bases for the application of the neo liberal school of thought from the 1980s to date are the World Bank and the International Monetary Fund (IMF) whose advice to developing countries (including Uganda) is to liberalise their economies involving: the elimination of market imperfections, by removing controls on the private sector; the privatisation of state assets and liberalisation of foreign investment regulations; the elimination of market inhibitory social institutions and practices, abolishing subsidies, and liberalising employment regulation; imposition of restrictions on government spending and abolition of tariff regimes (Preston, 1996). Such programmes of liberalisation have usually required parallel programmes of political repression (ibid).

Uganda like many African countries was ridden by crises in the 1970s and accepted the IMF/World Bank advice (Ochieng, 2000). The liberalisation package was implemented from 1987 with the commencement of the economic recovery programme. It is this economic liberalisation spirit that is replayed in the “private sector led growth” argument generally and Uganda’s ICT policy framework in particular. Liberalisation theory is based on assumptions of free market and perfect knowledge of the market. These assumptions are, however, known to be unrealistic. Although evidence shows that a market mechanism associated with increased competition and the lowering of prices has brought technology nearer to the poor rural women and men (for instance Hafkin, 2002), some other evidence is to the contrary. This contrary evidence argues that liberalisation, is often plagued by the chronic problem of “market failure” often premised on the fact that there is nothing like perfect this or perfect that (Preston, 1996). Moreover the interaction of the free market with gender yields outcomes that have gender implications as discussed below.

The operation of the liberalisation theory brings up three actors in the economy: the state, which must restrain its operations to the minimum, the private sector, which should lead the development process and the community often constituted of individuals and Civil Society Organisations (CSOs). CSOs ostensibly operate in the interest of the individuals particularly those said to be powerless and voiceless. Preston (1996) claims that usually the state acts in concert with the powerful private sector in the interests of the latter. Critiques of the liberal growth model point out that history indicates that often the bourgeoisie merchants push for state protection and help to do their work well. Preston (1996: 62) adds that “the pure market model is a sophisticated
intellectual construct and ‘the market’ is not a natural given of human life.” He adds that “when the pure market model is examined, it rapidly disintegrates: … individuals are created in society, they are not prior to it; … economic competition is always imperfect; … sovereign consumers are a myth…” (ibid)

The liberal economic theory perspectives should help to explain how the state is acting to ensure that women as well as men take advantage of liberalised ICT. Understanding how the people at the grassroots meet their ICT needs in the context of a liberalised ICT framework is critical for understanding its contribution to social change. Might there be corporate social responsibility or abuses meted on women and men by the powerful ICT companies with proxy support from the state? It is important to examine which liberal ICT policy mix yields benefits that are equitable from a gender perspective. However to understand how ICT, a Western technology, is being taken on, we utilise the Theory of diffusion and adoption of technology by Rogers.

Theory of diffusion and adoption of technology. Rogers (1995) proposed a theoretical framework to explain how innovations spread over time and the factors that affect the speed and intensity of their spread. His book identifies and examines innovations in communications. Rogers’ theory presents two core concepts: diffusion and adoption. Diffusion is a “process by which innovation is communicated through certain channels over time among members of a social system” (ibid: 10). According to Rogers, diffusion is a complicated process influenced by the cultural, technological, economic and social readiness rather than the relative merits of the innovation itself.

Adoption is a positive decision to employ an innovation. Rogers points out four characteristics of innovations that can be examined in relation to adoption: relative advantage; compatibility; complexity; trailability and observability (ibid: 15-17). Relative advantage is the extent to which an innovation is perceived as better than the idea it supersedes. Advantages can be measured through economic, social or technological variables as well as through such measures as convenience and satisfaction. Compatibility is the degree to which an innovation aligns with such elements as current practices, routines, infrastructure and resources. Complexity entails diverse variables as the difficulty of the adoption process or the individual or organisational learning process surrounding how to use an innovation. Trailability concerns the ability to experiment with an innovation before making a full commitment and finally, observability is the degree to which the effects of an innovation are visible to others. Empirical research carried out in various contexts has indicated that adoption and diffusion of ICT takes a culturally gender path (see for example Li and Kirkup (2005). This leads us to feminist thought to explain the variation.

Feminist Theory. Feminist theory provides a rich critique to the mainstream development theories and paradigms. To understand the relative contribution of ICT in a liberalised policy framework to women, its adoption and use requires using lenses provided by feminist theory. Feminist theory is a broad subject. Tong (1995) lists seven forms of feminisms namely liberal, Marxist, radical, psychoanalytic, socialist, existentialist and postmodernist. The paper will present a selection of perspectives from various feminist orientations.
Sandra Harding (1991) provides an extended critique of science in a book titled *Whose Science? Whose Knowledge?* Harding contends that feminist discussions of science, technology and theories of knowledge occur at a moment of rising scepticism about the benefits that science and their technologies can provide. This paper contends that understanding ICT (an applied science or technology) require that it should be placed in the broader context of science. Feminist critiques argue that science was developed to service the specific interests of Western bourgeoisie upper and political classes to the exclusion of, largely, women but also men of other categories/classes – who are in fact the majority. The central question is can a technology (in this case ICT) with such roots have relevance to Uganda especially for categories such as rural women? If yes, in what way is it relevant?

Harding (1991) further notes that Western sciences and their technologies have been regarded with both enthusiasm and dread. Some responsibility for high standards of living has been attributed to science. Food, clothing, medical treatment, cars, airplanes, computers, television sets and telephones have become available due to scientific and technological development. On the other hand, the responsibility for atomic bombs, industrial exploitation, polluted air, vast oil spills, dangerous contraceptives, health profiteering, inappropriate use of valium, high infant mortality rates, and famine have been attributed to science. Harding adds that “from a sociological perspective, it is virtually irresistible to regard contemporary science as fundamentally a social problem … that should be conceptualised as no different … from alcoholism, crime, excessive drug use and poverty” (ibid :2). She contends that “eliminating sexist bias in …the … sciences might require redefining objectivity, rationality and scientific method” (ibid: 19).

Apart from the preceding extreme views, there are other feminist theoretical perspectives. The liberal feminist perspective provides a beginning point in articulating women’s access to technology and ICT. The main thrust of this theoretical perspective is that female subordination is rooted in a set of customary and legal constraints that blocks women’s entrance and/or success in the so-called public world (Tong, 1995:2). Because society has the false belief that women are, by nature, less intellectually and/or physically capable than men, it excludes women from the public. As a result of this policy of exclusion, the true potential of many women goes unfulfilled. Tong adds that “if it should happen that when women and men are given the same educational opportunities and civil rights, few women achieve eminence in the sciences, arts and professions, then so be it” (ibid). According to liberal feminists, gender justice, requires first making the rules of the game fair and, second, to make certain that none of the runners in the race for society’s good and services is systematically disadvantaged (ibid).

Marxist feminists think it impossible for anyone, especially women, to obtain genuine equal opportunity in a class society where the wealth produced by the powerless many ends up in the hands of the powerful few. One underlying argument thus is that “capitalism itself, not just the larger social rules under which men are privileged over women, is the cause of women’s oppression” (Tong 1989:2). So the argument goes: “If all women … are to be liberated, the capitalist system must be replaced by a socialist system in which the means of production belong to one and all” (ibid).
The socialist feminist project can be understood as nothing less than the confluence of Marxist, radical and psychoanalytic streams of feminist thought (Tong 1989:173). Socialist feminists are convinced that living in a class society is not the only or even primary cause of women’s oppression – the grip of patriarchy is a key explanation. As socialist feminists see it, traditional Marxist Feminists are able to explain how capitalism caused the separation of the workplace from the homestead, and why homestead activities were gradually devalued, but they are unable to explain adequately why capitalism assigned women to the homestead and men to the workplace (ibid:174). It is argued that Marxist analyses “give no clues about why particular people fill particular places. They give no clues about why women are subordinate to men inside and outside the family and why it is not the other way around” (ibid).

Socialist feminists have developed two different approaches – the dual systems theory and the unified systems theory to “provide a complete explanation of women’s oppression” (ibid: 175). Dual systems theorists maintain that patriarchy and capitalism are distinct forms of social relation and distinct sets of interest which when they intersect, oppress women in particularly egregious ways. For women’s oppression to be fully understood both patriarchy and capitalism must be analysed first as separate phenomena and then as phenomena that dialectically relate to each other. Dual systems theorists describe capitalism as a material structure or historically rooted mode of reproduction or sexuality, or even as a non material structure largely ideological and/or psychoanalytic. Unified systems theorists attempt to analyse capitalism and patriarchy together through the use of one concept. To these, capitalism is no “more separate from patriarchy than the mind is from body” (ibid).

A critical analysis of the feminist theoretical perspectives shows that its key tenet is that science and in this case ICT is a masculine dominated field. There are also barriers to be overcome to enable a relative access by women to ICT in a class society whose systems privilege men over women. The extent to which these feminist assertions can be substantiated in Uganda is a subject of the proposed research agenda.

A glimpse in the literature

Information on gender and ICT in Uganda is largely anecdotal and literature on the topic is just beginning to emerge. Available evidence on gender and the ICT policy framework in Uganda points to the fact that there is inclusion of gender issues to an extent. For instance the national ICT policy (ROU, 2003) has provisions for gender. The rationale for the ICT policy was to “stimulate industrial growth, commerce, infrastructure and linkage of rural and urban communities as well as uplifting of disadvantaged groups, while taking care of gender balance” (ibid: Section 2.3). Further, policy objective 4.2(s) aims “to ensure gender mainstreaming in information and communication programmes and in ICT development.” Under strategies to ensure access to public domain information, the policy pledges to “ensure that facilities for communication are provided at levels of cost, which match the ability of their users to pay, so as to reduce gender and spatial disparities in information access” (Objective 7(e)). Strategies for gender mainstreaming are listed as: taking into account gender information needs and
interests of both men and women in all information and communication programmes; developing mechanisms of increasing women’s access to information (especially in rural areas) so as to reduce the gender information gap; use non-discriminative gender sensitive language in information and communication programmes; and ensure equal participation in all aspects of ICT development (ibid: objective 10 a-e).

Juxtaposing the provisions of the policy against some crude measure like the gender issues proposed by Hafkin (2002) below suggests that gender issues in ICT policy in Uganda have not yet been significantly covered. For a start, we could ask the nested question: Does the ICT policy address: 1) the gender distribution of ICT Infrastructure? 2) Concerns of women’s literacy and access to ICT? 3) Gender insensitive language in policy and ICT discourse? 4) Social and cultural stereotypes? 5) The gender-related differences in uses of ICT? 6) The status and position of women and men in the ICT labour and industry? 7) Gender distribution of power and decision making in ICT? 8) The privacy and security of women and men? To what extent can the “private sector led” stance advocated in ICT policy deliver outcomes that are gender sensitive?

A comprehensive gender review of Uganda’s ICT policy framework is yet to be done. However initial reviews of the Uganda’s national ICT policy indicate that the policy makes a superficial reference to gender mainstreaming initiatives (Madanda, 2006). Additionally, personal communication with selected policy makers and implementers in Uganda’s government institutions points to complaints related to lack of a clear policy framework with specific gender indicators to be achieved. This has left the implementation of gender related ICT initiatives to good will. Further, a look at The Rural Communications Development Policy for Uganda, (ROU, 2001) does not take gender into account. The policy takes a gender neutral stand, and has no mention of women or gender.

A study examining the prospects and problems of Internet use and access focusing on the profile of Internet café users in Kampala concluded as follows. “... while initiatives such as cyber cafes have brought Internet and ICT closer to the people in developing countries, the bad news is that these initiatives, especially when they are commercially based may only be increasing the digital divide within poor countries” (Mwesige, 2003:84). According to the survey results, the typical Internet café user in Uganda is a 25 year old single male with no children, who has completed high school at the very minimum (ibid: 93). Studies like this one raise a number of disturbing questions. Are ICTs deepening the gender differences and inequalities or might there be hope that ICT may reduce gender inequality in the different areas: rural/urban, literate/illiterate or in some sectors such as education?

The Women of Uganda Network (WOUGNET) report on Women and ICTs in Uganda by Nakafeero (2005) suggests that the rapid increases in ICT services in Uganda has had benefits for women as well as presented new forms of gender inequality. Nakafeero lists a range of ICT contributions to women’s empowerment. She notes that ICT has contributed to widening of women’s economic opportunities through providing new forms of employment, market access, distance education opportunities and sharing information; women have especially used mailing lists such
as the WOUGNET (www.wougnet.org) list that have facilitated women’s participation in policy processes; ICTs have been used to increase women’s capacities in health, education and information dissemination; and that ICTs have enabled women to share their local knowledge, improve their ability to communicate with the wider world and increase self awareness and confidence.

While a range of authors are hopeful that ICT brings about useful gains, some are pessimistic about the role of ICT in addressing the needs of the poor women and men (Hafkin, 2002; Mijumbi, 2002; Huws, 1995). The literature also indicates that access and utilisation of ICT by women and men is mediated by a number of factors. Further it demonstrates that a conducive national policy framework is crucial if meaningful ICT benefits are to be equitably distributed to women and men (Hafkin, 2002).

While available literature points to limited ICT use in terms of the elite and email application, the documentation of women’s gender experiences in accessing and utilising ICT in the urban and rural areas remains scanty. While men dominate usage, there are women users too. But what are their experiences as compared to men’s?

Hafkin has repeatedly pointed out a number of key gender issues that contribute to the gender digital divide in a general way, which require attention. These include infrastructure distribution that is often skewed against places dominated by women; language that limits those who are not knowledgeable in the key international languages especially English; cost that makes ICT out of reach to the poor who are mostly women; time limitations that constrain women more; cultural and institutionalized barriers that systematically exclude women; and decision making structures that are male dominated (Hafkin, 2002). To what extent do these issues manifest themselves in Uganda’s ICT sector?

A range of Non Governmental Organizations (NGO) reports tout the success of rural based telecentre initiatives (Mijumbi, 2002; Mayanja, 2001; Karamagi, 2006). Alternative research however indicates that these multipurpose telecentre create a sort of digital divide that they were supposed to over come (Minges (2001) cited by Mwesige (2003)). School based telecentres (SBTs) that aimed at addressing community needs have tended to exclude the surrounding communities. During school time, SBT access is limited to the evening, making it hard for users traveling long distances to use the facilities (Kawooya, 2004:427). The extent of gender exclusion however has not been the central locus of these studies and needs attention.

The Internet originated as an American technology. It has been argued that ICTs are racially white, Western, male artifacts and that the Internet itself overtly embodies American cultural qualities in terms of its language and technical users values. Culture is a critical factor in influencing people’s acceptance and use of ICT (Li & Kirkup 2005: 3). These authors refer to research carried out in UK and China in which significant gender differences were found. Men were more likely than women to use email or chat rooms. Men played more computer games than women, Chinese men being the most active games players. Gender differences were higher in the British group than the Chinese group (ibid). Since gender is culture specific, it means that these conclusions may not be generalized hence need for a Uganda context specific studies.
Research in education institutions has found that computer technology use is partly influenced by the level and length of adoption. Those with prior computer skills and access often use more than the late adopters. (Li & Kirkup, 2005; Madanda, Bantebya & Kabonesa, 2006)

Research on ICT in Uganda, apart from taking a largely gender neutral stance, has mainly focused on computer technology and especially Internet access. Another key ICT whose spread has outpaced computer penetration is the mobile telephone. It is important to look at how this is impacting social change and gender relations. At the moment, pioneer studies in this area are being formulated or just starting to be carried out. Finally, a closer look at the existing literature points to the need for research into gender and ICT.

**So what are the Research Issues?**

The preceding discussion presented the case for gender and ICT research in Uganda. A number of specific questions arise: How are grassroots women and men located in the largely peasant social cultural structure responding to the modern technologies of information and communication technologies? What is the level of their livelihood integration into the knowledge society? What are the explanations for the change, if any? Is the change a result of willed actions (active) or structural changes (passive)?

We noted before that ICT has largely been conceptualised as a technical area, where gender and sociological considerations are perceived as peripheral. There have been attempts of looking at ICT from a gender or sociological perspective. For example (Obbo, 2001), addressed the portrayal of gender in websites while APC, (2003:76-81) addressed women’s access to ICT. The areas noted by Adeya (nd: 73-5) still remain largely unresearched in Uganda. These are the social impact of ICT; the influence of attitudes, expectations, organisation and the management of ICT impact; the relationship between ICT and poverty; ICT curricula; the kind of technical problems women encounter and “what exactly is meant by ‘women-friendly’ systems are areas that have inadequately or not been studied” (ibid). Further, the cultural values and social positioning of women in the adoption and utilisation of ICT have been glossed over. Evidence, especially from Uganda, on how ICT is expected to promote development in a way that meets the differentiated needs of various categories of women and men is still scanty. Policy frameworks most appropriate for the equitable delivery of ICT benefits remain largely unknown. To summarise this section, some of the identified research gaps are:

- The nature of ICT policy models that will bring about transformations with equitable gender outcomes
- A holistic treatment of computer technology based applications and how these have contributed to gender equitable empowerment processes or not.
- The social impact that the use of mobile telephony has had on the urban and rural men and women of varied categorisations.
• The nature of women’s ICT information needs and how they can be met.
• The access and utilisation of ICT and resulting information by women and men and how this has influenced gender relations at the household and community levels.

Summary
The paper attempts to demonstrate that the link between gender and ICT, with reference to Uganda, still requires a lot of work to be done. It identifies research gaps and issues raised by the current ICT policy framework by utilizing arguments drawn from feminist theory, liberal economic theory, the theory of diffusion and adoption of technology and available literature. While there is some evidence to say that ICT is providing opportunities to women and men, there is also contrary evidence which indicates that ICT is introducing new forms of inequality in Uganda. But the extent and nature of these inequalities, and mechanisms for redressing them, requires adequate research.

References
Adeya, C. N (nd) ICTS and Poverty: A Literature Review, International Development Research Centre


APC (Association for Progressive Communications) - (2003) ICT Policy: A Beginner’s Handbook; Johannesburg: STE Publisher


Madanda, A. Ahikire J., Kwesiga J.C., & Tanzarn (2004) Gender Training and Skills Development Trainers Manual; Department of Women and Gender Studies, Makerere University, Kampala.


 PART EIGHT

Bio Informatics
Towards a Reusable Ontology Framework for Biological Information Integration

Gilbert Maiga, Faculty of Computing and Information Technology, Makerere University and Ddembe Williams, London South Bank University

Reusable knowledge structures are becoming increasingly important in the task of information integration. In biological and clinical research where vast amounts of data are being generated and need to be shared, the task of developing infrastructures for information integration is an important one. In this paper, we provide the research background and approach to an ongoing study that aims to develop a reusable framework for the integration of biological and clinical research data. Such a framework shall provide users with the ability to access dispersed sources of biological and clinical data. The theoretical basis for the reusable framework is given. The proposed approach for developing and validation of the framework using the Protégé ontology development environment is also outlined. The study contributes to scientific knowledge by defining ontology structure that integrates biological and clinical information. This framework will be useful for removing the barriers between clinical and biological information for clinicians and the biological research community. The validated tool may also be used for training and extension of skills by researchers and clinicians.

Introduction

In molecular biology and clinical medicine vast amounts of data are being generated. There is a need to share this data, and so the task of developing knowledge structures for integration of this data is an important one (Davidson et al., 2004). The resulting shift from a period of data starvation to a period of data overload requires knowledge structures to manage information in collaborative research. The integration of data generated at all levels creates synergy between biological and clinical information systems. This synergy requires a framework for sharing new knowledge generated from molecular biology and clinical medicine (BIOINFOMED, 2003). Such a framework can provide users with the ability to access dispersed sources of biological and clinical data. Existing frameworks lack features to manage and integrate these vast amounts of data (Garwood et al., 2004).

New approaches are required for data collection, maintenance, dissemination, query, and analysis (Huang et al., 2004). While research in molecular biology and clinical work are producing a lot of data that needs to be organized, queried, and reduced to useful scientific knowledge, existing data management technology is often challenged by lack of stability, evolving nature, diversity and implicit scientific context that characterize biological data (Davidson et al., 2004).
The large amounts of data generated by research and clinical work have created a need for advanced tools to help researchers and clinicians find relevant data from the different sources (Gilder et al., 2004). The diverse resources of clinical and biological information are distributed at several sites and their full benefit for biologists and clinicians cannot be realized without seamless integration of clinical and biological information. To facilitate scientific research, drug discovery, and clinical practice, information scattered in disparate sources needs to be integrated into a cohesive framework. Reusable knowledge structures are becoming increasingly important in this task of information integration. Ontology-based tools have been used as a possible solution for these integration systems. However, even where ontologies are used, there is still no standard framework for tools that provide mapping between the different information sources. Available tools used differ in function, the type of input required and outputs they produce. This makes comparison of their performance largely meaningless (Natalya and Musen, 2002).

In this paper we provide the research background and approach for an ongoing study that aims to develop a reusable ontology-based framework for the integration of biological and clinical data. The theoretical basis for the reusable framework is given. The proposed approach for developing and validation of the framework using the Protégé ontology development environment is also outlined.

The study has the potential of contributing to scientific knowledge by defining ontology structure that integrates biological and clinical information integration. This framework will be useful for shedding understanding on how the barriers between clinical and biological information can be removed. The validated tool may also be used for training and extension of skills by researchers and clinicians.

The Challenges To Biological Information Management

The need for wider and more systematic dissemination of experimental molecular biology data is widely recognized (Prince et al., 2004). According to Orchard et al., (2003), issues that need to be addressed include;

1. The nature and variety of information that should be recorded about molecular biology experiments

2. The functions that should be provided by repositories that make these datasets available

3. The computational architecture that should be used to provide these functions

4. The nature of the tools that should be developed for use with such a repository.

Building biological databanks is complex compared with data in most other domains, largely due to the large amount and variability in data, and the fast changes in schema. It requires specific methods to integrate the different conceptual models of different researchers and databanks (Lambrix and Vaida, 2004). Another challenge is the volume, complexity, and dynamic nature of the data being collected and maintained in heterogeneous and distributed sources. New approaches are needed for data collection, maintenance, dissemination, query and analysis (Huang et al., 2004)
To maximize the utilization of the available biological information for scientific research and discovery, information needs to be integrated into a cohesive framework in order to facilitate exploration, allowing biologists, researchers and clinicians to answer complex questions that involve querying multiple sources to discover interesting relationships between database objects, such as relationships among protein sequences, structures, and functions (Hongzhan et al., 2004). As yet, there is no formally accepted standard representation to support the systematic dissemination and sharing of biological data (Garwood et al., 2004). To fully explore these valuable biological datasets requires advanced bioinformatics infrastructures and specific tools for biological knowledge management. Present tools lack the necessary methods and features to effectively link clinical and health applications to existing genetic databases and so new tools, integrated in the clinical workflows are required to search, retrieve and relate genetic and clinical data (BIOINFOMED, 2003).

**Biological Information Integration**

The main integration approaches adopted include warehouse integration, mediator-based integration, navigational integration and the use of ontologies. In warehouse integration, data from multiple sources is built into a local warehouse. All queries are executed on the data contained in the warehouse rather than in the actual sources. Warehouse integration emphasizes data translation, as opposed to query translation in mediator-based integration (Sujansky, 2001). It allows the user to filter, validate, modify, and annotate the data obtained from the sources (Hammer, 2003), a very attractive property for bioinformatics. However warehouse integration has to regularly check throughout the underlying sources for new or updated data and then reflect those modifications on the local copy of the data (Davidson, 2004).

Mediator-based integration concentrates on query translation. A mediator in this context is a system that is responsible for reformulating at runtime a query given by a user on a single mediated schema into a query on the local schema of the underlying data sources. Unlike in the warehouse integration approach, none of the data in the mediator-based integration system are converted to a unique format according to a data translation mapping. Instead a different mapping is required to enable queries on the mediator to be translated to queries on the data sources (Sujansky, 2001). Navigational or link-based integration emerged from the fact that many web sources require users to manually browse several web pages and data sources to obtain the desired information (Davidson, 2004). This model allows the representation of cases where the page containing the desired information is only reachable through a particular navigation path across other pages.

**Integrating Biological Information with Ontologies**

Ontology as a branch of philosophy is the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality. Broadly it refers to the study of what might exist (Smith, 2003). A computer ontology is an agreement about a shared, formal, explicit and partial account of a
conceptualization (Ushold and Gruninger, 1996). Ontologies are explicit specifications of a conceptualization, which allow the formal specification of the terms in a domain and the relations among them (Gruber, 1995). They make possible sharing and reusing of knowledge, support the interoperability between systems and also allow inferences to be done over them. The use of ontologies is a possible approach to overcome the problem of semantic heterogeneity. To achieve semantic interoperability in heterogeneous information systems, the meaning of the information that is interchanged has to be understood across the systems. Semantic conflicts occur whenever two contexts do not use the same interpretation of the information (Goh, 1997). Many ontology-based approaches to achieve interoperability have been developed (Ushold and Gruninger, 1996). An asset of Ontologies is their relative independence of particular applications. An Ontology consists of relatively generic knowledge that can be reused by different kinds of applications (Jarrar et al., 2002). They minimize the data-interchange problem by controlling vocabulary and are a critical element of the computational infrastructure for communicating biological data and knowledge. They are an important step to solving the current inability to exchange models between different simulations and analysis tools since this problem is rooted in the lack of standard formats for describing models (Hucka, 2003).

Ontology mapping tools find use in integrating biological information. Ceusters et al (2003) have shown that ontology engineering is an essential part of any solution to the problems of medical terminology only if it is grounded in a sound ontological theory. They proposed the theory of granular partitions in an effort to build a more realistic, general and flexible, framework embodying the strengths of both set theory and mereology while at the same time avoiding their respective weaknesses. While ontology mapping has been used to support biological information integration, their use for intelligent information integration depends on approaches that still uses ad-hoc or arbitrary mappings especially for the connection of different ontologies (Wache et al., 2001). Even approaches that try to provide well-founded mappings- either rely on assumptions that cannot always be guaranteed or face technical problems. They suggest the need to investigate ontology mappings on a theoretical and an empirical basis (Wache et al., 2001). While Ontology mapping tools find use in biological and clinical information integration, Natalya and Musen (2002) contend that these tools are so different from one another that direct comparison may not be possible. The tools vary with respect to the precise task they perform, the inputs on which they operate and the outputs produced, thus making comparing their performance to one another largely meaningless (Natalya and Musen, 2002). A survey of ontology mapping works does not compare the ontology mapping works reported under any specific framework, simply because such a framework does not exist (Kalfoglou and Schorlemmer, 2003) and further observe that attempts made to provide such a framework are far from being standards.
Bio-Ontologies and Related Work

Existing bio-ontologies pertinent to current trends in bioinformatics and molecular biology include the RiboWeb ontology, EcoCyc, the Schulze-Kremer ontology for molecular biology (MBO, Gene Ontology (GO), and the TAMBIS Ontology (TaO) (Stevens et al, 2000). RiboWeb’s primary aim is to facilitate the construction of three-dimensional models of ribosomal components and compare the results to existing studies. The knowledge that RiboWeb uses to perform these tasks is captured in four ontologies, namely: the physical-thing ontology, the data ontology, the publication ontology and the methods ontology. EcoCyc, like RiboWeb, uses an ontology to describe the richness and complexity of a domain and the constraints acting within that domain, to specify a database schema. EcoCyc covers E. coli genes, metabolism, regulation and signal transduction, which biologists can explore and use to visualize information. EcoCyc’s use of an ontology to define a database schema has the advantages of its expressivity and ability to evolve quickly to account for the fast schema changes needed for biological information.

MBO attempts to provide clarity and communication within the molecular biology database community (Stevens, 2000). It has concepts and relationships that are required to describe biological objects, experimental procedures and computational aspects of molecular biology. GO, like MBO, has database annotation as its main purpose. GO has grown up from within a group of databases, rather than being proposed from outside. It seeks to capture information about the role of gene products within an organism. It is essentially composed of three hierarchies, representing the function of a gene product, the process in which it takes place and cellular location and structure. GO contains a wide range of concepts, and provides a rich level of detail in its three hierarchies. TaO uses an ontology to enable biologists to ask questions over multiple external databases using a common query interface. TaO describes a wide range of bioinformatics tasks and resources, and has a central role within the TAMBIS system. The TaO is available in two forms - a small model that concentrates on proteins and a larger scale model that includes nucleic acids (Stevens, 2000).

Ontology Development Theory

According to Ceusters et al, (2003) ontology is currently perceived as the solution of first resort for all problems related to biomedical terminology, and the use of description logics is seen as a minimal requirement on adequate ontology-based systems. Description logics are a family of knowledge representation languages, which can be used to represent the terminological knowledge of an application domain in a structured and formally well-understood way. However, description logics alone are not able to prevent incorrect representations because they do not come with a theory indicating what is computed by using them. The authors show that ontology engineering is an essential part of any solution to the problems of medical terminology only if it goes hand in hand with a sound ontological theory. The thesis of Ceusters et al, (2003) is that to deal with grammatically complex patient records and other documents in (multiple) natural languages the required ontological solution must at least contain; knowledge about
terms and how they are used in valid constructions within natural language; knowledge about the world, i.e. how the referents denoted by the terms interrelate in reality and in given types of contexts; an algorithm that is able not only to calculate a language user's representation of that portion of the world that is described in the pertinent utterances; track the ways in which people express what does not represent anything in reality. For all of these purposes the required solution must be grounded in an ontological theory.

**Ontology Mapping Frameworks**

Ontology mapping refers to the connection of an ontology to other parts of the application system. There are mappings between ontologies and the information they describe and mappings between different ontologies used in a system. An ontology mapping consists of a collection of functions assigning the symbols used in one vocabulary to the symbols of the other (Kalfoglou and Schorlemmer, 2003). Mappings support the integration process for ontologies that have to be linked to actual information. Their use for intelligent information integration depends on approaches that still use ad-hoc or arbitrary mappings especially for the connection of different ontologies.

Ontology mapping provides a common layer from which several ontologies can be accessed and exchange information in semantically sound manners. Developing such mappings has been the focus of a variety of works and frameworks including; Fernandez-Breis and Martinez-Bejar's Cooperative framework for ontology integration; the MAFRA framework for distributed ontologies in the Semantic Web (Kalfoglou and Schorlemmer, 2003) and the OntoMap framework for integrating upper level ontologies (Kiryakov et al., 2001). Ontology mapping tools available as plug-ins to the protégé editor have also been developed including SMART and PROMPT (Natalya and Musen, 2002)). Kalfoglou and Schorlemmer (2003) provide a comprehensive overview of ontology mapping works that reveals the problems faced including the lack of a standard terminology, hidden assumptions or undisclosed technical details, and the scarcity of evaluation metrics. Kalfoglou and Schorlemmer (2003) point out that the survey does not compare the ontology mapping works reported under any specific framework, simply because such a framework does not exist and further observe that attempts made to provide such a framework are far from being standards. They presented the IF-Map, an automated ontology-mapping framework based on channel theory of Barwise and Seligman (1997), a mathematical theory of semantic information flow. It is a mathematical model that aims at establishing the laws that govern the flow of information in a distributed system.

**Proposed Methodology**

This section outlines the theory, and methods to be applied to answer the research questions and achieve the specific objectives of the study.
Literature Review

A continuous review of literature from research journals, publications and biomedical databases shall be conducted:

- On biological and clinical information systems: It shall be conducted to get requirements for integrating biological and clinical information.

- On ontology development theories: Current theories used in biological and clinical ontology development are to be examined in order to identify quality metrics to be used as criteria for evaluating good biological and clinical ontologies.

- On ontology development tools: An evaluation of existing tools shall be done to assess their suitability for developing biological and clinical ontologies. A suitable tool shall be selected for use in developing an ontology.

- On ontology mapping theory, frameworks and tools: It shall be used to establish requirements (tasks, inputs and outputs) and metrics in a new framework for mapping between good quality biological and clinical ontologies.

Interviews and Questionnaires

Data on requirements for the new framework shall also be gathered from research scientists, biologists, clinicians and health care workers in research institutes, health departments and hospitals. It shall be used to build the clinical and biological ontology.

Ontology-Based Approaches to Integration of Information

There are single, multiple and hybrid approaches to ontology based integration (Wache et al., 2001). Single ontology approaches use one global ontology providing a shared vocabulary for the specification of the semantics (Figure 1a). All information sources are related to one global ontology. Each source is simply related to the global domain ontology.
In Multiple ontologies, each information source is described by its own ontology where the semantics of an information source are described by a separate ontology (Mena et al., 1996). In principle, the source ontology can be a combination of several other ontologies but it cannot be assumed, that the different source ontologies share the same vocabulary. The advantage of multiple ontology approaches is that no common and minimal ontology commitment about one global ontology is needed (Gruber, 1995). Each source ontology can be developed without respect to other sources or their ontologies. Hybrid approaches overcome the drawbacks of the single or multiple ontology approaches. Hybrid approaches are similar to multiple ontology approaches, the semantics of each source are described by its own ontology. But in order to make the local ontologies comparable to each other they are built from a global shared vocabulary (Wache et al., 2001).

**Data Schemas Vs Ontology Models**

Ontologies and data models, both being partial accounts of conceptualizations consider the structure and the rules of the domain that one needs to model (Guarino and Giarretta, 1995). But, unlike task-specific and implementation-oriented data models, ontologies, in principle and by definition should be as generic and task-independent as possible. Reusability and reliability are system engineering benefits that derive from the use of ontologies (Ushold and King, 1995). Another difference between an ontology model and a data schema is one of purpose (Zhan et al., 2002). An ontology is developed in order to define the meaning of the terms used in some domain whereas a schema is developed in order to model some data. Although there is often some correspondence between a data model and the meaning of the terms used, this is not necessarily the case. Both schemas and ontologies play key roles in heterogeneous information integration because both semantics and data structures are important. A schema, whether specified using XML or some database schema language, needs an associated formal ontology in order to make the semantics of the resource clear. When the meaning of data and schemas is made explicit using an ontology, programs can be designed that exploit those semantics.
The Proposed Ontology Framework

The proposed framework borrows from the thesis of Natalya and Musen (2002) that it is not possible to compare ontology mapping tools as they differ according to the precise tasks they perform, the inputs on which they operate and the outputs that they produce. Wache et al. (2001) identify the single, multiple and hybrid approaches to ontology integration based on content explication (Figure 1). In the single ontology model, there is a global shared ontology for mapping between the information sources. In the proposed theoretical framework, the thesis of Natalya and Musen (2002) is combined with the single ontology-mapping model described by Wache et al. (2001) for content explication. The functions, inputs and outputs that should be provided for ontology integration systems in large-scale biological and clinical information derived from the literature review and interviews are to be incorporated into this framework (Figure 3).

Figure 3: The Proposed Framework

Validation of the Framework

Ontology validation checks for consistency and quality (Gomez-Perez, 1994). According to Maedche et al., (2003) the ontology validation process should answer the following questions; does the ontology properly represent users’ requests and ontological elements from the real world? Is the knowledge represented in the ontology truthful and consistent? In the proposed study, validation aims to check whether the framework: (1) consistently represents different biological and clinical information sources; (2) supports mappings between the different biological and clinical information sources. The deliverables in the framework validation process are a global ontology mapping between biological and clinical databases. It shall be checked for consistency using theoretical biological and clinical test data.

Ontology Development

This study shall follow the well known ontology construction guidelines developed by Gruber, (1995), Ushold and Gruninger (1996), Natalya and McGuinness (2001) that
encourage the development of more reusable ontologies. The overall process, however, shall move through the following steps (Natalya and McGuinness, 2001): determine the domain and scope of the ontology; consider reusing existing ontologies; enumerate important terms in the ontology; define the classes and class hierarchy; define the properties of classes - slots; define the facets of the slots; create instances. The Protege ontology development environment is proposed for this study (Protege, 2005). Data for developing and testing the ontology shall be extracted from publications, biomedical databases, research institutes, health departments and hospitals. More data shall also be obtained from clinical and biological research staff using interviews and questionnaires.

Conclusion

The paper identifies the challenge faced in biological and clinical research where vast amounts of data are being generated and need to be shared. Developing knowledge structures for information integration is shown to be important in meeting this challenge. The paper provides the research background and approach to an ongoing study that aims to develop a reusable framework for the integration of biological and clinical research data. The methodology outlines an approach towards such a framework and its theoretical basis.

The literature also helps to identify some unresolved issues. While the important role of ontologies in structuring knowledge for information integration is recognized, the literature points to a lack of standards for comparing ontology mapping systems (Kalfoglou and Schorlemmer, 2003). There are no metrics for comparing ontology mapping works. Ontologies are seen as useful in representing medical terminology only if their construction conforms to sound ontological theory and how well existing bio-ontologies do conform to good theory remains in doubt (Ceusters et al, 2003).

References


Lambrix, P. and Vaida, J. (2004). Towards transparent access to multiple biological databanks Department of Computer and Information Science Linkopings universitet, Sweden


