The ICT Age

The ICT Age

Edited by

Anton Ravindran and Edmond Prakash

Cambridge
Scholars
Publishing

The ICT Age

Edited by Anton Ravindran and Edmond Prakash

This book first published 2016

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

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

STEERING INNOVATION, SERVING SOCIETY

ISBN (10): 1-4438-8714-5
ISBN (13): 978-1-4438-8714-4

TABLE OF CONTENTS

Editorial ... vii

Contributors ... ix

Cloud Computing Adoptions in Public Sectors ... 1
Fan Zhao

Cloud Computing Paradigm & Changing Roles of CIO ... 35
Anton Ravindran and Andy Koronios

Procedural 3D Caves, Clouds and Architecture Generation Method based on Shape Grammar and Morphing ... 87
Tomasz Zawadzki, Slawomir Nikiel and Eraldo Ribeiro

Jav-Atar: A Video Game Contest to Promote Computer Science ... 125
Andrés Adolfo Navarro-Newball

Automated Dance Choreography ... 141
Nigel Gwee

Internet of Things (IoT) and Cloud Computing ... 167
Huifang Deng, Caifeng Zou and Nazia Anwar

Cognitive Rehabilitation Gaming System ... 189
Ahmed Mohammed Elaklouk and Nor Azan Mat Zin

Context-Aware E-Learning Infrastructure ... 221
Stanimir Stoynov, Hussein Zedan, Emil Doychev, Veselina Valkanova, Asya Stoyanova-Doycheva and Vladimir Valkanov

Automating Helpdesk in the ICT Age ... 281
Nithya Kannan and Edmond Prakash

Editorial

We are living through exhilarating times and constant changes. Notwithstanding challenges faced, the world is on an exceptional ride towards advancement in every facet of life largely driven by the unprecedented developments in ICT over the past 3 decades. This book comprising 9 chapters covers the latest developments in cloud computing, game-based learning and virtual environments. We hope these chapters will help the reader to get a grasp of the ICT age in these 3 areas, as well as provide the background knowledge to extend this to other areas that are relevant to the reader. As this book is a collection of state-of-the-art research findings which were developed into chapters, it is not for programmers or developers or technologists looking for technical guidance or programming tid-bits. The intended audience for this book are academics, researchers including PhD students and practitioners.

First this book presents developments in Cloud computing technology and how we process and communicate such information. Cloud is increasingly being used in high-performance computing, where computer servers can be accessed and used over the cloud. Files and backup services are another area. We can store, modify and retrieve information anytime, anywhere, both securely and in a scalable manner and these are a few benefits mentioned as the vision of cloud computing to be the fifth utility. Challenges in cloud technology and its impact on the ICT age are emphasized in this book.

Next the emphasis is on game based learning in the ICT Age. Traditionally, game based learning and gamification, had little or no place in the learning curriculum. However, today, in classrooms, a range of game based technologies have found their way into the learning space, and are now being used to help make teaching and learning experiences that were considered demanding and challenging to become more appealing to the learners. This is transforming the way we teach and engage with learners, as young and old learners find game based learning more motivating. This is because of staged learning through different levels, through problem solving as well as engaging in critical thinking and decision making. Furthermore, in game based learning through ICT, learners engage actively through community development, not only within the classroom, but through global interaction.

The book also discusses virtual environments. ICT helps represent the virtual worlds through shape models, their behaviour and through interaction. The aim is to create a real life experience, but through the medium of ICT. Models can be created that mimic the real world objects, or even objects that cannot be seen in the real world.

We thank all the authors for the chapters they have contributed to this book. We also thank the staff at GSTF and CSP for all their hard work in making the coming together of this book on ICT Age.

We hope the book will help to widen the understanding of cloud computing, gaming and virtual environments and for readers to become more involved in these technologies whether as researcher, developer, practitioner, user or in any other capacity.

Dr. Anton Ravindran CEng, FBCS and Prof. Edmond Prakash

CONTRIBUTORS

Dr Huifang Deng is a full Professor at South China University of Technology in Guangzhou, China. He received his BSc and MSc in Physics in China and his PhD degree in Modelling of 2D TFT-LCD systems by finite elements from University College London (UCL), University of London, UK. From the year of 1989 to 2004, he was studying, working and living in the UK (for more than 15 years). From the year of 2001 to 2004, he served the Sunrise Systems Limited as a Chief Technical Officer and Chief Scientist in Cambridge, UK. In September 2004, he took the position as the Dean of Software School at South China University of Technology. His research interests include RFID technology and applications, Internet of things, cloud computing, data science, information diffusion theory in social networks, social computing, computer modelling, SOA, service computing, high-performance computing, scientific computing, and CAE/CAM, etc. So far, he has got over 100 papers published and co-chaired more than 10 International Conferences.

Asst Prof Emil Doychev received his doctoral degree from the University of Plovdiv, Bulgaria. Currently he is an Assistant Professor at the same university. Also, he is the Head of the Development Department in a private software company in Bulgaria. He is interested in: eLearning tools and environments, agile software development, and service-based architectures.

Dr Nigel Gwee is an associate professor of computer science at Southern University, Baton Rouge, Louisiana, U.S.A. He holds doctoral degrees in both musicology and computer science from Louisiana State University, Baton Rouge. He has published papers on software engineering and complexity analysis, and has developed computer programs for music and dance.

Nithya Kannan received her B. Tech degree in Information technology from Anna University, India, in 2011 and Masters in Information Management and Security from University of Bedfordshire, UK, in 2012. She is currently working in Accenture UK as a Software Engineer. Her research interests include network-level security, OS-level security and physical security (bio-metrics).

Prof Dr Andy Koronios is the Head of the School of Information Technology & Mathematical Sciences, at the University of South Australia. Andy holds academic qualifications in Electrical Engineering, Computing and education and a PhD from the University of Queensland. Andy has extensive experience in both commercial and academic environments and has research interests in information management & governance, e-commerce and the management and strategic exploitation of information as well as the changing role of the IT Leader. Andy has established two University Research Labs and a funded Research Centre and is currently the Director of the Strategic Information Management Research Lab in the Advanced Computer Research Centre. He has worked both as a consultant as well as a professional speaker on IT issues in Australia and South East Asia and has over twenty years' experience in the academic environment. He is a member of the Board of Governors of Global Science & Technology Forum, a Fellow of the Australian Computer Society, and a Founding Fellow of the International Institute of Engineering Asset Management, Editor-In-Chief of the International Journal of Information Quality and Associate Editor of the ACM Journal for Information Quality.

Assoc Prof Nor Azan Mat Zin is an Associate Professor at the Software Technology and Management Research Centre (SOFTAM), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia. Her research interests include Serious games, e-learning Technologies and Adaptive Systems.

Ahmed Mohammed Elaklouk is an Assistant Professor at the University College of Science and Technology in Palestine. His research interests include Multimedia systems development, Serious game design and Human-Computer Interactions.

Dr. Andrés Adolfo Navarro-Newball: Computer Scientist from the GSTF's partner university, Pontificia Universidad Javeriana, Cali (1994); MSc in Computer Graphics and Virtual Environments from the University of Hull, UK (1998); MEng in Networking from the ICESI University, Cali (2001); PhD in Computer Science (Graphics Lab) from the University of Otago, New Zealand (2010); beneficiary of the Coimbra Scholarship at the University of Siena, Italy (2006); beneficiary of the COLFUTURO and

University of Otago Scholarships (2009-2010). Dr Navarro-Newball was a Visiting Research Fellow at Manchester Metropolitan University, UK (2009) and a founding member of the Colombian Telemedicine Centre (2004-2007). Currently, he is a lecturer at the Computer Science Department, Pontificia Universidad Javeriana, Cali, where he is also part of the DESTINO research group, which is recognised by COLCIENCIAS (Colombian research authority). Finally, he has been a committee member for the CGAT and for the International Conference on Games and Virtual Worlds for Serious Applications. Dr. Navarro-Newball currently collaborates with the Museum i+D+C (Complutense University) research group.

Dr Slawomir Nikiel is currently the Professor at the Institute of Control and Computation Engineering, Department of Electrical Technology, Computer Science and Telecommu¬nication, University Of Zielona Góra, Poland. In 1995 he was awarded a Ph.D. degree in Computer Science in the field of Image Processing at the Technical University of Wroclaw, Poland. In 2009 he was awarded a D.Sc. degree in Computer Science in the field of computer graphics, at the Technical University of Wroclaw, Poland. His current research interests include virtual reality systems, game programming and multimedia systems. He has published over 60 papers in international conference proceedings and journals.

Dr Edmond Prakash is Professor and Head of Creative Technologies, Westminster School of Media, Art and Design at the University of Westminster, United Kingdom. He currently leads the Technology Enhanced Facility and has a key responsibility for promoting and embedding the use of digital technologies in research, teaching and learning across the faculty that includes photography, music, animation, television, film, media, journalism, fabrication and fashion. His research interests lie in the intersection between digital technologies and the professional learning. He is recognized internationally for his ground-breaking work in the field of 3D modelling, graphics, animation and games technologies. He has held positions at the Supercomputer Centre at the Indian Institute of Science, India; Nanyang Technological University, Singapore; Manchester Metropolitan University, UK; University of Bedfordshire, UK and Bournemouth University, UK. Over the past years, he has collaborated extensively with international academics and industry researchers. He has held visiting positions at NCSA, University of Illinois at Urbana Champaign, USA; CVC at Stony BrookUniversity, USA; Touch Lab, MIT, USA, Beijing Normal University, China and Pontificia

Universidad Javeriana, Cali, Colombia. He is currently member of the board of advisors at Global Science & Technology Forum.

Dr Anton Ravindran, is currently the CEO and Founder of RapidStart, President of Global Science and Technology Forum (GSTF) and an Adjunct Professor at the Faculty of Information Systems at Bina Nusantara University, Jakarta, Indonesia. He was a visiting Prof and Researcher at the Institute for Research in Applicable Computing (IRAC) at the University of Bedfordshire and Adjunct Professor, Dept of Management, Birla Institute of Technology & Science, Pilani – Dubai Campus. He is currently the Managing Editor of Journal on Computing (JoC) and Global Business Review (GBR) published by GSTF in print as well hosted and published on Springer's Open Access Platform – Global Science Journals (GSJ). He is the author and co-editor of the book "Business Review: Advanced Applications, Forging Connections between Computational Mathematics and Computational Geometry (published by Springer) as well as "Innovations in Technology Enhanced Learning". In the industry, he worked for several major MNCs in technical and management capacities including Computer Associates, Sun Microsystems, Oracle and Singalab (former IBM & NCB JV). Dr Ravindran founded and spearheaded 4 IT companies and has won several entrepreneurship accolades and awards including Singapore ASME Top Entrepreneur for two consecutive years (2005 & 2006) and winner of Singapore Indian Entrepreneur for large enterprises (2006) amongst others. He has been part of various professional bodies and committees including National Internet Advisory Committee (Singapore); Chair, Working Group for Cloud Computing Professional Development(SCS-IDA); President, Cloud Computing Chapter (SCS); Vice Chair Electric, Electronic and Allied Industry Group (EEAI of SMF); and President, Singapore Education International (SEI) amongst others. He holds a BSc, an MBA, and a Doctorate degree in IT Management from the University of South Australia, having been awarded the International Information Technology Management Academic Merit Scholarship. He is a fellow of the British Computer Society and a Chartered Engineer (UK).

Dr Eraldo Ribeiro is an Associate Professor of Computer Sciences at Florida Institute of Technology. In 2001, he was awarded a Ph.D. degree in Computer Vision in the Department of Computer Science at the University of York, U.K. under the supervision of Professor Edwin R. Hancock. Prior to this, he gained a Master of Science Degree with distinction in Computer Science (Image Processing) at the Federal

University of Sao Carlos (UFSCar- SP), Brazil in 1995. He has published over 50 papers in refereed conferences and journals.

Prof Stanimir Stoyanov received his doctoral degrees from the Humboldt University of Berlin, Germany, and the De Montfort University of Leicester, UK. Currently, he is a Professor at the University of Plovdiv, Bulgaria, where he is also the Head of the Department of Computer Systems. His research interests include: context-aware and adaptable software architectures, intelligent agents and multi-agent systems, agent- and service-based architectures, middleware, eLearning tools and environments, and mobile eLearning services.

Assoc Prof Asya Stoyanova-Doycheva received her doctoral degree from the University of Plovdiv, Bulgaria. Currently, she is an Associate Professor at the University of Plovdiv. Her research interests include: refactoring of software systems, ontologies, software technologies, and eLearning tools.

Asst Prof Vladimir Valkanov received his doctoral degree from the Bulgarian Academy of Sciences, Sofia, Bulgaria. Currently, he is an Assistant Professor at the University of Plovdiv, Bulgaria. His research interests include: context-aware and adaptable software architectures, intelligent agents and multi-agent systems, agent- and service-based architectures, refactoring techniques, eLearning tools and environments, and mobile eLearning services.

Dr Veselina Valkanova received her doctoral degree from the Bulgarian Academy of Sciences, Sofia, Bulgaria. Currently, she is a Senior Teacher at the Sofronii Vrachanski School, Plovdiv, Bulgaria. Her research interests include: eLearning tools and environments, aspects of creativity and teaching methodologies for the secondary school.

Asst Prof Tomasz Zawadzki received his Ph.D. degree and specialization – in Computer Graphics in 2014 at the Faculty of Computer Science, Electrotechnics and Automatics at the University of Zielona Góra, Poland. Currently he is working as Assistant Professor at the Faculty of Architecture, Design and Fine Arts at Girne American University, Cyprus. In 2011 he received a grant from Polish Ministry of Science and Higher Education and in 2012 from European Union. In 2013 he was visiting scholar at Florida Institute of Technology, USA. He has published over 25 papers in international journals and conference proceedings.

Prof Hussein Zedan received his PhD in Mathematics from the Bristol University, UK. He was a Research Professor and Technical Director of the Software Technology Research Laboratory at the De Montfort University, UK. Currently, he is a Professor at the Applied Science University of Bahrain. His research interests include: software engineering techniques, IT strategy and architecture, formal semantics, computer security, and critical systems.

Dr Fan Zhao is an assistant professor in the Lutgert College of Business at Florida Gulf Coast University, USA. He received his PhD degree in management information systems from University of Nebraska-Lincoln, USA. He has published extensively in various research areas of Information Systems, such as enterprise resource planning (ERP), data mining, human–computer interaction (HCI), e-commerce, and customer relationship management (CRM).

Caifeng Zou is serving Guangdong Mechanical & Electrical University as a lecturer and, meanwhile, pursuing a PhD degree in Computer Software and Theory at South China University of Technology in Guangzhou, China. She received her MS in Computer Science from Sun Yat-sen University in Guangzhou in 2008 and her BS in Computer Science from Shanghai University in Shanghai in 2006. Her research interests include internet of things, data mining, and cloud computing etc.

CLOUD COMPUTING ADOPTIONS IN PUBLIC SECTORS

FAN ZHAO

Abstract

Despite the number of benefits cloud computing technology has to offer all organizations, both public and private, risk elements to consider still remain. Previously, these elements were used in evaluating if an organization would move towards the cloud or not. Today, it is no longer a question of if, but a question of when. The task of evaluating the benefits and risks is now part of the gap analysis used to determine which vendor to contract with. In the public sector, government entities face a different set of rules and higher scrutiny when it comes to data security and budgeting. By evaluating what steps early adopters have taken in their investigation and analysis processes prior to choosing a vendor, this paper explores the essential information about cloud computing adoptions in public sectors and provides some guidelines that governments can use in their decision making process when considering cloud services.

Keywords: Cloud Computing, Public Sectors, e-government

1. Introduction

Cloud computing represents a change in the way organizations do business with regards to technology. It provides flexibility and cost savings to an organization in that they no longer need to invest large amounts of capital to get a software project up and running.

With its growing popularity over the last few years, many people have offered a variety of definitions for the "Cloud" (Buyya, 2008; Wang, 2008; Vaquero, 2009; Geelan, 2009). Some of the key words and phrases used to describe the cloud pulled from such definitions include: on-demand, pay-as-you-go, virtualization of resources, outsourced, web-based and elastic. In an effort to provide potential consumers with a "tool" to compare cloud services and deployment strategies, in 2011, the National Institute of Standards and Technology (NIST) published the following working definition of cloud computing (Brown, 2011):

> "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management efforts or service provider interaction (Mell & Grance)."

The definition goes on to define the essential characteristics of the cloud deployment models and the cloud delivery models.

1.1 Cloud Benefits and Concerns

Following the rush to define the cloud, the literature shifted to a review of its benefits and shortcomings/concerns in terms of organizations moving to the cloud. A survey of 137 federal government technology professionals revealed that among the participants, the number one business driver of cloud computing was lowering IT costs, while the top concern was cloud security (Biddick, 2011). The literature indicates that some of the key advantages of cloud computing include lower up-front costs to implement new technology, rapid implementation, scalability, low to no maintenance costs, and energy efficiency/green technology (Gruman & Knorr, 2008; Armbrust, 2009; Martson, 2011; West, 2010). Areas of concern included security, lack of control (data and applications are stored off site), compliance, integration, and reliability (Armbrust, 2009; Paquette, 2010; Martson, 2011; Schwartz, 2011). Many of these security issues are now being addressed (Kundra, 2010).

1.2 Moving to the Cloud

As the benefits become too great to ignore and the shortcomings are worked out as the technology evolves, current literature focuses on how organizations, both public and private, can take advantage of what the cloud has to offer through case study analysis and predictions for the future of cloud computing. In early 2011, realizing that the size of the federal IT budget could be reduced drastically by utilizing cloud computing, the U.S. Chief Information Officer released his "Federal Cloud Computing Strategy" report which discussed the newly implemented "Cloud

First" policy. This policy requires federal agencies to "evaluate safe, secure cloud computing options before making any new investments (CIO Council, 2011). The strategy sped up the process of federal government agencies moving toward the cloud, thus provided researchers many case study opportunities (Wyld, 2009; Biddick, 2010; Kundra, 2010; West, 2010). Many of the cloud case studies that have been previously published are reviews of what the organizations have implemented and how they have benefited. The case study literature lacks in reviewing the steps taken in the analysis and investigation stage of the process, prior to choosing a vendor. This paper aims to fill that gap.

Cloud computing is adopted based on the essential characteristics of the cloud deployment models and the cloud delivery models. The delivery models include Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). IaaS allows consumers the ability to deploy and maintain their software and operating systems using equipment housed at the provider's site. Consumers also have the ability control select networking elements. This eliminates the need for the consumer to purchase and maintain the equipment that runs their applications. In the PaaS model the hardware and operating systems are housed and maintained at the service provider's site while the consumer controls their deployed applications. This allows consumers the ability to develop and test new software applications without investing a large amount of capital for the hardware to run those applications. In the SaaS model, consumers access applications through a web environment. The service provider houses and manages the entire system and related software. Because the application is hosted, there is often nothing to install on the local machine thus alleviating the burden of maintaining the software the consumer would otherwise face. Web-based email is an example of SaaS. This paper will focus on the Software as a Service (SaaS) model which has been referred to as the core of cloud computing (Erdogmus, 2009).

Cloud computing is an emerging business and technology concept to support an on-demand delivery of computing, storage and applications over the Internet. A recent IDC report shows global revenue in cloud software market reaching \$22.9 billion and it will grow to \$67.3 billion in 2016 (Mahowald & Sullivan, 2012). This projection includes revenue generated by the shift from on-premise to on-demand providers as well as by the planning and architecture behind the shift. Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of applications and resources.

1.3 Structures of Cloud Computing

As high-speed internet has become more accessible to organizations, cloud models have helped organizations with limited IT resources take advantage of technology to improve business processes. Typically, an organization would rent licenses or access the software from an application service provider (ASP) that actually runs on servers or devices owned and maintained by the ASP. These large datacenter facilities with redundant layers of power and data security are often too expensive for many organizations to have under the on-premise model. Also, the ASP is generally responsible for maintaining and updating the software, and often includes some level of support for users in the monthly fee. This model has certain benefits for organizations that can attenuate some of the problems associated with maintaining purchased software in-house. For example, SaaS allows ASP to maintain their offerings consistently by automating testing, monitoring, maintenance and upgrades without sending out constant updates that need to be applied by end users. Also, SaaS allows smaller organizations with limited (or no) IT staff to benefit from the economies of scale and efficiencies implemented by the ASP. Equally important, SaaS allows organizations to pay to use the software they need, without making a huge investment in IT infrastructure for servers, software, etc., by "renting" access to what they need and paying monthly, quarterly or annually.

For enterprises, Cloud Computing can be adopted as one of the following services:

- Software as a service (SaaS): providing software subscription services
- Storage as a service: providing remote storage resource services
- Database as a service: providing remotely hosted database services
- Information as a service: providing remotely hosted information services
- Process as a service: providing business processes based on remote resources
- Application as a service: also known as SaaS
- Testing as a service: providing testing services for local or remote systems

- Platform as a service (PaaS): providing a complete platform to support application development, interface development, database development, storage, information and testing
- Infrastructure as a service (IaaS): providing a service to access computing resources remotely
- Security as a service: providing core security services remotely over the Internet
- Integration as a service: providing a complete integration stack service

According to Beaubouef (2011), there are three Cloud Models of ERP adoption:

- Software as a service: a subscription model for small customers who share hardware.
- Hosted ERP: a typical solution for large customers who have separate hardware and instances.
- Hybrid ERP: a combination solution that maintains on-premise software as well as integrated a degree of on-demand services.

Additionally, according to Gartner's report (2012), more and more companies are considering on-demand services in different applications (Figure 1.1).

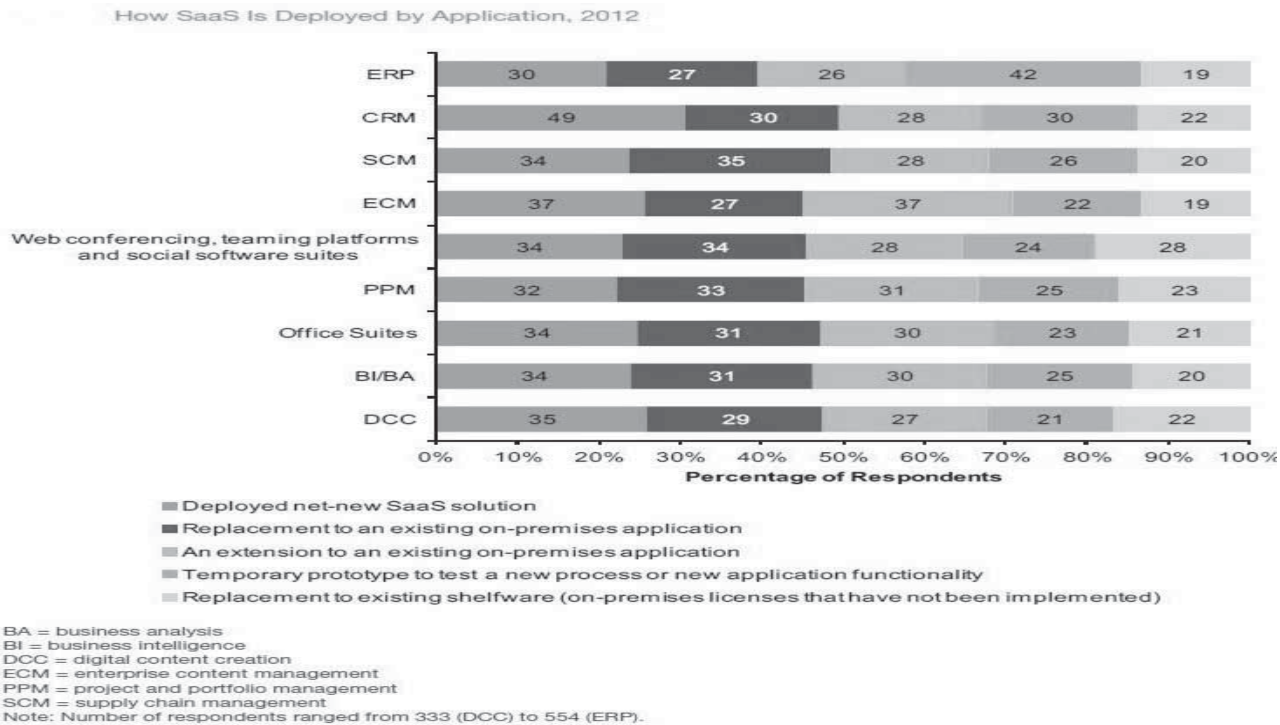


Figure 1.1. How SaaS Is Developed by Application

As the benefits become too remarkable to ignore and the shortcomings are worked out as the technology evolves, current literature focuses on how organizations, both public and private, can take advantage of what the cloud has to offer through case study analysis and predictions for the future of cloud computing. In early 2011, realizing that the federal IT budget could be reduced drastically by utilizing cloud computing, the U.S. Chief Information Officer, Vivek Kundra, released his "Federal Cloud Computing Strategy" report which discussed the newly implemented "Cloud First" policy. This policy requires federal agencies to "evaluate safe, secure cloud computing options before making any new investments (CIO Council, 2011). The strategy sped up the process of federal government agencies moving toward the cloud, thus provided researchers many case study opportunities (Wyld, 2009; Kundra, 2010; West, 2011). Many of the cloud case studies that have been previously published are reviews of what the organizations have implemented and how they have been benefited. The case study literature lacks in reviewing the steps taken in the analysis and investigation stage of the process, prior to choosing a vendor. This paper aims to fill that gap.

2. Literature Review

Cloud computing is an emerging business and technology concept to support an on-demand delivery of services over the Internet. To break down what cloud computing is, we need to start from the concept and structure of the cloud. We will start with the cloud architecture where cloud computing is divided into three types: private, public, and hybrid. A public cloud is described as resources dynamically provision on a fine-grained, self-service basis over the Internet, via web applications/web services, from an off-site third-party provider who shares resources. In less technical words, it's an outsides vendor that provides on-demand service that stores and process data for business; the business has less control but it can also have its benefits, will explain later. While, a private cloud is describe as data and process managed within the organization without the restrictions of network bandwidth, security exposure and legal requirements that using public cloud services across open, public networks might entail (Hoefer & Karagiannis, 2010). A private cloud is basically a public cloud, except it is made within the business resource and takes any risk of security breach or failure. A Hybrid cloud is the environment consisting of multiple internal or external providers; it's the combination of both private and public cloud, something that many companies are really considering moving towards.

The virtualization management is the technology that abstracts the coupling between the hardware and operating system (Hoefer & Karagiannis, 2010). By separating the logical resources away from their underlying physical resources in order to improve agility, flexibility, and reducing cost to enhance business value. There are different types of virtualization that suits a dynamic cloud infrastructure. One of them is server virtualization, the mapping of single physical resources to multiple logical representations or partitions (Hoefer & Karagiannis, 2010). It's what dynamically create, expands, shrunk or moved data as demand varies. This is why virtualization provides great advantage towards sharing, managing, and isolating the information in the cloud.

Now that we have a bright idea of what cloud computing is built upon, there is still a few questions to ask. What if there's a malfunction or failure with the server? Well, they are has been some outage in the past that it really causes problem to business and the vendors. For example, Microsoft Azure had an outage of 22 hours in March 13th-14th of 2008 this one of thousands of other failures that has happen before. Microsoft Azure did apologize and recognize the outage, but still believe and stands by its quality service. They know things like this happen affecting their customers but they provided a 33% credit to all customers for the month of the when the outage occur (Miller, 2012). A good solution for the servers is that they now back-up the information not only once but twice in different storage servers providing a more secure trust in with your data. There will a hot backup instance of the application which is automatically ready to take over without disruption also known as the failover. Security is also main concern because by accessing Cloud Computing, you're giving the vendor the power to handle your data or information. But securing cloud computing data is a contractual issue as well as technical (Hoefer & Karagiannis, 2010). In other words, you tend to trust the vendor to keep your data safe, by always checking on them and making sure that they are actually keeping the data safe, based on the contract signed. It also depends on which Cloud your business is running on, many say, a public cloud is less safe than a private cloud. But it all depends how well you take care of your cloud and must be really on top of your vendors to keep them secure.

Other issues that concern adapting the cloud computing are load balancing and scalable data storage. Load balancing is used to implement failover which continues the service after the failure of one or more components.

Where components are monitored manually, and when a server does not work properly, the load balancer is informed to stop sending traffic to the outage component. This feature is inherited from the grid-based computing for cloud based platform (Hoefer & Karagiannis, 2010). With the proper load balancer you will reduce the server cost and conserve some energy, becoming a greener enterprise.

The greater issue that we should be concerned about is the scalable data storage. Meaning that when data is thrown into the cloud, it should properly be organized otherwise there would be some trouble accessing the data. Vertical scalability is related to resources use, much like the old mainframe model (Hoefer & Karagiannis, 2010). If an application does not follow the vertical scalability will have trouble accessing the data causing to increase the cost because of the increase of demand to compute the data.

It is important for a company to understand the benefits and risks before they decide if cloud computing fits their company needs. As mentioned before the main benefits are sustainability, scalability, infrastructure, flexibility, lower cost, and availability. According to Gheorghe and Lupasc (Gheorghe & Lupasc, 2012), cloud ERP solution will save about 30% project time after analyzing the critical path of both classic and cloud ERP solutions. Table 2.1 summarizes the benefits derived from cloud computing ERP systems.

Table 2.1. Benefits of On-demand ERP solutions

Benefits of on-demand ERP	Reference
Short implementation cycles	(Gheorghe & Lupasc, 2012); (Karabek, et al., 2011);
Low entry costs	(Marston, et al., 2011); (Karabek, et al., 2011); (Bhardwaj, et al., 2010)
Reduced demand for own IT resources	(Karabek, et al., 2011); (Bhardwaj, et al., 2010)
Elasticity/Flexibility of services	(Marston, et al., 2011); (Karabek, et al., 2011); (Rittinghouse & Ransome, 2009)
Scalability	(Marston, et al., 2011); (Gartner, 2012)
Focus on core business	(Karabek, et al., 2011)

Governments, who are planning to implement e-government strategies, have begun to adopt cloud computing technologies in their IT systems. According to the report from Federal Chief Information Officer (Kundra, 2010), more than 24 government departments and state government agencies in United States adopted cloud computing solutions in various IT services. Paquette (2010) summarize four sections of current adoption of cloud computing technology by the United States federal government as following:

1. Early Use: the early adoption of cloud computing starts from informal use of personal applications, such as instant message and portal services by government employees, agencies, departments, and contractors. At this stage, government is trying to define the cloud computing technologies in public sector environment and identify the essential characteristics of the cloud computing applications (Wyld, 2009).
2. Formal, strategic direction: IT directors at different government levels realize the benefits of cloud computing and try to explore cloud computing as a strategic component in government transformation. President Obama (2009) emphasized the adoption of cloud computing by the Federal IT for a more transparent government to the public, advanced technology of government IT, and a better innovation environment.
3. Current applications for information sharing: today, cloud computing applications adopted by public sectors are more focused on information sharing and communications, such as using YouTube hosting services and mobile cloud services.
4. Applications and information processing: recently, the federal government is attempting to utilize cloud computing applications in data/information processing rather than just data storage or data sharing. For example, The Air Force is a 3-D virtual recruiting and training application on cloud (O'Hara, 2009).

Government acceptance of cloud computing has been growing fast. Wyld (2009) lists 10 benefits of cloud computing on e-government, such as rapid scalability, low maintenance/upgrades cost, improved resource utilization, improved economies of scale, improved collaboration capabilities, usage-based pricing, reduced IT infrastructure needs,, computational power, green-friendly factor, and improved disaster recovery capabilities. Pokharel and Park (2009) also identify "Expertise" as one of the benefits of cloud computing on e-government. At the same time when we talk about the advantages of cloud computing on public sectors, there are researchers who concern the risks of the technology. Outage/Accessibility is rated one of the highest factors (Kim, 2009). It is obvious that system outage is closely correlated with on-deman adoptions. After the outage issue of Microsoft Azure, CIOs realized the importance of system accessiblity for on-demand solutions. Security is another emphasis concerned by several researchers in their studies of cloud computing (Kim, 2009; Paquette, et al., 2010; Saeed, et al., 2011).

Most of the current studies about cloud computing applications in e-government have been focused on benefits, concerns and challenges of cloud computing adopted by government. There are several studies emphasizing on architectures of cloud computing in e-government (Zissis & Lekkas, 2011) and policy issues of cloud computing applications (Shin, 2013). However, relatively little research attention has been given to the early stage of cloud computing adoption by public sectors, such as assessment and planning processes. This study seeks to provide a better understanding of system analysis and selection during the early age of cloud computing adoption for public sectors.

3. Research Methodology

Carr et al (1996) propose a useful system adoption model for a successful Information System. They advocate a four-step process which is designed to assess the present position, decide on an appropriate change process, establish a sound theoretical framework for the change and ensure that aims are shared and personnel are involved and committed. This is achieved through the stages of:

- Assessment: justification, objectives and broad characteristics.
- Planning: the entire change process is laid down.
- Action: commitment, dissemination, training, change.
- Renewal: monitoring, feedback and evaluation.

In this study, we believe the first phase is extremely useful to explain and guide what we need to complete in system analysis and selection of cloud computing for public sectors.

In *Assessment phase***,** for the manager, the change process begins when questions are asked about what the originators of the proposal actually want to do. At this stage, no one is looking for answers to all of the questions, and the process should not begin with consideration of the change itself. It begins with a general review of the organization, and it is relevant to organizational health, which is itself to do with motivation. Starting with examining motives, managers should identify both positive and negative reasons for introducing change by asking all kinds of questions related to the change, such as what are the desired outcomes? What are the problems? How does the project fit with the strategy of the organization? What is the likely effect on the organization? What is to be the role of the manager? The organization has to investigate the details of the proposed change:

- Identify what changes are required
- Analyze changes
- Identify resources required

In the case study analysis in section V, we will adopt this model to discuss and identify the issues and challenges during the cloud computing adoption processes in the two case studies in section IV.

A Case study is a qualitative and descriptive research method to explore issues and factors from an individual or small number of participants and draw conclusions only about those participants or group in that specific context (Yin, 2014). This method is an ideal methodology when a new concept or system needs holistic and in-depth investigation (Feagin, et al., 1991). It has been proven a useful research tool in investigating system adoption studies (Kerimoglu, et al., 2008; Mackrell, et al., 2009). Therefore, in this study, we use the case study method to identify important factors during cloud computing adoption processes in public sectors from the following two cases.

3.1 The City and County of San Francisco

3.1.1 Background

The City and County of San Francisco, CA was established by charter in 1850 and is the only legal subdivision in the State of California with the governmental powers of both a city and a county. A Board of Supervisors exercises the City's legislative power and the Mayor exercises the executive power (Rosenfield, 2011). In 2008, the Citywide IT Plan: Current State Assessment for the City and County of San Francisco reported that the City consisted of more than 50 departments in addition to other organizations. It also reported that the city had seen a movement away from a central IT shop with a number of departments maintaining their own IT staff. In addition to the complex technology organizational structure, the city had a complex technology governance structure as well with a Committee on Information Technology (COIT) made up of 11 members and 4 subcommittees made up of 37 members, charged with providing the necessary technology policy, procedures, and oversight to ensure that the City meets its goals and objectives.

3.1.2 Feasibility/Strategic Email Study

The 2008 COIT Email Policy stated that the Department of Technology and Information Services (DT) was in charge of managing all email systems with the exception of the City Attorney's Office. Six other departments were noted as maintaining their own e-mail systems. Following the adoption of the email policy, the city began a Citywide Strategic E-Mail Study with three goals: (1) confirm the City policy of single e-mail standard, (2) determine which City entity should manage the central system and (3) select a single standard. At the time the city email setup consisted of a mix of Microsoft Exchange and IBM Lotus Notes email systems. The citywide email options outlined in the study consisted of:

- maintain the current hybrid approach
- migrate all departments to Lotus Notes
- migrate all departments to Microsoft Exchange
- migrate all departments to a hosted solution

The hosted solution was listed as a future option. After reviewing the study, the COIT Director stressed that the fundamental issue at hand was whether there should be a single citywide email system or whether departments should be allowed to use independent email systems. The City CIO asked the Architecture and Standards Sub-Committee (ASSC) to make recommendations for the city's e-mail platform based on the four options listed above.

3.1.3 Investigation

The Citywide Strategic E-Mail Study provided the City with four email options to choose from, one of them being to continue operating as they had in the past. In order to make a recommendation to the COIT the ASSC had to develop system functional requirements and a performance metric. To accomplish this, the Committee sent out an e-mail requirements survey to all the city departments. Based on the responses of 16 departments, representing approximately 15,000 users, the ASSC developed a list of tentative requirements on each of the survey items. They then compared the list of recommendations to the offerings of the potential service providers, IBM and Microsoft. The combined results indicated that there were not many functional differences between the competing vendors. The ASSC also developed an Email Evaluation Metrics based on the survey responses. The Metrics showed that, on average, the ability of the selected email systems to meet the functional requirements was most important, followed by the ability to meet the service requirements, annual costs and finally start-up costs. The ASSC then proceeded to obtain cost estimates for the following options: (1) Lotus Notes maintained by the City; (2) Lotus Notes hosted by IBM; (3) Microsoft Exchange maintained by the City; (4) exchange hosted by Microsoft and (5) Google's enterprise email system hosted by Google. Initial cost analysis indicated that the hybrid environment was the most expensive option. Cost analysis for the Google option was not presented at that time.

3.1.4 Recommendations

Based on their overall analysis, the ASSC provided the following recommendation to the COIT:

- All City Departments should be on one e-mail platform
- The platform should be managed by a central entity

At the following meeting on February 25, 2009, after watching presentations delivered by IBM and Microsoft and detailed discussions of both systems, the ASSC unanimously voted in favor of recommending a hosted Microsoft Exchange system. The sub-committee members’ average ranking for each option is located below in Table 3.1.

Table 3.1. Committee Rankings

Committee Rankings (ASSC, 2009e)

	Lotus Notes		Exchange		Current
	On-Prem	Hosted	On Prem	Hosted	Hybrid
Arch SC Member Group average	3.5	3.1	2.8	1	

Following that recommendation, the COIT determined that "within the next 2 years, general and non-general fund departments (will) move to a hosted Exchange solution" (COIT, 2009). In a February 2010 meeting the justification for choosing Microsoft was explained by the Acting City CIO, Jon Walton in response to questions from the public.

Once the decision had been made to go with a hosted Exchange solution, the focus switched to a needs analysis in an effort to decide who would host the system. In November, 2009, the ASSC decided to request a quote for consulting services that would provide the email technical business requirements necessary to construct the Request for Proposal (RFP) for the new hosted exchange systems. The basic scope of the project was defined to include email only for approximately 22,000 users with 25GB of space per box and needed to include training and current mailbox conversion. They had specifically decided against any advanced collaboration feature as in the LA Google Deal.

In the requirements and gathering stage, the COIT worked closely with the staff that managed the behind-the-scenes technical requirements on email in order to get their input. They wanted to make sure that the email administrator understood what systems they were getting or giving up. The COIT also decided to create a pilot program that the Police Department would be a part of in order to identify security issues. The goal was to have the pilot program for the conversion up and running by the end of the fiscal year. The Police Department was later pulled out of the pilot and replaced by the Department of Technology.

3.1.5 Hosting

In a March 2010 meeting, a proposal was made to have the Airport host the citywide email system creating a Center of Excellence. The following arguments were made in favor of this proposal:

- Quicker implementation in that they could avoid the RFP Process and start immediately
- Reduction in costs compared to outside vendors
- No consulting or professional services required
- High probability of success due to the fact that the Airport had previously migrated from Lotus to Exchange
- Existing investments in software and hardware protected

The following month, in analyzing the Airport Hosted option vs. the vendor Hosting option, the committee discussed the costs associated with each option as well as the capital involved. It was mentioned that many of their current systems had no recovery or refresh plans and that the aged equipment they were running had more downtime. The COIT Director also pointed out that one of the reasons a hosted model was considered was to get out of the business of owning equipment and licenses.

When comparing the two options side by side, the committee summarized that the Airport option relied more on in-house staff to implement the projects and had a large capital investment that required ongoing equipment replacement charges, while the vendor option relied more on professional services and had had a relatively constant cost over time. In addition, a side by side comparison of per mailbox cost indicated that the vendor option would be more cost effective over time. Industry standard was also considered noting that more governments are going with hosted vendor options.

Based on their analysis, the COIT decided to begin the RFP process in the search for a vendor hosted solution. The Airport was still able to compete in the RFP process.

After months of security testing and working with Microsoft to amend their standard agreement, the City and County of San Francisco announced on May 18, 2011 that they had signed a contract with Microsoft to host their Exchange solution.

3.2 The City of Los Angeles

3.2.1 Background

The City of Los Angeles, CA was established and incorporated in 1850 after California became a U.S. state. Since then, Los Angeles has become the second most populated city in the U.S. with a population of nearly 3.8 million according to the 2010 United States Census. The City of Los Angeles has 44 different departments, ranging from the police and fire departments to the department of transportation. The majority of the department's information technology services are maintained by the Information Technology Agency (ITA). The role of the ITA is to manage enterprise applications, infrastructure, telecommunications, and the e-mail system that is used across all departments.

3.2.2 Feasibility

The Information Technology Agency was in control of approximately 34,000 e-mail accounts spread across the 44 departments of the City of Los Angeles. The e-mail system that was being used was the GroupWise e-mail system. There was a general dissatisfaction among the majority of employees due to two main reasons: e-mail did not universally work on all devices and the e-mail inbox storage limit was too small. Employees were constantly peaking on their storage limit. A major factor contributing to the adoption of the new e-mail system for the City was their $400 million deficit. Therefore, they decided to adopt a new system that met the following three goals: *Automation, Efficient,* and *Cost Effective*.

The ITA determined that the new system to be put in place would immediately replace the legacy system of between 17,000 and 30,000 user e-mail accounts. The only exclusion would be the Los Angeles Police Department (LAPD) for security reasons. The entire solution would be implemented across all departments, including the LAPD, several months after the initial implementation. The implementation would require the migration of all historical data (archive and backup) from the existing e-mail system. In 2009, the ITA worked with the Computer Sciences Corporation (CSC), a Nevada based corporation, to determine the best vendor that would fit the City of Los Angeles' needs.

3.2.3 Investigation

The City of Los Angeles used the GroupWise e-mail system across all departments. According to the Information Technology Agency, the system did not meet the needs of the majority of departments and employees. The system was not portable and universal between devices, the storage space was too small, there was minimal archiving and collaboration tools for employees to share documents, and no disaster recovery plan. The general dissatisfaction with the GroupWise e-mail system was the reason why the Information Technology Agency worked with the Computer Sciences Corporation to determine a new vendor for their City wide e-mail necessities. Because E-mail was vital to all departments of the City the ITA and CSC determined the following items would be taken into consideration when choosing a vendor: (1) Minimal training required; (2) Intuitiveness of the system; and (3) Ease-of-use.

As far as the technical aspects were concerned, the ITA and CSC determined that the requirements for the new platform would include the following: (1) Mobility, portability, and universal to all devices; (2) 25GB storage space; (3) Synchronization; (4) Delegation; (5) Collaboration; (6) Remote access; (7) Disaster recovery; (8) Archiving.

The ITA and CSC compared and contrasted three different platforms for their project. They looked at their current system's vendor (Novell GroupWise), Google, and Microsoft.

3.2.4 Analysis and Recommendations

The ITA and CSC decided on the platform Google Apps due to its perfect fit with the scope of the project and the cost effectiveness of the application compared to its competitors. This was one of the main concerns since the City is already in a $400 million deficit. Google Apps could provide all of the implementation and technical requirements while also fulfilling the scope requirement of 34,000 accounts. Google Apps provides an e-mail client with 25GB storage, file sharing services with shared documents and editing, video chat, mobility, archiving and disaster recovery/backup systems, and is portable to all devices.

Google offered different pricing structures for the Google Apps implementation. There were two options in each structure: the number of accounts (17,000 or 30,000) and whether email-archive migration was or was not included in the implementation.

The ITA estimated that the overall cost savings for the five year contract would be $5.5 million and the project will have an ROI of upwards of $20 million. Between 65% and 80% of all of the City of Los Angeles' employees would have their technical productivity needs met solely by Google Apps.

4. Discussion

Following by Carr et al.'s model (1996), in the Assessment phase, both cities went through all the processes. They identified changes (new system to solve current dilemma), analyzed those changes, and identified the resources they need for the changes, such as budget for each alternative solutions. Table 4.1 summarizes the processes for both cities.

Table 4.1. Summary of Processes in Assessment Phase

Process	The City of San Francisco	The City of Los Angeles
Identify changes	Need a new email system based on the three goals of Citywide strategy: (1) confirm the City policy of single e-mail standard, (2) determine which City entity should manage the central system and (3) select a single standard	Need a new email system to meet three goals: *Automation, Efficient,* and *Cost Effective*
Analyze changes	Analyze the four options: • maintain the current hybrid approach • migrate all departments to Lotus Notes • migrate all departments to Microsoft Exchange • migrate all departments to a hosted solution Then, analyze the two options of hosting: Airport Vs. vendor hosting	Analyze the three options of platforms: current system's vendor (Novell GroupWise), Google, and Microsoft
Identify resources	Estimate the budget	Estimate the budget

Additionally, we found several interesting issues in the two cases. First, when we compare the two cases, though both of the adoptions are successful, we found a critical difference between them. With 50 departments and 22,000 users, city of San Francisco spent about 3 years before their final decision of vendor selection while city of Los Angeles, with 44 departments and 34,000 users, only spend about a year in decision making. We compared the steps each city completed and found that San Francisco had three more steps: 1. Vote for hosting system between IBM and Microsoft; 2. Develop a pilot program for Police Department; 3. Propose Airport hosting systems competing with vendor hosting systems. Besides the three extra steps, San Francisco also spent longer time on the following steps: 1. analyzing system functional requirements and cost metrics; 2. requesting a quote for consulting services; 3. analyzing the options between Airport and vendor hosting.

Second, many of the differences between the two cases can be linked to cultural differences between the two governmental entities. San Francisco did not have a centralized IT shop at the time of the study; conversely, the city of Los Angeles has an Information Technology Agency to centralize the IT solutions for each department at an enterprise level. Though not enough information is available to assert this claim, the researchers believe that this distinction affected the selection of vendors within the two entities. A collection of decentralized IT sub-departments within San Francisco's many departments would be difficult to coordinate for the successful completion of this system. This could be one of the reasons to explain why San Francisco spent about 3 years to complete the Assessment phase before they actually signed the contract with Microsoft.

Third, when selecting solutions, the City of Los Angeles focused more on cost as a priority. Therefore, they quickly worked with CSC and finalized the best solutions they needed. However, in San Francisco case, the city didn't develop a clear vision at the beginning, so they struggled with the initial four options first, then they came out another options, airport vs. vendor hosting, when they finally decided to go with a single host system. This considerably slowed their processes. This could be another reason to explain the lengthy adoption of San Francisco.

5. Conclusion

In past decade, cloud computing has gradually change software infrastructure, services and business models running on the Internet by providing much lower initial cost and extremely supportive data storage and processing functions. This technology attracts all types of organizations, including public sectors, to adoption on-demand solutions in daily business processes. Cloud computing is at an all-time high demand and has recently been helping several companies gain a competitive advantage in the global market by managing their data and processes securely. By implementing cloud computing it gives the company a benefit of storing all of their information on a cloud with easy accessibility and flexibility, as well as scalability. Some important factors that benefit a company using cloud computing as mention before are costs, uptime, privacy, internet reliability, and manageability. This study analyzed two cases of public sectors, who adopted cloud computing email systems successfully. First, from the project management perspective, all the steps that the two cities completed during cloud computing adoptions are corresponding to the processes proposed by Carr et al.'s model (1996). Therefore, we believe that the Assessment phase is very useful to guide public sectors when they are adopting any cloud computing systems. Second, after comparing the two cases, we found out several key factors that may impact the speed of the system adoption during the Assessment phase, such as centralized IT control, and clear vision. Third, simplifying the processes by reducing the steps or shorten each step in Assessment phase will save much time during the adoption project.

This study has implications for both the research and practice of cloud computing adoption in public sectors. Researchers could use these results to expand valid indicators of successful cloud computing adoptions in their empirical studies. Professionals, such as IT managers, project managers in public sectors, could use the results in their decision making during the cloud computing adoption projects. Vendors and consultants of cloud computing services could offer more effective services for their clients in the system adoption projects.

In the future, we will study more cases and summarize more key factors before we can generalize any theme to enrich the research foundation in this area.

Appendix

A. Appendix A

CCSF Email High Level Requirements				
Item	**Requirement?**	**Microsof t** (ASSC, 2009g)	***IBM*** (ASSC, 2009h)	**Comments (CCSF)**
1.0 Integrated IM including Searchable Chat History	Yes	Yes	*Yes*	
2.0 Built-in Archive Capability	Yes	Yes	*Yes*	
3.0 Built-in Single Attachment Storage	Yes	Yes	*Yes*	
4.0 Group/Team Calendar Overlay	Yes	Yes	*Yes*	
5.0 Native Client Support for Windows, Linux, and Mac	Yes	Yes	*Yes*	
6.0 Collaboration Integration	Yes	Yes	*Yes*	
7.0 Calendar Federation (federate multiple types (Google/iCal Feed)	No	Yes	*Yes*	Nice to have.
8.0 Built-in RSS Feed Reader	No	Yes	*Yes*	Nice to have.
9.0 Single Sign On Capability	Yes	Yes	*Yes*	
11.0 Anti-spam/anti-virus	Yes	Yes	*Yes*	Essential for hosted solution.
12.0 Encryption (Internal and External)	Yes	Yes	*Yes*	
13.0 Integration with Active Directory	Yes	Yes	*Yes*	
14.0 Smartphone Integration	Yes	Yes	*Yes*	
15.0 E-Discovery	Yes	Yes	*Yes*	Essential for hosted solution.
16.0 Multiple	Yes	Yes	*Yes*	

Addresses for Person in TypeAhead				
17.0 Application Integration Capability	Yes	Yes	*Yes*	
18.0 Browser Support for Multiple Browsers	Yes	Yes	*Yes*	Must support multiple browsers, including Internet Explorer, Firefox, Safari
10.0 Spell Check	Yes	Yes	*Yes*	
19.0 Platform Support: Windows, Mac, Linux, Blackberry, iPhone.	Yes	Yes	*Yes*	
20.0 Standard Mail Box Size	2.0 Gigabytes	Yes, 5GB	*Unlimited (onsite) 1GB hosted*	
21.0 Transport Layer Security for Messaging	Yes	Yes	*Yes*	
22.0 Security & Penetration Testing and Audits	Yes, for hosted solution	Yes	*Yes (hosted)*	
23.0 Intrusion Monitoring, Detection & Response	Yes, for hosted solution	Yes	*Yes (hosted)*	
24.0 Disaster Recovery & Restoral	12 hours		*DR set in SLA*	
25.0 Real time Failover Capability	Yes		*Yes*	
26.0 Service Uptime Metrics	99.99% uptime		*No. 99.9% is maximum for hosted. CCSF On-site TBD*	Cost for both 99.99 and 99.999%

B. Appendix B

Email Evaluation Metrics (ASSC, 2009i)				
1.0 Meets Functional Requirements	**2.0 Meets Service Requirements**	**3.0 Start Up Costs**	**4.0 Annual Cost**	**5.0 Your Department**
40%	30%	20%	10%	DHR
50%	20%	20%	10%	Public Health
30%	30%	10%	30%	Planning Department
20%	10%	30%	40%	SFO
40%	30%	10%	20%	DEM
30%	30%	20%	20%	Port
30%	30%	10%	30%	DT
34%	**26%**	**17%**	**23%**	

C. Appendix C

Option Cost Comparison (ASSC, 2009c)

	Lotus Notes		Exchange		
	On-Prem	Hosted	On Prem	Hosted	Current (Hybrid)
First Year	$2,100,000	$2,582,000	$3,885,000	$3,435,800	
Ongoing	$1,650,000	$2,357,000	$1,390,000	$2,260,800	
Five Year	$8,700,000	$12,010,000	$9,445,000	$12,479,000	$15,000,000

D. Appendix D

Airport vs. Vendor Option
Summary of Findings (COIT, 2010c)

	Airport	Vendor	Difference
Actual First Year Costs	$6,424,845	$2,607,840	$3,817,005
Project Expenditures	*$4,478,845*	*0*	*$4,478,845*
Intital Captial Investment	*$2,144,000*	*0*	*$2,144,000*
Average Cost Over 5 Years	**$3,581,582**	**$2,640,442**	**$941,140**

Total Cost, Year 1	$3,581,582	2,607,840	$973,742
General Fund Impact	**$1,738,870**	**$1,300,113**	**$438,757**

Initial Cost per Mailbox	$163	$119	$44
Beyond year 5	$126	$112	$20

E. Appendix E

Functional Requirements for Hosted Exchange Email for CCSP (COIT)

SaaS E-Mail and Collaboration Solution (SECS) Project RFP Proposer Requirements

REQUIRED FEATURES

E-mail

- Basic e-mail functionality, including but not limited to send, receive,
- format, and attachment

- Ability to create user defined e-mail groups or personal folders based
- on search criteria

- Ability to define rules for e-mail handling
- Ability to add both personal signatures and notes
- Ability to push contact lists and web links to mobile devices
- Ability to retain e-mail (List per-user limit, if any)
- Ability to copy, move, and store information to desktop or local storge
- Ability to print stored information locally
- Ability to scan or fax from multifunction devices to e-mail
- Work with City staff to establish remote printing to a City facility
- Ability to send, assign and delegate tasks
- Ability to utilize e-mail system remotely
- Ability to delegate e-mail functionality to another staff member (i.e., proxy assignments, including: mail/phone, appointments, reminder notes, tasks, etc.)
- Ability to define proxy access limitations (e.g., Read/Write; Subscribe to Alarms and Appointments, Modify Options, Rules, and Folders
- Retract and/or retrieve within City e-mail system

CONTACT MANAGEMENT

- Basic contact management functionality, including but not limited to last name, first name, middle initial, department, title, business address, contact log, notes, etc.
- Ability to synchronize contact information with desktop applications
- Ability to synchronize contact information with industry standard mobile devices
- Ability to share contact lists

CALENDARING

- Basic calendaring functionality, including but not limited to appointment, event, and sharing
- Ability to view multiple calendars at same time (both personal and global)
- Ability to schedule resources, including but not limited to facilities, conference rooms, and equipment management, set "view only" or "edit" rights, etc.) to another staff member
- Ability to print calendars locally in standard formats (such as daily, weekly, monthly, Franklin format, etc.)
- Ability to view and schedule from "free-busy" information
- Ability to view or hide appointment details

ARCHIVE AND BACKUP

- Ability to store and retrieve all live e-mail data for a minimum of two years (2) with capability for up to ten (10) years. 90 days available to user and 90 additional days available to System Administrators before data is automatically processed for long-term archive.
- Ability to archive data based on content, sender, recipient, and/or other metadata with different archival periods per City policy or legal

requirements

- Ability to retrieve or e-Discover archived data based on content, sender, recipient, and/or other metadata with different archival periods
- Ability to view, and perform all normal e-mail functions on archive by an e-mail administrator without having to restore
- Ability to restore archived e-mail data to "live" status

<u>SOLUTION ADMINISTRATION</u>

Ability, from the Administrative console, to:

- Fully manage all City accounts within the City network, including but not limited to addition, deletion, manipulation and suspension
- Fully manage SaaS identity and user accounts
- Control SPAM or provide anti-spam
- Control virus or provide anti-virus (including spyware)
- Apply content filter
- Ability to apply policies in managing solution
- Review restricted e-mail
- View all calendars and appointments
- Print historical, statistical and usage reports locally
  Prioritize e-mail accounts
- Manage attachment size
- Setup mail routing
- Manage multiple separate Global Address Lists (GALs)
- Use “Whitelist”, “Blacklist”, and aliases
- Ability to manage optional solutions as cited in Section II.B below

- Ability to use all domain names utilized within City as e-mail extensions Administration
- Ability to synchronize e-mail identities with identities that are managed in our internal authentication directory Administration
- Ability to control Blackberry, Treo, iPhone and other such mobile/smart devices, including the ability to synchronize calendar, contacts and e-mail (e.g., Blackberry Enterprise Server, etc.)
- Ability to integrate with internal applications using e-mail, specifically using SMTP, IMAP, SOAP, POP3, etc.
- Ability to manage DNS
- Ability to migrate historical or user archives from the current proprietary format to proposed solution after implementation
- Extent to which administration can be implemented in a distributed manner to different departments.

OPTIONAL SOLUTIONS REQUIREMENTS

Instant Messaging

- Ability to IM Internally;
- Ability to IM Externally;
- Ability to Track IM.

E-DISCOVERY

Ability to search based on the following criteria:
eDiscovery ,Content

- Sender and/or recipient
- Date range
- Meta-data.
- Ability to store search results with any metadata
- Ability to add and delete from search results to create an e-Discovery set.

COLLABORATION

- Ability to share data and files stored within the solution
- Ability to have multiple staff members work on common files at the same time from different or separate City work locations
- Ability to collaborate with staff members that are telecommuting or otherwise away from a City facility
- Availability of a Wiki type solution for collaboration that allows changes to be tracked by user
- Ability to maintain version control (i.e., who, when, what)

F. Appendix F

Microsoft BPOS bid, July 26, 2010

(CCSF Director of Technology Procurement, 2010)

	MICROSOFT ENTERPRISE E-MAIL SERVICES for the City and County of San Francisco Department of Technology						
	The purpose of this bid request is to obtain a vendor managed and hosted scalable Microsoft Exchange email service that can support the City's email volumes, availability, security, and functional specifications. Primary core services requested from this bid request include email, calendar and scheduling (free/busy) and contacts.						
SKU			**Quantity**	**Three Years**	**Five Years**	**Six Years**	**Comments**
74G-00002	***ExchngHstdArchv ALNG SubsVL MVL PerUsr***	***Exchange Hosted Archiving***	**23,201**				
TRA-00046	***ExchgOnlnStd ShrdSvr ALNG SubsVL MVL PerUsr for SA***	***Exchange On-Line service***	**23,201**				
UT4-00011	***SubsVL MVL PerUsr BlckbrryLic***	***Blackberry support per user***	**653**				
381-01587	***ExchgStdCAL ALNG LicSAPk MVL DvcCAL***	***Exchange CAL***	**15,776**				
	Legal Archiving and e-Discovery		1				
	Collaboration Tools		1				
	Instant Messaging		1				
	Presence		1				
	Web Conferencing		1				
July renewal date/co-term for all year options							
	Please provide payment options for e-mail rollout:						
	Monthly						
	Quarterly						
	Annual						

	Please describe the differences between the 3, 5 & 6 year options: (limit description to one page per year)						
Implementation Plan							
	What services does your company offer/provide as part of this bid requirement: (limit description to one page)						
	What are the additional fees, if any for the services above:						

Acknowledgements

Author would like to acknowledge the valuable contributions of Sandra D. Gaw, Nicholas Bender and Daniel T. Levy. Their efforts and expertise are the backbone of this chapter.

References

Anderson, E., Britz, B., & and Bell, W. (2012). *User Survey Analysis: Using Cloud Services for Mission-Critical Applications*. Retrieved Feb. 3, 2013, from *http://www.gartner.com/id=2182216*

Armbrust, M. F. (2009). *Above the Clouds: A Berkeley View of Cloud Computing.* Berkeley: University of California at Berkeley.

Beaubouef, B. (2011). *Cloud ERP – New Dog, Same Fleas*. Retrieved Feb. 5, 2013, from *http://gbeaubouef.wordpress.com/2011/05/22/cloud-erp/*

Bhardwaj, S., Jain, L., & Jain, S. (2010). Cloud Computing: A Study of Infrastructure as a Service (IaaS). *International Journal of Engineering and Information Technology*, 2(1), 60-63.

Biddick, M. (2011). *Federal Government Cloud Computing Survey.* InformationWeek.

Brown, E. (2011, October 25). *Final Version of NIST Cloud Computing Definition Published.* Retrieved from NIST: http://www.nist.gov/itl/csd/cloud-102511.cfm

Buyya, R. Y. (2008). Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities. *2008 10th IEEE International Conference on High Performance Computing and Communications*, (pp. 5-13).

Carr, D., Hard, K. and Trahant, K. (1996). *Managing the Change Process: A Field Book for Change Agents, Consultant, Team Leaders, and Reengineering Managers.* New York: McGraw-Hill.

COIT. (2009). *Meeting Minutes 3/26/09.* Retrieved from City & County of San Francisco Committee on Information Technology: http://www.sfcoit.org/index.aspx?page=105

COIT. (2011). *About Us: Committee on Information Technology*. Retrieved from City & County of San Francisco Committee on Information Technology: http://sfcoit.org/index.aspx?page=603

Erdogmus, H. (2009). Cloud Computing: Does Nirvana Hide behind the Nebula?, *IEEE Software,* 26, 4-6.

Feagin, J., Orum, A. and Sjoberg, G. (1991). *A Case for Case Study*. Chapel Hill, NC: University of North Carolina Press.

Gartner, Inc. (2012). Global IT Spending Forecast. Available at: *http://www.gartner.com/ newsroom/id/2055215*

Geelan, J. (2009, January 24). *Twenty-One Experts Define Cloud Computing.* Retrieved from Cloud Computing Journal: http://cloudcomputing.sys-con.com/node/612375?page=0,0

Gheorghe, G., & Lupasc, I. (2012). Crises Solutions for SMEs Case Study: ERP Cloud Vs. Classic Solution. *International Conference "Risk in Contemporary Economy" SIIIth Edition, (pp. 149-156)*. Galati, Romania.

Gruman, G., & Knorr, E. (2008, 04 07). *What Cloud Computing Really Means*. Retrieved from InfoWorld: http://www.infoworld.com/print/34031

Hoefer, C. & Karagiannis, G. (2010). Taxonomy of cloud computing services. *GLOBECOM Workshops (GC Wkshps)*, IEEE.

Karabek, M., Kleinert, J., & Pohl, A. (2011). Cloud Services for SMEs - Evolution or Revolution? *Business + Innovation*, 26-33.

Kerimoglu, O., Basoglu, N. and Daim, T. (2008). Organizational adoption of information technologies: Case of enterprise resource planning systems, *The Journal of High Technology Management Research,* 19, 21-35.

Kim, W. (2009). Cloud Computing: Today and Tomorrow, *Journal of Object Technology,* 8, 66-72.

Kundra, V. (2010). *State of Public Sector Cloud Computing.* Washington D.C.: CIO Council.

Mackrell, D., Kerr, D. and Hellens, L. (2009). A qualitative case study of the adoption and use of an agricultural decision support system in the Australian cotton industry: The socio-technical view, *Decision Support Systems,* 47, 143-153.

Mahowald, R. & Sullivan, C. (2012). Worldwide Saas and Cloud Software 2012-2016 Forecast and 2011 Vendor Shares. International Data Corporation (IDC). Available: *http://www.idc.com/getdoc.jsp?containerId=236184#.USTXoh3kseB*

Martson, S. L. (2011). Cloud computing - The business perspective. *Decision Support Systems, 51*, 176-189.

Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing - The business perspective. *Decision Support Systems*, 51(1), 176-189.

Miller, R. (2012). Data Center Knowledge: Data Centres Industry News and Analysis about Datacenters. Microsoft Offers Credit for Azure Cloud Outage » Data Center Knowledge. Available at: http://www.datacenterknowledge.com/archives/2012/03/12/microsoft-offers-credit-for-azure-cloud-outage/

Obama, B. (2009). Revocation of certain executive orders concerning regulatory planning and review. Available at: http://www.whitehouse.gov/the_press_of fice/Revocation-Of-Certain-Executive-Orders-Concerning-Regulatory-Planning-And-review/

O'Hara, C. (2009). Virtual learning gets second wind from Second Life. Available at: http://fcw.com/articles/2009/05/04/feature-virtual-learning.aspx_

Paquette, S. J. (2010). Identifying the security risks associated with governmental use of cloud computing. *Governement Information Quarterly, 27*(3), 245-53.

Pokharel, M. & Park, J. (2009). Cloud computing: future solution for e-governance. *ICEGOV*, (pp. 409-410). Bogota, Colombia.

Rittinghouse, J. & Ransome, J. (2009). *Cloud Computing: Implementation, Management, and Security*. CRC Press.

Saeed, I., Juell-Skielse, G. and Uppstrom, E. (2011). Cloud Enterprise Resource Planning Adotions: Motives & Barriers. *International Conference on Research and Practical Issues of Enterprise Information Systems*.

Schwartz, K. (2011). *Like it or not, cloud computing is here to stay.* Retrieved from Federal Computer Week: http://fcw.com/microsites/2011/cloud-computing-download/cloud-computing-application-development.aspx

Shin, D. (2013). User centric cloud service model in public sectors: policy implications of cloud services, *Government Information Quarterly*, 30, 194-203.

Vaquero, L. M.-M. (2009). A break in the clouds: towards a cloud definition. *ACM SIGCOMM Computer Communication Review, 39*(1), 50-55.

Wang, L. T. (2008). Scientific Cloud Computing: Early Definition and Experience. *High Performance Computing and Communications, 2008. HPCC '08. 10th IEEE International Conference* (pp. 825-830). Dalian: HPCC.

West, D. (2010). *Saving Money Through Cloud Computing.* Washington D.C.: The Brookings Institution.

Wyld, D. C. (2009). *Moving to the Cloud: An Introduction to Cloud Computing in Government.* IBM Center for The Business of Government.

Yin, R. (2014). *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage Publications.

Zissis, D. & Lekkas, D. (2011). Securing e-Government and e-Voting with an open cloud computing architecture, *Government Information Quarterly*, 28, 239-251.

Cloud Computing Paradigms and the Changing Roles of the CIO

Anton Ravindran
and Andy Koronios

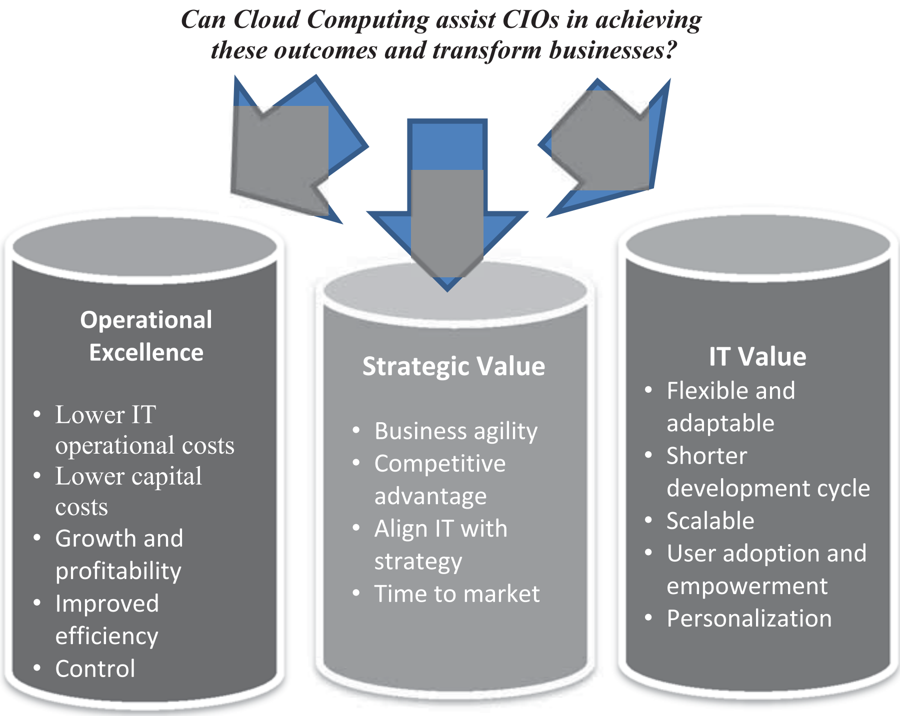
Abstract

The ever-increasing rate of growth, change and significance of Information Technology (IT) shows no sign of abating. Compared to the roles of other senior executives in enterprise, the role of the Chief Information Officer (CIO) has evolved quite significantly over the past few decades. Cloud Computing has emerged as a new approach for computing, with its data centres growing at an unprecedented rate. This has attracted a lot of attention from the scientific, research and business community. Cloud technology represents an evolution and a shift from the traditional IT paradigm. Many scholars have argued that the cloud model is evolutionary, while a few have suggested that it is revolutionary and that computing will become the fifth utility, thus leading to the eventual demise of the CIO. However, few empirical studies have been done to investigate the impact of Cloud Computing technology on CIO's roles and the future of the CIO.

This study investigates how the emerging cloud technology is reshaping the role of the CIO, and what skills will be required of the modern CIO to deliver value and a sustainable competitive advantage to the enterprise. This chapter also introduces new transformational roles for CIOs that they will have to play in the emerging Cloud Computing era, in addition to their existing responsibilities within the traditional IT function. The research and the conclusions of this chapter are based on extensive literature reviews and case studies of nine leading CIOs from large, multinational organizations.

Key Words: CIO**,** Cloud Computing**,** Business Strategy, IT Strategy, Strategic Alignment, Transformation, Innovation, Value Enablement, Disruptive Technology, Utility Computing, Change Agent, IT Innovation

Benefits and Risks of Cloud Computing: There is no argument about the growing worldwide adoption of the Cloud, which offers a slew of benefits to users, both individuals and enterprises. According to an IDC report (commissioned by Microsoft), Cloud Computing will help companies create 1.2 million jobs by the end of 2015 in North America and 14 million jobs worldwide. Further, IDC reported that IT innovation created by Cloud Computing could lead to $1.1 trillion a year in new revenues. While these are only projections, the impact of Cloud Computing on businesses has already demonstrated a paradigm shift in the way we compute. Like most things in life, the benefits of the Cloud model come with challenges and pose risks. In embracing and deploying Cloud Computing, CIOs must leverage from the benefits, while ensuring that the risks and limitations are addressed. The diagram below depicts three key areas: a) Operational Excellence b) Strategic Value and c) IT Value where CIOs can potentially leverage by embracing the emerging cloud model.



1. Key Sets of Values to CIOs

The Cloud model eliminates significant capital expenses for computing infrastructure, upfront license costs, maintenance costs, migration costs and upgrade costs (Brynjolfsson, Hofmann & Jordan, 2010; Hofmann & Woods, 2010; Wyld, 2009). Cloud applications can be deployed instantly and simultaneously to thousands of geographically dispersed users. Developers can experiment and test new and innovative ideas without having to invest in infrastructure. Cloud Computing offers scalability and elasticity to meet on-demand peak and seasonal computing resources (Hofmann & Woods, 2010). Thus, the Cloud has opened up a world of opportunities for big and small enterprises alike, including individuals, by providing access to computing resources without having to implement, administer and maintain it, or buy it – and has increased business agility. This will have an impact on IT departments and the role of CIO is to manage IT infrastructure and IT services of the enterprise.

Impact and Implications of the Cloud on CIOs: Carr (2005) argues that in the future like electricity virtually all IT services will come from a public cloud rather than from private or on-premise data centres. As a result, would IT services become a commodity and adopt a utility model? What is the impact on CIOs? Will it lead to the demise of CIOs and the extinction of IT departments? The increasing proliferation of IT has also led to significant investments in the implementation and maintenance of computing systems. IT has been one of the key disruptive technologies for many decades as it has changed how individuals and businesses function. The advances in computing technology have enabled businesses to automate and innovate, providing them with a competitive edge. A study of IT investments in businesses would reveal that the rapid rise of US-based organizations was due to their ability to quickly leverage IT for their business needs. IT investments in an average US company grew from $3,500 per worker in 1994 to about $8,500 in 2005, and during this period productivity doubled (Hofmann & Woods, 2010). On the other hand, organizations from other parts of the world were relatively slow in leveraging from IT.

With the emergence of the Cloud, the computing landscape and the CIO role is geared for another evolution. Early adoption of the Cloud can provide organizations with an opportunity to transform their business models and gain a competitive advantage. The real value of Cloud Computing to organizations is that it can be a catalyst for more innovation

(Brynjolfsson, Hofmann & Jordan, 2010). Organizations would be able to concentrate on their core business while leaving the task of running IT infrastructure to the Cloud service providers. Further, through the adoption of the Cloud, organizations can become more nimble and agile. Based on this literature study, while the impact of the Cloud Computing model on businesses is well researched and published, its impact on the role of CIOs and IT departments has not received much attention.

The Cloud Computing technology is a means to an end, not the end itself (Muller, 2012, Wyld, 2009). Today's CIOs are expected to deliver more than developing and implementing the latest technologies, they are expected to deliver measurable business value. Technology is increasingly becoming a commodity. CIOs need to transition from a technology guru to a business partner which is not easy (Muller, 2012). Though there are business drivers for the increased adoption of the Cloud Model, it needs to be managed. The role of a CIO continues to rapidly evolve from being a technical leader to a business leader. In 2007, a study by the IBM Centre for CIO Leadership found:

- 80 per cent of CIOs believe that they are valued members of the senior leadership team of the organization
- 69 per cent reported significant involvement in strategic decision-making; and
- 86 per cent of the respondents believed that their industry leaders are using IT to a large or greater extent to create competitive advantage.

Goldstein (2008) offers a very thought provoking and probably a very pragmatic view with respect to the Cloud model – he highlights the fact that the Cloud model is able to free CIOs from having to continuously worry about IT investments and then to manage and operate commodity services. This offers an opportunity for CIOs to focus on more strategic issues and to play a different role by devising more novel ways to take advantage of Cloud technology. However, Goldstein (2008) states that while it is unwise to ignore the provocative thoughts of Carr as 'the potential to completely transform the set of services that IT organizations provide is very real', this does not signal the demise of the CIO's role in importance and influence, but believes that the 'cloud is a force for change'. The benefits of the Cloud to enterprises are significant, however Muller (2012) has highlighted that unlike other utilities it is not a one-size-fits-all solution. Hence, companies and CIOs will have to develop a unique strategy in line with their business objectives to fully leverage from the Cloud model.

As a result, the CIO's role and skill-sets continue to evolve in tandem, in order to effectively harness the technology. One thing is evident from the literature research that there is universal agreement about the potential power of Cloud Computing and its impact on enterprises in spite of the concerns about privacy, security, control and lack of standards. Cloud Computing will uncover and unleash new growth and revenue opportunities for businesses. CIOs who don't embrace Cloud Computing and don't acquire new sets of skills will be at a competitive disadvantage. Cloud Computing is not a new technology and not revolutionary, but a new paradigm which will enable new business models, new opportunities, new revenue streams, and with added capabilities, some challenges, largely because of security, privacy and compliance.

Evolving Roles of CIOs: The role of Information Technology has evolved from a back room (machine room) support function to a strategic function (boardroom) and provides competitive advantage to organizations. Gottschalk and Taylor (2000) state that 'the CIO role emerged in the 70s as a result of the increased importance placed on IT'. Since then, according to Benjamin et al. (1985), the role of CIO has undergone significant changes since the concept was initially introduced nearly four decades ago. Broadbent and Kitzis (2005) define the CIO as 'the most senior executive responsible for identifying information and technology needs and then delivering services to meet those needs'.

According to Stephens (1993, cited in Gottschalk & Taylor, 2000), the CIO function is a continuously evolving role. This change and the corresponding importance of the role of the CIO has largely been fuelled by the unrelenting advances in and increasing reliance on technology, coupled with the constantly changing and increasingly uncertain business environment, driven by globalization, deregulation of markets, and increasing focus on corporate governance. Furthermore, because of the increasing adoption of IT by organizations, as information becomes pervasive and ubiquitous across all functions of business, the CIO role has continued to evolve and has become increasingly strategic and vital to the success of today's highly competitive and turbulent markets. The role of the CIO is at a crossroads as it continues to evolve, because of technological developments that will reshape the use of IT in the midst of growing economic uncertainty, escalating competition and lower customer loyalty (Broadbent, 2002).

Origins and Mission of CIOs: Understanding the Future Through the Past: Rockart et al. (1980) was one of the first efforts directed in defining and clarifying the role of the Information Systems (IS) manager. According to them, the role of the IS executive is shaped by management environment, relevant trends in a business environment and changing technology. Since Synnott coined the term CIO in 1981 (cited in Stephens et al., 1992), the role of the CIO has evolved from managing a data centre and other operational and line activities (a technical manager role – responsible for processing data) to a strategic and executive function, and has grown in importance today – no longer is the role of the CIO driven by the business, but rather it is seen as also driving the business (Basselier & Benbast, 2004; Broadbent & Kitzis, 2005; Lane & Koronios, 2007; Ross and Feeny, 1999; Grover et al., 1993). The modern CIO's mission is no longer about aligning business strategy and IT, but about redefining them. It is no longer about finding and implementing technology that fits into the business strategy of the enterprise, but about using technology to define new strategies.

Need for CIO: Over the past few decades the role of CIO has been evolving from technical manager to tactical, to strategic, to more of transforming the business by leveraging from IT. Since the 80s, streams of researchers have been studying the inception, need and development of this role as it became necessary for organizations. Jones and Arnett suggested that the CIO role was created in order to assign corporate responsibility to manage the information needs of the organization, while Stephens and Loughman argued that the CIO role was created to address the gap between the organizational and IT strategies (cited in Weiss & Anderson, 2004; Hunter 2010). Another reason given by Earl and Feeny (1995) for the creation of the CIO role is to align and integrate the IT function with business goals to ensure business survival and to attain competitive advantage, while Robbins and Pappas (2004, p1) opined that the CIO role was initially introduced in the 80s to support maintaining the IT infrastructure of the enterprise, and has grown in significance and importance over the years and evolved. Peppard & Breu (2003) state 'The evolution of the CIO role has mirrored the dependencies of business upon information technology'.

Gartner (2012) reports, more than two thirds of the chief executives in charge of large companies with $500m or more in annual revenues plan to increase their IT investments this year to drive growth, despite their concerns about the uncertain economic outlook. According to a survey conducted by the London School of Economics, IT investments in some industries represent more than 20 per cent of total costs (CMA Management/Magazine, 1998). A key factor in determining whether an organization's IT investment reaches its full potential may be tied to the role a CIO plays in an organization (Earl & Fenny, 1995). The person in the organization who knows the most about the organization's system, its adoption and uses, and technology, overall, is the CIO (Cha-Jan Chang & King, 2005; Applegate & Elam, 1992). However, for an organization to leverage from IT, its CIO must be able to align IT and strategy – technology and business (Stephens, Ledbetter, Mitra & Ford, 1992).

The drastic surge of importance that is being accorded to the CIO position has been a double-edged sword for most CIOs (Beatty et al., 2005). With the increased visibility and importance being extended to the CIO position, comes a corresponding increase in new job responsibilities and accountabilities. The CIO of today is required to play the role as a high-level corporate liaison and manager between the organization's business functions and the IS function, while also having to make ICT technology decisions that will impact the future competitiveness, scalability and sustainability of the business.

Beatty et al. (2005) further state that 'an analysis of the academic and practitioner publications reveals that there are generally four high-level responsibilities or issues facing the CIO. 1: Meeting the Changing Technical Needs of the Organization; 2: Building a Reputation as a Knowledgeable Business Executive; 3: Orchestrating the Successful Implementation of the IS Strategy; and 4: Maintaining competent IS Staff'. Information management initially and often focused on a technology perspective (the 'how' of technology), but has now evolved and has increasingly become focused on the importance of the strategic impact (the 'what' of management/business) of information. The modern day CIO is considered a business executive first, and a technologist second (Blair, 2005; Rockart, 2000 cited in Hodgson & Lane, 2010). Esat, VP and CIO at Coca-Cola (cited in Muller, 2012) says 'the success of the business depends on the success of IT. The CIO needs to translate technology into business advantages that will differentiate the company from its competitors. CIOs who cannot do that will not be successful – and their companies will not be successful'.

Critical Success Factors (CSFs) for IS Executives

At the beginning of the electronic data processing (EDP) era in the 1960s, the EDP manager's responsibility was to manage the 'machine room' – the data centre. As a result, required skills were purely technical expertise. In the 1970s, the role evolved to encompass some business responsibilities in its primary role as a technical manager, and was known as a Management Information System (MIS) executive. There are four critical success factors for IS executives, which are the first generation of IT leadership to be successful (Rockart, 1980):

(i) **Service:** ensuring the effective and efficient performance of IS and creating positive user perception of IT operations;
(ii) **Communication**: understanding the world of key users and top executives and helping them understand the IS environment;
(iii) **IS Human Resources**: assisting executives in finding IS talent to develop and use information databases; and
(iv) **Repositioning the IS function**: managing the technical, organizational, psychological, and managerial aspects of the firm's IS.

Nearly two decades later, Periasamy & Seow (1998) identified the following five CSFs for the CIO, which are second generation IT leaders, to successfully deploy IT and to deliver optimal value (cited in Lane and Koronios, 2007): (i) business knowledge; (ii) being nimble and versatile; (iii) technical competence; (iv) ability to manage, interact with and motivate staff; and (v) relationship management with internal and external stakeholders. In the following sections this study will examine the CSF for modern day CIOs in the Cloud model.

Four Eras of Computing and IT Leadership: One of the key factors for the changes in the roles and responsibilities of the CIO is the advances in technology. Since the 1960s, by all accounts, computing has evolved through three key eras – computing started with the introduction of mainframe computers in the late 1950s. From then to the late 1970s, it was known as the centralized mainframe computing era. This was followed by the decentralized computing of the mainframe, which led to the distributed computing era characterized by integrated networks of workstations of PCs, mini computers and mainframes (Ross and Feeny, 1999). From the mid-90s, it evolved into the web-based era, because of the growing emphasis of the Internet and the advances in telecommunications, wireless and satellite technologies.

The Ross and Feeny model (1999) identified the evolution of the CIO roles across the above three main technological phases – mainframe, distributed and web-based era. According to Andrews & Carlton (1997), the role of CIO has evolved in four distinct stages: initially, as a ‘glorified EDP manager’ largely supporting a function; followed by CIOs as ‘technocrats’ representing the technical expertise and systems integration; then as ‘business executives’ in the third stage of evolution; and in the fourth stage of evolution CIOs are characterized as ‘technocrats and business executives’ (cited in Lane and Koronios, 2007; Hunter, 2010). Andrews and Carlson (1997) also suggested that the CIO role is now in the fourth wave of change, where the CIO plays the combined role of both business executive and technocrat.

Mainframe Era – EDP Manager: During the mainframe era, the primary benefit of IT to organizations was in automating clerical tasks, thus reducing cost and time, and eliminating human error. Fundamentally, it was about cost reduction. As such, the role of the IT executive, which was generally known as Electronic Data Processing (EDP) manager, was operational in nature and as a technologist in ensuring the delivery of timely data and reports to internal departments. The role was operational with a delivery orientation and was responsible to ‘develop new systems to time and budget, and to operate existing ones to a high level of reliability’ (Ross and Feeny, 1999).

As the era evolved and the computing power continued to increase, the EDP department undertook larger projects with bigger budgets and expectations, which also led to delayed projects and cost and time over-runs. During this period, the EDP manager, the first generation IT leader, remained a middle level manager in the organization. But arguably, as technology continued to advance and, with bigger projects, bigger budgets and bigger expectations, there was growing demand for more technical skills, as well as the need for improved project management and communications skills, until the end of this era. This also led to the compelling need for integration to create seamless operations and reliability of the systems. This was a wake-up call to recognize the need for more authority for the IS department to manage the integration across different functions of the organizations. IT was now beginning to be recognized as an organizational effort, which could become a competitive advantage. This resulted in organizations having to develop an IT strategy for the first time.

Distributed Computing Era – CIO as a Technocrat: From the beginning of the early 1980s, distributed systems became part of the computing architecture of the organization. This was largely due to the advances in networking and computing technologies and partly a reaction to the lack of responsiveness from the centralized IS function of the past era. The introduction of desk top computers and personal productivity software, such as spread sheets and word processing, led to tensions between the corporate direction and local/departmental needs, which resulted in 'islands of computing' in organizations. The increasing need for robust, reliable, seamless and secure IT infrastructure resulted in the increasing importance of the role of the IT executive. With these changes and advances in hardware, software and networking technologies, the CIO's responsibilities expanded and they had to learn and acquire multiple skill-sets to meet the demands of both technology and the business.

In this new role, the CIO had to devise an IT strategy that aligned business and technology, which satisfied corporate-wide computing needs (Ross & Feeny, 1999). The CIO now had to develop and retain talent who had the technical skills in hardware and software, as well as networking. The CIO also needed to recruit talent with knowledge and expertise about the business domain and processes as Enterprise Resource Planning (ERP) became an integral part of the IT landscape in most of the organizations. The CIO was also now responsible for scanning the myriad of emerging technologies. Finally, given the increasing IT budget and its importance, as well the IT landscape of highly complex heterogeneous environments involving multiple vendors, the CIO is required to possess supplier management, relationship and negotiation skills. According to Rockart et al. (1996) 'the objective of the CIO's role as Technology Adviser was to achieve strategic alignment between business and technology' (cited in Ross and Feeny, 1999).

Internet Era (Web 1.0) – CIO as a Senior Strategic Executive: With the commercial introduction of the Internet, web-browsers, open technologies and wide-spread deployment of outsourcing marked the beginning of the web-based era in the early 90s. The Internet began to fundamentally change existing business models and spawn new business models. The key difference between the decentralized/distributed era and the web-based era is that the focus shifted from the need to align IT with business strategy to leveraging IT to create new business opportunities. With this the CIO landed securely in the boardroom and the role required a visionary who can develop new business processes and new business models to leverage from Internet-based technologies. Ross and Feeny (1999) suggested that the CIO in the web era will take increasing responsibility for defining an organization's strategic future (a) as a Strategic Thinker who leads the executive team in developing a business vision (b) maybe even as an Entrepreneur who will introduce new business initiatives. The evolution of the CIO role over the last three decades through the three eras cemented its growing importance and significance at a senior strategic executive level in the organization.

Post-Dot-Com Era (web 2.0) and CIO as Value Enabler: The post-dot-com era is the Web 2.0 era. The term Web 2.0 was officially coined by O' Reilly, who stated that Web 2.0 is the 'network as platform', spanning all connected devices, delivering software as a continually-updated service which gets better the more people use it through an 'architecture of participation', and goes beyond the page metaphor of Web 1.0 to deliver rich user experiences (cited in Dooley et. al., 2012). This era is characterized by the advances in communications and computing technologies, social media and the recent emergence of cloud computing. Furthermore, the post-bubble era, which started after 9/11 and with the financial crisis in 2001, has played a part in reshaping the role of the CIO because organizations became very conservative in spending and in observing regulatory compliance requirements. During this period new business models evolved and CIO's were facing increasing pressure to deliver value on IT investments. The CIO also became a key participant in governance, internal control, compliance, business continuity and risk management processes of the organization. Broadbent (2002) says 'IT governance specifies the decision rights and accountability framework to encourage desirable behaviour in the use of IT'. This has led to challenges from a governance and compliance perspective, which requires the CIO to play a pivotal role in policy, guidelines and monitoring to ensure compliance and enforcement.

Changing Roles and Responsibilities of the Modern CIO: The following three sections discuss the distinct types of CIOs, CIO roles in different business environments, and will outline global roles of CIOs.

1. Distinct Types of CIOs: Though there is a plethora of research publications and other literature on the roles and responsibilities of a CIO, there is no conclusive set of roles and responsibilities that have been universally adopted. Peppard et al. (2010) state that although much has been written about the role of CIO for the past 30 years 'an unambiguous description has yet to emerge'. Indeed, it is apparent that the role of the CIO is an ambiguous one. This is because a) there still seems to be a debate as to whether IT is an administrative expense where costs can be reduced, while the proponents argue that IT offers strategic benefits to organizations, and b) there is lack of clarity about what exactly a CIO is. This lack of clarity is because of unclear distinctions between corporate and business unit CIOs, inconsistent use of IT Director and CIO titles, and a mismatch between expectations and performance metrics of CIOs. Peppard et al. (2010) further state that because of the outdated view of the role of the CIO, some CxOs regard the CIO role to be 'primarily concerned with technology and, in some cases, operational processes'. Thus, there has been ongoing research and attempts to define the roles and responsibilities of the modern CIO.

The roles and responsibilities of CIOs may vary from organization to organization, depending on the size, stage of maturity, and the industry. Five distinct types of CIO have been (Peppard, Edwards & Lambert, 2011):

(i) **Utility IT director:** where the CIO role is technology and operationally focused
(ii) **Evangelist CIO:** where the role is to increase the profile of information within the organization
(iii) **Innovator CIO:** where the role is to leverage from IT and provide competitive differentiation to the organization over its competitors in the market
(iv) **Facilitator CIO:** where the role is to empower and enable the business with information capability
(v) **Agility IT Director/CIO:** where the role is to develop an agile IT infrastructure and to support the organizational and technological requirements.

2. Different CIO roles in Different Business Environments: The role of the CIO has become as diverse as the business models in place today. Robbins and Pappas (2003) have identified the following business environments, and describe the roles of the CIO in each environment as follows:

(i) **The Startup IT Environment:** In the startup business model, the CIO is required to understand the business strategy, business demands and be able to enable the organization to leverage from IT.

(ii) **The High-Tech Environment:** The CIO must be well grounded in technology products and services provided by the organization and be aware of the technical challenges faced by the development environment, as well as the supporting functional departments such as marketing, sales and support.

(iii) **The Brick-and-Mortar Environment:** According to Robbins & Pappas (2003), the brick-and-mortar businesses are largely traditional in their business approach and place financial, healthcare and government institutions in this category. The roles of CIOs can be varied based on their respective models.

(iv) **The Internet Environment:** The CIO in this environment is more engaged in business processes than in the brick-and-mortar and the high-tech environment (Robbins & Pappas 2003).

(v) **The Established Business Environment:** In established organizations where there are established business models, the CIOs have a more traditional business focus and may focus on streamlining operations and reducing overall costs, while their direct reports will focus on the specific needs of the divisional or functional business units.

During the past few decades there have been ongoing discussions and extensive research about the evolving role of IT in organization, based on the belief that IT plays different roles and functions in different organizations (Teo, Wong & Chia, 2000). Similarly, McFarlan, McKenny and Pyburn (cited in Teo, Wong 7 Pyburn, 2000) opined that the role of IT varies in importance among different firms. Another reason for this ongoing debate and research is the notion that advances in IT coupled with the changing landscape of competition continues to drive the evolution of the role of the CIO (Gottschalk & Taylor, 2000). The following paragraphs present some of the key findings in the literature with respect to roles and responsibilities at different periods.

Smaltz et al. (cited in Chen & Preston 2007) have identified six primary roles and responsibilities:

(i) Strategist where the CIO is heavily involved in shaping the overall vision of the organization and in strategic planning and decision-making;

(ii) Relationship Architect where the CIO is responsible for optimal strategic partnership with vendors;

(iii) Utility Provider where the CIO is responsible for the operational efficiency and effectiveness of the IT department;

(iv) Integrator where the CIO is responsible for integration within and between a business unit's enterprise and its partners;
(v) Information Steward where the CIO is responsible for developing the IT architecture, motivating and developing IT talent (talent manager), aligning IT plans with business plans and protecting the organization's information assets (security); and
(vi) IT Educator where the CIO is the champion for improving IT literacy in the organization and in providing valuable insights about the current and emerging technologies.

In another study, Laplante and Bain (2007) have identified five primary roles for CIOs (cited in Lane & Koronios, 2007):
(i) Business Strategist;
(ii) ICT strategist;
(iii) Change Agent;
(iv) Technology Advocate; and
(v) Functional Leader.

This is similar to the findings of Smaltz et al. as both studies have identified four roles that are common. Literature studies indicate mixed findings, whereby only certain roles are commonly identified by researchers and this is in agreement with the observations of Teo, Wong and Chia (2010) that the roles and responsibilities of CIOs will continue to evolve over time, provoked by the advances in technology and organizational needs. These findings are particularly relevant to the research objective given that organizations are now experiencing the introduction of newer technologies, such as Cloud Computing, which will have an impact on the roles and responsibilities of CIOs.

3. Global Roles of CIOs: According to Maes (2007), we can derive the following global roles for the CIO:
(i), **Information Strategist**: Central to this role is defining, developing and deploying the corporate information strategy in line with the business requirements of the organization to create competitive advantage for the business.
(ii) **Co-creator/Advisor business strategy**: As a member of the key strategic management team of the business, the CIO plays a pivotal role in advising the key management team members of the opportunities and risks of ICT.
(iii) **IT Portfolio Manager**: The CIO will be responsible for managing the relationships with the external and internal ICT

providers. Also they would be responsible for defining a long-term strategy for ICT services and be in charge of defining and managing delivery timelines, managing costs, service level agreements and compliance.

(iv) **Enterprise Architect:** The CIO is responsible for developing and managing an information architecture that is flexible and scalable, and developing a blueprint for the information/communication power of the organization.

(v) **Business Advisor:** In this role, the CIO is a liaison between the business unit heads and the information managers. They should be able to inspire the ICT team and assist in redesigning the business process in order for the business to leverage from technology. The CIO and team should be considered 'as part of the business and not as separate'. The CIO must also play a role in talent acquisition and talent retention, including managing skills transfer and training.

(vi) **Trend Watcher:** In this role, the CIO is responsible for keeping abreast of the developments in technology and the external world, and evaluating the value proposition towards organizational perspective.

A more recent study by Peppard (2010) also identified the following seven roles and key competencies of CIOs:

Role	Responsibility
Leadership	• Driving the organization forward in the use of IT • Creating a set of value expectations shared across all areas of business in relation to IT • Influencing key stakeholders • Growing and developing own leadership
Visionary	• Envisioning options and opportunities (both operational and strategic) • An advocate for new technology
Strategic Thinker	• Holistic view of business • Contributing to strategy discussions
Relationship builder	• Expressing empathy, listening, and demonstrating passion
Diplomat	• Collaborating with colleagues to achieve 'win-win' • Building personal networks across the organization • Creating the right impression
Deliverer	• Achieving credibility with both business and technical people through successful delivery of projects and programmes • Maintaining cost efficient IT operations and services • Meeting expectations
Reading the market	• Using the marketplace appropriately for sourcing • Commercial acumen • Networking externally with peers

2. CIO Competencies – Adapted from Peppard (2010)

The studies of Smaltz (2000), Lane and Koronios (2007), Laplante and Bain (2005), and Peppard (2010) mostly corroborates with the roles and responsibilities presented by Gootschalk and Taylor (2000), who combined the three critical management attributes (interpersonal, informational and decisional) developed by Mintzberg (1990), with six organizational development roles identified by the Computer Science Corporation (1996), to develop a set of CIO roles as follows, which this research finds to be the most comprehensive:

3. Six CIO Leadership Roles

Role	Responsibility
Chief Architect	Designs future possibilities for the business
Change Leader	Orchestrates resources to achieve optimal implementation in the future
Product developer	Helps define the company's place in the emerging digital economy
Technology Provocateur	Embeds IT into the business strategy
Coach	Teaches people to acquire the skills sets they will need for the future
Chief Operating Strategist	Invents the future with senior management

4. Six Leadership Roles of CIO – Adapted from Gottschalk and Taylor (2000)

It is evident that most of the literature if not all, agrees that the role has evolved from an operational (automating existing business processes) to a strategic executive position (enabling new processes) and there are strikingly similar roles and responsibilities identified by the various researchers. But there seems to be lack of clear consensus in defining a set of roles that has been universally accepted. This is due to the fact that there are different types of businesses which will need distinct types of CIOs.

Further, the previous studies have identified what has changed with the advances in technology. However, there has been very little research that has actually attempted to identify what will occur in the future for the CIO with the changing business environment and the emerging Cloud Computing technology.

The review of academic literature clearly indicates that today's CIO is responsible for aligning ICT with the business goals of the enterprise, and leveraging from ICT to achieve the strategic vision of the organization. There is a general consensus among researchers that the business role of the CIO will continue to increase. These point to the trend that the role has been shifting towards a leadership role that requires organizational, technical, interpersonal and business skills from a general perspective, and the responsibilities will vary depending on the business environment and the CIO type.

An important conclusion can be drawn: it is now evident that areas of the modern CIO have expanded to encompass all three vital areas of any enterprise – technical, business and organizational competencies.

5. Emerging CIO Roles and Responsibilities in the Cloud Model

As described and chronicled in the previous sections, though the roles and responsibilities of the CIO as a leader (Broadbent & Kitizis, 2005; De Lisi et al.,1998; Feeny, Edwards and Simpson, 1992; Peppard, 2010; Peppard et al., 2011; Onan et al., 2001; Karahanna & Watson, 2006; Remenyi et al., 2005) and as a business strategist (Carr 2003, Lane and Koronios, 2007; Laplante and Bain 2005) have been a topic of significant research interest for the past few decades, few have researched in the context of the latest developments of Cloud Computing. Hunter (2012) states that we can be certain that 'for the vast majority of businesses and organizations, cloud represents an evolutionary step away from traditional IT paradigm and something new'.

The following section will further expand on the previous research by focusing on the impact of the new generation of disruptive technologies and their impact on the changing roles and responsibilities of CIO. As such, it will attempt to answer the research questions in specific detail a) Will the emergence of Cloud Computing affect the role of the CIO? b) If so, how will Cloud Computing affect the role of the CIO and c) Will it lead to the demise of the CIO role?

Evidently the role of the IT leader has been in a state of flux and the literature clearly indicates that the role has undergone considerable evolution, and is still evolving. In addition, because of the challenges, constant changes and the difficulty in ensuring ROI from IT, there have been debates and opposing views as to whether 'IT matters' to businesses anymore (Carr, 2003; 2005). Technology changes much faster today. Two decades ago, most hardware infrastructure lasted nearly five years, whereas today operating systems for tablet PCs and mobile devices are updated every six months. As such, IT will be asked to find ways to support upgrades and migration every six months.
According to Evans (2012), the new technological developments such as mobile applications, social networks and Cloud Computing are providing new end-user experiences and the key features are: social, mobile and Cloud enabled. End users are increasingly becoming demanding for

advanced functionality, simplicity and ease of use that has become part of the consumer market. Evans (2012) further suggests that only when strategy, process and technology integration efforts are all combined for the benefit of the end-user experience that the whole will be greater than the sum of the parts. This will become a force multiplier for business. The unrelenting exponential growth of mobile devices is driving the growth of social media, which has resulted in the ongoing exponential growth of data. Hunter (2010) maintains that 'there's no place for all this new data to go except the Cloud'.

Byrnes (2005) opines that because of the ongoing advances in technology, today's CIOs have far more opportunities than their predecessors. He further states that the key is to actively participate in the business, partner with counterpart business leaders throughout the company, and drive towards a 'paradigmatic change'. Byrnes (2005) concludes that 'more than ever before, the CIO's effectiveness will determine the destiny of the company'. Internet, mobile applications, and social networks have no doubt provided access to technology, and placed more information in the hands of the consumer.

6. The Role of the CIO in the Cloud

Cloud Computing is a game-changing technological paradigm. It is not a new technology, but a new paradigm that encapsulates all facets of computing, including software, hardware, storage and even processes. Rehman & Sakr (2011) state that 'Cloud computing is a disruptive technology, one that embodies a major conceptual shift and is rapidly changing the way users, developers, and organizations work with computing infrastructure'. Menzel et al. (2010) also state that 'Cloud computing technology, representing a new model for Information Technology (IT) solution engineering and management that promises to introduce significant cost savings and other benefits'. This will inevitably alter the roles of CIOs. Cloud services such as SaaS are gradually removing the IT organizations from day-to-day support and operational functions. As such, CIOs are beginning to shift IT resources to other value-added activities that could potentially deliver higher levels of value to the business – IT playing a key role in innovation and transformation of the business. Wyld (2009) opines, 'as we look out over the immediate horizon, one can well imagine that in five to ten years, we may not speak of the Cloud Computing anymore'. This is because using what we now regard as Cloud Computing services – for applications, for storage, for email, etc. – will transform how we utilize technology, and may simply become the way things are done, just as today we can no longer tell if a phone call is coming from an analogue or cell phone. In all, unlike the traditional CIO whose primary focus was on deploying technology to achieve competitiveness through efficiencies, the transformational CIO in the Cloud model will focus on value creation by leveraging from technology.

In the Cloud, CIOs will play a more active role in developing sourcing strategy and procuring services, as opposed to purchasing assets, and in developing service level agreements and performance standards, security, compliance and liabilities (Milroy, 2012). According to the Nolan Norton Institute (2001, cited in Hunter, 2010), the senior information technology executive is moving away from the 'technology' and focusing more on 'information', and the CIO is becoming more of a 'prophet' and will leverage technology to transform the enterprise. Goldstein (2008) opines that it would be naïve to assume that there will not be any risk of loss of stature during the transition and feels it is very likely that the CIO will increasingly gather power, authority and stature from their ability to be business agile and being knowledgeable conveners of stakeholders around a host of issues related to management of technology.

Transformational CIO: Gartner Executive Programs Worldwide Survey of more than 2,000 CIOs identified Cloud Computing as a top technology priority for CIOs in 2011. Hodgson and Lane (2010) have stated that 'the business environment has changed significantly in the last decade in response to globalization of modern economies and increased reliance of organizations on information and communication technology (ICT) has become both an enabler of key business processes and as a source of competitive advantage through technological innovation for organizations'. According to Wood (2011), 'Cloud computing is capturing the imaginations of CIOs and with good reason' and further states that 'success in the future will depend upon how CIOs will approach investments in technology'. Hence, increasingly, the most senior IT executive, the CIO in medium to large organizations has a key role to play in contributing to the performance and strategic direction, innovation and transformation of an organization, and will play the following roles below.

Influencer and Innovator: With constant and increasing cost pressures on businesses in general and specifically IT investments, CIOs are beginning to leverage from virtualization, infrastructure convergence and cloud mobility for their business needs. CIOs recognize the disruptive impact and the strategic value of Cloud Computing. As a result of the ongoing advances in Cloud Computing, some have even suggested that the acronym should be redefined to mean ‘Chief Innovation Officer’. Radjou (2008) found ‘CIOs who excel – not only as IT leaders but as Chief Innovation Officers as well’ – and in their organizations the IT department was responsible for driving innovation. Commercial service providers are beginning to offer the entire IT needs of an organization, including hardware and software infrastructure, middleware platforms, software services and turnkey solutions, thus providing scalability and elasticity of computing. The Web continues to mature, and smart phones are fast becoming pervasive. The congruence of the advances in smart phones and Web technologies will have a dramatic impact on the way we work, live and play. As stated by McFarlan (2002, cited in Henderson et al., 2003), ‘the firm must do things differently with IT not just the same things better’. As such, he opines that the emerging CIO will be featured more from a position of ‘influencer’ rather than a ‘chief’.

Change Agent: The continuously emerging need for business agility, flexibility and personalization has driven organizations to seek new and innovative approaches for information systems development. The modern day CIOs have been called to drive business transformation and to create new and innovative business models by leveraging from the ongoing advances in information and communication technologies. It is evident that the role of the CIO is rapidly changing from ‘enabler of change’ to ‘shaping change’. In the information economy, efficiency, consistency and reliability may not be adequate to differentiate from competitors. The Cloud model relieves the CIO from having to worry about keeping the system up and running and allows the CIO to focus on innovation. But this will require a change in mindset from ‘efficiency based’ to ‘innovation based’ enterprise. Peck, CIO of Levi Strauss (cited in Hunter, 2011) says, ‘don’t fight the change, and don’t fight the cloud. Figure out how to embrace it’. McDaniel provides a proactive view that Cloud Computing will lead to transformation in organizations, which will require CIOs to engage in ‘creative destruction’ – which will require the CIOs to ‘reimagine the role of IT and their organizations’. According to McDonald and Aron (2011), this ‘creative destruction’ or ‘destructive creation’ will require CIOs to dismantle existing computing infrastructure, and to embrace new resources ‘in ways that redirect and liberate resources to deliver greater innovation and value’ (McDaniel, 2011).

Value Enabler: The agility and speed-to-market capabilities of Cloud Computing are enabling new business models. According to Muller (2012), the CIO is now expected to help the business generate value and grow business, and being perceived as a 'value enabler' is no longer merely a compliment – it is rapidly becoming a basic part of the CIO's mandate. According to Goldstein (2008), the emergence of a Cloud Computing paradigm may diminish the set of traditional services provided by the IT organization, while presenting more complex sourcing options for services. He further states that the 'fluid nature of the environment also suggests that the decisions may not have terribly long shelf lives'. Gold, CIO of Avaya (cited in Muller, 2012) states that three or four years ago a CIO's mandate was biased towards efficiency, cost savings, optimization, and streamlining of business operations, both inside the organization and across the enterprise. But today, the role has evolved where CIOs are still expected to drive efficiency through innovation, but they're also expected to drive revenue and deliver value for the businesses.

Today, technology enables business to fundamentally do new things such as personalization where the best customer is identified, and goods and services are tailor-made to their requirements and sold directly. At the same time, businesses have to change the way they sell to other customers, or stop selling to them completely. ICT capabilities today are much more powerful by sheer orders of magnitude than they ever were. This requires a paradigm shift in terms of the skills of the IT leader, to plan and execute new systems and to develop new business models. Hunter (2010) states that the Cloud has introduced a whole new way of looking at IT – 'the old way was about processes; the new IT paradigm is about results'. The traditional CIO's (technology focused) focus was on how to leverage from the technology, but in the Cloud model the transformational CIO will focus what can be accomplished by leveraging from the technology. With increased visibility and importance being given to the modern CIO, the roles and responsibilities have been broadened as part of the evolution from pure technology based skills and competencies to include business strategy, leadership and organizational skills.

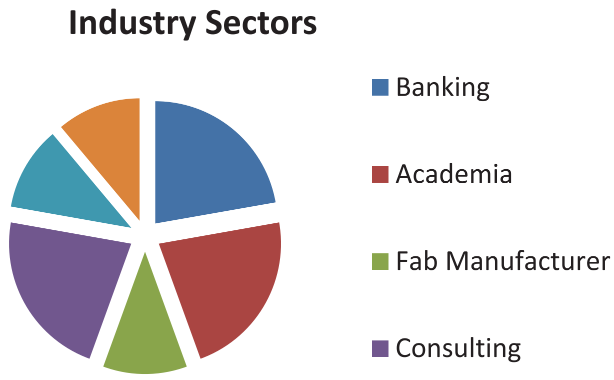
New innovations in ICT and business models lead to further innovations. For example, the continuous and rapid increase in computational power and storage capacity at reduced cost produces a demand for new applications, which in turn creates the demand for more storage and processing power. Computers are used to create yet another new generation of more powerful computers. ICT is inherently dynamic and it is not surprising that the role of CIO is continuously in a state of flux. More than ever before, the competencies and skills of the CIO will determine the destiny of organizations, as information management has moved from the wings of company operations to centre stage. It is also evident from the lack of consensus from the literature that the CIO role is far from being standardized, but none would argue that it is not a strategic position. However, depending on the industry, the culture of the organization, and the stages of growth, the organizational structure among other factors, the roles and responsibilities of the CIO will vary from organization to organization. This vast scope of responsibility also presents numerous challenges and complexities for the CIO and the organization. In conclusion of this chapter, this research makes the following inferences from the literature studies.

Operational ❯Tactical❯(Modern CIO) Strategic❯(Emerging CIO) Transformational

- **Operational to Transformational:** ̈The role of CIO has evolved from the initial DP Manager, where it was largely if not solely operational to a strategic executive who has a pivotal role to play in the transformation of the enterprise business. The transformational CIO will be an Innovator and Influencer, Change Agent, and Value Enabler.
- **Growing Significance of the Role:** The role of CIO has gone through a few key phases and has expanded rapidly both in scope and magnitude. Hence, it has become more complex, vital, dynamic and challenging. The life span of technology and products are decreasing, while enterprises are becoming increasingly reliant on information to make decisions.
- **Increasing Value of Information:** The role of the CIO will become more focused on the quality and compliance of the information of the enterprise. In today's globalized information economy, where information about the product and services are perceived to be more valuable than the product or service itself, the role of the CIO is no longer about provisioning of IT services, but about providing information.
- **Roles of Transformational CIO:** CIO roles have moved from being heavily 'technology' focused to 'strategic alignment and business focused', and now 'business transformation and information focused'. The emerging roles of the transformational CIOs **are Innovator, Change Agent and Value Enabler.**

The review of the academic literature on Cloud Computing and the evolution of the CIOs' roles have revealed that the Cloud model is a paradigm shift from traditional computing and presents both opportunities and challenges for CIOs and businesses. The literature study also suggests that the roles of the CIO continue to evolve as a result of the emerging Cloud Computing model. However, the literature review is not conclusive as to whether the Cloud model is revolutionary or evolutionary, and whether computing will become a utility as suggested by Carr (2005), eventually leading to the demise of the CIO. Further, some scholars have argued that it is an evolutionary model and do not envision the demise of the CIO role as advocated by Carr. From the literature it is undoubtedly evident that the emerging Cloud Computing model has begun to reshape the role of the CIO and its benefits to enterprises are well documented. However, the above referenced studies in the literature do not provide a clear and specific answer to the question as to how it affects the role of the CIO, whether it will lead to the demise of the CIO role, and what the roles and responsibilities of the future CIO to meet the challenges and opportunities presented by the emerging Cloud Computing model actually are. The case studies intend to answer these specific questions and will present roles and responsibilities of the CIO in the emerging Cloud Computing Paradigm.

The research methodology includes, in addition to the research based on secondary sources from the literature, interviews with CIOs to complement the findings of the academic research and to leverage from practical experiences of the interviewees. Therefore, the case studies in the following sections will build from the literature review and further investigate the impact of the emerging Cloud model on the roles of CIOs.



The semi-structured interview technique was used to gather information and to explore with the interviewees of their views, experiences and insights with respect to Cloud Computing and its impact on the role of the CIO and the enterprise. All interviews were scheduled in advance. All nine interviews were conducted in person and permission was sought from the interviewees to have their interviews recorded and all agreed to it. This also ensured that the same topics were covered in each of the nine interviews. Each interview lasted at least an hour, and was recorded and transcribed. The transcribed version was sent to all nine CIOs to seek feedback, for any corrections and to seek validation from them about the information gathered. All nine CIOs were senior IT professionals with advanced degrees with more than 20 years of experience, except one who had more than a decade. They all had similar status, though the IT budget and staff strength varied from one to another. All nine CIOs have extensive responsibilities and managed a large team of IT professionals and multi-million dollar IT projects on an ongoing basis.

Being a semi-structured interview, the questionnaire was grouped into four major segments: These are:

1. Questions about the role of CIO and its challenges;
2. Questions for CIO whose organization has either adopted Cloud or is planning to adopt Cloud;
3. Questions for CIO with respect to changing roles and skills; and
4. Concluding questions.

The nature of emerging technologies means that there is limited historical information and data. As such careful extrapolation of the information gathered is important to the success of the study. Christensen (2005) states that in every event studied about the industries and companies displaced by disruptive innovations, the industry knew about the technology involved for the disruption. As such, it is safe to conclude that predictive information exists if the researcher knows how to investigate, uncover and gather information and data.

A key factor in determining whether an organization's IT investment reaches its full potential is tied to the role of CIO and the person who knows the most about the company's IT system, adoption and uses, and technology overall, is the CIO (Earl & Fenny, cited in Cha-Jan Chang & King 2005). This study did not include discussions with CEO and business leaders, but this did not compromise the research and the outcome of the research findings in any way as the unit of analysis is CIO. Yin (1994) suggested that the unit of analysis as the actual source of information might be a country's economy, an industry or an individual.

The interview questionnaire and an information sheet were sent in advance to all the interviewees. After each interview, the transcribed version was sent to all interviewees to seek additional feedback and to preserve the essential information from each interview. The response to each question in each of the four sections of the questionnaire was analysed individually after each interview, first to identify core themes, responses and observations. The large amount of information gathered at each interview (lasting about an hour or more for each) was reduced to an appreciable level since the administered questionnaire was categorized into four sections. At the end of all nine interviews, further analysis was done to establish key themes and common observations, as well as to identify any exceptions amongst the nine interviewees. Each of the four sub-sections of the questionnaire were tabulated and presented in the following section, with the analysis discussions and findings for each of the four sub-categories.

This approach allowed the researcher to meet, discuss, investigate and test the theory with the nine CIOs from large and reputable organizations, from four different geographical areas, who have been IT leaders with extensive experience in the IT industry for many years, if not decades. All nine CIOs have experience with Cloud Computing technology. This empirical part of the research is supported by the extensive and comprehensive theoretical literature-based research. The interviews with the leading CIOs have been used to examine, further develop and reinforce the findings of the literature-based research.

The literature has supported the fact that the CIO role continues to evolve, as well as having an impact on the role of the CIO going forward. This section presents the findings of the case study research.

7. Analysis of the Empirical Data and Findings

The following section describes the nine CIOs involved in the case study. This is followed by an in-depth analysis of the interview findings for each of the four key sections of the interview questionnaire, namely:

1. Key roles of modern CIOs
2. Major Challenges faced by CIOs today
3. Future of the CIO in view of the emerging Cloud model.

For the purpose of anonymity, pseudonyms of the CIOs and companies will be used for all nine case studies. The pseudonyms are names of colonial personalities in Singapore. The nine interviewees involved in this study have worked as CIOs for their respective companies for different lengths of time ranging from six months to more than ten years. All nine CIOs have been in leadership positions for more than a decade. The nine CIOs interviewed have combined ICT industry experience of more than 250 years. Each interview was studied and reviewed very closely. Once the review was completed for each individual case study, a cross-case study analysis was performed to identify common aspects and the relationships/links between them as well to uncover differences if any. This allowed for a deeper and comprehensive understanding from the diverse backgrounds of the nine CIOs. All nine CIOs interviewed confirmed that the role of the CIO is evolving and that it will continue to evolve towards a more strategic role in managing the ongoing technological developments and business requirements. All CIOs also agreed that the CIO roles will change because of the emergence of the Cloud Computing paradigm.

All nine interviewees also confirmed that enterprises are feeling the pressure to better manage the investments in IT and to find new ways to support the business objectives of an organization. Therefore, CIOs are challenged to pursue innovative strategies in a much more aggressive manner than in the past. Eight of the nine interviewees also confirmed that they have implemented the Cloud in their organizations. There were different views to the extent of the impact of Cloud Computing, both to the organization and the role of CIO. Amongst the nine CIOs, one foresees the demise of the CIO in about 10 years and ICT becoming a utility and commodity. Carr (2005) has argued the scenario of computing as a disruptive technology that will result in IT becoming a commodity and a utility, while most of the other researchers (Hofmann 2010; Hofmann & Woods 2010; Brynjolfsson, Hofmann & Jordan, 2010; Muller 2012; Wyld 2009) have disagreed with these views and are of the opinion that the Cloud is a paradigm shift that will lead to the evolution of the CIOs role and more innovation in enterprises. This is further substantiated by the views expressed by the other eight CIOs who are of the opinion that Cloud Computing is evolutionary and that the role of the CIO will continue to evolve.

8. Key Roles of Modern CIOs

The nine CIOs interviewed have different levels of technical, business and managerial skills, and backgrounds. It was observed that these skills coupled with other organizational factors such as nature of the business had an influence in their views about the CIOs role and the emerging Cloud Computing paradigm. None of the nine interviewed were involved in operational functions, and all had limited involvement in hands-on technical work, but all are in leadership positions.

Name	Industry	Title	# of years in IT industry	Territory	Location
Maxwell	Airport Authority	Group CIO	➢ 20 Yrs	Singapore	Singapore
Cecil	Education	CIO	➢ 20 Yrs	Singapore	Singapore
Stamford	Property & Real Estate Development	VP & Head of IT	➢ 20 Yrs	Global	Singapore
Cavanaugh	Banking & Finance	CIO & MD	➢ 20 Yrs	Asia Pacific	Singapore
Elgin	Education	CIO & PVC	➢ 20 Yrs	Asia Pacific	Australia
Anderson	Semi-Conductor	Head of IT/CIO	➢ 20 Yrs	Asia	Singapore
Fullerton	Property Development & Diversified Industries	CIO	➢ 20 Yrs	Asia Pacific	Hong Kong
Clifford	Consulting	Senior VP	➢ 20 Yrs	Global	London
Anson	Banking & Finance	MD & CDO	➢ 10 Yrs	Global	Singapore

9. Demographics of the CIOs (Interviewees)

In order to understand the impact of the emerging Cloud model on organizations and CIOs, each interview was started with a lead-in question by asking the CIOs to identify three current roles that they think is key to the success of their responsibilities and their organization. Overall, all nine interviewees confirmed that CIO is a strategic position, which requires leadership and organizational skills. There was also an agreement amongst all interviewees that the CIO must have a good understanding of the business, need to focus on ICT strategy and business strategy alignment, as well as have an eye on IT investments and its value to the business.

All nine CIOs also agreed that the CIO should be able to anticipate change and must have an ‘eye’ on the future landscape of technology. Mr. Cecil commented that ‘CIO as the chief architect is responsible for the alignment business and ICT strategy and for transformation of the business'. Two of the CIOs identified 'talent hunting and mentoring' as a requisite skill of a CIO. All in all, seven key roles were identified by the nine interviewees, as listed in the table below. It must be noted that the interviewees were asked to identify at least three key roles of a modern CIO and all of them identified more than three.

Roles	Maxwelll	Cecil	Stamford	Cavanagh	Elgin	Anderson	Fullerton	Clifford	Anson	Total Nominations
Leadership	X	X	X	X	X	X	X	X	X	9
Strategy Alignment (Business & ICT) – Proactive Business Partner	X	X	X	X	X	X	X	X	X	9
Talent Development, Mentorship & Coach	X	X	X							3
Managing IT budget & ROI	X	X	X	X	X	X	X	X	X	9
Vendor Management	X						X			2
Futurist					X		X			2
Innovation & Transformational Capability	X	X								2
Risk Management, Compliance & Contingency Planning				X					X	2

10. Case study – CIO Leadership Roles

In summary, the following important conclusions can be drawn: All nine interviewees were unanimous and in agreement that a) CIO is a strategic leadership role b) CIO is responsible for the alignment of business strategy and ICT strategy, and c) for managing ICT budget and for its ROI.

Mr. Maxwell stated:

> *'Today's CIO is a full-fledged senior executive of the organization with deep understanding of the business and technical know-how'.* He also described being a *'pro-active business partner' as a key role of the CIO.*

Mr. Elgin described that the CIO should have *'transformational qualities and skills'*.

These indicate the continued evolution of the role from technical to strategic business executive, to proactive business partner, to transformational leader. These observations are supportive of the findings of the literature-based research; namely, the modern CIO is a leadership position with strategic responsibilities (Blair, 2005; Rockart, cited in Hodgson & Lane 2010) and is on the verge of further evolution (Goldstein, 2008; Krotowski, cited in Muller, 2012).

11. Major Challenge(s) faced by CIOs Today

All in all, during the interviews with the nine CIOs, eight major challenges were identified as depicted in the table below. Speed of change, managing IT budget, and aligning IT strategy and business strategy are the three key issues identified. Five of the nine CIOs noted that the speed of change and advances in technology are the key challenges, followed by five CIOs who noted that alignment IT strategy and business strategy remains a challenge, while nine CIOs also indicated that managing IT budgets remains an ongoing challenge. Interestingly, though researchers and IT leaders have long spoken about the need to align business strategy and ICT, this still remains a challenge.

Mr. Maxwell stated:

> *'developing the team to be more business oriented, getting them to collaborate and to get them to appreciate the expectations of both internal and external customers is a major challenge'.*

Mr. Cecil stated:

> *'the new IT environment entails the IT support staff to be more resourceful and capable in handling both IT skills and people skills. For example, a good systems analyst has to be equipped with business analyst skills as well'.*

Mr. Stanford stated:

> *'IT employees lack business and leadership skills. Skills are needed to influence others in the organization to think, work and act differently in order to fully leverage from ICT – skills are needed to communicate this'.*

Mr. Elgin stated:

> *'Today, if IT ceases to function, the organization ceases to operate'.*

Challenges	Maxwell	Cecil	Stamford	Cavanaugh	Elgin	Anderson	Fullerton	Clifford	Anson	Total Nominations
Speed of Change	**X**	**X**				**X**		**X**	**X**	5
Managing User Expectations				**X**				**X**		2
Managing IT Budget	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	9
Aligning IT Strategy & Biz Strategy	**X**	**X**	**X**	**X**			**X**			5
Talent Acquisition & Skills			**X**		**X**		**X**			3
Vendor Management				**X**	**X**					2
Increasing Volume of Data				**X**					**X**	2
Security, Confidentiality & Compliance				**X**	**X**				**X**	3

12. Case Study – Challenges faced by CIOs

It is evident that IT is strategic to the t survival and future of the enterprise, CIOs role is to ensure that there is alignment between business and IT strategy and to effectively manage the IT budget given the dynamic nature of the business environment. It is also evident that there is an emerging need to focus on security and confidentiality of the data which continues to grow in size and reach (global and virtual).

13. CIOs Views on Outsourcing and Cloud Computing Technology

In order to fully leverage from the emerging Cloud Computing technology it is necessary to understand as to whether and how the Cloud Computing model will reshape the role of the CIO and whether it will lead to the eventual demise of the role of CIO. The case study involving the nine CIOs allowed me to investigate the research questions in depth and to observe and understand from natural settings. The interview questions explicitly asked to express their opinion with respect to the research questions and the findings are discussed in detail in the following sections. All interviewees have had experience in outsourcing. As discussed in earlier sections, both outsourcing and Cloud Computing procure services from suppliers. Driven by the need to be agile, to be efficient and to manage cost, enterprises have been leveraging from outsourcing services. In the outsourcing model, businesses delegate some of their non-mission critical and/or non-core business functions to an outsourcing service provider. Six CIOS opined that outsourcing is beneficial to the organization, while two were neutral and the ninth CIO was not supportive of outsourcing at all. The six CIOs with positive experience expressed that outsourcing has freed them from the burden of infrastructural and resource management and has enabled them to focus more on strategic and value added activities.

Mr. Maxwell said that though it is beneficial, it is also complex, and that it is about ‘right sourcing’. He also opined that ‘he has learnt not to outsource the problem’. All CIOs said that they will not outsource applications that are mission critical and/or applications that require significant levels of customization to meet specific organizational and business needs. With the exception of one, all the CIOs who are supportive of outsourcing acknowledged that it has a positive impact on the role of CIO.

Mr. Cecil opined that:

> *‘outsourcing has placed more focus on vendor and project management and the CIO is now tasked to become more involved in looking at innovative use of technology in business process that meet the business objectives’.*

Mr. Stanford states:

> *'the only problem is that most outsourcing vendors come in from a technical solution perspective rather than address the business need, problem or an opportunity'.*

CIO	Outsourcing	Impact on CIO roles	Cloud computing Disruptive?	Implemented Cloud?	Will cloud eliminate CIO role/IT department?
Maxwell	Beneficial	Positive	No, evolutionary	Yes	No
Cecil	Neutral	Positive	No, evolutionary	Yes	No
Stamford	Beneficial	Positive	Disruptive	Yes	No
Cavanaugh	Beneficial	Positive	Disruptive	Yes	No
Elgin	Beneficial	Positive	Disruptive	Yes	No
Anderson	Beneficial	No major impact	No, evolutionary	Yes	No
Fullerton	Not good	NA	Disruptive	No	Yes
Clifford	Neutral	Positive	No, evolutionary	Yes	No
Anson	Beneficial	Positive	No, evolutionary	Yes	No

14. Case Study – CIOs views on Cloud Computing Technology

With respect to the Cloud Computing model and its impact, four of the CIOs felt it is a disruptive technology, and the other five CIOs felt it is a new computing paradigm which is evolutionary. Except for one CIO (who opined it is disruptive), all the other CIOs have implemented Cloud Computing in their organization and all eight felt that it will not eliminate the IT department and will not lead to demise of the CIO role. The lone CIO who opined that it is disruptive, has not implemented the Cloud, and also felt that the Cloud will lead to the demise of the role of the CIO, feels that ICT will eventually become a commodity. He suggested that computing as a utility model will become pervasive in about ten years.

Seven of the interviewees were of the opinion that outsourcing has a positive impact on the role of CIO, while one interviewee opined that outsourcing does not really impact the CIO but the middle management, and the last interviewee stated that outsourcing has a negative impact. The latter does not support outsourcing and expressed that 'it disengages the IT department from the internal stakeholders'.

All nine interviewees suggested that Cloud Computing will become a utility model for small and medium enterprises (SMEs), resulting in eventual elimination of in-house IT departments, while eight of the interviewees were of the opinion that large enterprises will adopt a hybrid model.

Mr. Maxwell stated:

> *'if there is anything I learnt over the past 20 years in this business, it is to be adaptive and choose whether to be a fast follower, an early adopter or just go along for the ride'*

while Mr. Fullerton commented:

> *'the role of CIO will gradually and but surely diminish and I foresee the eventual demise of the role as I think Cloud Computing will achieve its vision as a utility resulting in ICT becoming a commodity'.*

Mr. Cecil opined CIOs in large enterprises will never face off as long as they continue to renew their mindsets to innovate and contribute to the core business's outcomes. All nine interviewees were of the view that ICT is rapidly developing from an asset that was procured into a service that will be purchased and consumed on-demand, and this is beginning to take foothold.

15. Emerging New Roles of the CIO – Cloud Model

From the literature study and from the findings of the case studies of the nine CIOs, it is evident that the CIO role is indeed changing to fully leverage form the emerging Cloud Computing model. During the case studies, all nine interviewees were asked specific questions to investigate and seek their views with respect to the impact of Cloud Computing on their roles and responsibilities. The following sections will present the findings of the nine case studies. All CIOs without exception confirmed that the role of the CIO is changing. As such, all nine interviewees were asked to identify the key roles of CIOs for the Cloud model. Their responses are shown in the table below. All interviewees suggested that the CIO will play an innovator role as the Cloud is a catalyst for innovation. In addition, eight felt that the CIO will also play a pivotal role in the transformation of the organization in the Cloud model.

CIO	Innovator	Transformational	Trusted Advisor	Translator/Broker	Change Agent	Value Enabler	Data Security Expert/Evangelist
Maxwell	✓	✓	✓	✓	✓	✓	
Cecil	✓	✓	✓	✓	✓	✓	
Stamford	✓	✓	✓	✓	✓	✓	
Cavanaugh	✓	✓	✓	✓	✓	✓	✓
Elgin	✓	✓	✓		✓	✓	
Anderson	✓	✓		✓		✓	
Fullerton	✓	✓	✓				
Clifford	✓	✓	✓	✓	✓		
Anson	✓			✓	✓		✓
Total Nominations	9	8	7	7	7	6	2

15.1 Emerging CIO Leader: New Roles

The majority (seven) of the interviewees opined that the CIOs in the future will play an additional role as an advisor and as a change agent, while six of the interviewees also suggested another role as value enabler. Amongst the nine interviewees, seven identified translator/broker role for the emerging CIO.

The Cloud Computing model extends opportunities for enterprises to innovate as it has lowered the infrastructure cost, while providing scalability and elasticity with respect to computing resources. This is the true value of Cloud Computing as it is best suited for enterprises to innovate, evaluate, test and launch new products and services. The emerging CIO's role will shift from reactively supporting the IT needs of the enterprise to spearheading structural and mindset changes within the IT department, in order to drive innovation that supports transformation of the business model for the enterprise. Thus, the CIO will play the role of a business enabler rather than a supplier of IT services as in the past, working side-by-side with the users and businesses as a partner. As a result of these developments, the CIO will be in a position to predict the changes and play a key role in motivating the rest of the organization, including the executive management, to change. In fact, the CIO will no longer be just the change agent, which has been a well presented idea, but the change predictor and the leader in shaping the change.

Mr. Stamford stated that the CIO for the new era should focus on transformation efforts and should look beyond job security. Look beyond hiring internal resources and building a kingdom but focus on collaborative and innovate models by leveraging technology to serve business objectives. Mr. Cavanagh suggested that going forward CIOs must focus on innovation to increase business agility by leveraging from disruptive technologies as the conventional approach will not be adequate to remain competitive. As such, CIOs should play the role of a thought leader of ICT for the organization and play the role of an advisor with respect to articulating the trends and technology developments to the organization and support business transformation. In the New Economy, the CIOs role is no longer about provisioning of IT services but about making sense of the growing volume of data and to provide information for the enterprise to create value through innovative products and services. CIO’s are increasingly becoming responsible for leveraging from technology to increase profitability, improve competitiveness, attract new customers and serve new markets. In the Cloud Computing paradigm, CIOs focus on technology will continue to shrink, thus enabling them to focus more on innovation by leveraging from the technology.

Interestingly the two interviewees from the banking and financial sector identified the role of Data evangelist and Data security expert. Security is one of the key challenges in the Cloud Computing model. Risk of unauthorized access to sensitive data is of serious concern to the banking and financial industry. In the Private Cloud, the CIO will play a key role in security. However, running applications in the Public Cloud will result in ceding control over the protection and location of data to the service providers. In the Public Cloud, the customer has nil or limited control of the data and will have to trust the service provider. Generally, Cloud Service Providers work with several third party vendors and this has increased the potential hazard of security breaches. In many cases, users of

Public Cloud services don't know where their data is held. Data that might be secure in one jurisdiction may not be secure in another. In recent years, the banking and financial industry has been under enormous pressure, because of the increasingly complex regulatory and statutory compliance requirements as it is one of the most regulated industries. In the financial industry, improper access to data can be a major operational risk, as it directly translates to financial losses. Hence, the heightened need for CIOs from the banking sector to focus on security. Thus, the two CIOs from the banking industry have adopted the Private Cloud with dedicated data centres, which are managed by service providers.

16. Conclusion

From both the literature-based research (theory) and the qualitative case study based research (practice), it is possible to conclude that the role has continued to evolve for the past four or more decades from an (technical) operational role to tactical and to a strategic role. It is also evident the rate of change and the ubiquity of computing shows no sign of slowing. Exponential growth of mobile applications and social media are generating new data and information at an astonishing speed and unimaginable magnitude. According to Gartner reports, the Cloud Computing services market will reach over $150 billion by 2013. The Cloud model as a new technology paradigm has been widely accepted in the literature and it is fast gaining prominence and importance.

For the past five decades, the role of CIO has been an important area of IS scholarly research. Since 2005, there has been enormous interest and increasing scholarly debate and research on the emerging Cloud Computing technology. Despite this, there has been little theory based research, or empirical study that examines the impact of Cloud model on the CIO's role. Though there has been ongoing research for the past few decades and a plethora of literature on the evolution of CIO's roles, not much work has been done in the context of the emerging Cloud Computing model. This research contributes to theory by identifying the emerging roles of the CIO for the Cloud Computing era, in order to drive transformation and value creation for the enterprise.

The transformational CIO will focus on using IT resources in the Cloud to accomplish something that has never been done before. As discussed, Cloud Computing technology offers several benefits to enterprises, including lower and more predictable IT costs, pay per use, access to computing resources on-demand, less time to roll out applications, elasticity and scalability of computing resources. This will allow enterprises to accelerate their business processes and improve the time to market when introducing their products and services. Thus, Cloud Computing has removed barriers to entry and has thereby created new business models. The transformational CIO in the Cloud model will be an innovator, value enabler, change agent, and data evangelist. However, to liken the Cloud Computing model to other utilities such as electricity is over simplifying the real situation. As discussed, technical challenges and drawbacks such as latency, scalable distributed databases, security, privacy and compliance remains unresolved, hence the idea of a utility model is far from becoming a reality. From both the literature-based research and the case study, it is evident that the Cloud model can be revolutionary for start-ups and SMEs. In large enterprises, it will spawn new business models and new business opportunities by adopting new ways of applying technology to do business.

The Cloud Computing model represents a fundamental change to the economics of computing – the procurement and deployment of computing services. It is evident from the literature as well the case study, that in practice Cloud deployment often starts when organizations initially use it for part of their non-mission-critical application or as a resource for test projects. This confirms the findings of the empirical study. However, the benefits of the Cloud model is that it creates enormous economies of scale, thus bringing costs down while providing scalability and elasticity to computing resources. And it eliminates the technical complexities of designing, architecting, implementing, testing, maintaining, upgrading and migrating computing infrastructure. Given the tremendous ease, access and power it provides, it is no surprise that the Cloud model is catching on in the business world as it has a profound effect on the way IT is delivered and consumed. The Cloud model has begun to transform the information technology landscape, as they shift form procuring IT from vendors (which were considered capital expenditure and company assets) to IT services on a demand basis, based on a pay-per-use model. Cloud computing is beginning to reshape the computing paradigm and the CIO's role continues to evolve. For practitioners, this research has identified the emergence of four roles the transformational CIO should play to fully leverage from the Cloud Computing model. These roles are discussed in detail in the following section.

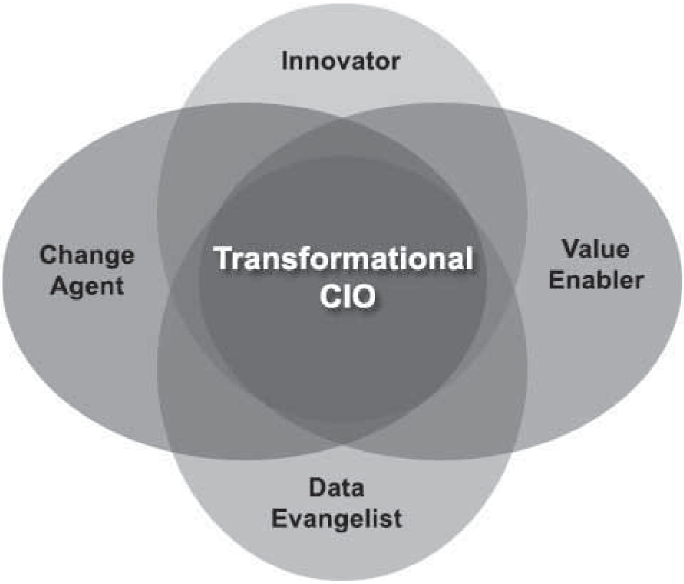
17. Evolving CIO Roles

Cloud Computing is a new paradigm, which is beginning to be widely accepted and deployed. Eight of the nine CIOs think it is evolutionary and only one CIO thinks it is disruptive. In any account, it has begun to alter the market place and business models. It has dismantled barriers to entry by eliminating huge capital investments in IT infrastructure and the need to maintain an in-house IT team to manage, support and maintain the IT investments. Furthermore, it has extended scalability and elasticity of computing on-demand and pay-per-use. The findings of the empirical study support the literature-based research.

As discussed in the earlier sections, it is evident from both the literature study and from the analysis of the case studies that the role of CIO continues to evolve and there is very little evidence to suggest the demise of the role in the near future. The findings of this research support this proposition. All nine CIOs agreed that Cloud Computing is a game changing technology and this confirms the observations of the literature review. Amongst the nine CIOs interviewed only one suggested the potential demise of the CIO on the basis that ICT will become a utility and a commodity. This observation corroborates the findings of the literature-based research where many believe that cloud model is evolutionary which will continue to transform the way we compute resulting in the evolution of CIOs role.

As such, it is safe to conclude that the CIO role is here to stay for the near future, but will continue to evolve and transform. This research has identified four key emerging roles of the transformational CIO which will become critical to be successful in the Cloud Computing paradigm: Innovator, Change Agent, Value Enabler and Data Evangelist.

18. The Emerging (New) CIO Leader



19. The Transformational CIO in the Cloud Model

The CIO who helps the enterprise to leverage from IT to constantly innovate and renew itself into an agile and smarter enterprise is a Transformational CIO. Most if not all the CIOs of today have mastered the art of leveraging ICT to achieve competitiveness. The CIO of today is a strategic executive with leadership skills and a good understanding of their organizational business needs, who is not only delivering operational excellence but leveraging ICT to optimize and create greater efficiencies in order to be competitive. Cloud Computing has begun to reshape the landscape of competition and has proven to be a commercial success – it is beginning to unleash new growth and business opportunities.

Cloud Computing is an evolutionary step forward from the traditional computing paradigm. For the near future it is far more likely that a hybrid model – a mix of Public Cloud and Private Cloud – will be the most popular Cloud Computing model. Though it is a compelling value proposition for SMEs to embrace the Public Cloud, it is evident that large enterprises and particularly for some sectors of the industry such as banking, will have to adopt a hybrid model.

Until now, the CIO's mandate has been to leverage ICT for greater efficiency, cost savings, optimization, and streamlining business operations, both inside the organization and across the entire value chain. With the emergence of Cloud Computing, the winds have begun to shift again. As discussed earlier, CIOs are still expected to drive efficiency, but they are also expected to drive innovation and value creation. IT has become the key enabler in integrating various functions such as HR, sales, marketing, inventory, product development, distribution and supply chain not only internally, but across the organization and its vendors, customer and partners. The success of business now largely depends on the success of IT. Hence, the emerging CIO, while retaining the existing roles, will be also responsible for transforming the enterprise through innovation and value creation.

20. An Innovator

One of the key strengths of Cloud Computing is that it is a catalyst for more innovation. It is evident from both literature study and case studies that in order to embrace change and to innovate, there is a need to move beyond traditional roles and responsibilities and embrace new technologies and business models to be competitive. While retaining most if not all the existing traditional roles, the CIO of tomorrow will have to expand his portfolio of skills to be successful. There is little doubt that ICT is the lifeblood of modern business activity. As such, when adopting Cloud Computing, the CIO must remain focused and retain the traditional roles as discussed in the earlier sections, including financial ROI imperatives and ICT value delivery. Often innovation and traditional roles require a balancing act. The Innovator CIO will create new opportunities, but must continue to keep the business humming, while turning the attention towards the future.

In order to innovate, CIOs must actively integrate business and IT across the entire organization and not on a piecemeal basis or in silos. Cloud

Computing frees up a lot of time spent on managing ICT infrastructure and legacy systems, ICT teams and resources, thus allowing more time to focus on innovation, business integration and value creation. The Cloud is also an ideal platform to innovate, test and incubate new ideas quickly and at a fractional cost, as the CIO does not have to invest in new infrastructure and software. The Innovator CIO will be responsible for creating plans for enhanced competitiveness of the enterprise.

21. A Value Enabler

Cloud Computing has resulted in increased technology and business agility as it allows computing services to be purchased on the fly, based on-demand and pay-per-use. Thus it has increased business responsiveness and has enabled real time data streams and information sharing. In the Cloud model, as CIOs do not have to worry much about the ICT infrastructure management, they are able to spend more time in creating better customer experiences, both for internal and external, in the products and services offered, by fully leveraging from data and information from globally available Cloud Computing infrastructure. A Savvy Value Enabler CIO will not merely respond to business requirements and customer needs, but create new services, products and experiences. In the past, it was IT that provided and fulfilled what the customer wanted. In the future, CIOs will be proactive in defining new solutions and new channels to distribute, and create superior customer experiences. As such, the CIO will play the roles of ‘influencer’, ‘advisor’, ‘articulator’ and ‘broker’ of technology and its value in order to transform the organization. In a world where technology has become increasingly pervasive and commoditized, the role of the CIO continues to evolve from technology provider to Value Enabler. For many CIOs, this represents a difficult transition from a technology guru to a business partner. In the new economy and the emerging Cloud model, we are no longer focused on productivity and bottom line, but about the top line as well.

22. A Change Agent

As discussed in the earlier sections, the emerging CIO in the Cloud will be responsible for playing the role of a Change Agent. Inevitably, because of the fear of the unknown, there will be always resistance to change. It is not enough to conceive innovative ideas and plan for innovation, but the CIO must build a solid foundation to drive innovation. And, in order to innovate and transform the business, the CIO must champion innovation within the organization, sell the (new) technology vision, sell the importance of technology innovation, educate the business executives, evoke excitement, push business and technology integration to achieve the corporate vision. The Change Agent CIO will also play the role of a mentor in motivating and cultivating internal IT talent to strive for excellence through innovation and learning new technologies and skills, and as ‘broker’ and ‘collaborator’ in managing relationships with vendors and Cloud Service Providers. In the past, CIOs have made it a priority to build internal IT skills, but there has been little effort in developing domain expertise and business acumen. Going forward, in the Cloud model, this would be essential as the job roles of internal IT are likely to evolve with the evolving role of the IT department and the CIO.

23. A Data Evangelist

Businesses depend on its customer information, appointment information, purchase orders, invoices, accounting reports, or inventory to operate. Low quality data can disrupt the business activities leading to financial losses and lost customers, as well as reputational and legal risks. Information has become the key asset of enterprises as there is a growing emphasis and reliance on information rather than technology. Enterprises also have been suffering from data silos – data residing in pools that are not integrated. Furthermore, enterprises are also beginning to suffer from data deluge. Driven by social networks and mobile applications in the Cloud, data volume continues to grow exponentially.

However, organizations have not fully leveraged from the access to data and information. The CIO must lead the initiatives to evangelize the currency of the corporate data, implement business intelligence applications and educate the business users about the value of analytical decision-making. The CIO must play a role in leveraging and gaining value from the proliferation of big data which will be increasingly stored in data centres in public and public clouds. The CIO as a Data Evangelist will be responsible for ensuring data is protected and is fully integrated into decision-making. The CIO will be responsible in answering the question: Can the users and decision makers make sense of the data? Can you make critical and strategic decisions based on the information? The **DATA** has evolved to **INFORMATION,** which is about providing **INSIGHTS** for decision-making and **TRANSFORMING** enterprises.

Furthermore, security, privacy, integrity of data, and Intellectual Property (IP) in the Private Cloud are key challenges faced by the CIO. There is growing demand for statutory compliance of data in enterprises. Given data and information will now reside across multiple applications, platforms and service providers, the CIO should play the lead role in educating about data governance and will have a greater responsibility to steer initiatives to design, implement and manage frameworks, processes and technologies to ensure that the data is secure, the integrity and privacy are maintained, and ensure that it is in compliance. The success and future of the business will depend on how well the CIO and the IT department adopt, implement and analyse data. In summary, based on theory and the empirical data from the case studies, this chapter makes the following conclusions:

	• CIO role has evolved from an operational role to that of strategic, and continues to evolve.
	• The impact and repercussions of the Cloud model is just beginning. The ways we are used to doing business based on traditional models of the past will not be able to deliver value in the future. It requires a new way of thinking, living and working.
	• Traditionally, when it comes to IT applications, CIOs had to decide whether to ‘buy’ or ‘build’ based on business requirements. The Cloud model has presented with an additional and lucrative option to ‘rent’ based on need//demand and ‘pay’ based on ‘consumption/usage’.
	• Cloud Computing model is beginning to impact the role of the CIO and the IT departments.
	• CIOs from large enterprises agree that their role will continue to evolve, and will involve more focus on innovation, business transformation, enabling value to the enterprise, data and information evangelism, as well as Change Agent, while retaining their current traditional roles.
	• From the literature review and the case studies, there is evidence that Cloud Computing is likely to lead to the demise of the IT department in Small and Medium Enterprises (SMEs). Further research focusing on SMEs is required to be conclusive. In the Cloud model, the roles played by the IT staff in large enterprises will evolve/change. The IT organization will become leaner and more business focused than its predecessor.
	• CIOs of large enterprises will play a strategic and transformational role going forward as the Cloud model matures.
	• The CIO role will now require transformational skills, in addition to traditional skills of CIOs, which include technical, organizational and strategic. The new CIO roles identified are innovator, Change Agent and Value Enabler and Data Evangelists in order to be a Transformational CIO.
	• There is no evidence to believe the demise of the CIO role will occur in large enterprises.

24. Discussion and Conclusions

In conclusion, different industries and different businesses have different strategies and different operating models. The study indicates that the Cloud is yet another key stage in the evolution of computing technology and the evolving role of the CIO. Acquiring and developing the new skill-sets and mindsets will take time. The study suggests that the Cloud model is driving change and becoming a catalyst for innovation. CIOs must make responsible and timely decisions about how (which Cloud deployment model), when, and what applications and services are to be turned to Cloud, as no one can deny that the Cloud is transforming the industry. The Cloud model enables the CIO to focus on what technology can do for the enterprise and worry less about how technology works. Because of this paradigm shift, the study indicates that new roles of CIOs are evolving, which would be more apt for the new Cloud model. Further, the study suggests that the CIO role is not necessarily threatened from moving IT services out of the organization as this is not necessarily going to make the role less important or lead to the eventual demise of their role.

The Cloud Computing technology will continue to gain prominence and importance and will change the way we do business. The Transformational CIO will continue to play a crucial role in leading the enterprise in accomplishing or even exceeding its business goals. Thus, it is in the interest of the CIO to define the future of the transformational role rather than have someone-else do it for them. The success of the CIO in the future will depend on his/her ability to anticipate and address the ongoing changes and uncertainty, in both business and technology. The CIO must be able to deal with management challenges to anticipate, initiate and to adopt change (change agent) – encourage and embrace market dynamics and technological advancements (evangelist/innovator) – i.e. to take charge of disruptive changes, not react to it. As suggested by Krotowski – CIO Global Upstream at Chevron (cited in Muller, 2011), there is a critical difference between the traditional CIO and the transformational CIO in the Cloud Computing era. The traditional CIO is a technical expert, whose primary role was keeping the lights-on as opposed to the transformational CIO, who is not necessarily a technical expert, but a strategist who will play a leadership role in transforming the organization by leveraging on technology, and will play a primary role in helping the senior executive team guide the enterprise.

24.1 Limitations and Future Research

This study was limited to CIOs from very large enterprises. There is a degree of loss of external validity because it is a case study research of just nine CIOs from large enterprises. These CIOs were of the opinion that Cloud Computing, particularly the Public Cloud is a far more appropriate and compelling value proposition to small and medium enterprises than to large enterprises. Further research, similar in nature could be conducted in order to find out from small and medium enterprises how it impacts their businesses and IT departments. Further, this study could also prove useful to investigate the notion that CIO and IT leaders of medium enterprises may become extinct as the Cloud models mature. Though the unit of analysis is an individual, the CIO, the case study research could have involved a cross section of the enterprise. As such, another potential similar study can also be conducted in order to investigate how the business leaders (CEOs) perceive the emergence of the Cloud Computing model, its impact on the enterprise and on the roles of CIOs. Also, this chapter did not concentrate on any specific industry. Perhaps future research can be done to examine the configuration of roles for the transformational CIO in selected industries where Cloud Computing is expected to have the greatest impact, as well as research to investigate whether there are any differences between industries. Lastly, large scale quantitative research across different industries and geographies would complement and further advance the research conducted in this study to investigate the roles of the transformational CIO.

References

Andrews, P & Carlson, T 1997, 'The CIO is the CEO of the Future', CIO Conference, Naples, Florida, <http://www.cio.com/conferences/eds/sld018.htm>.

Applegate, LM & Elam, JJ 1992, 'New information systems leaders: A changing role in a changing world', Journal of MIS Quarterly, vol. 16, no.4, pp. 469-490.

Bassellier, G, & Benbasat, I 2004, 'Business competence of information technology professionals: conceptual development and influence on IT-business partnerships', Journal of MIS Quarterly, vol. 28, no. 4, pp. 673-694.

Beatty, RC, Arnett, KP & Liu, C 2005, 'CIO/CTO Job Roles: An Emerging Organizational Model', Communications of the IIMA, vol. 5, no.2.

Benjamin, RI, Dickinson, C Jr. & Rockart, JF 1985, ‘Changing Roe of the Corporate Information Systems Officer’, Journal of MIS Quarterly, vol. 9, no.3, pp. 177-188.
Broadbent, M 2002, ‘CIO Futures – Lead with Effective Governance’, Refereed proceedings of the ICA 36th Conference, Singapore, October 2002.
Broadbent, M & Kitzis, E 2005, ‘The new CIO leader: setting the agenda and delivering results’, Harvard Business School Publishing, Massachusetts.
Brynjolfsson, E, Hofmann, P & Jordan, J 2010, ‘Economic and Business dimensions and Electricity: Beyond the Utility Model’, View Points, vol. 53. no.5, pp.32-34 DOI:10.1145/1735223.1735234.
Carr, NG 2003, ‘IT Doesn't Matter’, Harvard Business Review May 2003, Harvard Business School Publishing Corporation.
Carr, NG 2005, ‘The End of Corporate Computing’ Journal of MIT Sloan Management Review, vol.46, no.3, pp.67-73.
Chang, JCJ & King, WR 2005, ‘Measuring the performance of information systems: A functional scorecard’, Journal of Management Information Systems, vol.22, no.1, pp. 85-115.
Chen, DQ & Preston DS 2007, ‘Understanding CIO Role Effectiveness: The Antecedents and Consequents’, Refereed proceedings of the 40th Annual Hawaii International Conference on System Sciences, 2007.
CMA Magazine 1998, ‘CEOs unable to justify IT spending’, June 1998, vol. 72, no. 5, pp. 28-29.
Christensen, CM 2005, ‘The Ongoing Process of Building a Theory of Disruption’, Journal of Product Innovation Management, January 2006, vol. 23, no. 1, pp. 39-55.
De Lisi, P, Danielson, R & Posner, B 1998, ‘A CEO’s-Eye View of the IT Function’, Business Horizons, vol. 41, no. 1, pp. 65-74.
Dooley, JA, Jones, SC & Iverson, D 2012, ‘Web 2.0 Adoption and User Characteristics’, Web Journal of Mass Communication Research, June 2012.
Earl, MJ & Feeny, DF 1995, ‘Is your CEO adding Value?’ The Mckinsey Quarterly, 1995 Number 2, pp.144-161.
Evans, N 2012, ‘2012 IT Predictions: Disruptive IT Trends Move Into the Fabric of Mission-Critical Systems’, viewed 27 June 2012, http://blogs.unisys.com/disruptiveittrends/2012/01/17/2012-it-predictions-disruptive-it-trends-move-into-the-fabric-of-mission-critical-systems/.
Feeny, DF, Edwards, BR & Simpson, KM 1992, ‘Understanding the CEO/CIO Relationship’, MIS Quarterly, vol. 16. no.4, pp. 435-448.

Gartner Dataquest's Statistics 2012, ‘How many computers are there in the world?’ 8 September 2011, viewed on 12 July 2012, http://www.worldometers.info/computers/

Goldstein, P 2008, ‘The Tower, the Cloud, and the IT Leader and Workforce’, pp. 238-261, viewed 12 September 2012. http://net.educause.edu/ir/library/pdf/PUB7202x.pdf

Gottschalk, P & Taylor, NJ 2000, ‘Strategic Management of IS/IT Functions: The Role of the CIO’, Refereed proceedings of the 33rd Hawaii International Conference on System Sciences-2000, vol.7, pp. 7066.

Grover, V, Jeong, SR, Kettinger, WJ & Lee, CC 1993,‘The Chief Information Officer: A Study of Managerial Roles’, Journal of Management Information Systems, vol. 10, no. 2, pp. 107-130.

Henderson, JC, Kulatilaka, N, Venkatraman, N & Freedman, J 2003, ‘Riding the Waves of Emerging Technologies: Opportunities and Challenges for the CIO’, Society of Information Management (SIM), April 2003.

Hodgson, L & Lane, MS 2010, What Are the Key Job Functions and Critical Competencies Required for the Role of the CIO in Achieving Superior Organizational Performance?’ Refereed proceedings of Informing Science & IT Education Conference (InSITE,) 2010, vol 10, pp. 257-266.

Hofmann, P 2010, ‘The Limits of Public Clouds for Business Applications - An overly simplistic reliance on the utility model risks blinding us to the real opportunities and challenges of cloud computing’, The 2010 International Conference on E-Business Intelligence, Atlantic Press 2010.

Hofmann, P & Woods, D 2010, ‘Cloud Computing: The Limits of Public Clouds for Business Applications’, IEEE Computer Society, vol.14, no. 6, pp. 90-93.

Hunter, G 2010, ‘The Chief Information Officer: A Review of the Role’, Journal of Information, Information Technology, and Organizations, vol. 5, pp.126-143.

Karahanna, E, & Watson, RT 2006, ‘Information Systems Leadership’, IEEE Transactions on Engineering Management, vol. 53, no. 2, pp 171-176.

Laplante, P A & Bain, DM 2005, 'The Changing Role of the CIO: Why IT Still Matters', IT Professional, vol. 7, no. 3, pp. 45-49.

Maes, R 2007, ‘An integrative perspective on information management’, in A. Huizing & E.J. De Vries (eds), Information Management: setting the scene, Oxford, UK; Elsevier, pp. 11-26.

McDaniel, E 2011, 'Making the Case for to Business Leaders', viewed on 28 September 2012, http://c4i.gmu.edu/eventsInfo/reviews/2011/papers/15-McDaniel-paper.pdf

Menzel, M, Schonherr, M, Nimis, J & Tai, ST 2010, '$(MC^2)^2$: A Generic Decision-Making Framework and its Application to Cloud Computing', International Conference on and Virtualization (CCV 2010), Singapore, May 2010.

Milroy, A 2012, 'BLOG: The incredibly shrinking IT department, viewed 4 October 2012, http://www.cio-asia.com/resource/featured-blogs/blog-the-incredibly-shrinking-it-department/

Muller, H 2012, On Top of the Cloud: How CIOs Leverage New Technologies to Drive Change and Build Value Across the Enterprise, John Wiley & Sons, Inc., Hoboken, New Jersey.

Onan, M & Gambill, S 2001, 'Small Company CIOs', Journal of Computer Information Systems, vol. 42, no. 2, p.90.

Peppard, J 2010, 'The New Conversation of the Chief Information Officer (CIO)', 15th February, version 1.1, pp. 1-10 <http://www.som.cranfield.ac.uk/som/dinamic-content/media/ISRC/The%20new%20conversations%20of%20the%20CIO.pdf>

Peppard, J 2010, 'Unlocking the Performance of the Chief Information Officer (CIO)', California Management Review, vol. 52, no. 4, pp.73-99.

Peppard, J & Breu, K 2003, 'Beyond Alignment, A Coevolutionary View of The Information Systems Strategy Process' , Refereed proceeding of the Twenty-Fourth International Conference on Information Systems 2003, pp. 743-750.

Peppard, J, Edwards, C & Lambert, R 2011, 'Clarifying the Ambiguous Role of the CIO', MIS Quarterly Executive, vol. 10, no.1, pp.31-44.

Periasamy, KP & Seow, A 1998, 'CIO: Business Executive or Technical Expert?', SEARCC'98 Conference paper, National University of Singapore and Housing Development Board, Singapore

Radjou, N 2008, 'Why Indian CIOs Excel at Driving Business Innovation', HBR Blog Network, Harvard Business Review, viewed on 27 June 2012, < https://hbr.org/2008/05/why-indian-cios-excel-at-drivi/ >.

Rehman, MS & Sakr, MF 2011, 'Teaching the cloud-experiences in designing and teaching an undergraduate-level course in cloud computing at the Carnegie Mellon University in Qatar', Refereed

proceedings of Global Engineering Education Conference (EDUCON), 2011 IEEE, pp 875-879.
Remenyi, D, Grant, KA & Pather, S 2005, 'The chameleon: a metaphor for the Chief Information Officer', Journal of General Management, Spring 2005, vol. 30, no. 3, pp 1-11.
Robbins, S & Pappas, A 2003, 'Within and Beyond: Understanding the Role of the CIO', in D Lane (ed), CIO Wisdom: Best Practices from Silicon Valley's Leading IT Experts, Prentice Hall PTR, Upper Saddle River, NJ, Chapter 1, pp. 1-16.
Rockart, JF & Treacy, ME 1980, Executive Information Support Systems, Center for Information Systems Research, MIT, Sloan School of Management, pp. 1-42.
Ross, JW & Feeny, DF 1999, The Evolving Role of the CIO, Chapter 19, Center for Information Systems Research, Sloan School of Management, Massachusetts Institute of Technology, pp. 385-402.
Smaltz, DH 2000, 'The Elevation of CIO Roles: Organizational Barriers and Organizational Enablers', Journal of Healthcare Information Management, vol. 14, no.1. pp. 81-91.
Stephens, CS, Ledbetter, WN, Mitra, A & Ford, FN 1992, 'Executive or Functional Manager? The Nature of the CIO's Job', Journal of MIS Quarterly, vol. 16. no.4, pp. 449-467.
Teo, TSH, Wong, PK & Chia, EH 2000, 'Information technology (IT) investment and the role of a firm: an exploratory study ', International Journal of Information Management 20 (2000), pp. 269-286.
Weiss, JW & Anderson, D 2004, 'Aligning technology and business strategy: issues & frameworks, a field study of 15 companies', Refereed proceedings of the 37th Annual Hawaii International Conference 2004, vol. 8, pp. 80220.3.
Weiss, JW & Anderson, D 2004, 'CIOs and IT Professionals as Change Agents, Risk and Stakeholder Managers: A Field Study', Engineering Management Journal, vol. 16, No. 2, pp. 13.
Wood, L 2011, 'Emerging Technology', smartenterprisemag.com, viewed 21 June 2011, <http://www.smartenterprisemag.com/showArticle.jhtml;jsessionid=WF5HBCKDV51M>.
Wyld, DC 2009, 'The Utility of Cloud Computing As A New Pricing - and Consumption - Model for Information Technology', International Journal of Database Management Systems (IJDMS), vol. 1. no.1, pp. 1-20.
Yin, RK 1994, 'Discovering the Future of the Case Study Method in Evaluation Research', Evaluation Practice, vol. 15, no.3, pp. 283-290.

PROCEDURAL 3D CAVES, CLOUDS AND ARCHITECTURE GENERATION METHOD BASED ON SHAPE GRAMMAR AND MORPHING

TOMASZ ZAWADSKI, SLAWOMIR NIKIEL AND ERALDO RIBEIRO

Abstract

Complex dependencies of a scene structure, and the intricate geometry of models, make modeling of large-scale virtual environments a challenging task, which is complicated further by the compulsion of having to manually create the geometry to ensure uniqueness of objects. Although the process can be greatly simplified by employing multiplied geometric primitives, automated or semi-automated procedural shape synthesis pose extremely interesting alternatives that can create a wide variety of procedurally generated objects, which are admirable under stochastic conditions. However, the resultant output could possess unsatisfactory features, e.g. awkwardness of geometry, unrecognizably deformed or biased shapes, or undesirable visibility of artifacts. The present study aims to gain control over the process of procedural synthesis, by designing and developing a series of presets for retaining differently appearing shapes within the same topological set. The present investigation entails a new approach to a procedural three-dimensional (3D) model-construction algorithm that combines the benefits of both discrete and continuous modeling perspectives. The algorithm models in the study blend scene components such as caves, architectural structures and clouds, by amalgamating the discrete descriptiveness of shape grammars with the continuous flexibility of shape morphing. This step evolves a modeling approach that can be regulated by a morphing parameter to yield several types of geometry. Though the research effort focuses on describing the algorithm, it also enumerates results from several challenging models that can contribute to the generation of complex virtual scene components.

Keywords: algorithms, procedural modeling, shape grammar, computer graphics, feature-based models, morphing.

1. Introduction

Complex structures are necessary elements of visually convincing virtual scenes. Systems based on automated shape construction can be used for the modeling of buildings (Wonka *et al.,* 2003), whole urban structures (Parish & Muller, 2001; Greuter *et al*., 2003), terrains (Peytavie *et al*., 2009), clouds (Bouthors & Neyret, 2004; Schpok *et al*., 2003), plants (Prusinkiewicz & Lindenmayer, 1991; Prusinkiewicz, Lindenmayer & Hanan, 1988), or caves (Am Ende, 2001; Boggus & Crawfis, 2009; Boggus & Crawfis, 2010; Schuchardt & Bowman, 2007; Johnson *et al*., 2010). Algorithms that enable full automation of the modeling process are both time-saving and econoimical to the digital media production process. Procedural systems are also used in CAD setups such as: City Engine (procedural cities), Houdini (procedural animation), Terragen (procedural landscapes), and Art of Illusion (procedural textures). There is a constant development of new methods, and a merging of technology and dynamical systems (Clempner & Poznyak, 2011; Si Trapani & Inanc, 2010). The challenge of automated shape modeling constitutes an important aspect of computer graphics activity, and has attracted the attention of the digital media industry for several years. Digital movies have created an insatiable demand for spectacular visual effects in 3-D graphics. Besides entertainment applications, shape modeling has several practical uses ranging from CAD engineering applications, through scientific visualization, to advanced game programing and Virtual Environments. Meeting the ever-growing demand in this field becomes next to impossible without the use of procedural modeling systems, as long as real-time simulations are considered. The present undertaking aims at proposing to extend the set of currently available procedural methods, with a hybrid of shape grammar and morphing, thereby offering improved performance and greater versatility.

2. Review of Literature

2.1. Virtual Caves

Existing efforts in the virtual construction and visualization of 3D cave structures include the use of scanning hardware to obtain accurate spatial data of original cave structures (Am Ende, 2001), which can be used to visually reconstruct the geometry of a real cave. This physical approach to 3D mapping of caves is, however, an extremely laborious and time-consuming process. Schuchardt and Bowman (2007) employed traditional means like 2D cave maps to test whether immersive virtual reality provided better spatial understanding of structures that defy mental visualization; the cave model followed in their work was derived from the cave survey and measurement data obtained in an actual cave.

The procedural creation of fully synthetic 3D cave models has also been researched by Boggus and Crawfis (2009). Their investigation focused on the procedural generation of solution caves that are formed as a result of dissolution of rock by acidic water. These researchers applied knowledge about the formation of solution caves to create cave models for virtual environments. Their method involved approximating water transport to develop a coarse level of detail models for a cave passage (Boggus & Crawfis, 2009; Boggus & Crawfis, 2010). They also demonstrated ways of generating 3D cave models with cave patterns, and suggested the use of techniques like bump mapping and displacement mapping for mapping surface detail. Johnson *et al*. (2010) examined the possibility of using a cellular automata-based algorithm for the real-time generation of 2D infinite cave maps, which could represent cave levels in video games, but left the generation of 3D cave mapping for the future. Peytavie *et al*. (2009) presented a framework for representing complex terrains including caves, using a volumetric discrete data-structure (Fig. 2.1). Additionally, they proposed a procedural rock generation technique to automatically generate complex piles of rocks; their aim was to display physically plausible scenes created without physically-based simulations. Their approach mainly emphasized a unique data-structure for efficient sculpting, editing and reconstruction in customized high-level terrain authoring tools, as opposed to a purely procedurally driven approach.

Fig. 2.1. Left: a canyon with boulders detached from the cliffs; Middle: a rock arch, Right: a cave

2.2. Virtual Cities and Buildings

Greuter *et al*. (2003) worked on the creation of virtual cities, for which an algorithm was designed to simplify scene elements and generate speed. His method succeeded in creating a variety of buildings made of simple solids, where details of windows and doors were portrayed with differing image textures, which could be freely explored in first-person perspective. Buildings and streets were created on-the-fly according to the observer's current field of vision, and the shapes of the buildings were determined by their location on the mesh (Greuter *et al*., 2003). The design involved a purely procedural approach, used pseudo-random numbers and special algorithms, and resulted in a high degree of building shape diversity, flexibility and a pseudo-infinite virtual world (Fig. 2.2).

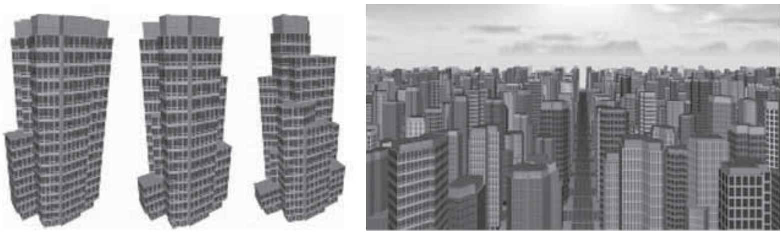


Fig. 2.2. Left to right: Buildings generated by Greuter (steps 1, 4, and 8); Right: generated city

Another example of the formalism to generate virtual urban areas was presented by Parish and Müller (2001). The system they evolved could create a virtual city by using a relatively small amount of statistical and geographical input. Moreover, the user exerted a high degree of control. City creation was limited to generating roads and simple buildings without additional detail. The map of the city was derived from geographical data that took the form of aerial photography. The system consisted of several different tools that helped to make the urban agglomeration a step by step procedure, as shown below in Figs. 2.3 and 2.4 (Parish & Muller, 2001).

Fig. 2.3. Illustration of building generation by the Mueller-Parish method.

Fig. 2.4. City generationusing the Parish-Mueller method, where the distribution of buildings is meaningful.

Wonka *et al* (2003) developed an alternative method of generation based on the grammars of division, or, to be more explicit, a novel parameterized grammar technique founded on general shape grammars. Matching attributes system and control grammar were developed which offered a diversity of styles and building types. The Müller and Greuter method yielded buildings that were randomly generated solids covered with textures, whereas Wonka *et al*. (2003) designed a grammar that resulted not only in solids, but also produced geometric details such as windows, doors, roofs and so on. The idea of PL-systems was abandoned by him, since these were based on the process of growth in open space. Such buildings imposed a lot of restrictions on space and location. The main difference between Wonka systems and previously mentioned approaches is the availability of a huge database of grammatical rules for implementation, thereby eliminating the need for individual grammars for each building. The grammar described in their work was complicated by the unclear selection of possible matching rules (Fig. 2.5).

Fig. 2.5. Buildings generated by Wonka from the same sort of rules with different spacing.

2.3. Procedural Clouds

The methods followed for the representation of cloud shapes are few and far between, and include the following: volumetric clouds, either explicit or implicit (Nishita *et al*., 1996), procedural clouds (Ebert, 1997; Schpok *et al*., 2003), billboards (Harris & Lastra, 2001; Dobashi *et al*., 2000), and surfaces (Gardner, 1984; Gardner, 1985). Volume approaches are yet unsuitable for real-time rendering of clouds in spite of being fast, because the resolution required for a cloudy sky, which is both large and very detailed, will before long be too demanding for graphics hardware. Billboards and planes are the current popular solutions in computer games (Harris & Lastra, 2001), but need a small number of slices to keep an interactive pace -- the problem is in the huge number of pixels rendered rather than in the number of polygons. The surfaces permit very efficient rendering, but look too crude to represent ethereal cloud shapes. Nevertheless, Gardner was able to make ellipsoid surfaces look volumetric in the scope of ray-tracing, thanks to view dependent transparency shaders evaluated at each ray. Fortunately, recent hardware has *pixels shader* functionality, making this approach feasible for real-time rendering. An early attempt in that direction was made by Elinas and Sturzlinger, (2000). Another approach to the modeling of clouds was the method based on quasi-spherical particles or blobs (Bouthors & Neyret (2004) (Fig. 2.6).

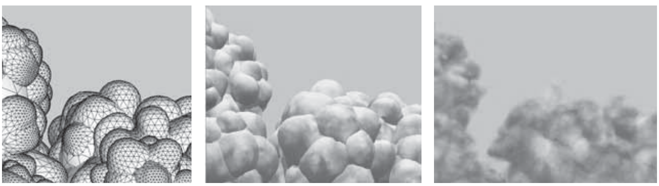


Fig. 2.6. Left: a zoom on the mesh of the generated blobs; Middle: the same model with the Perlin texture; Right: the same model with the Gardner shader.

2.4. Shape Grammars

Stiny and Gips are forerunners of the shape grammars. Their work used a "linguistic model of the generational system" to support the design process (Stiny & Gips, 1972; Stiny, 1975). Definition 1 is a basic formula of shape grammars that is analogous to the one of formal grammars. It is graphically expressed in terms of words composed of symbols and a set of grammatical rules called productions (Stiny, 1980).

Definition 1. Shape grammars are defined as follows:

$$SG = < V_T, V_M, R, I > \tag{2.1}$$

where
$\boldsymbol{V_T}$ is a set of terminal shapes,
$\boldsymbol{V_M}$ is a set of non-terminal shapes ($V_T^* \cap V_M = \emptyset$),
$\boldsymbol{R}$ is a finite set of ordered pairs *(u, v),* and
$\boldsymbol{I}$ is an initial shape.

Pairs *(u, v)* are such that *u* is a shape consisting of an element of V_T* combined with an element of V_M, and *v* is a shape consisting of (a) the element of V_T* contained in *u,* or (b) the element of V_T* contained in *u* combined with an element of V_M, or (c) the element of V_T* contained in *u* combined with an additional element of V_T* and an element of V_M.

2.5. Morphing

The idea of shape metamorphosis, commonly known as morphing, forms a wide and important area of computer graphics. Generally, morphing can be defined as both a continuous (i.e., over time) and smooth process of transformation of one shape into another. So-called key shapes (by analogy to key-framed animation) may have different topologies, and transformation smoothness does not have to display homeomorphism (Martyn, 2004). Typically, morphing presents problems at two stages: firstly, it requires identification of ‘features’ of key shapes which are to be morphed, and secondly, the interpolation of shapes according to ‘trajectories’. The second step results in intermediate shapes that have some topological characteristics of both the key shapes i.e., the ‘beginning’ and the ‘final’ shapes. The present paper makes use of morphing intermediate shapes as an alternative to CSG Boolean productions e.g., the ‘beginning’ *and* ‘final’ shapes.

Several methods exist for metamorphosis of 2D shapes, represented largely by polygons (Alexa, Cohen-Or & Levin, 2000), and to a lesser extent, by images (Wolberg, 1998). The problem has also been investigated in a 3D domain, and according to Lazarus and Verrous (1998) they comprise procedures based on polygonal mesh representation (Kent, Carlson & Parent, 1992; Lee *et al*., 1999), as well as modes using voxel representations (Turk, 1999).

3. SG-M Hybrid Algorithm

The present research study employed a hybrid method, which combined two independent approaches (a) a hierarchical shape representation, and (b) a set of operations that were applied to the shapes (i.e., shape rules). A continuous morphic shape rule was incorporated in addition to discrete shape operations.
The representation consisted of three main elements: the *root* – a place where the modeling process ends; the *nodes* – which store information about actual shape appearance; and the *leaves* – which determine the possibility of rules application (Fig. 3.1).

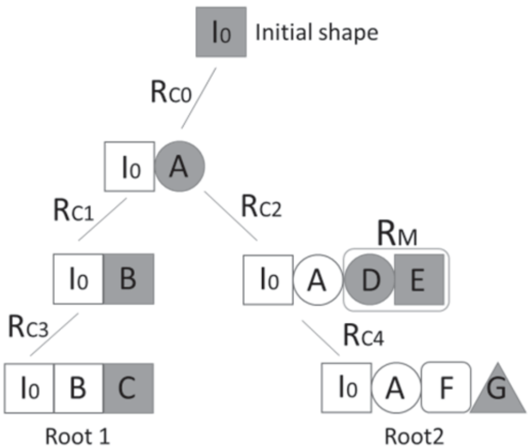


Fig. 3.1. Hierarchical representation with two possible modeling ways.
Specific terminal shapes are shown in red. Non-terminal shapes are shown in blue.

The example shown in Fig. 3.1 can be described as follows:

$$R_0: I_0 \Rightarrow I_0 \cup \mathrm{A} \tag{3.1}$$

$$R_1: I_0 \cup \mathrm{A} \Rightarrow I_0 \cup \mathrm{B} \quad (3.2)$$

$$R_2: I_0 \cup \mathrm{A} \Rightarrow I_0 \cup \mathrm{A} \cup (D\ morph\ \mathrm{E}) \quad (3.3)$$

$$R_3: I_0 \cup \mathrm{B} \Rightarrow I_0 \cup \mathrm{B} \cup \mathrm{C} \quad (3.4)$$

$$R_4: I_0 \cup \mathrm{A} \cup (D\ morph\ \mathrm{E}) \Rightarrow I_0 \cup \mathrm{A} \cup \mathrm{F} \cup \mathrm{G} \quad (3.5)$$

$$R_0 \Rightarrow R_1 \Rightarrow R_3 \quad (3.6)$$

$$R_0 \Rightarrow R_2 \Rightarrow R_4 \quad (3.7)$$

Shape rules contain Boolean operations such as sum, difference or union, which are termed *classic shape rules* (R_C). Another type of rule, known as the morphic shape rule (R_M) is provided, and the latter works by morphing two input shapes into one output shape with the help of a linear-interpolation parameter (shown in Fig. 3.2). Fig. 3.3 is a flow diagram of the modeling process. The investigation revealed that matching rules could be applied only to non-terminal shapes that were recognized during the modeling process (Fig. 3.4).

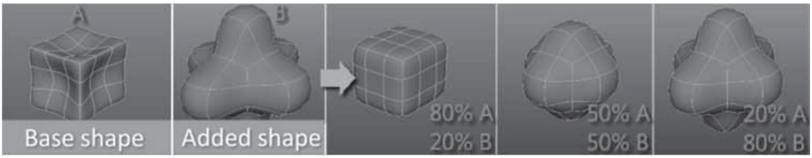


Fig. 3.2 An example of morphing rule (R_M) for another contribution of shape A and B

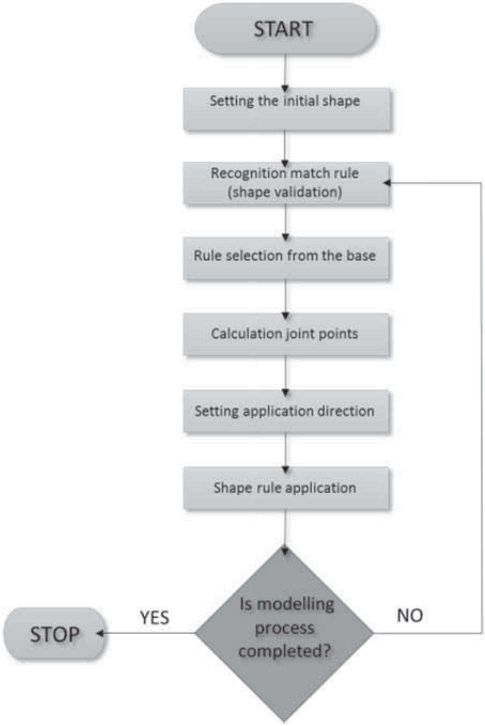


Fig. 3.3 Main SG-M block diagram.

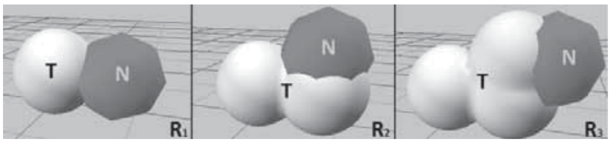


Fig. 3.4 Terminal and non-terminal shapes on simple rule example: sphere $_{(T)}$ ∪ sphere $_{(N)}$ ⇒ sphere $_{(T)}$ ∪ sphere $_{(T)}$ ∪ sphere $_{(N)}$ directly from editor

3.1. Selective Algorithms for Caves, Clouds and Architecture Modeling

The results produced by the SG-M algorithm were random. Minor changes were necessary to restrict them to a specific class of structures. The alterations affect the selection of appropriate shapes, rules for modeling mode, the combining method and the range of used operations (Fig 3.5).

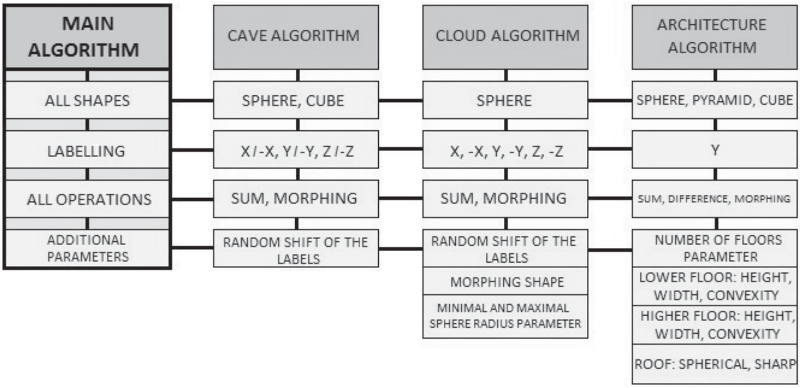


Fig. 3.5 Selective algorithms based on the main shape grammar and a morphing hybrid oriented towards caves, clouds and architecture modeling

3.2. Procedural cave generation using the SG-M concept

The number of iterations was equal to the sum of all rules R_C and R_M. In each rule, the scene base shape was recognized and subsequently added to another shape formed by morphing a sphere and a cube (Fig. 3.6).

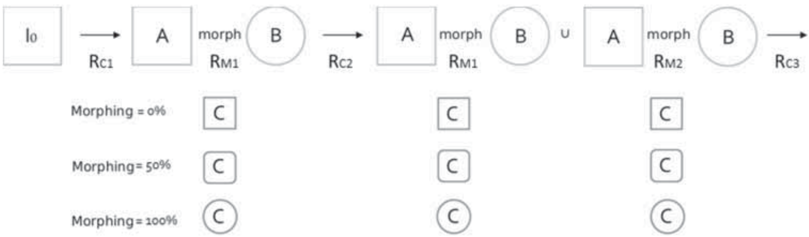


Fig. 3.6 Application of shape rules and controlling the modeling process by morphing the parameter in cave mode. Letters A, B, C define shapes

Fig. 3.7 Controlling the modelling process using the S_J parameter – shift of the joint points for shapes.

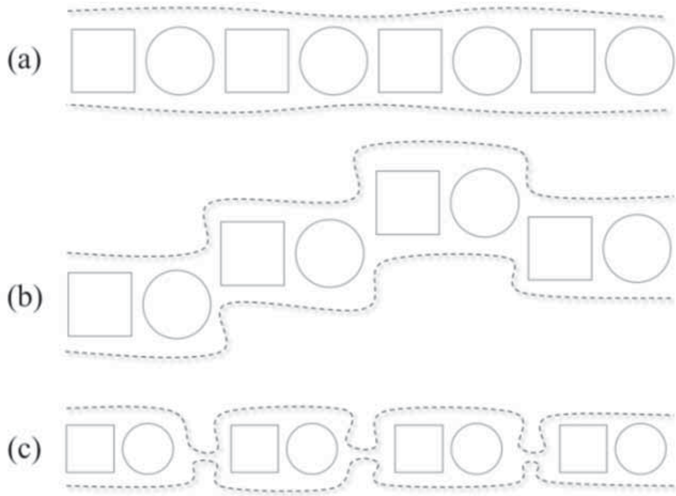


Fig. 3.8. S_J parameter characteristic: (a) regular cave, (b) non-regular cave in Y axis, (c) non-regular cave in Y or Z axis.

It is possible to select the direction of the joint points under which new rules will be added (+X/-X, +Y/-Y, +Z/-Z). It is also possible to control the S_J parameter (Figs. 3.7 and 3.8).

3.3. Procedural Clouds Generation using the SG-M Concept

Cloud modeling can be performed by summing many spheres in successive rules constrained to specified directions. The present method created base shapes and shape rules that were to be used during the modeling process (Fig. 3.9).

The number of rules was calculated as the maximum of the value *Max* for each direction, i.e. Max (X, -X, Y, -Y, Z, -Z). For each rule, all *Max* values were reduced by 1. Next, new connections were created for those rules that had a positive value. In the *nth* rule, n spheres were added to the object. The radius of the sphere in the *nth* iteration was interpolated between the initial and the final radius. In this step, the amount of required shape rules was determined. For shapes connecting points, some random offset was introduced. After performing all classic rules, the morphic rule was used to morph the structure with the selected shape i.e., sphere, torus) (Fig. 3.10).

3.4. Procedural Architecture Generation using the SG-M Concept

First of all, it was assumed that the buildings were multistoried. The process was initiated by setting the first rule for the ground floor (which was the initial shape); the last rule was laid down for the topmost floor. Intermediate rules were created using linear interpolation. For each level, characteristics such as height, width, and convexity were determined, e.g., a floor was a cube that may be convex (in which case four sum rules was applied), or concave (here, four difference rules was applicable). After the application of the last rule, modeling of the roof shape was achieved by morphing from a sharp shape (i.e., a pyramid) to a spherical shape (i.e., sphere). All modeling directions were selected automatically according to basic building principles, i.e., axes –X/X, -Z/Z were selected to adding/cutting sides, axis +Y was chosen for building up floors and adding a roof (Fig. 3.11).

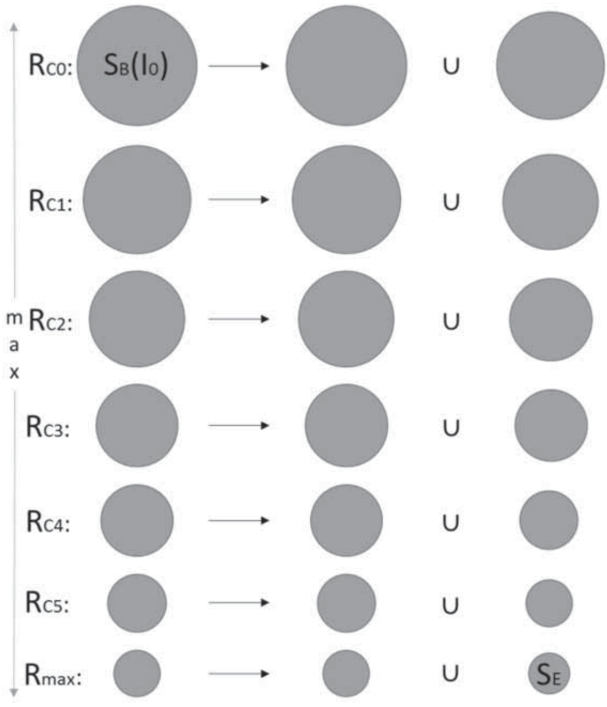


Fig. 3.9 Rule-creation process using the *Max* parameter for a *specified axis (+X, +Y, +Z, -X, -Y, -Z)*. S_B defines shape begin radius; S_E defines shape end radius.

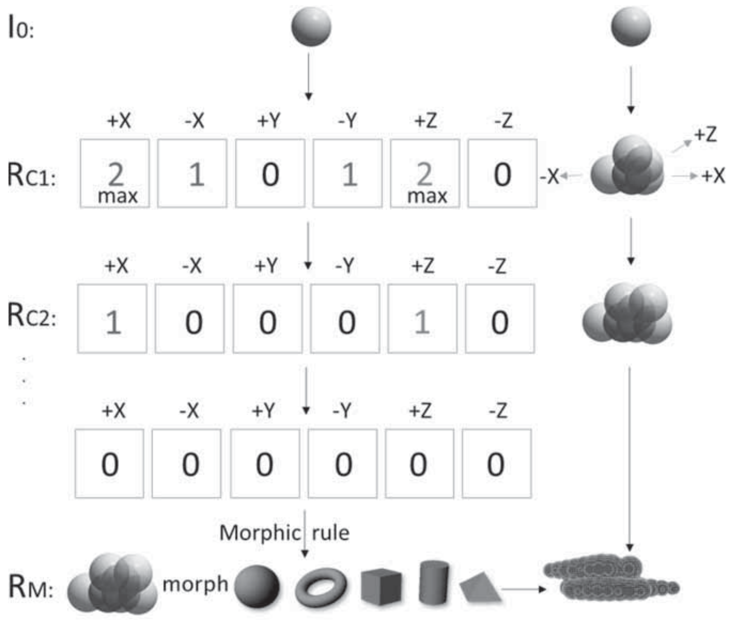


Fig. 3.10. Cloud-modelling algorithm. Colours show directions of adding spheres.

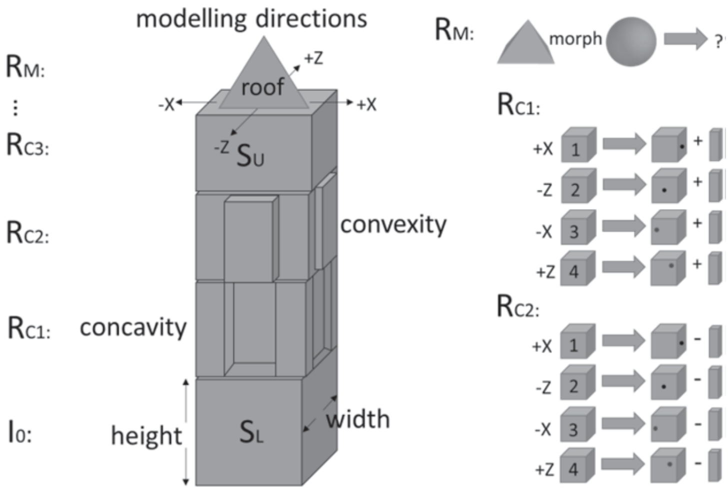


Fig. 3.11. Multistorey architecture construction with possible application rules. S_L defines the upper floor, S_L defines the lower floor.

3.5. Functional Description of Three-Dimensional Shapes

An alternative functional description of shapes was proposed along the following lines. In contrast to classical algorithms, emphasis was not placed on a polygon mesh object. Instead, consideration was given to the actual function that described space where solids were located. The algorithm performed operations only on functions or scalar functions describing the field for selected shapes.

Definition 2. A subset $A \subset R^n$ (in our case R^3) is called an implicit object if there exists a function f: $U \rightarrow R^k$, $A \subset U$, and a subset $V \subset R^k$, such that $A = f^{-1}(V)$ (Velho, Gomes and Figueiredo, 2002):

$$A = \{P \in U : f(P) \in V\} \tag{3.8}$$

Definition 3. The distance from a point p to a surface M in R^n (in our case R^3) is the minimum of the Euclidean distance d_E (P, s), where $s \in M$ [30]:

$$d(P, M) = \inf_{s \in S} d_E(P, s) \tag{3.9}$$

Definition 4. When f is a real-valued function, that is k = 1, then f is a point-membership classification function that returns a value according to the relationship of a point $P = (x_1, \ldots, x_n)$, given as its argument, with the implicit object A defined by f (Velho, Gomes and Figueiredo, 2002): i.e.,

$$\boldsymbol{f(P)} \begin{cases} > 0 \quad then\ P \notin A\ (point\ is\ outside\ surface) \\ = \mathbf{0}, \qquad then\ P \in A\ (point\ is\ on\ surface) \\ < 0, \quad then\ P \in A\ (point\ is\ inside\ surface) \end{cases} \tag{3.10}$$

Primitive shapes like the cube, sphere, torus and cylinder were defined using standard implicit objects based on functional description (Velho, Gomes & Figueiredo, 2002):

A sphere *f* is given by:

$$f(P) = P.x^2 + P.y^2 + P.z^2 - r \ , \tag{3.11}$$

A cuboid *f* is given by:

$$f(P) = \max(P.x - S.x, P.y - S.y, P.z - S.z) \tag{3.12}$$

A torus *f* is given by:

$$f(P) = (P.x^2 + P.y^2 + P.z^2 - (a^2 + b^2))^2 - 4 * a^2(b^2 - P.y^2) \tag{3.13}$$

A cylinder *f* is given by:

$$f(P) = \max(P.x^2 + P.y^2 - r, P.y - S.y) \tag{3.14}$$

A cone *f* is given by:

$$f(P) = P.x^2 + P.z^2 - r \tag{3.15}$$

where:
P is a point in $R^{3,}$
S.x, S.y, S.z is an extension in the x, y, z axes,
r is the sphere, cylinder and cone radius,
a is a top torus radius,
b is a bottom torus radius.

Definition 5. An implicit CSG solid is defined by any set of points in R^n (in our case R^3) that satisfies $F(x) \leq 0$ for some $F \in S_j$. (Velho, Gomes and Figueiredo, 2002). An example for a sum operation is given by:

$$f(P) = f \bigcup g = \max(f(P), g(P)) \tag{3.16}$$

A morphing controlled by a morphing parameter a is defined by the following linear interpolation:

$$f(P) = f * g = f(P) * (1 - a) + g(P) * a \tag{3.17}$$

Shapes are described by the final function as a composite of the above functions.

3.6. Joint Points, Bonds, and Directions

An important problem that needed to be solved was the determination of the joint points of the grid for the main shapes, and for the objects created during the modeling process.

Definition 6. The joint point J defines the way in which one shape can be combined with another. The joint point is described as a point P and a direction D in 3D space:

$$J = (P, D) \tag{3.18}$$

The *bond* B combines two shapes into one by using appropriately selected joint points. Two bonds make a joint point for overlapping points *P* and opposite directions *D*. This can be seen as the "gluing" of two walls facing each other in opposite directions. This operation helps avoid unnecessary connections (Fig. 3.12). The bond is given by:

$$B = (J_1, J_2) \Leftrightarrow (J_1.P = J_2.P \wedge J_1.D = -J_2.D) \tag{3.19}$$

where *B* is the bond, J_1, J_2 are joining points, *P* is a point in R^3, and *D* is a direction for the shape. If the bond does not have a direction, i.e., when *D* = (0, 0, 0), then a joint point can be created with any other point.

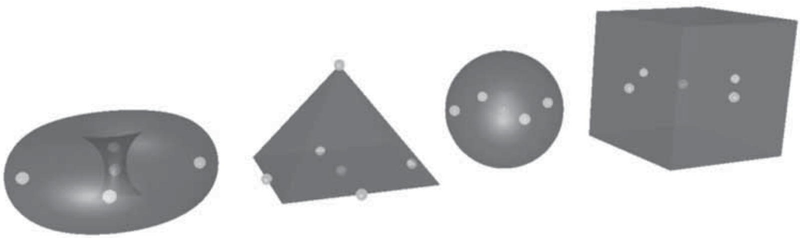


Fig. 3.12 Joint points for base shapes. Red colour shows possibility to join with central point of the shape; green colour shows other possible connections.

3.7. Grid Display Algorithm – Marching Cubes

The object surface can be described conceptually by a function called *density function*. For a point $P \in$ 3D, this function produces a single floating-point value. Density function can be either positive (when P is located inside the solid), negative (when P is located in the empty space), or zero (zero values are indicators of the surface of the object). To explain further, the difference between the locations of density function zero and the boundary between positive and negative values, forms the surface of the solid.

With the objective of constructing a polygonal mesh that spanned the object surface, GPU was used to generate polygons for a "block" of structures at a time. The blocks were further subdivided into 32x32x32 smaller cells, or voxels. Polygons (triangles) were constructed within these voxels to represent the solid surface. Application of the marching-cubes algorithm (Lorensen & Cline, 1987; Dyken *et al*., 2008; Goetz, Junklewicz & Domik, 2005) resulted in the generation of polygons within a single voxel, given as input the density value at its eight corners. As the output, this algorithm produced anywhere from zero to five polygons. An indication that the cell was entirely either inside or outside the solid, was given when all the densities at the eight corners of a cell had the same sign, thus implying that no polygons are defined. In all other instances, the cell lay on the boundary of the solid and one to five polygons were generated. The technique of space partitioning was employed when working with cubes. Reading of incidental edges, i.e., those colliding with the surface of the shape volume, was possible upon determination of the shape intersects of all the boxes. The use of linear interpolation facilitated the selection of the exact point of intersection of the solid with each cube edge. The density values at the eight corners were taken, assessed for positivity or negativity, after which each value was assigned one bit in a binary representation. The bit was set to either zero or one depending on whether the density values were positive or negative respectively. From the marching-cubes' pre-defined table of intersections (256 combinations), the system read the triangles that formed the points of intersection, and displayed them. Interpolation determined precisely where a vertex was placed along an edge. The vertex needed to be located where the density value was approximately equal to zero. For example, if density at the end *A* of the edge was 0.1 and at the end *B* it was -0.3, the vertex was placed at 25 per cent of the way from *A* to *B* (Fig 3.13).

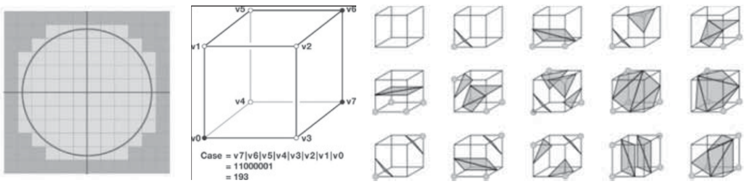


Fig 3.13 Left: cross sections of cubes intersected by a solid (green colour depicts the selected cubes); Middle: a single voxel with known density values at its eight corners; Right: the 15 fundamental cases in *Marching Cubes*.

3.8. Mapping a Texture on the Surface of a Three-Dimensional Object

In order to improve the realism of the modeled geometry, an additional 3D graphics software called 3D Studio Max was used to create the effect of textures on the final object. Based on the created module, the geometry of the developed model was exported to the *.obj format. then the file was loaded into an external environment. In the program, the file model lacked texture and its geometry did not have artifacts on the surface. The rendering and imposition of texture enriched the visual effect of the SG-M algorithm, resulting in visually pleasing forms.

4. Results

4.1. Steering Parameters

The modeling process implemented in the present study was based on the SG-M algorithm, and was affected by the following parameters:

- for the main SG-M algorithm - the level of detail of the mesh (L_D, 1 - 32), (for all modes)
- for caves: - amount of classic rules (R)
  - labelling (X/-X, Y/-Y, Z/-Z, L_B, 3 direction)
  - random shift of the labels (R_{SL}, 0 - 100),
  - amount of morphic rules (R_M)
  - morphing parameter (M_P, 0..100%)

- for clouds: - beginning and end sphere radius (20 - 200)
  - labelling (X, Y, Z, -X, -Y, -Z, L_B,6 direction) (amount of classic rules)
  - random shift of the labels (R_{SL}, 0 - 100),
  - cloud deformation with basic shapes (morphing parameter, M_P, 0..100%)

- for architecture: - number of floors (1..20) (amount of classic rules, R),
  - lower floor: height, width, convexity.
  - higher floor: height, width, convexity.
  - roof: from spherical to sharp (morphing parameter, M_P, 0..100%).

4.2. Clouds Customization

Random shift of labels parameter (R_{SL}) was found to alter topology of generated clouds from one solid model to a group of many dispersed smaller structures (Fig 4.1).

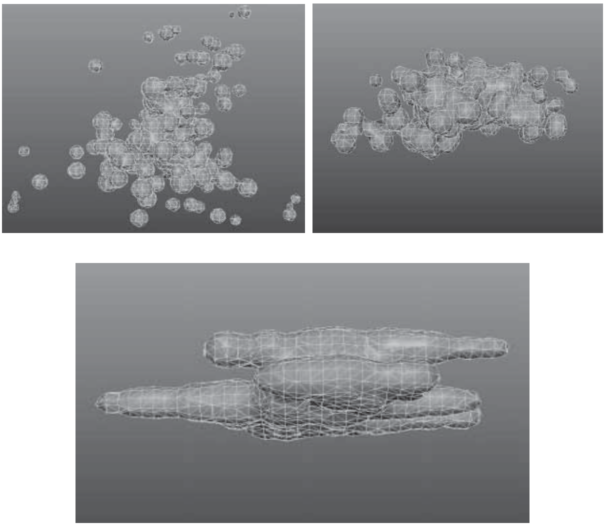


Fig 4.1. Random shift of the labels: top left $R_{SL} = 100$, top right $R_{SL} = 50$, bottom $R_{SL} = 0$.

Changing the radius of minimal and maximal sphere helped in obtaining other structures of clouds. Using different values for two input spheres yielded better results than using similar values, as is evident from Fig 4.2. The algorithm used in this study enabled greater control over the procedure for cloud modeling.

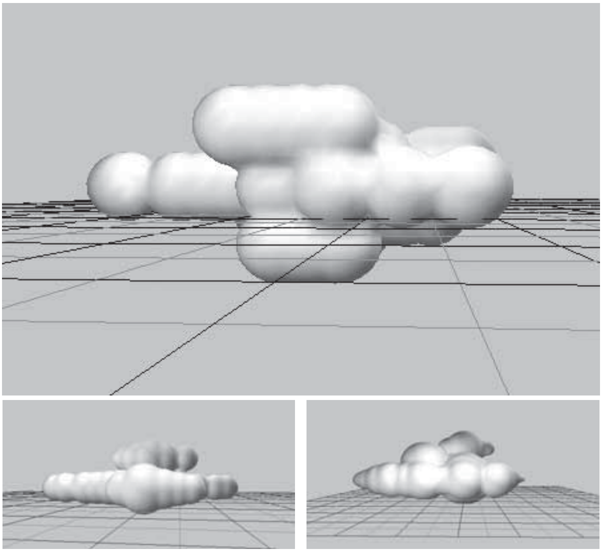


Fig 4.2 Another radius sphere size: top: same values; Bottom left: values with slight difference; Bottom right: values with marked difference.

4.3. Cave Customization

The algorithm implemented for cave customization (therefore.e., the cave moseling process), was influenced by the following parameters: the level of detail of the model grid (L_D, 1..32), classic rules (P_C) – operations (sum, difference, intersection), morphic rules (P_M) – operation morphing, direction of the labeling (L_B, *-X/+X, -Y/+Y, -Z/* +Z and random), random shift of the labels (R_{SL}, 0..100), morphing parameter (M_P, 0..100%), and production number for each shape (P_S, 0..10, this helps to assemble primitive shapes into more complex shapes).

Parameters of the experiment: $L_D = 10$, $P_C = 10$, $P_M = 2$, $L_B = -X/+X$, $M_P = 52\%$, $P_S = 3$.

Figs. 4.3 and 4.4 illustrate the resulting three-dimensional structure showing the topology of the caves.

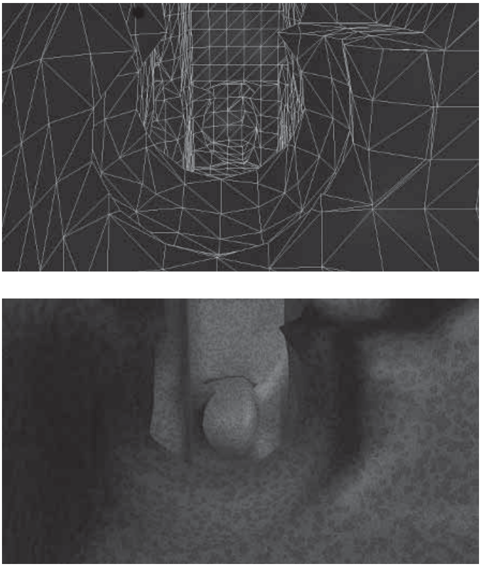


Fig 4.3 Cave structure obtained for L_D= 5. Top: grid model; Bottom: textured rendering

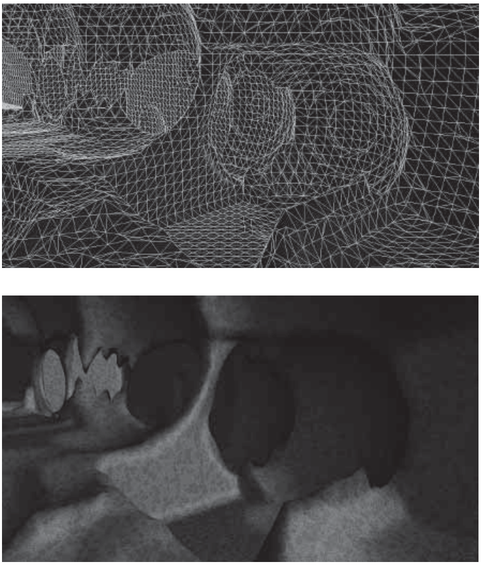


Fig 4.4. Cave structure obtained for L_D= 32. Top: grid model; Bottom: textured rendering.

4.4. Various Results

Figure 4.5 depicts some results obtained using the method developed in the present study.

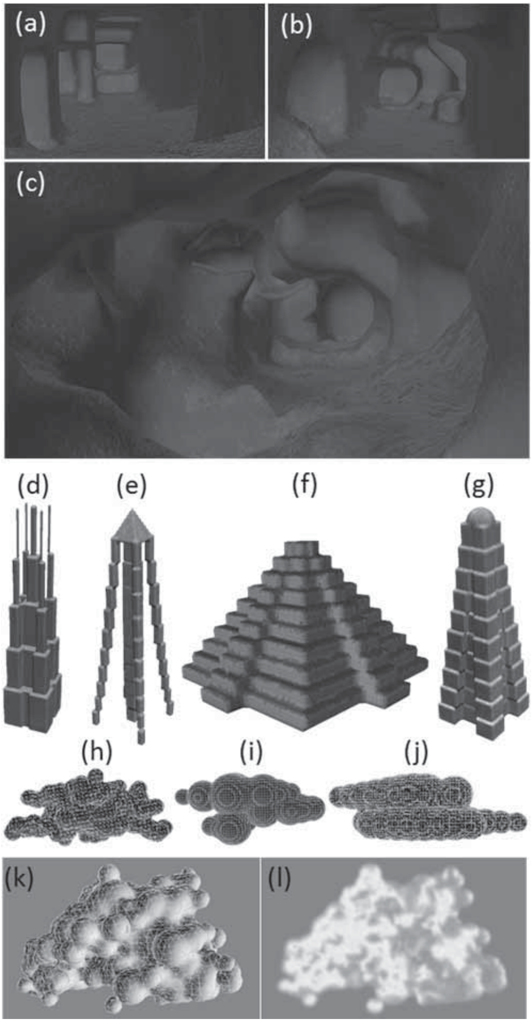


Fig. 4.5. Possible results (mesh and final renders): (a-c) caves, (d-g) – architecture, (h-l) clouds. For objects in (b), (c), (e), (g), (j), (k), and (l), a morphic rule was applied with a different value of morphing parameter.

The platform used for simulations consisted of - nVidia GeForce GTX 460M GPU, i7-2630QM CPU and 12 GB RAM. Processing times: from 500 ms to 10 sec (caves: 500 ms–2 secs, buildings: 3–10 secs, clouds: 1.5–9 secs)

5. Geometry Verification Methods

In the present investigation, consideration was given to the development of valuable verification methods, and several metrics were proposed as tools that can be used to verify the geometry of the model.

5.1. Surface Area of the Model

Generation of the grid was followed by formation of triangles that make up the model. Each of the triangles had three vertices (V_1, V_2, V_3) in the three-dimensional space described by x, y, and z coordinates. The given formula (based on the vector product), was applied for determining the surface of the triangle so described:

$$S = \tfrac{1}{2} * ((V_2 - V_1) \times (V_3 - V_1)) \tag{5.1}$$

The relationship between the parameters of the sampling density and the surface of the resulting structures is portrayed in Fig.4.5. The obtained surface structure of these objects was expressed in m^2, and ranged from 0 to some value, with the limit being infinity (Fig. 5.1).

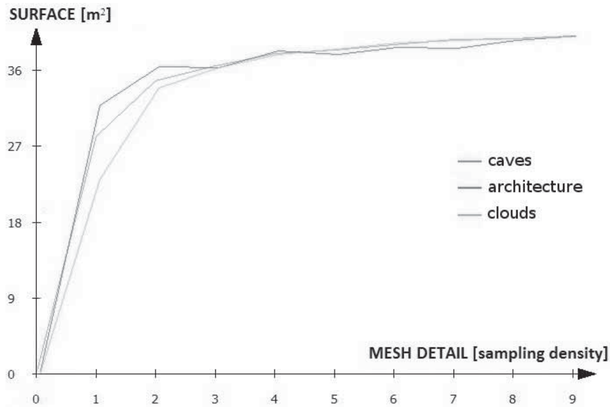


Fig. 5.1 Comparison surface vs mesh density example. Top: cave; Middle: cloud; Bottom: building (L_D, represented by sampling density).

5.2. Volume of the Model

Sampled cubes were obtained from existing productions during the mesh generation process. All the cubes were counted when their vertices were generated inside the solid. The interior was determined by the function describing the shape. Computation of f(p) -- which is a function of the distance from the point p to the cube surface – gave an indication that all eight vertices V1..V8 of the cube satisfied the (Vi) < 0 condition. The volume was calculated with the help of the following expression:

$$V_M = N_C * (C_E)^3 \tag{5.2}$$

where:
V_M is the volume of the model,
C_E represents cube edge *(C_E = 1/grid density),*
N_C equals numbers of cubes inside.

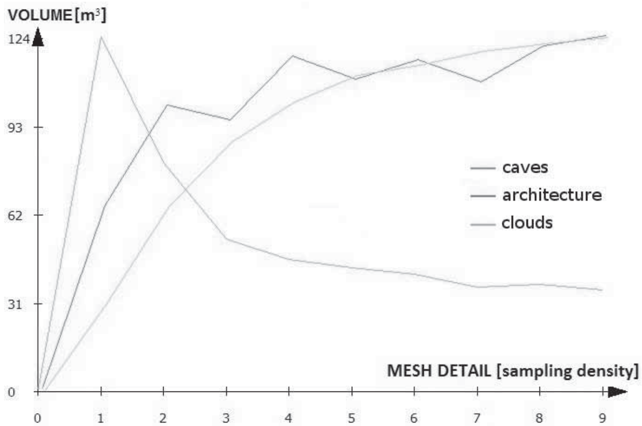


Fig. 5.2 Comparison volume vs mesh density example. Top: cave; Middle: cloud; Bottom: building (L_D, represented by the sampling density)

5.3. Concavity and Convexity Factors

All 3D structures exhibiting a convexity ratio below 100% displayed concavity on the grid, whereas structures with a convexity ratio exactly equal to 100% showed convexity of mesh, e.g. a cube. Concavity and convexity factors were defined by the following formula:

$$C_F \cong \frac{100 * T_G}{T_G + T_L} \tag{5.3}$$

where:
C_F stands for concavity and convexity factor,
T_G is the number of triangles between which the angle is greater than 0°,
T_L equals the number of triangles between which the angle is less than 0°.

Values obtained after application of the above formula ranged from 0 to 100%, and depended on the number of existing triangles with angles of *d* greater than or less than 0°. Values of d>0 and d<0 respectively indicated concavity and convexity. A summing up of the relationship between the edges of all triangles yielded information about the ratio of the solid (Fig. 5.3).

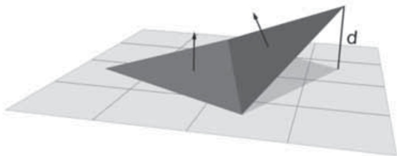


Fig 5.3. Pictorial representation of the calculation of convexity (d – parameter defining the angle between neighbouring triangles)

Research literature provides evidence that objects with the topology of tunnels/caves, clouds and architecture display ratios between 30--54%, 76—88%, and 81—100% in that order. Examples of the extreme values of C_F are shown in Fig. 5.4.

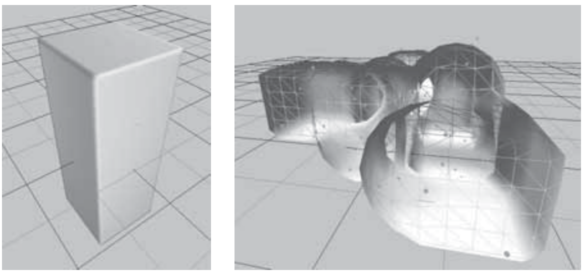
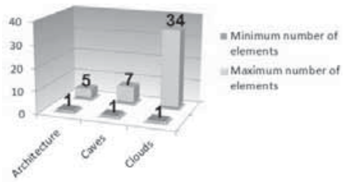


Fig. 5.4. Extreme values of C_F. Left: C_F =100 % - clear mesh (for very simple building, e.g a box); Right: C_F = 43 % - concaves on the grid (e.g., cave)

5.4. Independent Model Objects

All the triangles were viewed in the generated mesh, and new groups were created. Ovelapping of any of the vertices of one group with the tip of another group, led to a combination of the groups into one unit. It was found that the number of groups after reviewing all the triangles equalled the number of elements that made up the grid. Research has shown that objects of cave, architecture and cloud topology though made up largely of one type of element, also contain many other elements. The likelihood of the existence of a single type of object is most with architecture modeling, less for cave modeling, and least forcloud modeling (Fig 5.5).



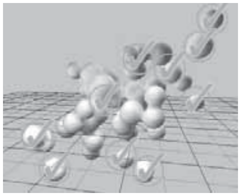


Fig. 5.5. Left: independent model objects for: architecture, caves and clouds. Right : example of the distributed cloud with 10 separate elements

5.5. Time Complexity

A very important consideration during the analysis of results is the time complexity or performance of the proposed method (Fig. 5.6):

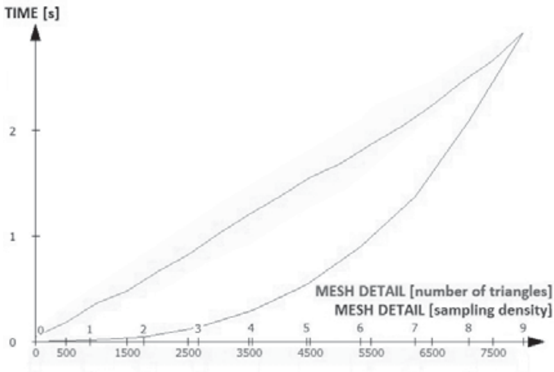


Fig. 5.6. Time vs mesh detail (L_D, represented by sampling density) and number of triangles for the cave example from Fig 13

5.6. Histograms

The methodology followed in the present study represented shapes using a histogram of inter-vertex angles. It was assumed that models comprise edges, each of them determined by two points (i.e., vertices). Further, the vertices were assigned normal vectors. Building up the representation was carried out by creating a set of edges of all triangles of the model. Here, edges belonging to two triangles were added twice, but perhaps with different normal vectors. Subsequently, a value in the interval [0, 1] was calculated for each edge: this was given by the difference of the two angles between normal vectors and the line containing the vertices (Fig. 5.7).

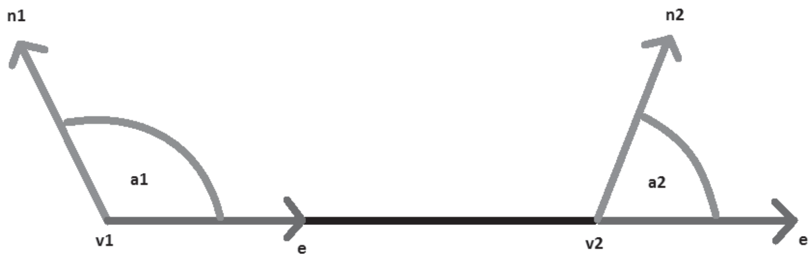


Fig 5.7. Calculation of the angles between the vertices.

$$e = \frac{(V_2 - V_1)}{|V_2 - V_1|} \tag{5.4}$$

$$a_1 = \arccos(e \bullet n_1) \tag{5.5}$$

$$a_2 = \arccos(e \bullet n_2) \tag{5.6}$$

$$d = a_2 - a_1 \tag{5.7}$$

$$v = \frac{d}{2(\pi + 1)} \tag{5.8}$$

where n_1 and n_2 are the normal vectors. For concave surfaces, this value belongs to the interval [0, 0.5). For convex surfaces, it belongs to the the interval (0.5, 1].

In the next step, a discrete histogram was created indicating the frequency of numbers lying in the given interval. This was done by calculating the hiustogram for each value of the set, using the formula given below:

$$i = \mathit{floor}\,(v * R) + 1 \tag{5.9}$$

where: R is the number of histogram intervals.

Next, the multiplicity of the i-th interval was incremented by one. After considering all values in the set, the histogram was normalized by dividing its compartments by the total number of edges. The normalization ensured that the sum of the ranges of the histogram equalled 1. For the purpose of classification, it was necessary to prepare the histograms corresponding to master object classes (Fig. 5.9).

Verification of the correctness of the model depended on the density of the grid. Best results were obtained for $L_D=10$ (Fig. 5.8).

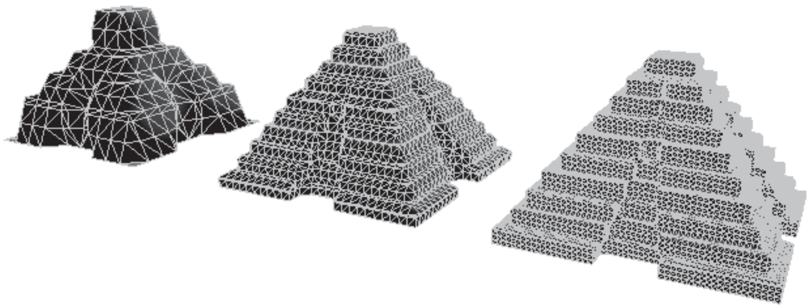


Fig 5.8. Level of detail (L_D), from left L_D=2, L_D=5, L_D=10

When density was large, the histogram was reduced to a single bar representing the most common angle. The user can control the histogram resolution (R) to improve object classification. Figure 5.9 shows a set of histograms of an object for varying resolutions.

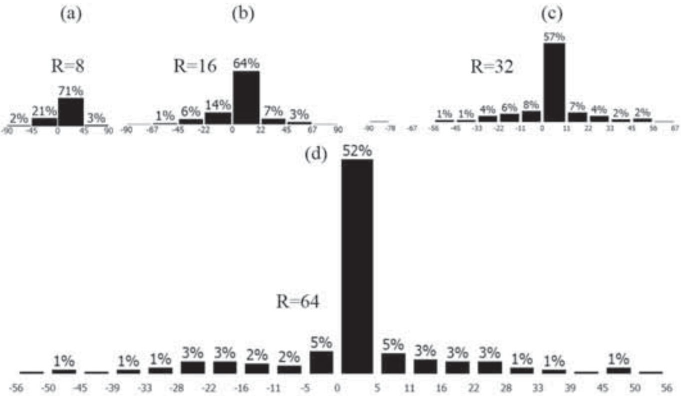


Fig 5.9. Varying resolutions for an object: (a) 8, (b) 16, (c) 32, (d) 64.

One stage of classification was to create a wide shape base that would be used for shape recognition.

From all objects selected to one class, average histograms were created for three main classes: architecture, clouds, and caves (Fig. 5.10).

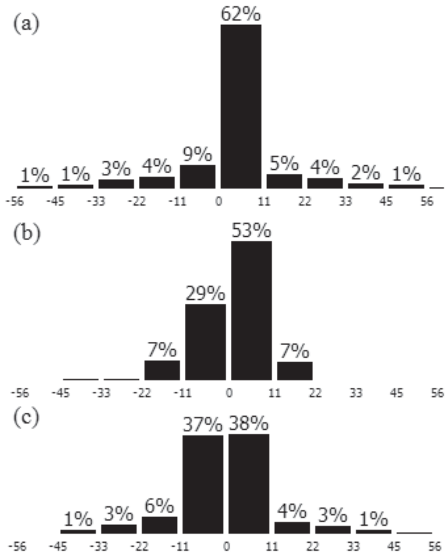


Fig 5.10. Main histogram classes: (a) architecture, (b) clouds, and (c) caves

The method followed in the present study produced different histograms for all of the tested classes, and in the next step, it helped in recognition based on comparison of main classes with randomly generated objects using a calculation of Mean Squared Error (MSE).

6. Performance of the Proposed Method – Parallel Implementations

Multiprocessor farms and multi-threading architecture have been in use in the digital industry since the invention of parallel computing. The main task has always been to speed up the media production process. System architecture can be adapted to a particular computational task, but in itself it is not a determinant of performance or efficiency. The 3D scene can be considered as a set of cubes covering the rendering space. The number y of cubes can be expressed as:

$$y_k(p) = F^l(u_k(p)) \,\|\, F^{l-1}(u_k(p)) \,\|\, ... \,\|\, F^1(u_k(p)) \tag{6.1}$$

Where:
y is the number of cubes inside,
k denotes the number of sequential cubes inside,
u *is* the cube edge,
p stands for- the point with coordinates (x, y, z) in Euclidean R^3
$\|$ is the symbol of executing operations simultaneously,
F^{δ} represents the parallelization operator, $\delta \in \{1,...,l\}$.

Many operations can be performed simultaneously during the position point's calculation of different productions. Dedicated computing parallelism is also used, although there are no streams of instructions. The parallelization is significantly facilitated by eliminating the communication overhead for the acquisition of stream commands. For the evaluation of computational tasks, performance in multicore systems speedup, the following formula is used:

$$S_m = \frac{T_1}{T_m} \tag{6.2}$$

where:
S_m is the parallel computation speedup,

T_1 = the execution time of the sequential algorithm,

T_m = the execution time of the parallel algorithm with *m* processors.

In multiprocessor systems, the actual value of the speedup is lower than the theoretical value (based on the number of processors), due to communication overheads and the need to share resources such as memory and buses. This property describes a model of efficiency, which is a measure of concurrent use of resources. Speedup efficiency is defined as:

$$E_m = \frac{S_m}{m} \tag{6.3}$$

where:
E_m is the speedup efficiency with *m* processors $E_m \in (0;1\rangle$.

Ideally,, the efficiency should be 1.0, which means that the speedup S_m is proportional to the number of processors m or computing elements.
The concept of throughput can be introduced, which corresponds to the number of data processed per time unit.

$$P = \frac{D}{T} \tag{6.4}$$

Where:
P – throughput,
D – the number of computed data,
T – the measurement time.
7. Results of analysis

The comparisons of computation time for both sequential and parallel algorithm are illustrated in Table 6.1. Tables 6.2 and 6.3 respectively depict the speedup amd throughout evaluations for the sequential/parallel algorithm for the scenes depicted in Fig. 6.1.

Sequential algorithm		
Caves	Clouds	Architecture
0.81[s]	0.62[s]	0.54[s]

Parallel algorithm (*m*=255)		
Caves	Clouds	Architecture
0.096[s]	0.125[s]	0.069[s]

Table 6.1. The computation time evaluation for the sequential/parallel algorithm

Speedup efficiency (*m*=255)		
Caves	Clouds	Architecture
0.028	0.015	0.027

Table 6.2. The speedup efficiency evaluation for the parallel algorithm

Throughput (*m*=255)	
Sequential algorithm	Parallel algorithm
0.8[GFlop/s]	20.61[GFlop/s]

Table 6.3. The throughput evaluation for the sequential/parallel algorithm

The platform used for simulations consisted of nVidia GeForce GTX 460M GPU, i7-2630QM CPU and 12 GB RAM.

Fig.6.4 shows the speedup of a parallelized shape morphing system and the shape hybrid system vs the number of measurements (data size). The speedups were evaluated based on measured run times.

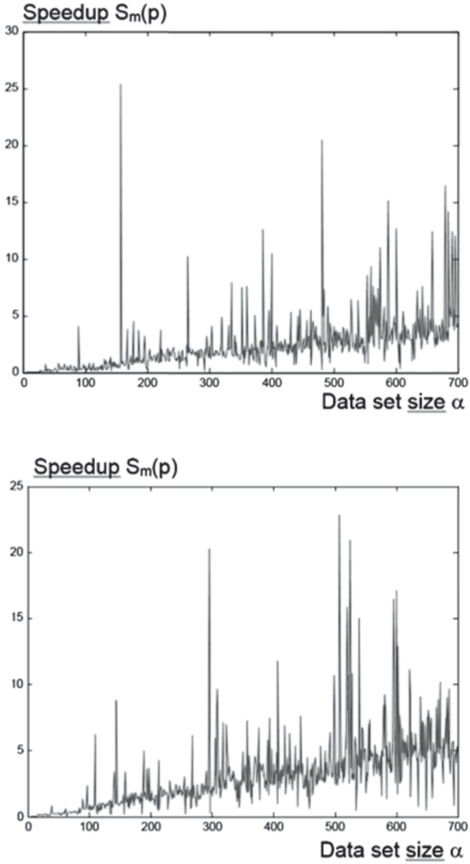


Fig. 6.4. Left: the speedup of the parallel shape morphing system; Right: the speedup of the parallel shape hybrid system.

The speedups, shown in Figure 6.4, were found to be equal to formally calculated speedups of parallel algorithms. it was also observed that parallel implementations ran approximately 5-6 times faster than their sequential versions.

8. Summary

Computer graphics systems are an essential component of modern information systems. Recent times have witnessed a noticeable trend in the application of new methodologies for modeling 3D objects in virtual-reality systems. This trend seeks to answer the growing market demand from the domains of digital entertainment, simulation for 3D gaming, and 3D VFX industries. The present research effort successfully attempted to introduce an innovative method for real-time procedural modeling of the three-dimensional geometry of caves, clouds, and buildings. The method simultaneously used morphing and the classical formalism of shape grammars. The new algorithm designed in the experiment enabled modeling of complex objects with characteristics that markedly influence visual realism viz., greater variety and geometric complexity. In addition to the advantages of shape grammar, the morphing parameter permitted the establishment of a continuous percentage contribution of two input shapes, to produce a single output object, and in turn aid in fine-tuning control of the modeling process. Future research will focus on the development of shape rules by providing an intelligent expert system for steering rule selection.

Parallelization of 3D scene rendering is a must for the digital entertainment industry. Recent advances in multiprocessor and multicore hardware architectures have greatly improved the process of 3D digital media production. The present paper puts forth analysis of performance for the sequential and parallel procedural shape hybrid system, based on shape grammar and morphing. The parallel version, has been proven to achieve performance foreseen in the speedup estimation, even for the most demanding case of procedural shape construction. The parallel SG-M algorithm opens doors to finer 3D scenes, thus extending current digital media production capabilities. It is planned to further develop methodology for supporting still more diversified 3D models, as well as models with dynamics.

References

Alexa, M., Cohen-Or, D., Levin D. (2000). As rigid as possible polygon morphing. *Computer Graphics*, (34), pp. pp. 157-64.

Am Ende, B.A. (2001). 3D Mapping of Underwater Caves, *IEEE Computer Graphics Applications*, 21(2), pp. 14-20.

Boggus, M., Crawfis R. (2009). Procedural Creation of 3D Solution Cave Models, *Proceedings of the 20th IASTED International Conference on Modeling and Simulation*, pp. 180-186.

—. (2009). Explicit Generation of 3D Models of Solution Caves for Virtual Environments, *Proceedings of the 2009 International Conference on Computer Graphics and Virtual Reality*, pp. 85-90.

—. (2010). A Framework for Interactive Modelling of ThreeDimensional Caves, *Proceedings of the 6th International Conference on Advances in Visual Computing (ISVC'10), Volume Part II*, pp. 213-221.

Bouthors, A., Neyret, F. (2004). Modelling Clouds Shape, Proceedings of EUROGRAPHICS (short presentations).

Clempner, J.B., Poznyak, A.S. (2011). Convergence method, properties and computational complexity for Lyapunov games. *The International Journal of Applied Mathematics and Computer Science,* 21(2).

Di Trapani, L.J., Inanc, T. (2010). NTGsim: A graphical user interface and a 3D simulator for nonlinear trajectory generation methodology. *The International Journal of Applied Mathematics and Computer Science*, 20(2).

Dobashi, Y., Kaneda, K., Yamashita, H., Okita, T., Nishita, T. (2000). A simple, efficient method for realistic animation of clouds. *Proceedings of ACM SIGGRAPH 2000,* pp. 19-28.

Dyken, Ch., Ziegler, G., Theobalt, Ch., Seidel, H.P. (2008). High-speed Marching Cubes using HistoPyramids. *Computer Graphics forum*, 27(8), pp. 2028-2039.

Ebert, D. S. (1997). Volumetric procedural implicit functions: A cloud is born. In *SIGGRAPH 97 Technical Sketches Program*, *ACM SIGGRAPH*.

Elinas, P., Sturzlinger, W. (2000). Real-time rendering of 3D clouds. *Journal of Graphics Tools 5*, (4), pp. 33-45.

Gardner, G.Y. (1984). Simulation of natural scenes using textured quadric surfaces. In *Computer Graphics,* (18), pp. 11-20.

—. (1985). Visual simulation of clouds. *Computer Graphics.* (19), pp. 297-303.

Goetz, F., Junklewitz, T., Domik, G. (2005). Realtime marching cubes on the vertex shader. *Eurographics 2005, Short Presentations*.

Greuter, S., Parker, J., Stewart N. and Leach G. (2003). Real-time procedural generation of pseudo infinite cities. *Proceedings of GRAPHITE 2003*, *ACM Press*, pp. 87-95.

Harris, M. J., Lastra, A. (2001). Real-time cloud rendering. *Computer Graphics Forum* 20(3), pp. 76-84.

Johnson, L., Yannakakis, G.N. and Togelius, J. (2010). Cellular Automata for Real-time Generation of Infinite Cave Levels. *Proceedings of the 2010 Workshop on Procedural Content Generation in Gam*es *(PC Games, 10),* pp. 1-4.

Kent, J.R., Carlson, W.E., Parent, R.E. (1992). Shape transformation for polyhedral objects. *Computer Graphics*, (26), pp. 47-54.

Lazarus, F., Verrous, A. (1998). Three-dimensional metamorphosis: a survey. *The Visual Computer*, 14(8-9), pp. 373-89.

Lee, A.W.F., Dobkin, D., Sweldens, W., Shroeder, P. (1999). Multiresolution mesh morphing. *Computer Graphics*, (26), pp. 43-6.

22. Lorensen, W.E., Cline, H.E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, (21), pp. 163 - 169.

Martyn, T. (2004). A new approach to morphing 2D affine IFS fractals. *Computers & Graphics*; (28), pp. 249-72.

Nishita, T., Nakamae, E., Dobashi, Y. (1996). Display of clouds taking into account multiple anisotropic scattering and skylight. *SIGGRAPH 96 Conference Proceedings,* Addison Wesley, pp. 379-386.

Peytavie, A., Galin, E., Grosjean, J. and Merillou, S. (2009), Arches- a Framework for Modelling Complex Terrains. *Computer Graphics Forum*, 28(2), 4 . pp. 57-467.

Parish, Y. I. H., Muller, P. (2001). Procedural modeling of cities. *Proceedings SIGGRAPH'01*, *ACM Press,* pp. 301–308.

Prusinkiewicz, P., Lindenmayer, A. (1991). *The Algorithmic Beauty of Plants*. Springer-Verlag.

Prusinkiewicz, P. Lindenmayer, A. Hanan, J. (1988). Development of Herbaceous Plants for Computer Imagery Purposes. *Computer Graphics*, 22(4), pp. 141-150.

Schpok, J., Simon, J., Ebert, D. S., Hansen, C. (2003). A real-time cloud modeling, rendering, and animation system. *Symposium on Computer Animation'03,* pp. 160- 166.

Schuchardt, P., Bowman, D.A. (2007). The Benefits of Immersion for Spatial Understanding of Complex Underground Cave Systems. *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology (VRST '07),* pp. 121- 124.

Stiny, G., Gips, J. (1972). Shape grammars and the generative specification of painting and sculpture. Information Processing, (71), *pp.* 1460-1465.

Stiny, G. (1975). *Pictorial and Formal Aspects of Shape and Shape Grammars*. Basel, Birkhauser Verlag.

—. (1980). Introduction to shape and shape grammars. Environment Planning, 7(3), pp. 343–361.

Turk, G., O'Brien J.F. (1999). Shape Transformation using variational implicit functions. Computer Graphics, (33), pp. 335-42.

Velho, L., Gomes, J., Figueiredo, L.H. (2002). *Implicit Objects Computer Graphics*, Springer; 1st edition.

Wolberg, G. (1998). Image morphing: a survey. *The Visual Computer*, 14(8-9), pp. 360-72.

Wonka, P., Wimmer, M., Sillion, F. and Ribarsky, W. (2003). Instant architecture. *Siggraph 2003*, *ACM Transactions on Graphics*, 22(3), pp. 669–677.

JAV-ATAR: A VIDEO GAME CONTEST TO PROMOTE COMPUTER SCIENCE

ANDRÉS ADOLFO NAVARRO-NEWBALL

Abstract

In this ICT age, advances in technology are making our lives easier, but a huge amount of technological progress is still needed. Paradoxically, in an era when computing has pervaded every aspect of our lives, Computer Science is not seen as a desirable career. Additionally, the number of students, women and minorities studying Computer Science is not enough for society's needs. We explore the causes of the decrease in interest in Computer Science and describe previous experiences that attempt to encourage young people to study this subject. Then, we describe an experience of our own promotional Computer Science event and its impact in one of our universities. We finish with a discussion of our results in relation to this ICT age.

Keywords: Computer Science, Information Technology, Enrolment, Motivation, Video game, Animation.

1. Introduction

An ACM blog (Guzdail, 2012) states that there is less Computer Science (CS) in US high schools. It is claimed that this is due to factors such as the relative difficulty of introducing CS into the typical high school curriculum; CS is not part of the common core that the US governors are promoting, and CS is battling issues of perception and access. The aforementioned claim reflects a paradoxical situation occurring worldwide. Indeed, in this ICT age, technological advances are making our lives easier, while CS is not perceived as a desirable career, although huge technological advances are still needed. For instance, CS and Information Technology (IT) enrolment has decreased (Heersink and Moskal, 2010). Despite the fact that computing has pervaded every aspect of our lives, students are not eager to study it, and CS programs suffer from low retention (Buckley, Nordlinger and Subramanian, 2008; Haungs et al., 2012). Additionally, the number of students, women and minorities studying CS is not enough for society's needs. This is probably because what CS has to offer is not communicated adequately. Among the causes as to why students do not wish to engage in a CS career are:

- They do not know what CS does.
- They feel that it is an isolated profession in which computing professionals spend most time programming and not problem solving (Heerskin and Moskal, 2010). Additionally, they perceive programmers as 'geeks' who cannot have a social impact. In particular, female students are concerned about the insularity of working alone over long periods (Berenson et al., 2004). However, contrary to what is believed, computing jobs require communication and teamwork skills (Estey et al., 2010).
- They are influenced by non-academic factors. For example, it has been shown that female enrolment can increase if non-academic factors (e.g. social support) are enhanced (Haungs et al., 2012).
- They perceive CS as not fun, 'nerdy', irrelevant and a male-dominated career.
- They perceive technology as exciting, but programming as boring (Heerskin and Moskal, 2010). Here, computing is perceived as programming.
- Poor academic performance causes procrastination, avoidance and dishonesty.
- Usually, courses focus on the mechanics of CS, forgetting its key applications. In fact, advanced CS courses focusing on programming fail to attract students, or provide motivating applications (Webb, Repenning and Koh, 2012).

For these reasons, there is a need not only of a new curriculum, but of a new vision of CS pedagogy.

2. Promoting Understanding of CS

Next, we review four approaches aimed to promote understanding of CS. Computing for a Cause aims to show how CS can be used to solve problems; Personal Robots Programming and Video Game Development aim to motivate students; and High School Teacher Education aims to change the perception of those who are in direct contact with young students in regard to CS.

2.1. Computing for a Cause

In this approach, CS is presented as a discipline that empowers students to solve problems of personal interest or on a social scale, thus improving the quality, quantity and diversity of students (Buckley, Nordlinger and Subramanian, 2008). This approach emphasises problem representation and modelling in a way that real world problems are formulated using CS tools. When successfully applied, it reinforces teamwork, interdisciplinary work, and positive attitudes towards the relevance of students' work. Also, it transforms programmers into citizens participating in a large community (Buckley, Nordlinger and Subramanian, 2008). Commonly, it follows a project-based learning methodology (Haungs et al., 2012).

Successful examples of this approach are: (1) the CS Olympiad (Kearse and Hardnett, 2008), where students are challenged in a competitive, creative and innovative environment that encourages problem solving in preference to programming; and (2) the CS0 course (Haungs et al., 2012), where students choose the applications of their projects following common learning objectives, structure and desired outcomes. Overall (Kearse and Hardnett, 2008; Haungs et al., 2012), participants learn about the breadth of CS, developing teamwork and time management skills. Students interact with professionals and perform better. Finally, they obtain exposure to the big picture of CS and learn how CS allows them to work in many industries and make real world contributions.

2.2. Personal Robots

In this approach, personal robots are used to learn programming and to increase motivation (attention, relevance, confidence and satisfaction). Robots allow the teaching of a variety of CS concepts. However, they are not yet relevant to students, and the challenge is greater than in software-based systems (McGill, 2012).

2.3. Educating Teachers

Education in CS is important. High school teachers need help to motivate students. In this approach, workshops are offered that provide material to emphasise computational thinking and the many possibilities of CS (Blum and Cortina, 2007). Workshops succeed when they change the teacher's perception of CS from being seen as problem-solving algorithms to CS as a tool for developing computational thinking skills for all aspects of life.

2.4. Game Courses

It is well known that many video game companies started from the passion of teenagers programming their own games. Recently, Hanna, a ten-year-old girl, won the Microsoft's Kodu Game Lab Contest: "Her game is called Toxic, and in it players collect coins and hearts while solving puzzles to help save the environment" (Swift, 2011). Indeed, there is a prevailing interest in computer games and entertainment and an increasing prevalence of computer game design courses in CS (Distasio and Way, 2007; Estey et al., 2010).

In this approach, video games are used to deliver core and advanced CS concepts, to promote computational thinking, and to make CS relevant through the development of serious games with social content. Here, students address real problems by participating in engaging learning experiences (Yongpradit, 2012; Webb, Repenning and Koh, 2012). Indeed, game development implies serious CS, such as developing skills in complex logic, object-oriented programming, algorithms, computer graphics, user interfaces, artificial intelligence and data structures (Yongpradit, 2012; Distasio and Way, 2007). Additionally, it promotes creativity, expression, and team and multidisciplinary work (Haungs et al., 2012; Yongpradit, 2012). However, even though gaming can be correlated to an interest in CS, it has lost some of its cultural significance and its introduction into an exclusive community (CS) (DiSalvo and Bruckham, 2009). Still, an important percentage of students see the relation of gaming and CS, especially when they understand that the games they play are computer software.

Successful examples of this approach are: (1) Labyrinth (Distasio and Way, 2007), a modular, flexible, easy-to-use and compelling framework to design and experiment with games. While it is utilised in many computing fields, it fosters interdisciplinary education in communication (storyline), art and design (look and feel), and business (marketability); (2) the scalable game design course (Webb, Repenning and Koh, 2012), which embeds existing CS and math education in a project by following a guided discovery approach, while reaching middle school students where there is a high percentage of women and underrepresented communities; and (3) industry experiences emulation (Estey et al., 2010) by integrating cooperative learning, group orientation, and peer review into the game design process following a studio-based learning approach. Here, the sense of community, motivation, excitement and perceived value of feedback are maximised, while core CS concepts are learnt.

3. A Decrease in CS Enrolment

There has been a sharp decrease in interest in CS. For example, Figure 1 shows the steep decline in enrolment figures at the Pontificia Universidad Javeriana, Cali – Colombia. Enrolments in 2013-1 represent only 22.6% of the enrolments in 2004-1. This is simply an example of what is happening worldwide. Even though there are additional causes (e.g. financial problems, poor performance, marketing), there is an evident lack of interest in CS.

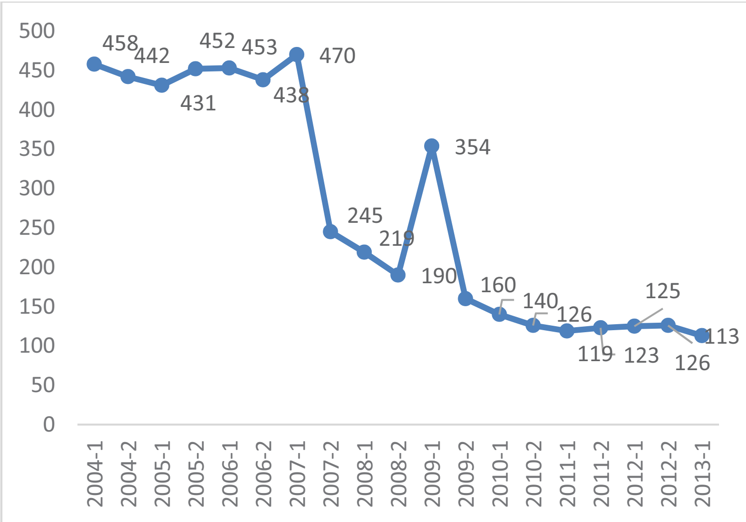


Figure 1. Decrease in CS enrolment at the Pontificia Universidad Javeriana, Cali – Colombia.

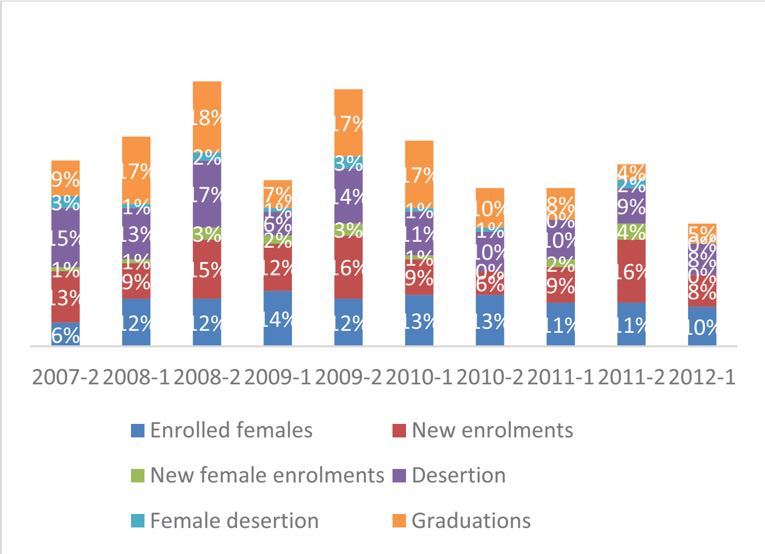


Figure 2. CS enrolment, desertion and graduation distribution at the Pontificia Universidad Javeriana, Cali – Colombia.

Figure 2 shows the distribution of enrolments, desertions and graduations from 2007 to 2012 at the Pontificia Universidad Javeriana, Cali – Colombia. Graduation represented a higher percentage until 2010-1; desertion has been steadily significant through the years; female enrolment represents only about 10% of the total, but female desertion is low. CS is evidently a male-dominated career.

4. Jav-Atar, Our Approach

Our CS program has had a strong decrease in enrolment and retention. Thus, this chapter's main author was told to promote it just a few days after he finished his PhD. His background being in Computer Graphics, he was expected to be a 'visual', motivating and persuading professor. He accepted the challenge back in 2010. All he knew then was that he wanted to show CS as a creative career where one has the chance to meet interesting people and uncover sources of information from many disciplines, depending on the application to be developed. It was decided in conjunction with our Office of Institutional Promotion, and the Head of Career that we would lead a video game and animation development contest for high school students. During the contest, the subjects of video games and animation would be contextualised within CS.

First, we had to come up with a name. We used Jav (from Javeriana University) and mixed it with avatar (a graphical representation of the user) to obtain Jav-Atar (by coincidence the motion picture *Avatar* was popular at the time). Then, we had to decide on a subject around which the contest would be developed. Finally, the logistics and stages of the contest had to be planned. One major difficulty was that local high school students perceived CS only as repairing computers or using productivity tools and they had no programming experience. The experiences of Jav-Atar 2011 and Jav-Atar 2012 are summarised next.

4.1. Jav-Atar 2011, the Taking of the Castle

This version of the contest was inspired by events from our country's history, the Battle of Cartagena de Indias. Participants were encouraged to visit libraries and interact with history teachers to document events and create the game plot. The battle "was an amphibious military engagement between the forces of Britain under Vice-Admiral Edward Vernon and those of Spain under Admiral Blas de Lezo. It took place at the city of Cartagena de Indias in March 1741, in present-day Colombia. The battle, although now largely forgotten in Britain, was the most significant of the War of Jenkins' Ear and one of the largest naval campaigns in British history. The war later was subsumed into the greater conflict of the War of the Austrian Succession. The battle resulted in a major defeat for the British Navy and Army. The battle marked a turning point in American history, as Spain preserved her military supremacy in that continent until the nineteenth century" (Wikipedia, 2013a).

The contest brought together 38 students (7 women) from 9 high schools, forming 14 video game studios with a maximum of 3 members per studio. Each high school sent an accompanying technology teacher. The contest was presented in October 2010. Next, our university offered the participants short courses on tools such as: Paint.NET, Audacity, Blender, XNA Game Studio and UDK. The courses took two weeks and took place in the afternoon slot during November 2010. The accompanying teachers were invited as observers. Then, discussion sessions with participating teams took place in January 2011. Video games were submitted to the Office of Institutional Promotion in April 2011 and were evaluated by three lecturers from the CS department.

The closing ceremony was inspired by the Oscars. High schools teachers, parents, families and friends were invited. There were interviews, speeches, emotions and anticipation. Categories awarded included: A) best script; B) best art; C) best development process; D) best game for 8th and 9th grade; and E) best game for 10th and 11th grade. The biggest prizes were given to D (an MP4 player for all members of the team) and to E (a 45% scholarship for the first term for all members of the team). All projects were described, and the features that made each one special were shown. The winner (Figure 3C) of the main category (E) was developed using XNA Game Studio by modifying the assets provided by the Platformer template and using a little programming. It showed an interesting balance of mechanics, path choice opportunities, increasing difficulty among levels, respectable art and overall an interesting level of abstraction. It was developed by a two-member group. Today, one of the participants is enrolled in our CS program.

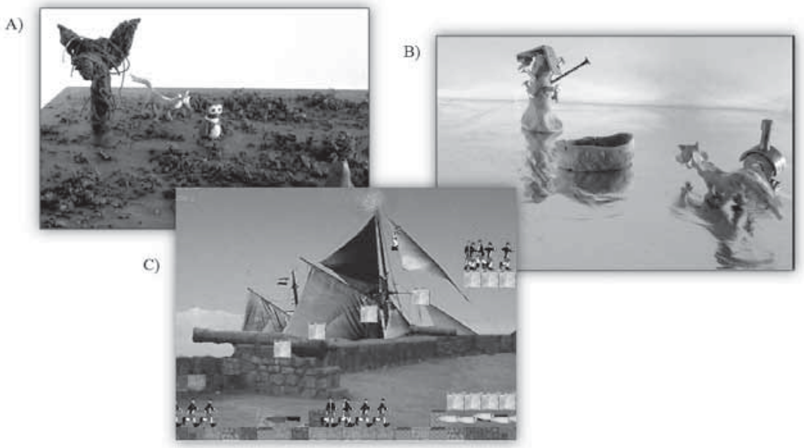


Figure 3. A) The beautiful dry forest. B) The dry forest after human intervention (animation by C. Gálvez and J.C. Garcés). C) Entering the battle after arriving in Cartagena (game by M. Rengífo and A. Velasco).

At the end of the contest, 38 participants were surveyed. Here, 30 students associated CS with creative development of innovative software; 26 with the development of software to help society; 17 with video games; 20 with multimedia; 7 with modelling systems; 14 with teamwork; 2 with math; and 6 with repairing computers. 18 expressed that they would definitely study CS, and 14 of whom would do so at our university. The majority of participants highlighted teamwork and creativity during the contest but would desire more short courses. Four would not participate again. All of them graded the experience as good or excellent. No female participants expressed interest in CS.

4.2. Jav-Atar 2012, Lost in the Forest

This version of the contest was inspired by the continuous destruction of the dry forest, mainly by human hands. We asked for an animation in order to avoid the logistical problems that occurred in Jav-Atar 2011 during the short courses, and presented the animation field as related to video games and as a field of study in CS. Participants were encouraged to interact with biology teachers, to visit the local Natural History Museum and to go to nature reserves in order to gather information to develop an animation. "The tropical and subtropical dry broadleaf forest biome, also known as tropical dry forest, is located at tropical and subtropical latitudes. Though these forests occur in climates that are warm year-round and may receive several hundred centimetres of rain per year, they have long dry seasons that last several months and vary with geographic location. These seasonal droughts have a great impact on all living things in the forest. Though less biologically diverse than rainforests, tropical dry forests are home to a wide variety of wildlife including monkeys, deer, large cats, parrots, various rodents, and ground-dwelling birds. Mammalian biomass tends to be higher in dry forests than in rain forests, especially in Asian and African dry forests. Many of these species display extraordinary adaptations to the difficult climate. Dry forests are highly sensitive to excessive burning and deforestation; overgrazing and exotic species can also quickly alter natural communities; restoration is possible but challenging, particularly if degradation has been intense and persistent. Degrading dry broadleaf often leaves thorny shrub lands, thickets, or dry grasslands in their place" (Wikipedia, 2013b).

The contest brought together 40 students (19 women) from 7 high schools forming 16 animation studios with a maximum of 3 members. Each high school sent an accompanying technology teacher. The contest was presented in October 2011. Next, our university offered the participants a one-afternoon presentation on tools such as Virtual Dub, Audacity and animation concepts. The accompanying teachers were invited as observers. Then, a discussion session with participating teams took place in January 2012. Animations were submitted to the Office of Institutional Promotion in April 2012 and were evaluated by three lecturers from the CS department. Following a suggestion from teachers participating in the previous Jav-Atar, animations were shown in a social setting before the final evaluation in a meeting with all the teams.

The closing ceremony was similar to the previous version. Categories awarded included: A) best script; B) best art; C) best development process; D) best animation for 8th and 9th grade and; E) best animation for 10th and 11th grade. The biggest prizes were given to D (a digital camera for all members of the team) and to E (a 45% scholarship for the first term for all members of the team). All projects were described, and the features that made each one special were shown. The winner (Figure 3 A and B), developed using a stop motion technique, used dry forest animals as protagonists and showed a deep ecological message within a very well developed story in a silent film style with beautiful music. It was developed by a two-member group. Today, one of them is enrolled in our CS program.

At the end of the contest, 33 participants were surveyed. Here, 23 students associated CS with creative development of innovative software; 19 with the development of software to help society; 10 with video games; 13 with multimedia; 9 with modelling systems; 6 with teamwork; 2 with math; and 9 with repairing computers. Six expressed they would definitely study CS, and all of whom would do so at our university. The majority of participants highlighted teamwork and creativity during the contest but would desire more short courses. One would not participate again. All of them graded the experience as good or excellent. One female participant expressed interest in CS.

5. Results

After our Jav-Atar experiences, we have had mixed results. It appears that we managed to influence the perception of CS among participants to some degree. Also, there seems to be more interest in animation among women. Still, the perception of CS as repairing computers remains strong (Figure 4). Additionally, we have had a slight but steady increase in new enrolments. Many of the participants who claimed that they would study CS in our university did not join our program (Figure 5). There is some evidence that this final decision is influenced by economic factors. Finally, after the biggest drop in 2010 and after Jav-Atar's appearance, the number of new enrolments seems to be increasing (Figure 6). This may be simply a coincidence, but 50% of the last group of interviewed candidates expressed that their interest in the career was due to Jav-Atar. Also, female enrolment seems to be increasing slightly.

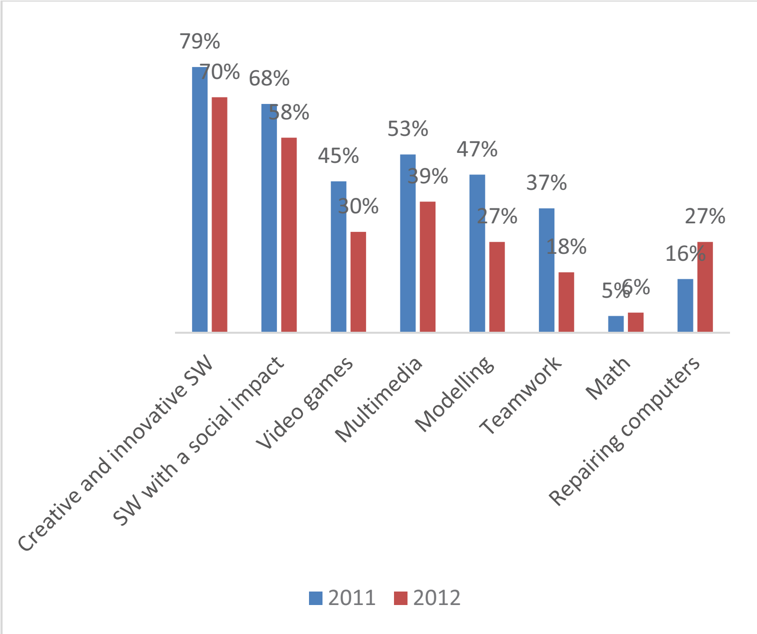


Figure 4. Perception of CS among participants.

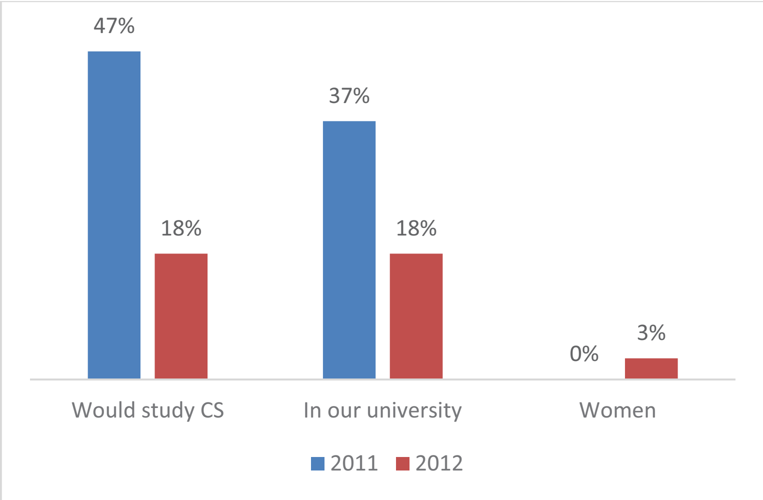


Figure 5. Interest in CS studies among participants.

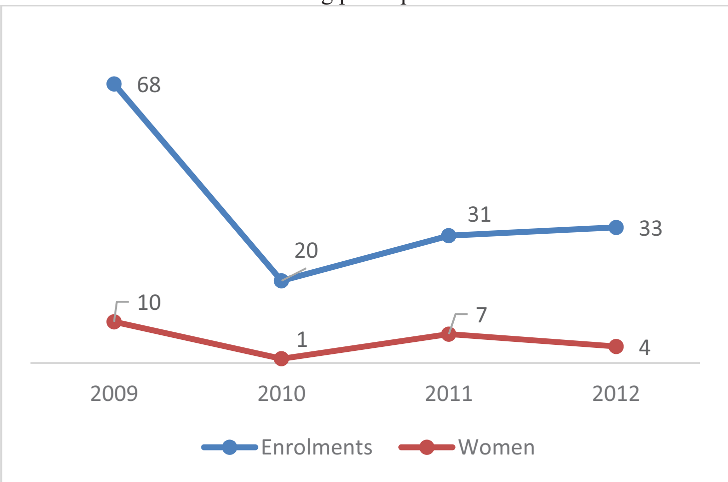


Figure 6. Recovering new enrolments figures in CS.

6. Discussion and Further Work

There has been a worldwide decrease in interest in CS and IT during the last few years. However, we cannot afford to claim this as an excuse. In this ICT age, action needs to be taken and the real role and need of CS needs to be evidenced to all. Indeed, there is a need not only for a new curriculum, but for a new vision of CS pedagogy and promotion, and the same has to be done for IT. We have explored innovative ways to promote CS. In particular, we decided to take a mixture of the Computing for a Cause, Educating Teachers and Game Courses approaches (Sections 2.1, 2.3 and 2.4). We were able to relate a video game contest to a relevant theme by utilising creativity among participants. Additionally, high school teachers (still observers) are becoming more interested every year. There are some promising preliminary results (Section 5), but we cannot claim victory. We expect to utilise and evolve Jav-Avatar for some years to come. Nevertheless, the video game subject may become extinct in the future, and we have to keep up with technological advances and fashion.

Other issues deserve further exploration. We believe in the actual power of human intellect as the ultimate technological advance; however, in the ICT age, we need to pay attention to technology as a fashion. Even though deeper studies are required, we perceive that technology is exciting to all but the development of it is not. Added to this, we may have reached a point of technological arrogance where we believe that computing devices do everything we need for us (e.g., they even replace creativity for some users). This and the fanaticism promoted by technological marketing (Sielger, 2010; Trenholm, 2011) can make us mistakenly believe that everything is done and that technology can do everything for us. Meanwhile, we may miss the real problems, questions and applications that are still required by CS. If we do not open our minds and show what CS and IT are about, then we may become like the society described by H.G. Wells (1895, p.80) in *The Time Machine* ("the balanced civilization that was at last attained must have long since passed its zenith, and was now far fallen into decay"), turning, in this way, our exciting ICT age into an intellectual ICE age.

6.1. Jav-Atar 2013 and beyond

For Jav-Atar 2013: Cyber Spatial Odyssey, we have been inspired by Ulysses' epic journey as described by Homer and all the great voyages from literature (e.g., Gulliver, and Verne). We are taking the video game approach once more and are offering short courses again, and we have involved teachers more. This time, we are taking a multidisciplinary approach in association with the Communications major in an attempt to better show the ubiquity of CS. This time, at least 88 video game studios are pre-registered. We know neither how many will submit a game yet, nor what will be the effect of working with another career, but we expect the effect will be positive for both programs. We have converted the contest into a creativity festival where everyone will be a winner, and we expect to create social impact in the region from 2014 by presenting the university and the involved programmes as a place where young people of all social strata can come and develop their ingenuity.

Acknowledgments

We are grateful for past and present participating high schools, teachers and participants. It has been a rewarding and joyful experience for us, and a display of creativity.

References

Berenson, S.B., Slaten, K.M., Williams, L. and Ho, C. (2004). Voices of women in a software engineering course: reflections on collaboration, *J. Educ. Resour. Comput.* (4) 1, Article 3 (March 2004).
DOI=10.1145/1060071.1060074
http://doi.acm.org/10.1145/1060071.1060074.

Blum, L. and Cortina, T.J. (2007). CS4HS: an outreach program for high school CS teachers, *Proceedings of the 38th SIGCSE technical symposium on Computer science education* (SIGCSE ’07), 19-23, New York, NY: ACM. DOI=10.1145/1227310.1227320
http://doi.acm.org/10.1145/1227310.1227320.

Buckley, M., Nordlinger, J., and Subramanian, D. (2008). Socially relevant computing, *Proceedings of the 39th SIGCSE technical symposium on Computer science education* (SIGCSE ’08), 347-351, New York, NY: ACM. DOI=10.1145/1352135.1352255
http://doi.acm.org/10.1145/1352135.1352255.

DiSalvo, B.J. and Bruckman, A. (2009). Questioning video games’ influence on CS interest, *Proceedings of the 4th International Conference on Foundations of Digital Games* (FDG ’09), 272-278, New York, NY: ACM. DOI=10.1145/1536513.1536561
http://doi.acm.org/10.1145/1536513.1536561.

Distasio, J. and Way, T. (2007). Inclusive computer science education using a ready-made computer game framework, *Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education* (ITiCSE ’07), 116-120, New York, NY: ACM. DOI=10.1145/1268784.1268820
http://doi.acm.org/10.1145/1268784.1268820.

Estey, A., Long, J., Gooch, B. and Gooch, A.A. (2010). Investigating studio-based learning in a course on game design, *Proceedings of the Fifth International Conference on the Foundations of Digital Games* (FDG '10), 64-71, New York, NY: ACM. DOI=10.1145/1822348.1822357 http://doi.acm.org/10.1145/1822348.1822357.

Guzdail, M. (2012). BLOG@CACM; Why Isn't There More Computer Science in U.S. High Schools? Available at: http://cacm.acm.org/blogs/blog-cacm/156531-why-isnt-there-more-computer-science-in-us-high-schools/fulltext.

Haungs, M., Clark, C., Clements, J. and Janzen, D. (2012). Improving first-year success and retention through interest-based CS0 courses, *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (SIGCSE '12), 589-594, New York, NY: ACM. DOI=10.1145/2157136.2157307 http://doi.acm.org/10.1145/2157136.2157307.

Heersink, D. and Moskal, B.M. (2010). Measuring high school students' attitudes toward computing, *Proceedings of the 41st ACM technical symposium on Computer science education* (SIGCSE '10), 446-450, New York, NY: ACM. DOI=10.1145/1734263.1734413 http://doi.acm.org/10.1145/1734263.1734413.

Kearse, I.B. and Hardnett C.R. (2008). Computer science Olympiad: exploring computer science through competition, *Proceedings of the 39th SIGCSE technical symposium on Computer science education* (SIGCSE '08), 92-96, New York, NY: ACM. DOI=10.1145/1352135.1352167 http://doi.acm.org/10.1145/1352135.1352167.

McGill, M.M. (2012). Learning to Program with Personal Robots: Influences on Student Motivation, *Trans. Comput. Educ,* (12) 1, Article 4 (March 2012), 32 pages. DOI=10.1145/2133797.2133801 http://doi.acm.org/10.1145/2133797.2133801.

Sielger, M.G. (2010). Fox News: Apple Is The New Religion And The Pope Is Scared. Available at: http://techcrunch.com/2010/07/29/apple-religion/.

Swift, J. (2011). Kodu Cup winners announced. Available at: http://www.i-programmer.info/news/98-languages/2601-kodu-cup-winners-announced.html.

Trenholm, R. (2011). Apple stimulates brain's religious responses, claims BBC. Available at: http://crave.cnet.co.uk/gadgets/apple-stimulates-brains-religious-responses-claims-bbc-50003807/.

Webb, D.C., Repenning, A. and Koh, K.H. (2012). Toward an emergent theory of broadening participation in computer science education, *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (SIGCSE ’12), 173-178, New York, NY: ACM. DOI=10.1145/2157136.2157191 http://doi.acm.org/10.1145/2157136.2157191.

Wells, H.G. (1895). *The Time Machine,* London, UK: William Heinemann.

Wikipedia (2013a). Battle of Cartagena de Indias. Available at: http://en.wikipedia.org/wiki/Battle_of_Cartagena_de_Indias.

—. Tropical and subtropical dry broadleaf forests. Available at: http://en.wikipedia.org/wiki/Dry_forest.

Yongpradit, P. (2012). Fun, phone, and the future: Microsoft XNA game studio, Windows phone, and Kinect SDK, *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (SIGCSE ’12), 321-322, New York, NY: ACM. DOI=10.1145/2157136.2157232 http://doi.acm.org/10.1145/2157136.2157232.

AUTOMATED DANCE CHOREOGRAPHY

NIGEL GWEE

Abstract

The commercial world is teeming with books and videos on dance instruction, but these are relatively fixed and inflexible. We can take advantage of the computer's capabilities by providing an on-demand service for dance choreography. Computationally, however, dance choreography is a non-trivial problem: it is NP-hard, often requiring heuristic methods. Also, while dance notation exists, it is extremely difficult to understand because dance is a three-dimensional activity.

We have developed Terpsichore©, a multimedia dance choreography software package that implements the International Ballroom Dance Syllabus. Here we describe its various features, focusing on important user-oriented design and human-like interface principles.

Using the software as a tool, we investigate the conditions under which exhaustive searching is feasible as opposed to when heuristic algorithms become necessary. Parameters include number of figures to be used, and choreography length. Various fitness functions that enable pruning of the search tree in heuristic methods are suggested.

We also describe instructional modules for the software, including interactive choreography creation and test modules, in which user-interface issues become important considerations. We explore real-time delivery of the completed choreography in a multimedia environment using audio and video. We touch on platform-dependent implementation issues, including the use of programming languages that best exploit various multimedia resources.

Finally, we discuss unique copyright issues involved in software development, a relatively recent issue in copyright law. We highlight the

need for more appropriate representation of software and measures of its originality.

Keywords—*software*; *NP-Complete*; *NP-Hard*; *ballroom*; *dance*; *software engineering*

1. Introduction

Instructional videos for ballroom dancing, among other performing arts, are very numerous these days. Beginning with the VHS videocassette, for which lots of rewinding was necessary, this instructional medium has been upgraded to DVDs and even Blu-Ray disks, where near-instant access greatly facilitates repetition, a key facet of kinesthetic learning. Throughout its existence, however, one of the drawbacks of all recorded performances is the fixed nature of the content. Fortunately, present-day technology gives us vastly enhanced options, and we describe here a software approach that delivers a much more flexible learning experience. What we shall call *automated dance choreography* is a resource that offers almost endless possibilities in the teaching of the art of ballroom dance. While no substitute for personal teaching, automated dance choreography does provide an invaluable supplement to traditional pedagogical methods. In the following, we discuss first a limited amount of NP-Complete and NP-Hard theory that forms a basis for the dance choreography problems we shall be tackling. We shall define dance choreography problems that we show are NP-Complete, and optimization versions which are NP-Hard. We then discuss various solution algorithms that can be used (§2). Next, we introduce our dance choreography software, Terpsichore©, and highlight its features and design techniques, including the user-interface (§3). We also mention some issues that we encountered when we copyrighted this software (§4).

2. Complexity Theory and Automatic Dance Choreography

Creating automatic dance amalgamations runs into the problem of exponential complexity. In order to provide a theoretical basis, we explain this complexity in terms of NP-Complete and NP-Hard complexity theory.

2.1 NP-Complete and NP-Hard Complexity Classes

We briefly describe the concept of NP-Complete and NP-Hard complexity categories, and distinguish between the two. An NP-Complete problem is a decision problem that (a) is solvable by a polynomial-time non-deterministic algorithm, and (b) to which every other polynomial-time non-deterministic algorithm can also be reduced in polynomial time. An NP-Hard problem is one to which an NP-Complete problem is Turing-reducible. For our purposes, we can say that an NP-Hard problem, among many others that can be identified as such, is an optimization problem that corresponds to an NP-Complete problem. In other words, the difference between the two is that NP-Complete problems are exclusively decision problems, whereas NP-Hard problems need not be.

A decision problem is defined as one that has a binary (e.g., yes/no) answer. Some of the dance choreography problems described below are examples of decision problems. Such problems are solvable by a polynomial-time non-deterministic algorithm if a correct, non-deterministically provided answer can be verified by an algorithm that takes no more than polynomial time to execute. A polynomial-time algorithm is one whose running time increases by no more than a polynomial factor, with respect to the size of the input. Polynomial reducibility is, as the name implies, a polynomial-time algorithm. An optimization problem is one that does not necessarily produce a binary answer, i.e., it is not necessarily a decision problem. Again, the corresponding optimization dance choreography problems described below are examples of NP-Hard problems.

An important milestone in complexity research was reached when Stephen A. Cook posited the first NP-Complete problem (Cook, 1971). This was the CNF Satisfiability problem (SAT). Cook's seminal result paved the way for the ongoing discovery of very many NP-Complete and NP-Hard problems, including those we introduce below. Thanks to Cook, in order to prove that a given problem is NP-Complete, "all" we have to do now is show that it is solvable by a polynomial-time non-deterministic algorithm, as per the above definition of NP-Completeness, and, significantly, that *some* known NP-Complete problem can be reduced to it in no more than polynomial time. The task of proving NP-Completeness is simplified from having to show that *all* polynomial-time non-deterministic algorithms are reducible (clearly impractical, if not impossible) to having to show that just *one* NP-Complete problem is reducible. As a result, whenever a researcher wishes to show that a particular problem is NP-Complete (or NP-Hard), there exist very many prior such problems from which to base a proof.

Apart from facilitating the categorization of similarly complex problems, NP-Complete theory allows us to link a vast set of complex problems within a cohesive algorithmic framework. The implications of this theory show that an efficient solution for any one NP-Complete/Hard problem automatically produces efficient solutions for all NP-Complete/Hard problems. Thus, all NP-Complete/Hard problems stand on an equal footing. Just imagine: an efficient solution to dance choreography gives us an efficient solution to more world-shaking problems, such as job scheduling and optimization, and project planning.

2.2 Complexity of Dance Choreography Problems

In this section, we identify and define dance choreography problems, and show how we can prove that they are NP-Complete. We shall define two groups of problems based on two decision problems: a simple, relatively specialized, version of the dance choreography problem, and a more general version of this problem (DANCE CHOREOGRAPHY and DANCE CHOREOGRAPHY 2, respectively). From these we derive more general decision problems, and the corresponding optimization problems. These problems are related through a transformation chain employed in their NP-Completeness and NP-Hardness proofs.

The NP-Completeness proofs assume solvability and verifiability by a polynomial-time non-deterministic algorithm, which can be easily worked out by the reader. For the reduction portion of the proofs, we use the technique of "restriction," whereby we show that the problem in question is a more general case of a known NP-Complete problem. Having shown that certain problems are NP-Complete, we define the optimization versions. These latter versions are NP-Hard, and the proof of NP-Hardness uses the concept of "Turing reducibility." We begin with the simplest, DANCE CHOREOGRAPHY (DC):

DANCE CHOREOGRAPHY (DC)

INSTANCE: A finite set *Figures*; a set *Follow* of tuples $\langle f_i, f_j \rangle$, $f_i, f_j \in$ *Figures*; $n \in \mathbb{N}$; $f_{start}, f_{end} \in$ *Figures*; a set *Compulsory* $\subseteq$ *Figures*.

QUESTION: Is there a sequence *Amalgamation* = $\langle f_1, f_2, \ldots, f_n\rangle$, $f_i \in$ *Figures*, $1 \le i \le n$, and $\langle f_j, f_{j+1} \rangle \in$ *Follow*, $1 \le j < n$ such that $f_1 = f_{\text{start}}$ and $f_n = f_{\text{end}}$, $\{f_i : 1 \le i \le n\} \supseteq$ *Compulsory*?

This first decision problem (DC) is proven to be NP-Complete by choosing another, previously proven, NP-Complete problem and showing how it can be transformed into DC in polynomial time. We use the NP-Complete problem known as HAMILTONIAN CIRCUIT (HC) (Karp, 1972). Specifically, we show that DC is a more general version of HC (Gwee, 2012b). The following is the definition of HC, followed by the proof that DC is NP-Complete.

HAMILTONIAN CIRCUIT (HC)

INSTANCE: A graph *G*.

QUESTION: Does *G* have a Hamiltonian Circuit (simple cycle containing all the vertices)?

THEOREM 2.1: DC is NP-Complete.
PROOF: Restrict to HC: Make $n = |Figures| + 1$, *Compulsory* = *Figures*, $f_{\text{start}} = f_{\text{end}}$. ♦

We now define the optimization version of DC, which is DANCE CHOREOGRAPHY OPTIMIZATION (DCO). In the latter problem, we modify the decision problem as follows:

DANCE CHOREOGRAPHY OPTIMIZATION (DCO)

INSTANCE: A finite set *Figures*; a set *Follow* of tuples $\langle f_i, f_j\rangle$, $f_i, f_j \in$ *Figures*; $n \in \mathbb{N}$; $f_{\text{start}}, f_{\text{end}} \in$ *Figures*; a set *Compulsory* $\subseteq$ *Figures*.

QUESTION: What is the optimal *Amalgamation* = $\langle f_1, f_2, \ldots, f_n\rangle$, $f_i \in$ *Figures*, $1 \le i \le n$, and $\langle f_j, f_{j+1} \rangle \in$ *Follow*, $1 \le j < n$, in the sense that $|\{\text{distinct elements } f_1, \ldots, f_n\}|$ is maximal, that satisfies $f_1 = f_{\text{start}}$ and $f_n = f_{\text{end}}$, $\{f_i : 1 \le i \le n\} \supseteq$ *Compulsory*?

In this definition, the optimality measure is that of set cardinality: the more varied the figures used in the amalgamation, the better. Informally, the optimization problem is one of maximizing the variety of figures used. This problem is NP-Hard.

To facilitate proving that this is so, we define a new decision problem, which is a modification of DC by introducing the notion of the distinctness of the figures used in the amalgamation. The modified problem, DANCE CHOREOGRAPHY k (DCk), is defined below, and its NP-Completeness proof follows.

DANCE CHOREOGRAPHY *K* (DCK)

INSTANCE: A finite set *Figures*; a set *Follow* of tuples $\langle f_i, f_j \rangle$, $f_i, f_j \in$ *Figures*; $n \in \mathbb{N}$; $f_{\text{start}}, f_{\text{end}} \in$ *Figures*; a set *Compulsory* $\subseteq$ *Figures*; $k \in \mathbb{N}_0$.

QUESTION: Is there a sequence *Amalgamation* = $\langle f_1, f_2, \dots, f_n \rangle$, $f_i \in$ *Figures*, $1 \leq i \leq n$, and $\langle f_j, f_{j+1} \rangle \in$ *Follow*, $1 \leq j < n$ such that $f_1 = f_{\text{start}}$ and $f_n = f_{\text{end}}$, $\{f_i : 1 \leq i \leq n\} \supseteq$ *Compulsory*, $k \leq |\{\text{distinct elements } f_m : m = 1..n\}|$?

THEOREM 2.2: DCk is NP-Complete.
PROOF: Restrict to DC: Make $k = 0$. ♦

We can now use DCk to show that DCO is NP-Hard. This is done by showing that DCk is “Turing-reducible” to DCO, as follows.

THEOREM 2.3: DCO is NP-Hard.
PROOF: DCk is Turing-reducible to DCO in the following manner: Let *max* be the optimal solution for an instance of DCO, where $max = |\{f_m : m = 1..n\}|$. Then, the answer for an instance of DCk is “yes” if $k \leq max$, “no” otherwise. ♦

In DCO the only optimality measure is the cardinality of the set of figures used in the amalgamation. Naturally, the problem can be generalized by generalizing the optimality measure itself. For example, other optimality measures could be included, such as the simplicity (or difficulty) of the figures used, or the inclusion of certain popular sub-sequences; these measures could then be integrated within a fitness function $F: \{\langle f_1, \dots, f_n \rangle\} \rightarrow \mathbb{N}_0$. The practical version of the DCO, using various optimality measures, is the one whose solution implementations are discussed in the next section. This version, DANCE CHOREOGRAPHY OPTIMIZATION 2 (DCO2), is also NP-Hard.

DANCE CHOREOGRAPHY OPTIMIZATION 2 (DCO2)

INSTANCE: A finite set *Figures*; a set *Follow* of tuples $\langle f_i, f_j \rangle$, $f_i, f_j \in$ *Figures*; $n \in \mathbb{N}$; $f_{\text{start}}, f_{\text{end}} \in$ *Figures*; a set *Compulsory* $\subseteq$ *Figures*; a function F: $\{\langle f_1, f_2, \ldots, f_n \rangle\} \rightarrow \mathbb{N}_0$.

QUESTION: What is the *Amalgamation* $= \langle f_1, f_2, \ldots, f_n \rangle, f_i \in$ *Figures*, $1 \leq i \leq n$, and $\langle f_j, f_{j+1} \rangle \in$ *Follow*, $1 \leq j < n$, that satisfies $f_1 = f_{\text{start}}$ and $f_n = f_{\text{end}}$, $\{f_i : 1 \leq i \leq n\} \supseteq$ *Compulsory* for which $F(\langle f_1, f_2, \ldots, f_n \rangle)$ is maximal?

THEOREM 2.4: DCO2 is NP-Hard.
PROOF: An algorithmic solution with specialization of the evaluation function F of DCO2 to compute set cardinality provides the solution to DCO. ♦

Along the same lines, we now define a group of decision and optimization problems that includes the specification of a predefined subsequence that the solution amalgamation must contain: the subsequence may contain wildcard figures. This group of problems forms the basis of the instructional modules whose implementation will be described in the following sections. We shall define the decision problems called DANCE CHOREOGRAPHY 2 (DC2), DANCE CHOREOGRAPHY 2 *k* (DC2k), and the corresponding optimization problems DANCE CHOREOGRAPHY 2 OPTIMIZATION (DC2O), and DANCE CHOREOGRAPHY 2 OPTIMIZATION 2 (DC2O2). For completeness, we also outline the appropriate NP-Completeness and NP-Hardness proofs.

DANCE CHOREOGRAPHY 2 (DC2)

INSTANCE: A finite set *Figures*; a set *Follow* of tuples $\langle f_i, f_j \rangle$, $f_i, f_j \in$ *Figures*; $n \in \mathbb{N}$; $f_{\text{start}}, f_{\text{end}} \in$ *Figures*; a set *Compulsory* $\subseteq$ *Figures*; a sequence *Fixed* of figures $f_i \in$ *Figures* $\cup$ $\{any\}$, $1 \leq i \leq n$.

QUESTION: Is there a sequence *Amalgamation* $= \langle f_1, f_2, \ldots, f_n \rangle$, $f_i \in$ *Figures*, $1 \leq i \leq n$, and $\langle f_j, f_{j+1} \rangle \in$ *Follow*, $1 \leq j < n$ such that $f_1 = f_{\text{start}}$ and $f_n = f_{\text{end}}$, $\{f_i : 1 \leq i \leq n\} \supseteq$ *Compulsory*, $f_i = Fixed_i$ if $Fixed_i \neq any$?

THEOREM 2.5: DC2 is NP-Complete.
PROOF: Restrict to DC: Make $f_1 = f_n$ and $f_i = any$, $1 \neq i \neq n$. ♦

DANCE CHOREOGRAPHY 2 *K* (DC2K)

INSTANCE: A finite set *Figures*; a set *Follow* of tuples $\langle f_i, f_j \rangle$, $f_i, f_j \in$ *Figures*; $n \in \mathbb{N}$; $f_{\text{start}}, f_{\text{end}} \in$ *Figures*; a set *Compulsory* $\subseteq$ *Figures*; a sequence *Fixed* of figures $f_i \in$ *Figures* $\cup$ $\{any\}$, $1 \leq i \leq n$; $k \in \mathbb{N}_0$.

QUESTION: Is there a sequence *Amalgamation* $= \langle f_1, f_2, \ldots, f_n \rangle$, $f_i \in$ *Figures*, $1 \leq i \leq n$, and $\langle f_j, f_{j+1} \rangle \in$ *Follow*, $1 \leq j < n$ such that $f_1 = f_{\text{start}}$ and $f_n = f_{\text{end}}$, $\{f_i : 1 \leq i \leq n\} \supseteq$ *Compulsory*, $f_i = Fixed_i$ if $Fixed_i \neq any$, $k \leq |\{\text{distinct elements } f_m : m = 1..n\}|$?

THEOREM 2.6: DC2k is NP-Complete.
PROOF: Restrict to DC2: Make $k = 0$. ♦

DANCE CHOREOGRAPHY 2 OPTIMIZATION (DC2O)

INSTANCE: A finite set *Figures*; a set *Follow* of tuples $\langle f_i, f_j \rangle$, $f_i, f_j \in$ *Figures*; $n \in \mathbb{N}$; $f_{\text{start}}, f_{\text{end}} \in$ *Figures*; a set *Compulsory* $\subseteq$ *Figures*; a sequence *Fixed* of figures $f_i \in$ *Figures* $\cup$ $\{any\}$, $1 \leq i \leq n$.
QUESTION: What is the optimal *Amalgamation* $= \langle f_1, f_2, \ldots, f_n \rangle$, $f_i \in$ *Figures*, $1 \leq i \leq n$, and $\langle f_j, f_{j+1} \rangle \in$ *Follow*, $1 \leq j < n$, in the sense that $|\{\text{distinct elements } f_1, \ldots, f_n\}|$ is maximal, that satisfies $f_1 = f_{\text{start}}$ and $f_n = f_{\text{end}}$, $\{f_i : 1 \leq i \leq n\} \supseteq$ *Compulsory*, $f_i = Fixed_i$ if $Fixed_i \neq any$?

THEOREM 2.7: DC2O is NP-Hard.
PROOF: DC2k is Turing-reducible to DC2O in the following manner: Let *max* be the optimal solution for an instance of DC2O, where $max = |\{f_m : m = 1..n\}|$. Then, the answer for an instance of DC2k is “yes” if $k \leq max$, “no” otherwise. ♦

DANCE CHOREOGRAPHY 2 OPTIMIZATION 2 (DC2O2)

INSTANCE: A finite set *Figures*; a set *Follow* of tuples $\langle f_i, f_j \rangle$, $f_i, f_j \in$ *Figures*; $n \in \mathbb{N}$; $f_{\text{start}}, f_{\text{end}} \in$ *Figures*; a set *Compulsory* $\subseteq$ *Figures*; a sequence *Fixed* of figures $f_i \in$ *Figures* $\cup$ $\{any\}$, $1 \leq i \leq n$; a function F: $\{\langle f_1, f_2, \ldots, f_n \rangle\} \rightarrow \mathbb{N}_0$.

QUESTION: What is the *Amalgamation* $= \langle f_1, f_2, \ldots, f_n \rangle$, $f_i \in$ *Figures*, $1 \leq i \leq n$, and $\langle f_j, f_{j+1} \rangle \in$ *Follow*, $1 \leq j < n$, that satisfies $f_1 = f_{\text{start}}$ and $f_n = f_{\text{end}}$, $\{f_i$

$: 1 \leq i \leq n\} \supseteq$ *Compulsory*, $f_i = \mathit{Fixed}_i$ if $\mathit{Fixed}_i \neq$ *any* for which F ($\langle f_1, f_2, \ldots, f_n\rangle$) is maximal?

THEOREM 2.8: DC2O2 is NP-Hard.
PROOF: An algorithmic solution with specialization of the evaluation function F of DC2O2 to compute set cardinality provides the solution to DC2O. ♦

To summarize the contents of this subsection, we have defined several dance choreography decision problems, and their corresponding optimization problems. We did this in order to provide a basis for the following discussion of algorithmic solutions and implementations. We have also linked these problems conceptually to other significant problems, using NP-Complete theory. Figure 2.1 depicts the transformation sequence used in the proofs of the NP categories of the dance choreography problems defined here.

$$\begin{tikzcd} \mathrm{HC} \arrow[r, "\propto"] & \mathrm{DC} \arrow[r, "\propto"] \arrow[d, "\propto"] & \mathrm{DCk} \arrow[r, "\propto T"] & \mathrm{DCO} \arrow[r, "\propto T"] & \mathrm{DCO2} \\ & \mathrm{DC2} \arrow[r, "\propto"] & \mathrm{DC2k} \arrow[r, "\propto T"] & \mathrm{DC2O} \arrow[r, "\propto T"] & \mathrm{DC2O2} \end{tikzcd}$$

Figure 2.1. Transformation sequence of dance choreography NP-problems.

2.3 Exhaustive and Heuristic Algorithms for Dance Choreography

Because the general dance choreography optimization problems are NP-Hard, exhaustive algorithms will only work in practice for small instances. For more ambitious problem sizes, we need less "perfect" and more practical methods, such as heuristic algorithms.

The exhaustive algorithm, however practically unsuitable, is nevertheless essential to consider because it provides a benchmark to determine the effectiveness of more practical solutions. Hence, we begin with this brute-force approach. We shall be using our software implementation of the dance choreography problem, Terpsichore©, to perform demonstrative trials (beginning with Example 2.1). A more detailed description of the software itself is presented in later sections.

Example 2.1. As an indication of the exponential increase in computing time, we show what can happen as we attempt to create amalgamations of increasing length. We ran the exhaustive algorithm to search for amalgamations of increasing length from one to twelve figures. The figures were chosen from our proprietary Foxtrot syllabus. To ensure a uniform start, the opening figure in each trial was the figure called "Feather Step." The optimality measure was the degree of variety of figures: preference was given to as many different figures as possible; its values are normalized to the interval [0..10]. The experiment was performed on a 2.3 GHz Intel Core i5 processor, and processing time was measured in clock ticks (60 clock ticks equal one second). As Table 2.1 shows, processing time increased exponentially with increasing amalgamation length.

Table 2.1. Performance of the Exhaustive Algorithm (Foxtrot)

length	*time*	*optimal*	*suboptimal*	*total*
1	<1	1	0	1
2	<1	6	0	6
3	1	44	0	44
4	2	102	11	113
5	10	433	133	566
6	43	1188	1004	2192
7	192	3860	5267	9127
8	850	10276	26016	36292
9	3780	27146	124084	151230
10	16828	66810	539951	606761
11	72314	148646	2344870	2493516
12	316001	314738	9796400	10111138

A well-known heuristic approach is the so-called greedy algorithm. There is no moral connotation attached to the name of this algorithm: it merely stipulates that the search is guided by the concept of the "best" choice, with no intent to backtrack. In our dance choreography problem, we use the fitness function as defined in DCO. We implement this by several criteria, one of which is the notion of Figure Variety, used also in Example 2.1 to evaluate the amalgamations produced by the exhaustive algorithm (Example 2.2).

Example 2.2. The results of the application of the greedy algorithm to the same problem as Example 2.1 are shown in Table 2.2. The "time" column shows the clock ticks; the “fitness” column shows the optimality measure, with values in the range [0..10]. Note that the speed trade-off is in the quality of amalgamations: not all the amalgamations were optimal compared to the results obtained by the exhaustive algorithm. Nevertheless, an overall acceptable quality was obtained in this and several other trials not shown here. The running time was drastically shortened, however; in fact, none of the computations exceeded one clock tick in this experiment.

Table 2.2. Performance of the Greedy Algorithm (Foxtrot)

length	*time*	*fitness*
1	<1	10
2	<1	10
3	<1	10
4	<1	8.8
5	<1	10
6	<1	9.2
7	<1	7.9
8	<1	10
9	<1	8.3
10	<1	9.0
11	<1	9.1
12	<1	8.8

Besides the greedy algorithm, other heuristics-based algorithms could be used. One promising method is the genetic algorithm. We have used this approach in the NP-Hard problem of Species Counterpoint (Gwee, 2003). Smart modifications to the exhaustive algorithm itself are also possible: for example, the branch-and-bound technique to implement best-first searching.

3. Terpsichore©: Multimedia Dance Software

We now describe our dance choreography software, called Terpsichore©, from the point of view of software design and implementation. Our software generates ballroom dance amalgamations based on figures formalized by various International Standard dance associations: Dance Vision International Dance Association (DVIDA), International Dance Teachers’ Association (IDTA), and Imperial Society of Teachers of Dancing (ISTD) (DVIDA, 2008; Howard, 2007; ISTD, 1994). Based on structural and pedagogical considerations, we have compiled our own syllabus by grouping these dance figures into three categories: “Foundational,” “Intermediate,” and “Advanced.”

In §3.1, we describe the implementations of the NP-Hard problems described above. In §3.2, we discuss some interface issues in the design of the program. Finally, in §3.3, we discuss the programming languages used and the architecture of our software.

3.1 Implementation of the Optimization Problems

The various optimization problems defined above are implemented in our software under two modes, which we call “generative” and “interactive.” Under the generative mode, Terpsichore© generates amalgamations based on the input parameters of DCO2. The exhaustive algorithm that implements this problem is called “Thorough Search (with specified required figures)” (fig. 3.1). The user selects a finite set of *Figures*, specifies f_{start} and f_{end} from our proprietary syllabus (fig. 3.2), specifies the amalgamation length, n, determines the evaluation function F (fig. 3.3, showing the “Variety” criterion), and chooses a set of *Compulsory* figures (fig. 3.4). Examples 3.1 and 3.2 illustrate some results from this configuration.

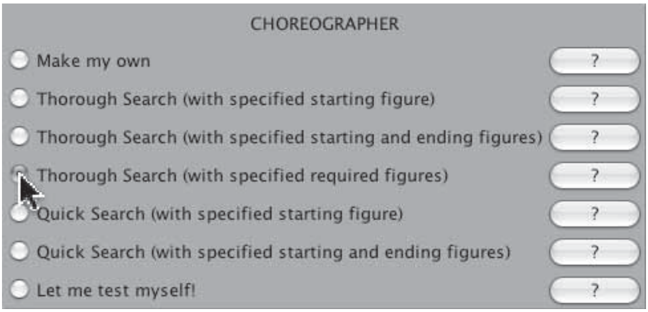


Figure 3.1. Implementations of dance choreography problems.

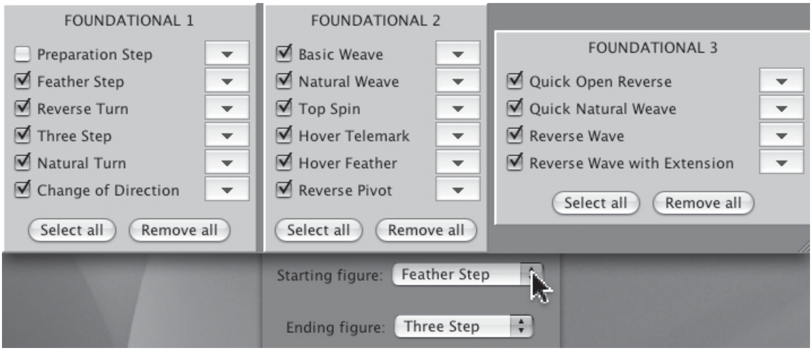


Figure 3.2. *Figures*, f_{start}, f_{end} selections.

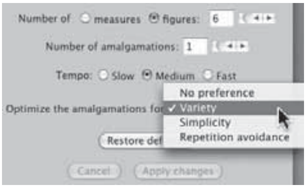


Figure 3.3. *n*, *F* selections.

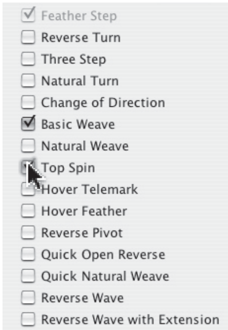


Figure 3.4. *Compulsory* figures selection.

Example 3.1. With the settings as shown in figs. 3.1, 3.2, 3.3, and *Compulsory* = ∅, 28 optimal and 36 suboptimal Foxtrot amalgamations were found (fig. 3.5 shows one of these optimal amalgamations). Note that fixing the ending figure, f_{end}, results in far fewer amalgamations, compared to the number found in the corresponding setting shown in Table 2.1. ♦

Example 3.2. With the settings as shown in figs. 3.1, 3.2, 3.3, and 3.4 (two figures chosen for the *Compulsory* set: Basic Weave and Top Spin), only 4 optimal and 0 suboptimal amalgamations were found (fig. 3.6). Not surprisingly, with the added restriction imposed by the compulsory figures, the overall number of optimal amalgamations is substantially lower than that obtained in Example 3.1. ♦

1. Feather Step.
2. Top Spin.
3. First three steps of a Reverse Turn.
4. Reverse Pivot.
5. Reverse Turn.
6. Three Step.

10

I found a total of 28 optimal and 36 suboptimal amalgamation(s).

Figure 3.5. Output for *Compulsory* = ∅.

1. Feather Step.
2. Top Spin.
3. Quick Open Reverse.
4. First 4 steps of a Reverse Wave.
5. Basic Weave.
6. Three Step.

10

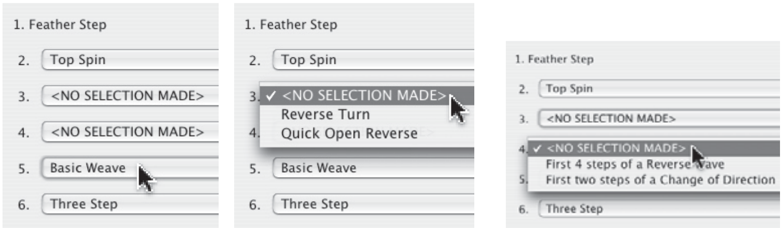
I found a total of 4 optimal and 0 suboptimal amalgamation(s).

Figure 3.6. Output for |*Compulsory*| = 2.

Under the interactive mode, the user collaborates with the software in the creation of amalgamations. In this mode, there are two implementations of DC202, called "Make my own" and "Let me test myself" (see options shown in fig. 3.1), respectively. In both implementations, the user selects a finite set of *Figures* from our proprietary syllabus, specifies the amalgamation length, *n*, determines the evaluation function *F* (fig. 3.3, showing the "Variety" criterion), and chooses a subsequence of figures (fig. 3.7). These implementations omit the specification for *Compulsory*, i.e., they assume *Compulsory* = ∅. Note that if the user specifies every figure in the subsequence, the algorithm amounts to a verification of the subsequence; if the user specifies none of the figures, the algorithm amounts to the generation of an amalgamation performed entirely by the software; if the user specifies some but not all of the figures, this amounts to the software "filling in the blanks" in producing an amalgamation.

The difference between the "Make my own" and "Let me test myself" options is that in the former the user is guided in the making of the subsequence, being shown only the figures that are possible at any particular stage of the amalgamation; in the latter, the user is not so guided, and is thus "tested" in his/her knowledge of the syllabus. Example 3.3 illustrates options given to the user in the "Make my own" interactive mode.

Example 3.3. Using the "Make my own" module, and with the settings as shown in figs. 3.2, 3.3, and 3.4, and the subsequence fixed as in fig. 3.7a, Terpsichore© gives two options for the third figure in the amalgamation (fig. 3.7b) and two more options for the fourth figure in the amalgamation (fig. 3.7c). In this instance, each of the two options for the third figure enables two options for the fourth figure, giving a total of four possible amalgamations, the same result as obtained with the settings in Example 3.2. ♦



(a) (b) (c)

Figure 3.7. Options offered in the "Make my own" interactive module.

3.2 User-Oriented Design

We describe a number of design decisions made in the development of Terpsichore©. The goal of these is to simplify the user interface, and to provide the user with the least stressful experience possible.

When the user launches the program, all that is shown is a start-up screen, with one button, which is therefore the default. For good measure, the cursor is even brought to this default button (fig. 3.8). Naturally, like most software, Terpsichore© has many features. But, as in most software, there is a primary task: in word processing, for example, it might be to create and populate a text document with words; in spreadsheet calculation, it might be to create and populate a spreadsheet with numbers. Most other functions are incidental to the main task. In our case, the main task is to create dance amalgamations. The user, especially the first-time user, should be granted immediate access to this primary feature. Others can be explored later. Too often, when launching a program, especially for the first time, one is confronted with a multitude of choices. The user either gets lost and does not know where to begin, or gets infuriated because, having acquired the program presumably in order to accomplish an immediate task, they find they will have to think too hard and too long before setting off to accomplish this immediate task. All the fancy features splashed out in an equally fancy interface will come to nothing for an impatient user. Apart from Google®, it is somewhat amazing that more software designers do not take simplicity more seriously.

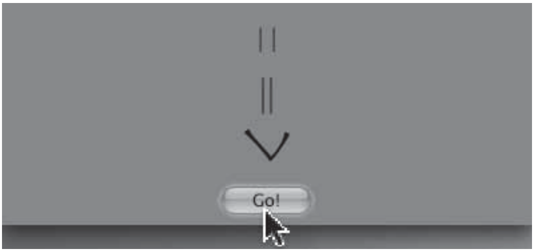


Figure 3.8. Minimalist window shown at program launch.

Another aspect that contributes to a more natural human–computer environment is the non-predictability of the computer response to requests. Human beings do not respond the same way to the same request every time: this applies to simple answers, as well as more complex responses. An element of unpredictability is always welcome, and makes the computing experience a fresh and invigorating one for the user, who sits down and never knows quite what to expect. Courtesies, like “thank you,” and “you’re welcome” never hurt, either. We think this makes our software fun to use (Example 3.4).

Example 3.4. Figure 3.9 illustrates the variety of responses Terpsichore© gives before generating an amalgamation. The balloons here represent visually what actually occurs aurally. In fig. 3.10, Terpsichore© warns the user that a certain operation may take a long time, owing to the exponential time required for an exhaustive search; this notice is triggered when the user has chosen an amalgamation length exceeding a threshold value, say of eight figures. Figure 3.11 illustrates the reading of certain Help texts, including the observance of common courtesies. ♦

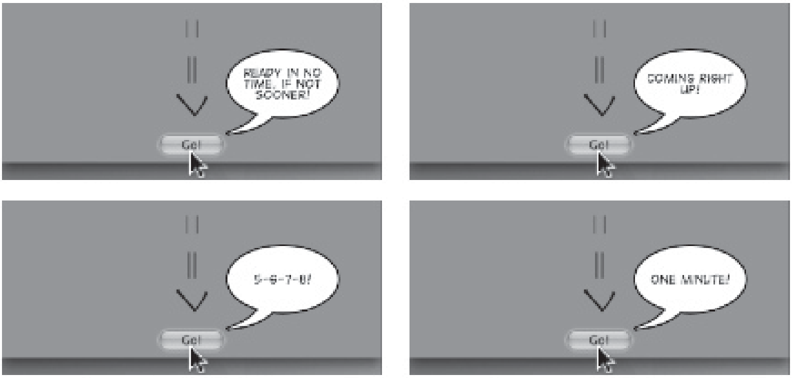


Figure 3.9. Terpsichore©'s response variety.

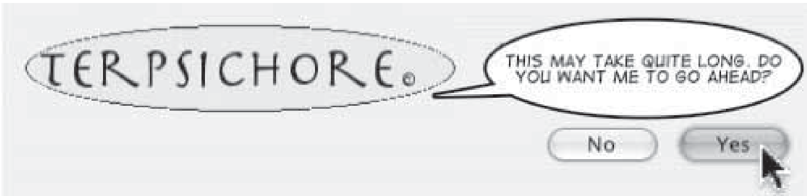


Figure 3.10. A gentle warning of a potential long computation.

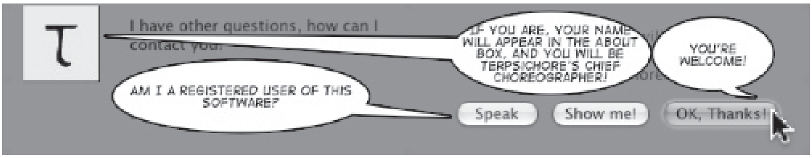


Figure 3.11. Help window courtesies.

Besides the interface, the output itself can be varied, since, as the experiments described above show, there is no shortage of amalgamations satisfying the same set of parameters. In our implementation, the search algorithms incorporate random choices in the various branches of the search tree (Example 3.5).

Example 3.5. With the settings as shown in figs. 3.2, 3.3, 3.4, and either "Thorough Search (with specified required figures)" or "Make my own," Terpsichore© gives any one of four possible amalgamations at various times (figs. 3.12a–3.12d). Compare these with those shown in Examples 3.2 and 3.3. ♦

1. Feather Step.
2. Top Spin.
3. Reverse Turn.
4. First 4 steps of a Reverse Wave.
5. Basic Weave.
6. Three Step.

10

I found a total of 4 optimal and 0 suboptimal amalgamation(s).

(a)

1. Feather Step.
2. Top Spin.
3. Reverse Turn.
4. First two steps of a Change of Direction.
5. Basic Weave.
6. Three Step.

10

I found a total of 4 optimal and 0 suboptimal amalgamation(s).

(b)

1. Feather Step.
2. Top Spin.
3. Quick Open Reverse.
4. First 4 steps of a Reverse Wave.
5. Basic Weave.
6. Three Step.

10

I found a total of 4 optimal and 0 suboptimal amalgamation(s).

(c)

1. Feather Step.
2. Top Spin.
3. Quick Open Reverse.
4. First two steps of a Change of Direction.
5. Basic Weave.
6. Three Step.

10

I found a total of 4 optimal and 0 suboptimal amalgamation(s).

(d)

Figure 3.12. Four possible amalgamations, any one given randomly at various times.

One of the basic principles of graphical user interface design is to allow the user to "call the shots." One user's desirable feature may be another's bane. One of the most annoying features of Microsoft's Word® program is that it sometimes appears too "intelligent," correcting the user's capitalization of the first letter of a sentence, for instance: this should really not be the default action. No adult user likes to be told what is good for them. Terpsichore© would like to remind the user to save changes, but some users do not want to be reminded constantly or, indeed, do not really care to save anything anyway. Therefore, the user is afforded a wide range of choices in preference settings (fig. 3.13).

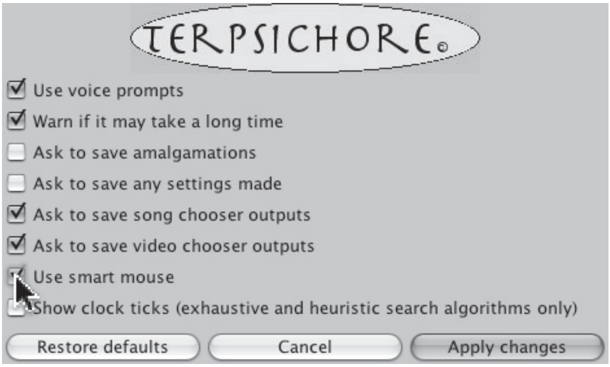


Figure 3.13. User calls the shots in the preferences window.

Different learners learn in different ways. One of the advantages of multimedia tools is the availability of a variety of learning approaches. After the amalgamations are derived, Terpsichore© offers a range of playback options, visual and aural, including a combination of both. The aural playback is timed to coincide with the duration of individual figures and foot movements, so that the user can practice the amalgamations in real time (Example 3.6).

Example 3.6. Figure 3.14 illustrates the cueing of the individual foot movements in the figure "Quarter Turn to the Right" of the Waltz. The cueing is timed to the duration of the individual movements, in this case in Waltz timing. Again, the balloons here represent visually the actual aural cues. ♦

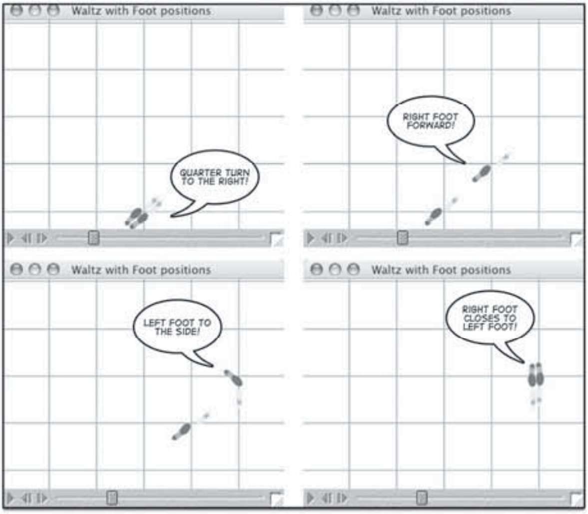


Figure 3.14. Cueing of foot movements in a Waltz figure.

3.3 Implementation of Terpsichore©

Despite claims to the contrary, no one general-purpose programming language is ideal for all purposes. Terpsichore© was developed using four general-purpose languages: C++, Java, Objective-C, and Objective-C++. Each of the languages was chosen because of special capabilities it provided. In addition, the scripting language AppleScript was used to automate several multimedia features.

C++ (ISO/IEC TR 19768:2007) was used to program the search engines implementing the dance choreography problems. Much use was made of the template feature of C++ that enabled many versions of the implementations to be realized economically. Thus, the Thorough Search involving the specification of the first figure, the first and last figures, subsequence of figures, and compulsory figures were all different versions of essentially the same implementation. The template feature facilitated the programming of all these versions with minimal programmer code duplication.

Java (J2SE 5.0) was chosen because it provided a convenient method to program the user interface. All the visible aspects of the program, such as the windows shown in several figures above, were realized by the Swing components provided in the Java language. Thus, consistency is assured across several platforms.

The Java Native Interface was used to link the Java and C++ components. The connection between the two components was the user choice of the various search algorithms, basically the Thorough Search and the Quick Search: the Java modules provided the interface shown in fig. 3.1, and the C++ modules provided the functionality. The “Make my own” and “Test myself” interactive modes used the Java modules to obtain user input, and the C++ modules to perform the search for suitable amalgamations and to verify user input. The combination of the C++ and Java modules, and the Java Native Interface glue mechanism, constitutes the portable aspect of the software implementation.

The Objective-C 2.0 and Objective-C++ languages were chosen to take advantage of the Macintosh® OS X Operating System’s advanced multimedia capabilities. Together with AppleScript, the audio and video components were realized. This part of the implementation is not portable. Figure 3.15 illustrates the various modules (represented as classes), and the languages used to program them.

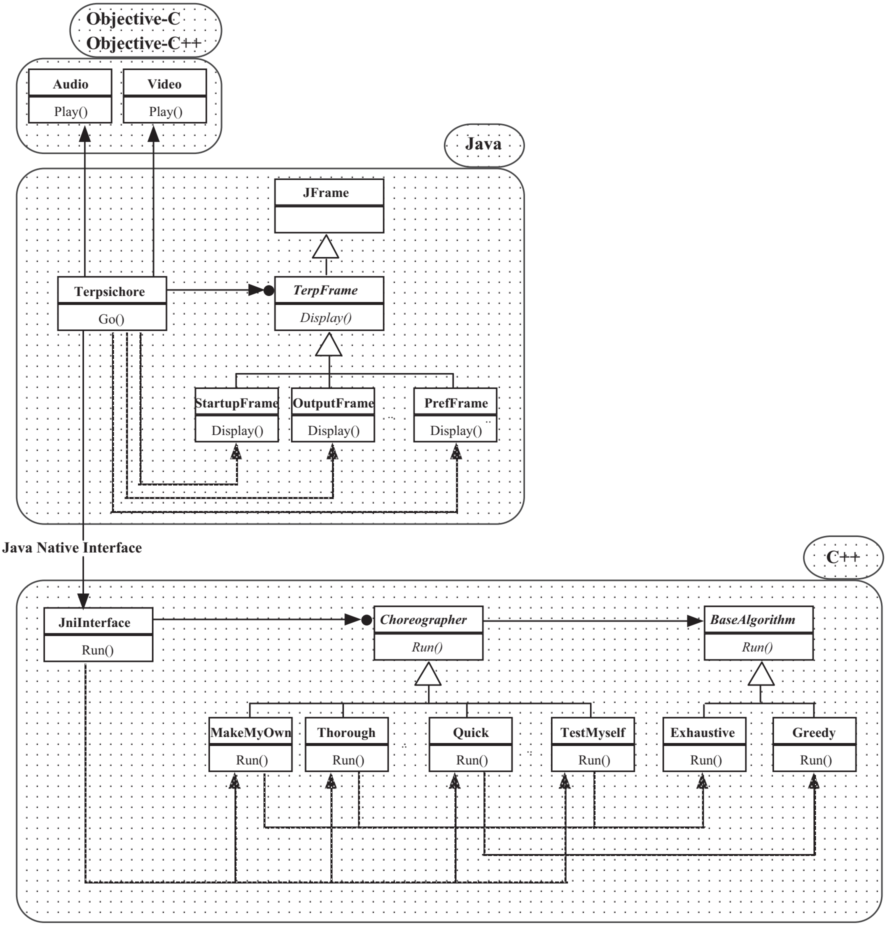


Figure 3.15. Classes and languages used in the software.

The architecture of this software enables modification in its interface to provide, for example, an on-demand service for dance choreography. The portable components of the architecture can be used almost unchanged, whereas the multimedia aspects need to be reworked to provide extensible services.

Dance is a three-dimensional activity; ballroom dance is not only that, but is also unique in the dance arts in that it is a partnering activity. One of the challenges in communicating dance choreography is to develop the notation required: online instruction is made difficult by the inability of current notation to convey completely and accurately the intentions of the choreographer, both the basic mechanical aspects as well as the interpretative character of the choreography. Nevertheless, the architecture makes it possible for future implementations to convey such concepts satisfactorily. Perhaps future enhancements of three-dimensional audio-visual rendering can be incorporated into this software to make it a viable instructional medium.

4. Software Copyright Issues

Terpsichore© was copyrighted on July 1, 2011 (Registration Number TXu 1-762-479) (Gwee, 2012c). A prototype was demonstrated at the 85th Annual Meeting of the Louisiana Academy of Science, 2011 (Gwee, 2011). The complete software was demonstrated at the Inaugural Meeting of the National Academy of Inventors, 2012 (Gwee, 2012a). We shall briefly describe the circumstances under which this software was copyrighted. First, we outline some landmarks in the development of software copyright in the United States, and then highlight copyright issues involved in our application for copyrighting Terpsichore©.

4.1 A Brief History of Software Copyright

Copyrighting software is a relatively recent area in copyright law. The legal profession is typically very conservative, and the glaring contrast between the march of technology versus the apparent insistence of the legal profession to cling to outdated methods of assessment is never more apparent than in the case of software copyright.

It was only as recently as 1974 that the Commission on New Technological Uses of Copyrighted Works (CONTU) decided that computer programs are copyrightable "… to the extent that they embody an author's original creation." The Copyright Act of 1976 states that "… original works of authorship fixed in any tangible medium of expression, …" are copyrightable ("Copyright Act," 1978). Although this act was effective as of 1978, in practice the notion of originality being linked to creativity was made explicit much later, in the landmark case Feist Publications, Inc. v. Rural Telephone Service Co., 1991 ("Feist v. Rural," 1991).

In 1998, a group of computer scientists (Hal Abelson, Roy Campbell, Randall Davis, Lee Hollaar, and Gerald J. Sussman) submitted an amicus brief in which they suggested ways to analyze a computer program with respect to copyright protection ("Amicus Brief," 1998). Their timely suggestions have, unfortunately, not enjoyed much consideration beyond the case for which they were written.

4.2 Copyrighting Terpsichore©

When we applied for copyright of our software, we tried to follow all the rules spelt out for us. Although it seemed easy enough to satisfy the requirements of the US Copyright Office, we felt that the requirements were inadequate in capturing the essence and uniqueness of our software.

In the United States and Canada, computer programs are considered literary works for copyright purposes ("Apple v. Franklin," 1983; Copyright Act of 1976). Legal authorities consider such an analogy only partially appropriate, however (Newman, 1999), a view with which we concur. The current application process for computer programs requires twenty-five pages of the first part of the source code, and twenty-five pages of the last part of the source code, and appears to be partly inspired by this imperfect analogy (United States Copyright Office, 2011).

We have argued that this requirement also ignores the fact that most computer programs are not read in linear fashion, as literary works are. Moreover, twice twenty-five pages hardly represents most computer code nowadays. We have suggested other metrics, such as lines-of-code, and other ways of demonstrating creativity, such as original adaptations of what are called "library code," no matter where they appear in the overall structure. A program such as Terpsichore© would be better represented with some of these changed metrics (Gwee, 2012c).

5. Conclusion

We have formulated several Dance Choreography problems and shown them to be variously NP-Complete and NP-Hard. These classifications connect the Dance Choreography problems directly to other problems in computer science and related areas of computational research. We described various algorithmic solutions to these problems, and demonstrated their implementation in our multimedia software, Terpsichore©. We also discussed and illustrated user-interface and portability issues in the design of the software.

Future work on Terpsichore© will also include the addition of more realistic computer animation including 2D- and 3D-modeling featuring partner dancing, a uniquely difficult but not insurmountable task (Herbison-Evans, Green, & Butt, 1982; Sealey, 1982). More sophisticated evaluation functions and algorithms to measure aesthetics in the various amalgamations produced will be another fruitful avenue of research. The extensibility and portability of the software promises more wide-ranging online features. Such future development of the software will certainly further challenge the established legal concepts of software copyright.

Acknowledgements

The author wishes to thank Attorney Frances Ball for legal advice in the copyrighting of the software. Mr. William (“Bill”) Davies, Dance Sport Adjudicator and Examiner, and former North American and United States Ballroom Dance Champion, gave invaluable artistic advice during the development of Terpsichore©, for which the author is also very grateful.

References

Amicus Brief in Support of Computer Scientists, Harbor Software, Inc. v. Applied Systems, Inc., (2nd Cir. 1998) (No. 97-7197).

Apple Computer, Inc. v. Franklin Computer Corp., 714 F. 2d 1240 (3rd Cir. 1983).

Cook, S. A. (1971). The Complexity of Theorem-Proving Procedures, *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, Shaker Heights, Ohio: Association for Computing Machinery, 151–158.

Copyright Act of 1976, 17 U.S. Code, ch. 1, § 101–102 (1978).

Dance Vision International Dance Association. (2008). *Bronze (Silver/Gold) Level International Style Standard Manual*, Las Vegas, NV: Dance Vision.

Feist Publications, Inc. v. Rural Telephone Service Co., 499 U.S. 340 (1991).

Gwee, N. (2003). Feasibility of Genetic Algorithms for Music Composition, *Proceedings of the International Conference on Artificial Intelligence*, Las Vegas, Nevada, 428–433.

—. (2011). Implementation of an NP-Complete Dance Choreography Problem, *Louisiana Academy of Science 85th Annual Meeting*, University of Louisiana at Monroe, Monroe, Louisiana.

—. (2012a). Copyrighting Terpsichore©: Multimedia Dance Software, *National Academy of Inventors Inaugural Annual Conference*, Tampa, Florida.

—. (2012b). Terpsichore©: From NP-Complete Problem to Multimedia Software, *Proceedings of the 3rd Annual International Software Engineering & Applications Conference*, Singapore, 189–193.

—. (2012c). Software Copyright: A Programmer's Perspective, *Technology and Innovation*, 14 (3), 237–247.

Herbison-Evans, D., Green, R. D., & Butt, A. (1982). Computer Animation with NUDES in Dance and Physical Education, *Australian Computer Science Communications*, 4 (1), 324–331.

Howard, G. (2007). *Technique of Ballroom Dancing*, Brighton, England: KenadS Design & Print.

Imperial Society of Teachers of Dancing. (1994). *The Ballroom Technique*, London, England: Lithoflow Ltd.

Karp, R. M. (1972). Reducibility among Combinatorial Problems. In: Miller, R. E., & Thatcher, J. W. (Eds.), *Complexity of Computer Computations* (pp. 85–103), New York, NY: Plenum.

Newman, J. O. (1999). New Lyrics for an Old Melody: The Idea/ Expression Dichotomy in the Computer Age, *Cardozo Arts & Entertainment Law Journal*, 17 (3), 691–703.

Sealey, D. (1982). Notate: Computerized Programs for Labanotation, *Journal for the Anthropological Study of Human Movement*, 1 (2), 70–74.

United States Copyright Office. (2011). Circular 61: Copyright Registration for Computer Programs. Revised 05/2011.

INTERNET OF THINGS (IOT), CLOUD COMPUTING, AND BIG DATA

HUIFANG DENG, CAIFENG ZOU
AND NAZIA ANWAR

"For us, nowadays ICT can be interpreted as the Internet of Things, Cloud Computing and the Theory of Big Data"

—Authors

Abstract

The Internet of Things (IoT) is the most important concept of the Future Internet for providing a common global IT platform to integrate seamless networks and networked things. As a rising industry, IoT has to deal with a number of difficulties and challenges; but, at the same time, cloud computing may facilitate the growth of IoT. IoT and cloud computing are the hottest topics of the Future Internet. Cloud computing not only provides facilities for the computation and processing of Big Data, but also serves as a service model. In this chapter we give the definition of IoT, Cloud Computing, and Big Data, discuss the key technologies involved, briefly introduce their applications, and address the existing problems and challenges. The primary goal of this chapter is to provide a very brief introduction to those interested in doing further research on these topics or exploring their various potential applications.

Keywords: Internet of Things (IoT), Cloud Computing, Big Data, Key Technologies, Future Internet, Smart City

1. Introduction to IoT, Cloud Computing, and Big Data

Currently, the Internet of Things (IoT) and cloud computing are the hottest topics regarding the Future Internet. IoT is the most important concept pertaining to the Future Internet for providing a common global IT platform to integrate seamless networks and networked things. Cloud computing provides backend solutions for processing huge data streams and the computational capability that will be needed once everything becomes connected to seamless networks, as is anticipated in the near future.

1.1 Internet of Things (IoT)

The Internet of Things, or IoT, is a network of interconnected items embedded with sensors, computing systems, and Internet connectivity capabilities. IoT is defined as the worldwide dynamic global autonomous network of interconnected, uniquely identifiable physical or virtual things, devices, or objects, based on standard network protocols. IoT refers to a revolutionary paradigm, consisting of a continuum of uniquely addressable intelligent objects capable of actively communicating and accessing information from their surroundings.

IoT is a grand internetwork, or an extended Internet. Through integrating sensing equipment/devices/technologies such as RFID technology, infrared sensors, GPS, GIS, or laser scanners, it connects the existing information-oriented Internet with any physical object in the world, and conducts information exchange and communication via sets of agreed protocols with a view to realizing the intelligent recognition, positioning, tracking, monitoring and management of objects. It is a conceptual network that enables end-users of the present Internet to interact directly with any physical object in the world which has the capability of communicating information. Essentially, IoT connects all physical objects to the existing Internet via sensing devices to realize their intelligent identification and management. IoT has great applications and raises great business opportunities.

The physical and virtual things or objects in IoT have unique identities, physical attributes, intelligent interfaces, and virtual characteristics, and are connected with the information, business, and social networks. These smart things in IoT actively interact with the environment through communicating data and information, while responding autonomously to the events by triggering actions.

IoT is an efficient infrastructure that integrates many services and applications, such as digitizing personal activities and creating smart cities. The scope of IoT is vast and affects every aspect of our lives, including industry, health, security, shopping, entertainment, transportation, and a lot more. IoT applications enable things, techniques, and devices to become smarter, more intelligent, more autonomous, more efficient, and more reliable.

1.2 Cloud Computing

According to the National Institute of Standards and Technology (NIST), the definition of *cloud computing* is "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" (NIST, 2011). Although many conflicting definitions of cloud computing have been proposed, the above definition is the most widely acknowledged. The main reason for the various ambiguous definitions of the term "cloud computing" is perhaps due to the interplay of infrastructure, architecture, software, storage, and service models deployed in many cloud computing applications.

Cloud computing has become one of the most powerful and economical paradigms in terms of technology, architecture, and IT services in modern ICT for performing complex computations in large-scale scientific and business applications. A large number of organizations have been shifted to the cloud computing environment. The benefits of cloud computing include virtualized resources, elasticity, scalability, flexibility of services, on-demand delivery of resources, efficiency, parallel processing, cost saving, and so on.

Cloud computing is a form of utility computing where IT resources are delivered on a Pay-As-You-Use basis, just like electricity is charged based on usage. Cloud computing has transformed the IT industry into a more economical and flexible model for the sharing of resources. The users of cloud computing no longer need to be worried about investing capital in buying infrastructure for the deployment of their services. Instead, they can easily lease resources from the cloud service providers in real time. Using cloud computing, large organizations can process their large-scale jobs more quickly, since using a hundred servers for an hour outweighs the benefits of using one server for a hundred hours. The paradigm of cloud computing has undoubtedly revolutionized the IT industry (Armbrust et al., 2010).

In cloud computing, users can run ubiquitous applications and store/access their data on an off-site, location-transparent, centrally managed, shared platform (Creeger, 2009). The resources provided by the cloud are usually run on virtual machines paralleled with the physical infrastructure.

The most common types of service models, based on the packages offered by the cloud providers, are outlined below:

(1) Infrastructure as a Service (IaaS): Products offered by the cloud providers in this service model include the delivery of infrastructure equipment (e.g., hardware components, virtual machines, computing servers, networking, storage devices, etc.). Clients are charged on a usage-basis. Examples of IaaS include HP Cloud (HP Cloud, 2015), Amazon Elastic Cloud Compute (EC2) (Amazon, 2015), and Windows Azure Virtual Machines (Microsoft, 2015).

(2) Platform as a Service (PaaS): Services offered by the cloud providers in this service model include computing platforms along with implicit infrastructure to host the resources of the platform. The clients create/run their applications on better-equipped platforms provided by the cloud providers, such as Hadoop MapReduce applications which can be deployed over the Amazon Elastic MapReduce service (Amazon Elastic MapReduce, 2015).

(3) Software as a Service (SaaS): In this service model, entire applications with their installation and configuration management are offered by the cloud provider through the Internet. The clients no longer need to be concerned about managing complex hardware and software. Examples of SaaS are Microsoft Office 365 (Microsoft, 2015) and Google Apps (Google, 2015).

1.3 Big Data

There has been a tremendous growth in the volume of structured and unstructured data produced by organizations, IoT, and social media. This fast and sudden growth of Big Data requires a correspondingly fast development of Big Data technologies provided through virtualization platforms. Early users of Big Data in the cloud environment were the clients who deploy Hadoop clusters in cloud environments, such as Amazon AWS, Microsoft Azure, and IBM.

Big Data refers to the huge increase in the volumes of data that needs advanced technologies and techniques to capture, store, process, distribute, manage, and analyze it. The terms *volume*, *variety*, and *velocity* were initially presented by Gartner to describe the challenges of Big Data. Gantz and Reinsel proposed that the elements of Big Data are characterized by 4Vs, namely, *volume*, *variety*, *velocity*, and *value* (Gantz and Reinsel, 2011).

(1) *Volume* refers to the continuously expanding magnitude of all types of data produced from various sources, such as IoT, social media, multimedia, and so on. The size of Big Data is often measured in multiple terabytes or petabytes. Data analysis techniques can be employed to create hidden patterns and information from the Big Data. Laurila et al. collected longitudinal data from smart mobile devices and presented interesting research such as visualization patterns for complex data and prediction of behavior patterns (Laurila et al., 2012). Collecting longitudinal data requires considerable effort and underlying investments.

(2) *Variety* refers to the various structures of data gathered through IoT, social networks, smart devices, or sensors. The data collected can be structured, semi-structured, or unstructured. Structured data consists of organized tabular data collected from databases and spreadsheets. Data collected from smart mobile devices, sensors, social media, blogs, videos, images, text, logs, online games, and audio is often unstructured. Data produced by the Web users is often semi-structured (Gandomi and Haider, 2015).

(3) *Velocity* refers to the rate of data generation and speed of data transfer and analysis. The Big Data streamed from smart devices and sensors require real-time instantaneous analysis and intelligence.

(4) *Value* is a crucial characteristic of Big Data. It is the process of rapidly determining hidden patterns or information from various types of Big Data (Gandomi and Haider, 2015).

1.4 Combination of IoT and Cloud Computing

The Internet of Things is a broad concept that visualizes “things or objects or devices” as being parts of a “smart world” having smart devices, smart buildings, smart cars, smart phones, smart cameras, smart RFID tags, smart traffic lights, smart multi-dimensional barcodes, smart cities, etc. This smart world involves greater connections with each other, in order to bring significant changes in our lives and work. For instance, nowadays many smart devices are connected to the Internet; smart buildings have multiple sensors to protect them from disasters or accidents and to save energy; and smart devices are used to support various fields including healthcare services, business, education, and industry. Huge volumes of data and information are generated by these devices which are connected directly or indirectly via the Internet.

Cloud-based platforms and applications help to connect objects in IoT at any place and at any time. Therefore, the front-end to access IoT is the cloud. The cloud provides the needed infrastructure, computation powers, storage and applications to interact with smart devices in real time (Rao et al., 2012).

Combining IoT with cloud computing results in exciting research dimensions. The huge amounts of data produced from the connection of the devices in IoT can only be captured, processed, stored, and transformed into valuable information through clouds. Clouds provide physical and virtual systems, applications, and tools to efficiently and intelligently process and analyze Big Data and IoT. Clients and users can access cloud computing utilities on-demand, anywhere in the world and at any time (Canellos, 2013).

1.5 Cloud-Centric Internet of Things

Integrated cloud computing and IoT need to be developed in order to build the smart world. The cloud provides a decentralized, scalable, and reliable environment for multiple users to access devices with unreliable connections and limited energy. The cloud infrastructures and platforms should support (1) the quality of service (QoS) requirements of multiple devices, users, or applications; (2) the quick development of applications and tools; (3) the reliable and rapid execution of heterogeneous jobs; (4) the rapid provision of resources in real time to process requests having critical priorities; (5) the reliable management of failures; (6) the optimization of QoS parameters; and (7) the timely replication of real-time applications in case of the failure of any resource to complete it (Gubbi et al., 2013).

1.6 Relationship between Cloud Computing and Big Data

Big Data and cloud computing are integrated to provide more efficient cloud platforms, such as Hadoop. Large Big Data sources are stored in the cloud through applications that present data visually for analysis and future decision-making. Big Data uses cloud storage instead of traditional storage devices. Big Data utilizes virtual machines on the cloud for fast processing. Hence, the cloud acts as a service model for Big Data storage and processing (Hashem et al., 2014).

2. IoT and the Smart City

Borgia presented an example of a smart city ecosystem in order to elaborate the concept of IoT. The city is the heart of every nation, providing resources like energy, water, gas, etc. to the population. Resource consumption is increasing day by day while resource production is decreasing, resulting in the deterioration of the environment. Crucial challenges like quality of life and safety in the city can be met by adopting a number of measures. A smart city operates concurrently at both the physical and virtual levels. ICT technologies enable the controlling of services, such as entertainment, health, transport, electricity, waste management, drainage systems, traffic control systems, and so on. Indeed, IoT has the potential to transform a traditional city into a smart city by providing tools like intelligent transportation systems, automatic traffic control sensors, intelligent garbage collection, energy management systems, and pollution control systems (Borgia, 2014).

The vision of the Internet of Things first became official in 2005, when a report on it was published by the International Telecommunication Union (ITU) (International Telecommunication Unit, 2005). The report aroused great interest among researchers and the business community.

More and more IoT applications are needed to develop fault-tolerant interaction with heterogeneous infrastructures. Future IoT applications will respond quickly to runtime internal or external events and will be able to engage in predictive analysis and intelligent decision-making.

2.1 Smart Grid

The smart grid is the foundation stone for managing and monitoring energy consumption in the development of the green environment. The term *smart grid* refers to the intelligent distribution of the electricity system for the transfer of energy from the producers of electricity to the consumers in a bi-directional mode. Electricity is generated by micro-grids through renewable energy sources like water, waste, wind turbines, solar panels, and so on. It is then delivered to the smart grid energy storage units. Smart grids can control, exchange, and monitor energy flows, quickly respond to failures, and undertake intelligent predictive analysis about the future demands of energy requirements due to smart switches, smart meters, smart appliances, smart applications, and smart automatic control devices (Ancillotti et al., 2014).

Smart grids produce information about the total power consumption and consumption patterns such as peak load consumption. The smart grid is integrated with other smart devices in the smart city. For instance, smart grids are integrated with smart charging stations to provide energy to electrical machines like electric vehicles. The energy transfer between smart charging stations and smart electrical machines will be bi-directional (Borgia, 2014).

The smart grid is the ultimate solution to save energy, enhance efficiency, minimize losses, and deploy renewable energy. The sensors in the smart grid measure, monitor, and communicate information for real-time intelligent decision-making, thus making the grid self-monitoring and so more efficient.

2.2 Smart Buildings

Smart automated buildings are equipped with numerous smart devices and sensors—for instance, e-meters, smart lights, smart phones, tablets, smart cameras, laptops, switches, sensors, and smart washing machines, etc.—and are connected through networks. These smart buildings can be monitored and controlled remotely through the Internet. Moreover, many smart applications have been developed to implement IoT capabilities, like intrusion detection systems, asset management systems, surveillance systems, and security systems (Borgia, 2014).

The appliances in smart buildings can be remotely switched on and off to save power and to circumvent accidents. The intrusion detection systems can detect the opening or closing of doors and windows to avoid burglars.

3. Key Technologies Involved in Cloud Computing

There are many technologies employed in cloud computing, such as virtualization, service models, programming models, deployment models, data processing and management, data storage, etc.

3.1 Cloud Computing Service Models

Cloud providers offer services at three levels, as shown in Figure 1.1. Users can lease resources simultaneously at infrastructure, platform, and software level. Cloud service models are SaaS, PaaS, and IaaS (NIST, 2011). SaaS provides access to cloud apps in addition to infrastructure. Clients can access the applications through interfaces via the Internet. PaaS provides platforms on the cloud infrastructure so that users can deploy the applications they require. The cloud provides user-friendly application deployment tools and an Integrated Development Environment (IDE) to the clients. As with SaaS, cloud users are not given access to the underlying cloud infrastructure. IaaS provides access to the infrastructure of the cloud at an abstract or virtual level. The physical infrastructure is managed by cloud providers, while the virtual infrastructure is managed by users. Clients can lease resources like cloud storage, networks, and servers.

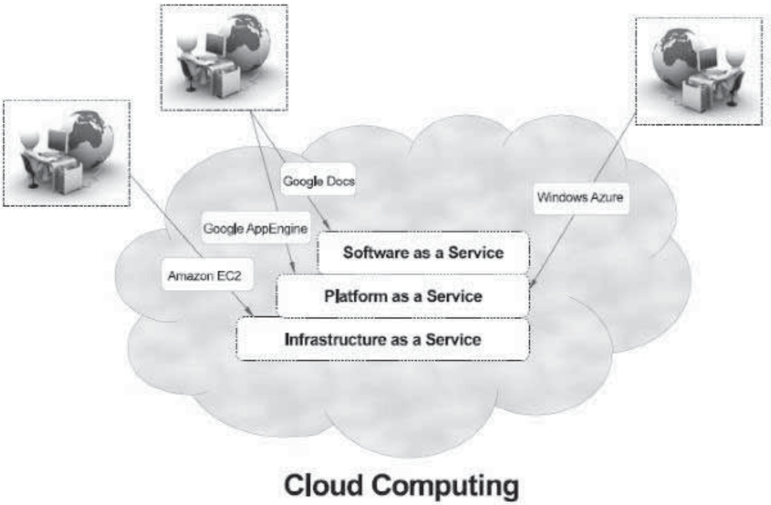


Figure 1.1 Abstraction layers of cloud services (Courtesy of NIST, 2011)

3.2 Cloud Storage Technology

Cloud users can store data online via networks on the virtualized cloud storage system. Clients can lease cloud data centers according to their requirements. Users can access and modify their data through a user interface provided to them via the Internet. Cloud data centers are distributed geographically over the world at heterogeneous locations. Furthermore, redundant copies of user data are stored in the cloud storage systems at geographically dispersed locations to ensure recovery of data in case of failure.

The cloud storage providers generally set up geographically distributed storage servers. This is to ensure that the data can be recovered successfully even in a time of natural disaster. Moreover, many copies of same data are stored in different data centers, to ensure the users can access their data any time, even if there is any failure in the servers. Compared with traditional storage systems, cloud storage also reduces the construction cost of storage systems (Divya & Jeyalatha, 2012).

3.3 Cloud Programming Models

Cloud users communicate with the cloud through high-level tools like Hadoop. This is in contrast to traditional systems, where users employ low-level tools like MPI for distributed memory systems and OpenMP for shared memory systems (Branch et al., 2014).

Cloud programming models transparently schedule parallel execution of processes. Cloud computing implements the MapReduce programming model designed by Google. This model schedules and allocates nodes by dividing the processes into multiple sub-processes, and through two phases (Map and Reduce). MapReduce thus works in two steps, Map and Reduce. It supports parallel execution, load balancing, and fault-tolerance. The user needs to write only two functions, namely Map and Reduce, for submitting their procedures to MapReduce, which in turn initializes data and calls the two functions (Padmini, 2015).

Many implementations of MapReduce have been proposed after Google introduced it. One widely used implementation of this is Hadoop (Apache Hadoop, 2015).

MapReduce is primarily used for mass data processing for Big Data. In contrast with traditional database models, MapReduce is more scalable and reliable. Moreover, traditional RDBMS have complex SQL queries, while MapReduce implementations such as Hadoop can execute processes more efficiently while meeting the scalability requirements.

3.4 Cloud Deployment Models

Cloud deployment models can be categorized into *private clouds*, *community clouds*, *public clouds*, and *hybrid clouds*.

Private clouds serve solely for a single organization. The platform of a private cloud is created by the organization itself or by some other stakeholder. The organization is responsible for managing the infrastructure and platform, which is complicated and expensive. Private clouds are less scalable and elastic, but more secure as only the organization's authorized personnel can access them.

Community clouds can be accessed by multiple organizations which have specific shared concerns, like security considerations, common goals, policies, and so on. Community clouds can be developed internal or external to the organization. The main issue for a community cloud is security, due to its shared nature. All the organizations sharing the same community cloud must abide by the shared policy of their cloud.

Public clouds provide access to the general public over the Internet. This is the most widely used type of cloud computing. In this model, the cloud provider (e.g. Amazon.com, Microsoft, and Google) offers infrastructure and services to a wide range of users on a pay-per-use and on-demand basis. Public clouds must be highly flexible and scalable to meet unpredictable client requests. Public clouds are less costly and may serve a large amount of customers in the market, but their security functionalities must be well protected.

Hybrid clouds can have the features of public, private, and community clouds. In a hybrid cloud, an organization can allocate workloads into distinct clouds based on their security, strategic, or corporate needs. Hybrid clouds offer greater elasticity and scalability. For instance, an organization can use public as well as private clouds to process confidential and critical data. Hybrid clouds require additional overheads for the management of heterogeneous resources. Moreover, they require proper security measures for their private data.

4. Classification of Big Data

Big Data classification is crucial due to the large volume of Big Data in the cloud. The five aspects of this classification are: (1) data sources, (2) content format, (3) data stores, (4) data staging, and (5) data processing (Hashem et al., 2014).

4.1 Data Sources

Social media: Social media produce information to share ideas, or information through the Internet among social networking communities like Twitter, Facebook, blogs, etc.

Machine-generated data: Machine data is the data automatically produced from electronic devices and their applications, like computers or smart phones.

Sensing: Sensing devices convert physical quantities into signals.

Transactions: Transaction data involves an event, like a financial spreadsheet.

IoT: IoT refers to objects or things having unique identifications and which are connected to the Internet, like smart apps, smartphones, laptops, tablets, smart digital cameras, and smart meters (Hashem et al., 2014).

4.2 Content Format

Structured: Structured data can be accessed and stored in RDBMS by using a query language like SQL. RDBMS is a database-management system used to access and control data. Structured data may be in the form of dates, numbers, words etc.

Semi-structured: Semi-structured data is not stored in RDBMS. An example of semi-structured data is a table.

Unstructured: Unstructured data have no particular format. The size of unstructured data is growing enormously day by day. Examples of unstructured data are pictures, text strings, audios, videos, multimedia, and social media data (Hashem et al., 2014).

4.3 Data Stores

Document-oriented: Document-oriented databases store document files or information and support complex data formats, like binary forms (e.g., PDF and MS Word) and XML (Extensible Markup Language).

Column-oriented: Column-oriented databases store data in similar-attribute columns, like BigTable.

Graph database: Graph databases store graphical data with nodes, vertices, edges, or links, like Neo4j.

Key-value: Key-value is a database that stores huge volumes of data (Hashem et al., 2014).

4.4 Data Staging

Cleaning: the process of separating meaningless data from data streams.

Transform: the process of converting data into a form suitable for analysis.

Normalization: the method of reducing redundant data from a database (Hashem et al., 2014).

4.5 Data Processing

Batch: MapReduce-implementations can be used to run batch applications, as they support scalable parallel executions.

Real time: The Simple Scalable Streaming System (S4) (Neumeyer et al., 2010) can be used to process real-time Big Data. S4 is a distributed platform that is used to process continuous unbounded streams of data (Hashem et al., 2014).

5. Applications of IoT, Cloud Computing, and Big Data

IoT, cloud computing, and the theory of Big Data (ICT) have the potential to generate a large number of applications, yet a great deal of research is still needed to develop those applications. Future ICT apps will help to improve our standard of living. A very wide range of applications can be deployed to enable the communication of ICT apps with each other and with the environment.

The integration of IoT applications, cloud computing, and Big Data is a big challenge for researchers and communities. The conjoining of IoT, cloud computing, and Big Data will provide powerful and intelligent next-generation systems. The information generated as a result of communication between IoT objects, such as sensors, smart phones, smart vehicles, smart cameras, smart grids, and smart meters, is enormous, and requires IoT apps for Big Data on the cloud. The integration of IoT, cloud, and Big Data is also referred to as the *Internet of Everything* (IoE).

5.1 Transportation System

The transportation system has become more automated, instrumented, and intelligent due to the presence of smart devices, network connections, sensors on roads, and sensors in vehicles. ICT has enabled better safety and navigation. Intelligent systems and sensors have enabled us to improve the food supply chain (Atzori et al., 2010).

5.2 Smart Medical Treatment and Healthcare

ICT has many applications in the healthcare sector. It can be used to enhance current assisted-living technologies. Medical sensors are used to monitor various parameters like blood pressure, body temperature, and heart rate. Some other sensors, such as accelerometers, are used to monitor patient activities. The vision of ICT is to collect and then send this information to remote centers. These remote medical centers will be capable of doing remote monitoring of patients and taking quick real-time response actions based on the Big Data collected. This can help the physician specialist to correctly diagnose a patient's illness.

Wearable sensors monitor the daily activities of patients. The sensors are equipped with apps which can provide suggestions on how to improve lifestyles and health (Miorandi et al., 2012).

5.3 Automation, Monitoring, and Control of Industrial Production Processes

Industrial automation involves a lot of services, tasks, and applications. Each of these tasks has different priority levels: for instance, to support alternative actions in case of emergency situations, to ensure smooth operation and workings of the plant and machinery, automated alerting, generation of up-to-date logs, automated control, and automated data communication and information transfer. Industrial automation systems are complicated and expensive. In the future, cost-effective and efficient automation may be realized with wireless sensor networks (WSNs), which will provide innovative solutions to current problems.

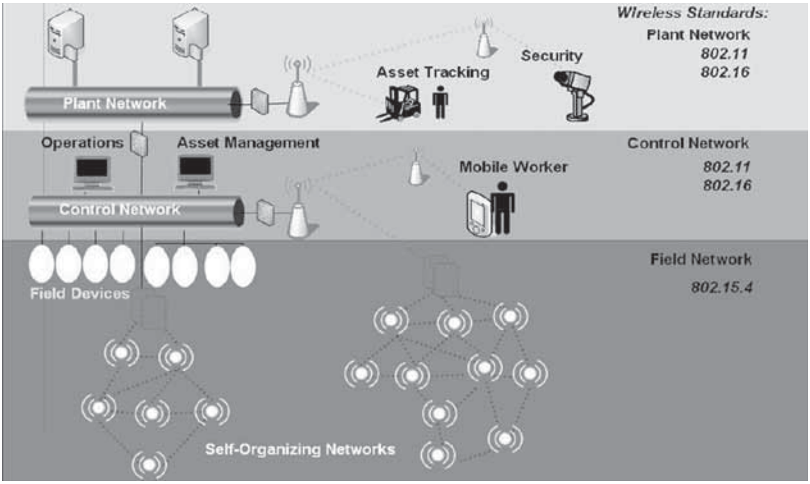


Figure 1.2 Industrial wireless network architecture (Courtesy of Zheng et al., 2011)

Figure 1.2 presents an overview of the industrial wireless network architecture. Wireless HART (Highway Addressable Remote Transducer) and ISA100.11a provide specifications for the support of wireless automation apps. Instead of multiple control rooms, a single control room can be connected with multiple subnets via wireless communication systems (Zheng et al., 2011).

6. Existing Problems and Challenges of IoT, Cloud Computing, and Big Data

Cloud computing brings many challenges for cloud service providers, infrastructure engineers, system administrators, software programmers, and system developers. Figure 1.3 pinpoints some of the challenges. Virtualization refers to the creation of independent virtual machines (VMs) for hosting services and applications. In order to meet QoS standards, effective management and creation of a large number of VMs is needed. Therefore, VMs are shifted to appropriate servers when the services have higher quality requirements, and later consolidated dynamically to a lesser number of physical servers (Buyya et al., 2009).

There are only a few tools available to meet the challenges of IoT and Big Data in cloud computing environments.

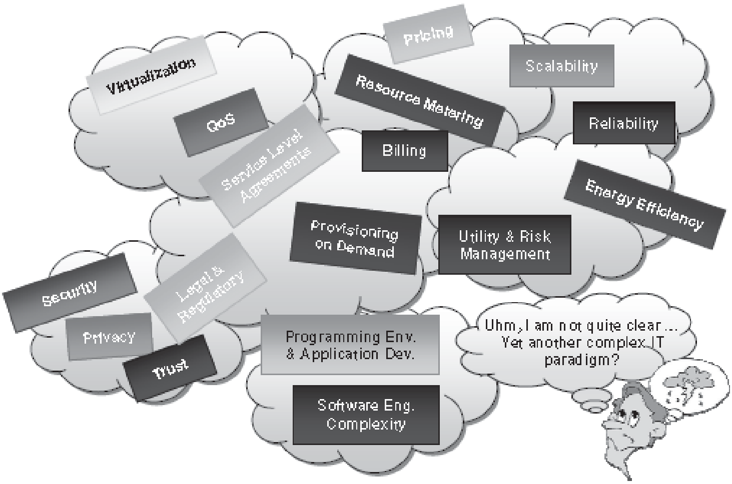


Figure 1.3 Cloud computing challenges. (Courtesy of Buyya et al., 2009)

6.1 Security, Privacy, and Trust

Security, privacy, and trust are among the biggest challenges of IoT and cloud computing. Security is indeed critical for cloud computing and a security breach can affect the entire cloud. Organizations host their confidential and important data on the cloud, which is mostly managed by cloud providers or some other third party. So cloud providers try to gain the trust of their clients by ensuring the desired level of privacy and security of their applications on the cloud (Buyya et al., 2009).

Physical control of the hardware and the ability to visually monitor and check data connections is needed to ensure security. Cloud providers offer various security solutions, such as cryptography, VMs support, and APIs standardization.

Organizations are also concerned about the privacy and security of their data hosted on clouds because the cloud service providers themselves can access their data. So cloud providers must also address security concerns to build trust in the cloud.

6.2 Data cleaning

Data cleaning is a challenging task due to the heterogeneous nature of unstructured data gathered from various sources.

6.3 Distributed Storage Systems

There are a number of solutions for the storage and retrieval of large volumes of Big Data in cloud computing environments. However, several challenges hinder the successful implementation of such solutions, including: the capability of current cloud technologies to provide the necessary capacity and performance to address massive amounts of data; the optimization of existing file systems for the volumes demanded by data-mining applications; and concern about how data can be stored in a manner such that they can be easily retrieved and migrated between servers (Hashem et al., 2014).

6.4 Bottlenecks of Big Data Transfer

With the passage of time, applications are becoming more and more data-intensive. If we assume to transfer applications from clouds to traditional disks, the data transfer and placement would be very complicated and expensive.

6.5 Big Data Analysis

Big Data analysis requires scalable and efficient analytic algorithms and tools to process data and generate rapid results (Talia et al., 2013). Big Data analysis is a complicated process, as data are gathered from heterogeneous sources which have different data formats.

6.6 Costing Model and Charging Model

Cloud users must be aware of the tradeoffs between computation, communication, and integration. Using the cloud can greatly reduce infrastructure costs, but can increase data communication costs. This can be an issue especially if the client is using a hybrid cloud. Similarly, SaaS users must be aware of the tradeoffs between cost-saving and multi-tenancy. Therefore, the choice of an appropriate charging model is crucial (Dillon et al., 2010).

References

Amazon. Amazon Elastic Compute Cloud (Amazon EC2). Available at: http://aws.amazon.com/ec2/ (accessed April 29, 2015).

Amazon Elastic MapReduce (Amazon EMR). Available at: http://aws.amazon.com /elasticmapreduce/ (accessed April 29, 2015).

Ancillotti, E., Bruno, R., Conti, M. (2014). Smoothing peak demands through aggregate control of background electrical loads. In: *Proceedings of IEEE PES Innovative Smart Grid Technologies Conference*, 1-5.

Apache Hadoop. 2015. Available at: http://hadoop.apache.org/ (accessed April 30, 2015).

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., and Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM*, 53 (4), 50-58.

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, 54 (15), 2787-2805.

Borgia, E. (2014). The Internet of Things vision: Key features, applications and open issues. *Computer Communications*, 54, 1-31.

Branch, R., Tjeerdsma, H., Wilson, C., Hurley, R., and McConnell, S. (2014). Cloud Computing and Big Data: A Review of Current Service Models and Hardware Perspectives. *Journal of Software Engineering and Applications*, 7, 686-693.

Buyya, R., Pandey, S., & Vecchiola, C. (2009). Cloudbus toolkit for market-oriented cloud computing. *Cloud Computing*, 24-44.

Canellos, D. (2013). How the “internet of things” will feed cloud computing’s next evolution. Available at: https://blog.cloudsecurityalliance.org/2013/06/05/how-the-internet-of-things-will-feed-cloud-computings-next-evolution/ (accessed April 29, 2015).

Creeger, M. (2009). Cloud Computing: An Overview. *ACM Queue*, 7 (5), 2.

Dillon, T., Wu, C., & Chang, E. (2010). Cloud computing: Issues and challenges. *IEEE International Conference on Advanced Information Networking and Applications*, 27-33.

Divya, K., & Jeyalatha, S. (2012). Key technologies in cloud computing. In: *IEEE International Conference on Cloud Computing Technologies, Applications and Management*, 196-199.

Gandomi, A., and Haider, M. (2015). Beyond the hype : Big data concepts, methods, and analytics. *International Journal of Information Management*, 35 (2), 137-144.

Gantz, J., Reinsel, D. (2011). Extracting value from chaos, *IDC iView*, 1-12.

Google. Google Apps. Available at: http://www.google.com/apps (accessed April 29, 2015).

Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29 (7), 1645-1660.

Hashem, I.A.T., Yaqoob, I., Badrul Anuar, N., Mokhtar, S., Gani, A., and Ullah Khan, S. (2014). The rise of "Big Data" on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115.

HP Cloud. URL: https://www.hpcloud.com/ (accessed April 29, 2015).

International Telecommunication Union. (2005). ITU Internet Report 2005: The Internet of Things, International Telecommunication Union, Geneva.

Laurila, J.K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T.-M.-T., Dousse, O., Eberle, J., Miettinen, M. (2012). The mobile data challenge: Big data for mobile computing research, Workshop on the Nokia Mobile Data Challenge. In: *Proceedings of the Conjunction with the 10th International Conference on Pervasive Computing*, 1-8.

Microsoft. Windows Azure Virtual Machines. Available at: http://www.windowsazure.com (accessed April 29, 2015).

Microsoft. Microsoft Office 365. http://www.microsoft.com/office365 (accessed April 29, 2015).

Miorandi, D., Sicari, S., De Pellegrini, F., and Chlamtac, I. (2012). Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, 10 (7), 1497-1516.

Neumeyer, L. B., Robbins, A. Nair, A. Kesari, (2010). S4: Distributed Stream Computing Platform. *IEEE International Conference on Data Mining Workshops*, 170-177.

NIST, (2011). The NIST Definition of Cloud Computing. Available at: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf (accessed April 29, 2015).

Padmini, K. (2015). Securing Data Management Based on Key Technologies in Cloud Computing. *International Journal of Advance Research in Computer Science and Management Studie*s, 3 (2), 165-172.

Rao, B.P., Saluia, P., Sharma, N., Mittal, A., Sharma, S.V. (2012). Cloud computing for Internet of Things & sensing based applications. In: *Proceedings of the IEEE International Conference on Sensing Technology (ICST)*, 374-380.

Talia, D., (2013). Clouds for scalable big data analytics, *Computer*, 46, 98-101.

Zheng, L., Zhang, H., Han, W., Zhou, X., He, J., Zhang, Z., & Wang, J. (2011). Technologies, applications, and governance in the Internet of Things. *Internet of things-Global technological and societal trends. From smart environments and spaces to green ICT*, 141-175.

Cognitive Rehabilitation Gaming System

Nor Azan Mat Zin and
Ahmed Mohammed Elaklouk

Abstract

Brain injury is the leading cause of long-term disabilities in most countries. The increasing rate of brain-damaged victims and the heterogeneity of impairments burden healthcare systems and result in a high cost of rehabilitation. On the other hand, traditional rehabilitation exercises, which are boring, cause patients to neglect the prescribed training required for their recovery. Currently, there are no specific standards and guidelines on designing a feasible and effective game-based intervention that can sustain patient motivation during cognitive rehabilitation exercises. Therefore, we have proposed a game-based approach to address these problems. This chapter presents a Rehabilitation Gaming System (RGS) designed for cognitive rehabilitation. The RGS is developed based on a proposed conceptual framework, which is also presented in this chapter.

Keywords: Serious Games, Cognitive Rehabilitation, Game Design, Brain Injury.

1. Introduction

Brain injury, such as traumatic brain injury (TBI) and stroke, is the leading cause of long-term disabilities in most countries. Brain injury results in cognitive impairments such as attention, memory, and executive deficiencies. These impairments dramatically affect people's lives and limit their ability to perform everyday activities (LOPresti et al. 2008). Rehabilitation includes therapeutic activities and techniques that aim to reduce the impact of deficits in a patient's abilities (Broeren et al. 2008).

However, rehabilitation is a long-term process that consumes massive training and financial resources (Fok 2009). Human resources and facilities face limitations in catering to the increasing numbers of brain-damaged individuals who have been involved in serious accidents and hence, healthcare systems become heavily burdened and overtaxed. For example, in the United States alone, 7% of the inhabitants – or approximately 20 million individuals –suffer from cognitive disabilities (Carmien et al. 2003). Additionally, the heterogeneity of impairments that individuals suffer from is a significant factor in the planning, development, and evaluation of rehabilitation. In other words, depending on the brain lesion (location, type, and size), the individual is then characterized by a specific combination of impairments. Hence, rehabilitation is a strongly personalized intervention instead of a generic one and since tasks must be tailored to each individual, the process will incur a high cost. Koenig (2012) reported that health care costs for brain-damaged patients are among the highest compared to all other healthcare services in many countries. For example, in the United States, approximately $62.7 billion is estimated as service costs for stroke patients, annually.

Brain-damaged patients frequently report that conventional rehabilitation exercises are boring due to their repetitive nature and this causes them to neglect the training required for their recovery (Burdea 2003). Current research revealed that after sustaining a brain injury, the brain has the ability to cure itself through repetitive, intensive, and task oriented exercises (Doidge 2007; Panic 2010a). Therefore, patient motivation is essential for rehabilitation success and is a determining factor in the outcome of a rehabilitation process (Maclean et al. 2002). The main problem facing therapists is how to encourage patients to actively participate in rehabilitation programs (Alankus et al. 2010; Burke et al. 2010).

Tagliaferri et al. (2006) reported that the majority of patients (75%) with traumatic brain injuries are those who are less than 35 years of age. This group is more inclined to play games on computers and/or handheld game devices compared to other age groups.

Therefore, with such a substantial effect it has on patients and on healthcare systems, it is crucial to develop an effective intervention that can engage patients. Hence, rehabilitation delivery, scenarios, and organization should be designed to achieve this goal. We believe that the use of serious games and advances in technology benefits rehabilitation and opens up a lot of new opportunities to meet stakeholder needs and expectations, which include (1) increasing the quality and efficiency of rehabilitation, (2) offering options that can sustain patient motivation and overcome the boring, repetitive nature of conventional rehabilitation, (3) provide an appropriate level of therapy, and (4) overcome the limited resources and facilities of conventional rehabilitation methods.

However, to the best of our knowledge, there are no existing standards and guidelines on how to design feasible and effective motivational games for rehabilitation. To this end, this chapter has defined a conceptual framework as a guide to design serious games for cognitive rehabilitation of brain-damaged patients. The framework can assist game developers and researchers in the rehabilitation field to design and effectively deliver games that are focused on rehabilitation.

This chapter starts off by reviewing relevant literature that involves contributions of serious games towards rehabilitation followed by the potential of, and requirements for game-based rehabilitation interventions. A conceptual framework and its components are discussed and finally, the Rehabilitation Gaming System (RGS) is described.

2. Serious Games for Rehabilitation

"Serious games" is a term that is used to describe games designed particularly for training and education compared to traditional (off-the-shelf) games, the primary purpose of which is to entertain. Rego et al. (2010) defined serious games as "computer games that allow the player to achieve a specific, non-entertainment purpose using entertainment and engagement components provided by the experience of the game." Prensky (2001) outlined twelve elements as to why games engage individuals, as shown in Fig. 1. Using games in the rehabilitation process provides for a meaningful and engaging experience.

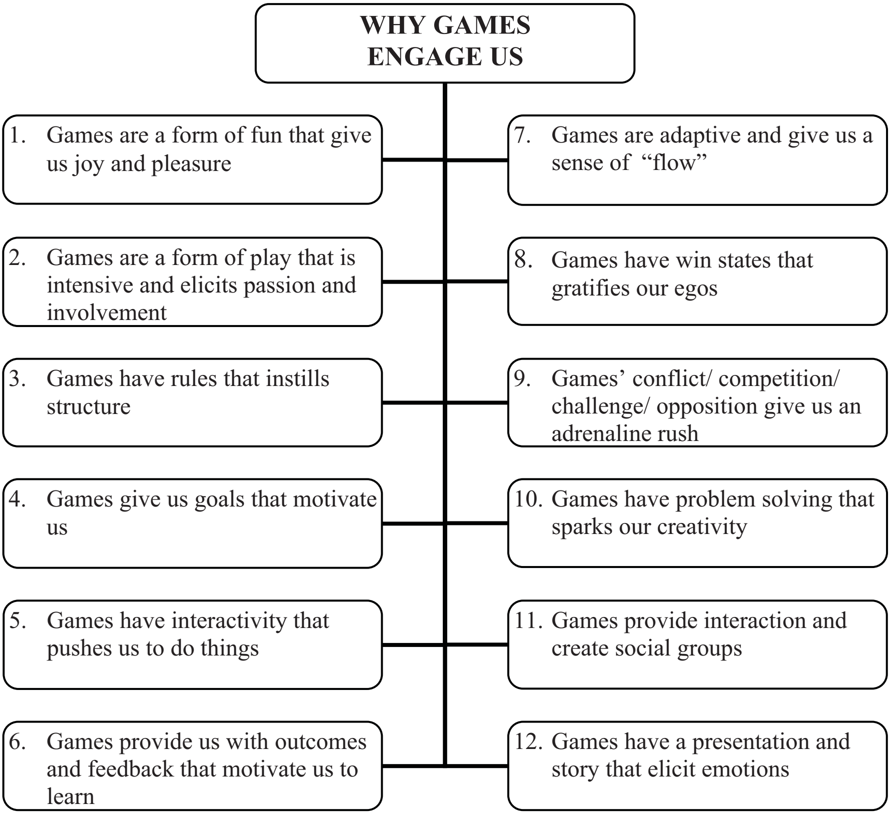


Figure 1. The twelve elements of engagement (Prensky 2001)

Numerous advantages are inherent in using games as rehabilitation interventions. A game-based rehabilitation has many advantages over traditional rehabilitation exercises. They can be more engaging than paper exercises and accommodate built-in reward and motivation systems instead of real world incentives as the sole motivation for completing the rehabilitation tasks (Bartle et al. 2010). In addition, games afford a virtual medium, which can be used as a safe environment that individuals with brain injury can explore at their own pace. They can become immersed in the game world and improve their abilities and knowledge without taking any risks (Torrente et al. 2012).

On the other hand, the most important sought-after advantage is the ability to distribute game systems using the Internet so that they can easily reach remote areas (Bartle et al. 2010). Consequently, the human resource costs per patient treated and the treatment time will decrease and home-bound patients can get easier access to the treatment (Cruz et al. 2013).

There are a few customizable game interventions that target people with cognitive disabilities. Research on therapeutic games for cognitive rehabilitation is in its infancy, compared to other types of disabilities (Torrente et al. 2012). A review of the current literature on the use of games for rehabilitation found some shortcomings, such as the small sample size involved, the limited time invested in usability and acceptance testing done on volunteers, testing was done with able-bodied and healthy users, and lack of regard for the therapist who is one of the principal end-users of the technological intervention (Broeren et al. 2008; Fok 2009; Burke et al. 2010; Rego et al. 2011). For example, Fok (2009) integrated a virtual telephone as well as a computer-interfaced prosthesis and developed Internet-enabled exercises that enable individuals with memory impairment to practice at home. However, only 40 healthy adults were involved in this evaluation study. In addition to this, Rego et al. (2011) reported the impact of using new forms of interaction in cognitive therapeutic games via development of a game prototype that enabled individuals to play using three interaction inputs i.e. mouse, sound, and/or motion. However, similar to the previous study, this was a small usability test, only involving 20 healthy users in the evaluation of the game prototype. Furthermore, Broeren et al. (2008) studied the effects of serious games' virtual reality on individuals with cognitive and physical impairments. This intervention only involved five patients.

As most of the studies have the tendency to avoid true end-users (i.e. patients and therapists), there still exists a lack of knowledge with respect to the actual requirements for the design of therapeutic games. Therefore, a prototype was used to elicit true requirements from actual users (i.e. patients and therapists). In our previous work (Elaklouk and Zin 2012a; Elaklouk and Zin 2012b), we determined game design principles that are critical for rehabilitation and developed a game prototype called the "Ship Game" as a cognitive rehabilitation intervention. This game tested both patients and therapists in order to determine the potentials of, and the requirements for therapeutic games.

3. Rehabilitation Gaming System Requirements

The “Ship Game”, as shown in Fig. 2 and Fig. 3, consists of a number of mini-games that are presented in Fig. 4 and Fig. 5. Each mini-game targets different cognitive skills with special focus on memory, attention, concentration, executive functions, and hand-eye coordination. The design principles of the “Ship Game” are identified as follows: (1) meaningful play, which stems from the link between a player’s action and the outcome of the game, created and maintained using both the game’s feedback and positive handling of failure, (2) the challenge, which lies in the balance between an individual’s skills and the challenges faced, (3) portability i.e. the system’s capability to be utilized anywhere (home, hospital or clinic), and (4) interaction technology i.e. the technology that the patient uses for system interaction.

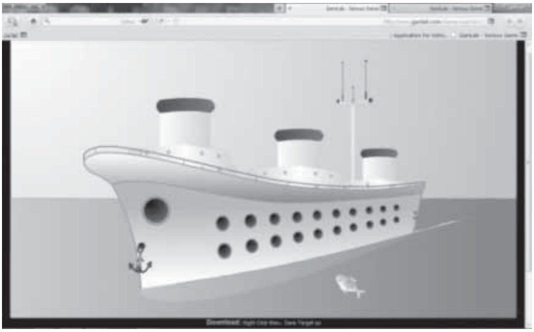


Figure. 2. The Ship Game’s main interface

Figure 3. The game menu ‘onboard the ship’

Figure 4. The first door game prototype

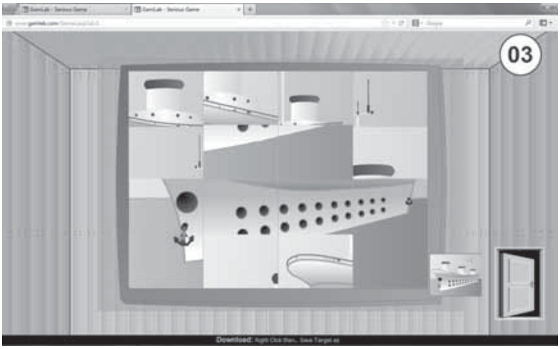


Figure 5. The second door game prototype

Based on the investigation into the "Ship Game" involving both patients and therapists, the potentials of, and requirements for designing rehabilitation games were identified:

a) *The game should be playable and accessible to a broad range of patients.* It is important to involve as many patients as possible because not every patient can play the same game due to their physical and/or cognitive limitation and this should not prevent them from playing the game in the first place. Adding an online gameplay option opens the opportunity for outpatients and home rehabilitation. Additionally, developing games with flexible input methods enables therapists to use these games for different rehabilitation purposes.

b) Any rehabilitation game should aim to *have the least number of rules*, the *maximum amount of fun*, and *handle failure in a positive way*.
c) *The games should be challenging.* Game activities should not be impossible to accomplish as this would result in frustration and they should not be too easy because this could result in boredom. Providing patients with the right amount of challenge and clear feedback can sustain their motivation and maintain their level of engagement in the gaming experience.
d) *Games should able to save data concerning the activities involved in playing games and present this data in a suitable way.* Brain-damaged patients want to be assured that their achievements in the games and the time they spent on this type of treatment are really contributing to their overall success and recovery. Moreover, records of patient activities allow therapists to monitor their patients' progress. Additionally, they can specifically help their patients more effectively outside of a one-on-one rehabilitation session.
e) *Therapists should have control over the game's therapeutic activities.* Therapists should experience and control the exercises themselves so that patients can actively practice the game exercises that are closely linked to their abilities, which is very important for their cognitive improvement.
f) *The integration between the gaming system and the rehab center's information system is required.* The rehabilitation process may be described as a multi-disciplinary process that encompasses various rehabilitation units. Patient recovery plans involve many professionals, including physical and occupational therapists who address the patient's motor and cognitive impairments and choose the best techniques to use for his/her progress. Hence, to effectively use games for rehabilitation, this integration is required.

4. A Conceptual Framework for Designing Cognitive Rehabilitation Gaming System

Based on the investigation into the "Ship Game" and from the revision of related literature, a conceptual framework for designing a game-based cognitive rehabilitation intervention was proposed and this is illustrated in Fig. 6. The study's aim is to establish a framework that can be used by game developers and practitioners and at the same time design a game system for cognitive rehabilitation. The proposed framework comprises of four components i.e. condition, process, activity, and outcome. Each component plays a crucial role in the design of an efficient and effective game-based cognitive intervention that can sustain patient motivation. The following sections explain these components in more detail.

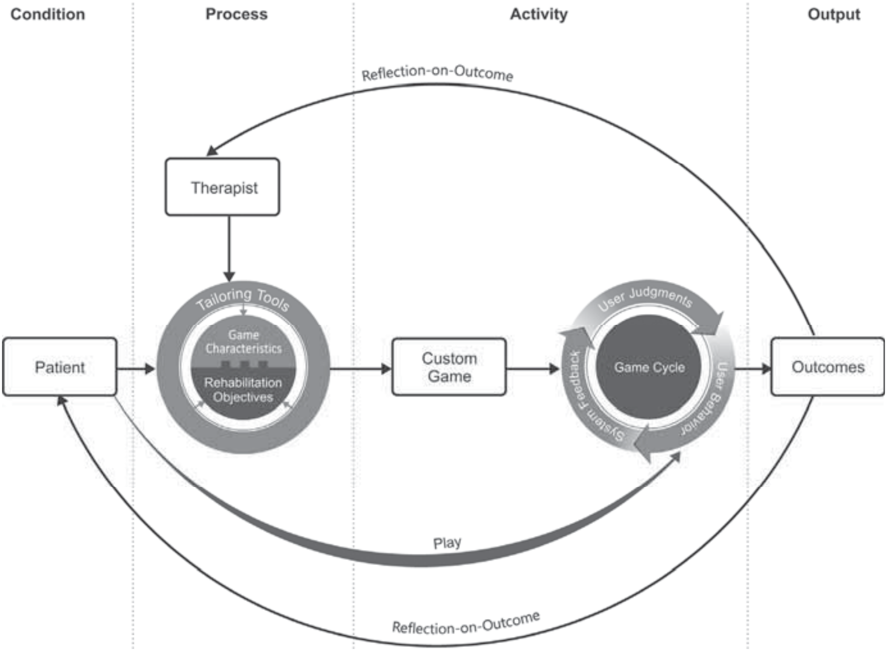


Figure 6. A Conceptual framework for designing a cognitive rehabilitation gaming system

4.1 Condition

Perry et al. (2010) reported that the contributions of therapeutic games are many. However, new games that better accommodate and address patient needs, from therapeutic as well as motivational perspectives, are still required. Hays (2005) argued that games will be instructionally effective if their specific characteristics are linked to instructional objectives. Moreover, if an individual has no interest in the instructional content, specific game features can be used to trigger his motivation/interest (Garris et al. 2002). Hence, in rehabilitation, the Effective Rehabilitation Game (ERG) can be described according to equation (1).

$$ERG = IRO + GC. \tag{1}$$

The IRO represents the Intended Rehabilitation Objective and GC represents Game Characteristics. The IROs represent the rehabilitation goals that must be achieved in the gameplay.

These goals are determined via brain-damaged patient evaluation and assessment according to the first part (condition) of the proposed framework shown in Fig.6.

A crucial aspect of patient rehabilitation is that the consequences of brain injury manifest differently in different patients. Additionally, although patients may have the same disorder, their responses to the disorder can be unique. Patient-centered practice is reported – by many studies – to improve health status and efficiency of health care (Oates et al. 2000).

Understanding patient needs, priorities, and preferences prior to technology selection can positively affect the matching between patient and technology. However, the main factor that accounts for the poor match of patient and potential technology intervention is the lack of adequate evaluation and assessment of patient needs and preferences (Scherer and Glueckauf 2005; LOPresti et al. 2008). The only fundamental issue in design is how it relates to targeted groups of patients. The complexity of matching a patient with technology emerges not only from his/her unique combination of cognitive, sensory, and physical abilities but also from his/her expectations and reactions to the new technology. These reactions arise from personal preferences, needs, abilities, and prior experiences with technology (Scherer and Glueckauf 2005).

Therefore, patient-centered evaluation is necessary in clarifying therapy goals prior to planning rehabilitation interventions. This orientation is consistent with rehabilitation models that consider the recognition of a patient’s problems, needs, and mediators relevant in targeting problems, as the first step before planning rehabilitation (Steiner et al. 2002). Hence, in the “condition” part of the framework, the patient assessments should be conducted by therapists. Also, factors that influence patient predispositions to the use of particular technology must be understood. Hence, the preferences, needs, and limitations of the patient as well as the environment in which the activity will take place must be understood so that the most appropriate training strategies can be designed.

Cognitive assessment practice normally begins with simple tests such as the Mini Mental State Exam (MMSE). In case of MMSE poor performance, it is essential to conduct in-depth evaluation using a Neuropsychological Assessment (NA) (Tsaousides and Gordon 2009). The NA is usually carried out to evaluate the extent of deterioration of a particular skill, in order to find out the damaged area after a patient has sustained brain injury. By applying NA tests, we can assess various cognitive domains such as attention, memory executive function, and other functions (ÉML. 2009). For example, the Wisconsin Card Sorting Test (WCST) is used to measure the "executive function" deficits that affect the individual's ability to initiate, plan, set goals, anticipate consequences, and respond adaptively. Additionally, direct observations and interviews offer a useful method for assessing cognitive deficiencies and skills. Therapists may collect information by interviewing family members and caregivers about their patient's everyday behavior and by also using direct observation valuing the behavior of the patient (Panic 2010b). In addition, the International Classification of Functioning, Disability and Health (ICF) can be used as a valuable framework to manage and classify patient information and help therapists to identify aspects of a patient's condition that affect his/her recovery (BELCHIOR 2007). The most significant advantage of the ICF is in expanding thinking further rather than fixing particular impairments. ICF describes "disability and functioning" as outcomes resulting from the interactions between "health conditions" and "contextual factors". The contextual factors are classified into personal factors and environmental factors. ICF personal factors, which have been formally acknowledged, recognize the importance of patient personal choices, interests, likes, and dislikes, all of which are being addressed in rehabilitation programs. (Rosenbaum and Stewart 2004; BELCHIOR 2007). Therefore, game intervention for cognitive rehabilitation is effective when it is shaped to meet patient needs and preferences and not when it is prescribed as an isolated intervention to address particular cognitive deficiencies. This process needs to begin with a comprehensive evaluation that gives therapists the opportunity to assess patient ability, needs, preferences, and expectations. The evaluation will produce information that can tailor the game to suit a variety of patients and hence enhance the quality of rehabilitation training.

4.2 Process

Garris et al. (2002) argued that, in game-based learning, success in pairing game characteristics with appropriate instructional practices can trigger an individual's motivation towards achieving the intended learning outcome. The challenge is to adapt these game characteristics/features for instructional purposes in order to enter a game-cycle that motivates and sustains a learner's engagement. Although numerous literatures have noted the potential advantages from incorporating game features/characteristics into instructional applications, the question that remains is "how will this incorporation (pairing) be done?"

Therefore, in a cognitive rehabilitation context, we will try to answer the question of "how to incorporate game characteristics into rehabilitation programs". In other words, based on our proposed framework (process part), we will answer the question of how rehabilitation professionals can provide patients with game interventions that are designed and customized to accommodate individual skills and deficits and sustain his/her motivation, but in a cost-effective way.

In rehabilitation, a user-centered design approach is crucial (Perry et al. 2010). While the evaluation from the "condition" part is concerned with considering a patient's context, ability, needs, preferences, and expectations, the end of the "condition" part defines realistic therapy goals. This in turn can help in planning the most appropriate game interventions. Therefore, rehabilitation professionals can now use "tailoring tools" to pair a game's characteristics with defined rehabilitation goals and hence, create and deliver the appropriate game for an individual patient that is tailored to his/her needs and expectations.

Many studies have discussed game characteristics (Garris et al. 2002; Wilson et al. 2009). For instance, Garris et al. (2002) described game characteristics in terms of six categories, which are shown in Fig.7.

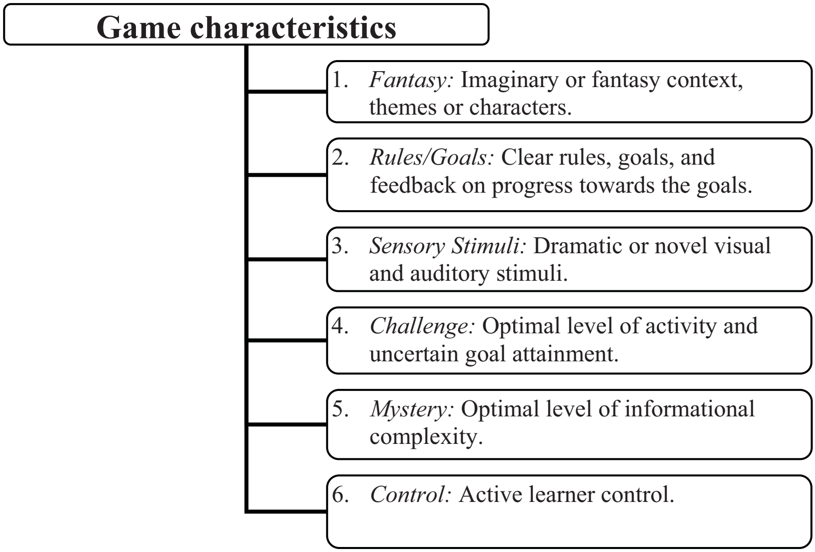


Fig. 7. Game characteristics (Garris et al. 2002)

A therapist's motivation for using games in the rehabilitation process is clearly a motivational benefit to the patient. Increased treatment adherence translates to increased functional outcomes (Perry et al. 2010). Many potential benefits will be gained from incorporating these game characteristics into instructional rehabilitation programs. This incorporation can provide a training environment in which the patient can practice rehabilitation tasks without real-world consequences specifically in case of failure, which is crucial in the context of rehabilitation. Moreover, games that can provide progressive difficulty levels allow the patient to gain familiarity and improve his/her skills gradually. This can enhance patient enjoyment and confidence. Patients with greater confidence as a result of applying skills they have learned in game-based intervention are more resilient to real-world difficulties.

Therefore, according to our proposed framework (Fig. 6), the "*Tailoring Tools*" are responsible for mapping a game's characteristics with the intended rehabilitation objectives. For example, a "challenge" is a crucial characteristic of instructional games. Providing optimal challenges for a specific patient means that the game difficulty must match the patient's ability in being neither too easy nor too difficult. To enable optimal challenges, it is necessary to continuously adapt the game or create new game levels using “tailoring tools” to match the patient’s current skills. This means “flow experience” will be increased, and the patient will strive towards the goal, and hence, the intended rehabilitation outcomes can be achieved. Adequate flexibility to tailor and adapt game difficulty to suit patient ability is required not only for functional improvement, but also as a motivational incentive for cognitive and motor engagement (Perry et al. 2010). The complexity of a game environment depends on its “tailoring tools” and the ability of these tools to map the intended rehabilitation objectives with various game characteristics (e.g., fantasy, rules/goals, sensory stimuli, challenge, mystery, and control) based on the needs and preferences of the patient.

The developments of individualized game-based rehabilitation tasks imply a higher cost. For widespread use in clinical settings, on-going costs for developing individualized game-based tasks must be kept low. Additionally, the timeframe of in-patient rehabilitation is short in most countries. For example, in the United States, stroke patients spend 15 days, on average, in in-patient rehabilitation (Koenig 2012). Hence, if the game development can be adjusted to the short period of stay of in-patient rehabilitation, game-based intervention could provide a solid alternative for individualized cognitive rehabilitation.

Moreover, over time, personal and environmental factors may change (Scherer et al. 2007); similar to therapy goals and demands throughout the process of rehabilitation (Koenig 2012). Hence, rehabilitation interventions should be adapted to fit these changes.

Adaptive training that automatically modifies the current difficulty level to suit a patient’s abilities may have had increments that were too large for some of the patients, and consequently, they may not have advanced as well as the others (Jaeggi et al. 2011). Furthermore, although a fully automated rehabilitation intervention that assesses the patient´s deficiencies and then uses this assessment to create an individualized rehabilitation plan is virtually possible, such an intervention has a lesser chance at being medically accepted (Perry et al. 2010). The rehabilitation professionals’ experience in formulating and determining rehabilitation objectives and selecting exercises and facilities to attain those objectives has to be recognized. However, compromising between a fully automated rehabilitation intervention and a fully therapist-dependent rehabilitation intervention is probably the best alternative (Perry et al. 2010). Reusable content that can be adapted to suit the patient seems to be a cost-effective rehabilitation intervention (Koenig 2012).

From a physical rehabilitation context, while some researchers have designed games to address particular deficits, these customized games take time to create, are expensive, and they only cater to a small number of brain-damaged patients. Environments with authoring tools decrease the time and expense associated with designing the game, and enable therapists to quickly create games tailored to individuals with brain injury (Tam 2010). However, in cognitive rehabilitation, where there is diversity and heterogeneity of cognitive impairments, an environment with "authoring tools" that create customizable games can be more cost-effective and provide games that can meet the specific needs of brain-injured individuals. A study in game-based learning emphasizes that instructor-oriented authoring tools that are user-friendly seem to be indispensable to educational game development. These tools extract technical aspects, reduce game development costs, and enable instructors to deliver game-based intervention that can fit the learner (Torrente et al. 2008).

Furthermore, Garris et al. (2002) in their study, argued that almost pure discovery-based learning does not exist. In other words, self-directed knowledge constructed by learners by themselves, is unrealistic. They further mentioned that, although the goal of instructional games is to attain self-directed, self-motivated learners, the role of the instructor is also critical in supporting the learner's knowledge construction and if this is overlooked, other supporting strategies such as online help should be available.

Therefore, similarly, in cognitive rehabilitation, while people have specific impairments and needs, it is critical that cognitive rehabilitation be paired with appropriate therapist support for effective rehabilitation to occur. If the patient does not receive the required support, the game-based intervention may become incorrectly used or not used at all. Hence, the therapist plays an essential role in patient motivation and game system development. To be successful and effective, game-based interventions should be incorporated into the therapist's daily clinical practice. After all, the therapist is the one who instructs, motivates, and assesses the patients and plays an important role in the patient's recovery.

However, rehabilitation professionals often do not possess advanced knowledge in game/software engineering to understand the underlying design and development. Hence, intuitive interactions within the game environment together with game tailoring facilities/tools should not require too much technical knowledge. Moreover, we want the rehabilitation professionals who will use the game system for their patients' cognitive rehabilitation to be accepting of it.

4.3 Activity

As shown in Fig. 6, the output of the process part is a "custom game", which is an input to the "activity" part and at this point the game is ready to be played by the patient. The game activity is tailored to keep the patients engaged. Retention of patient attention and his/her deep involvement depends on how effective these game activities have been tailored by therapists. The game should involve activities that are appropriate for the target patient. The affective judgments that are constructed from the beginning and on-going game play imply the direction and intensity of further behavior. Patients who have positive judgments are more actively engaged in gameplay, exert more effort and concentration, and are committed to continue the game activity and frequently return to play the game again. If the therapist succeeds in mapping the game's characteristics with the intended rehabilitation goal in the game, a repeating game cycle will be achieved, and this, in turn, may help motivate the patient to engage in the rehabilitation intervention, ultimately leading to specific cognitive and affective outcomes. The game cycle is the key component triggered by specific game characteristics/features, reflecting the user engagement manner in the game activity. The game cycle, as depicted in Fig. 6, consists of repeated loops of user judgment, user behavior, and system feedback. The game cycle is described by various literatures as loops of gameplay that lead to certain user judgments such as increased involvement, enjoyment or confidence. These judgments lead to user behaviors such as greater persistence or intensity of effort, and these behaviors result in system feedback, which again triggers user judgment (Garris et al. 2002; Mattheiss et al. 2009).

4.4 Output

The gameplay activities generate specific outcomes, which are indicated by game achievement i.e. the level of patient achievement in playing the game. This achievement can be in terms of game scores such as the total amount of assets collected, or the time taken to achieve the goal within the game. These game scores indicate the level of patient skills and abilities while he/she is playing the game and can also serve as a form of patient assessment. Hence, new gameplay activities should be modified and adapted to suit each patient's level.

In a game-based learning context, Kearney and Pivec (2007) introduced a time element to allow individuals to progress throughout the game, incrementing their knowledge and acquiring new levels of skill. This suggests that knowledge is acquired over time and an individual's abilities and/or skills are increased through experience. Therefore, from one day to another, progressing to a new game level and achieving small successes may motivate the patient. However, in the long run, patients would want to be assured that their achievements in the games and the time they have spent on this type of treatment are really contributing to overall rehabilitation success. Therefore, presentation of the patient's performances over time will motivate the patient to continue using the game. Furthermore, tracking how much time the patient spends on the game treatment gives therapists invaluable information about their patients' progress and how the games are actually helping them, particularly for those who are not in one-on-one rehabilitation sessions.

In addition, if patients can visually see their improvements over time on graphs or other visual statistical presentations, they will be more inclined towards using the game as a form of treatment and will continue using it. Therefore, outcomes can be extended to describe "changes in patient outcomes". In other words, automatic measures can be developed to track a patient's cognitive performance over time so as to measure improvements in a given cognitive function. However, information about the sensitivity and reliability of automatic measurement systems is still absent. This gap needs a complex debriefing process and a comprehensive empirical study to validate any proposed automatic measures scientifically. A better solution to measure "changes in outcomes" would be to use validated instruments such as sound questionnaires and standardized clinical instruments previously discussed in the "condition" part.

Furthermore, outcome generally plays a crucial role. Taking into account the short time frame of in-patient rehabilitation and the therapist's limited time, the individualized rehabilitation should effectively use this short span of time without the need for numerous supervised rehabilitation sessions. Therapists would like to have data regarding what their patients do when they practice, and use these as monitoring and tracking mechanisms. This promotes the possibility of unsupervised rehabilitation that can be continued after the patient is discharged from in-patient rehabilitation services. Thus, patients do not need long instructions and supervision from their therapists.

On the other hand, reflection is an important part of the rehabilitation process for the patient as well as for the therapist. According to the "active learning" concept, learning follows a cyclic loop instead of a linear pattern. For example, experiencing learning intervention results in reflections, and these reflections lead to certain conclusions, which in turn lead to construction of a plan for new actions and then back to experiencing again (Kolb 1984). Reflection is a crucial part of this cycle. Without reflection, the cycle cannot produce new conclusions and actions (Bizzocchi and Paras 2005).

In the proposed framework, the "reflection on outcomes" provides a link between the "outcome" and the therapist and also between the "outcome" and the patient. These are shown in Fig. 6. The reflection on outcomes has two objectives.

Firstly, in regard to the patient, it is vital to consider the patient's expectations of rehabilitation game activities and outcomes. People like to work at things that are important to them rather than at things that are important to others. Affective reactions comprise of self-efficacy, attitudes, and feelings of confidence, preferences, and dispositions. These reactions are recognized as one type of outcome. For example, attitude change is considered one of the training objectives of an instructional intervention (Garris et al. 2002). Moreover, besides the game-playing activity itself, characteristics of individual expectancies, interests, motives, outcomes, and consequences should be considered in the development of game-based learning (Mattheiss et al. 2009). Gameplay produces expectations in the individual in regard to the outcome of the game. These expectations affect individual motivation. In other words, a player who believes that he/she can reach certain outcomes and eventually does so is motivated to continue using the game and this can be an incentive for him/her to exert more effort to attain the intended goal of the game (Mattheiss et al. 2009). Hence, motivation and engagement in game-based trainings will be achieved, if patients believe in potential success during gameplay. This perception strengthens the patient's confidence, and is reflected through outcomes representing his/her performance in the game experience.

Secondly, in regard to the therapist, reflecting on outcomes can provide an assessment of the patient's progress and his/her expectation. After the game has been designed, delivered, and then played by the patient, the therapist needs to check goal attainment by comparing outcomes with intended rehabilitation objectives. This also proves useful in assessing the effectiveness of game interventions. Outcomes enable therapists to capture changes in skills; what the patient is able to do, their level of task performance, and affective reactions. Therefore, outcomes can reflect a patient's progress towards a goal and his/her expectation. Hence, rehabilitation effectiveness is achieved when therapists are guided by the outcome to continuously adjust and modify the game according to their patient's current skills and expectations via therapist-oriented tailoring tools offered by the game environment. Useful information that enables rehabilitation professionals to compare a patient's performance with normative data is also gained. Plus, the outcome of this reflection provides the opportunity to compare task performance across patients and rehabilitation sessions so as to quantify a patient's progress.

5. Rehabilitation Gaming System

To demonstrate the implementation of the proposed framework, a Rehabilitation Gaming System (RGS) prototype is developed for experimental use and evaluation. The first consideration for the prototype development process is to use the Web. Development of a Web-based rehabilitation platform is a very promising foreseeable solution. Web-based platforms can be used in internal networks as well as long-distance public networks for patients to remotely access them from home at no cost. The Web is a cross-platform environment that can be accessed from both desktop and mobile platforms regardless of the operating system used. Such systems could be utilized in clinical settings for in-patient and outpatient rehabilitation. The second consideration for the development process is to use Adobe Flash. This software is a multimedia platform that is well known for creating simulations and games that can be easily viewed on the Web, and this is considered as one of the positive factors of using this software in the development process.

The main interface of the RGS is shown in Fig. 8. The user (i.e., therapist and/or patient) can only advance on to the next level after filling in his/her user name and password fields.

Once the therapist logs into the RGS system, the therapist's main interface will appear as shown in Fig. 9. This interface is characterized by three buttons at the top of the screen, which are "main", "patient", and "sign out". The "main" button allows the therapist to go back to the main interface. The "patient" button allows the therapist to add new patients, as well as access and modify patient information. Clicking on the button marked "sign out" exits the user from RGS.

The middle of the screen shows a list of game levels that was previously created by the therapist. Therapists can edit and modify this list by clicking on the game's level name or by clicking on the "new level" button to create a new game level. The bottom of the screen displays patient results. Once the patient finishes the game exercises, therapists can login to the system, and see details such as patient name, the levels that were played by the patient, when the patient accessed the game system (time and date), the total amount of time taken by the patient to finish the game level, and the total correct answers. Therefore, therapists can easily track patient performance online, and this enables them to adjust and modify the game's tasks accordingly.

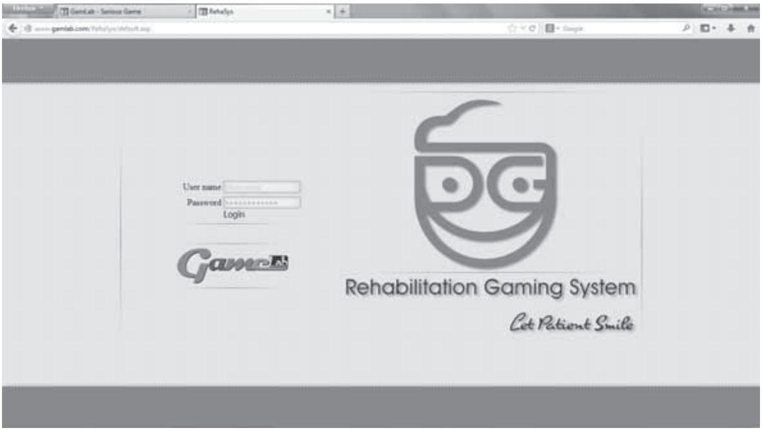


Fig. 8. RGS main interface.

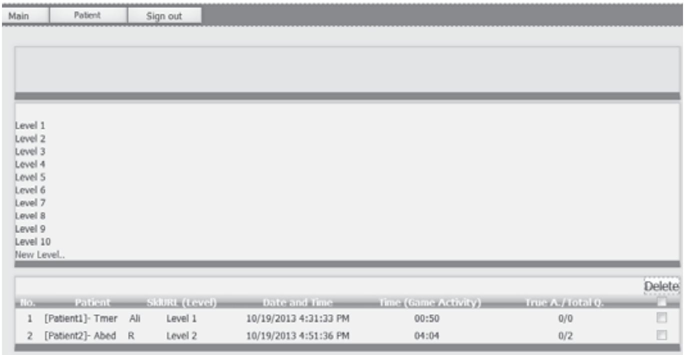


Fig. 9. Therapist's main interface.

The RGS patient editor, shown in Fig. 10, allows therapists to add new patient information such as their user name, password, first name, last name, email, phone number, gender, and case description. On top of that, the RGS patient editor allows therapists to assign game levels to the patients. Also, it allows therapists to search for certain patients when they select option/criteria and click on the button marked "search". The system will then generate a query, which will consult the RGS database for the desired information. A list of patient details will be presented so therapists can simply click on the "check box" that appears beside each patient's record, allowing the therapist to "delete" or "update" patient information. The aim of this research, in the end, is to integrate the RGS with the management information system of the respective rehabilitation hospital. Therefore, patient information will be captured from the hospital information system. However, this module (i.e. create and modify patient profile) was purposely developed so that RGS could also be used as a stand-alone system. Moreover, the generated game interface for patients is text-less. Also, therapists have the ability to select any language spoken by their patients. Hence, RGS can also be used worldwide. Furthermore, it offers a new feasible and cost effective alternative for rehabilitation.

The game design environment, shown in Fig. 11, lets therapists create the game and save it. To the right is an empty field surrounded by a border. On the left, there are tools that can be used by therapists to build and tailor the game within the field on the right. There is no limit to the number of mazes that therapists can create. Therapists can simply click on a particular tool to activate it, and then start the game design process on the empty field. In case of an error, like putting the wrong game object in the wrong place, the therapist can simply select the correct one from the "tools" panel and drop it over the wrong one. The game object will instantly turn into the correct choice.

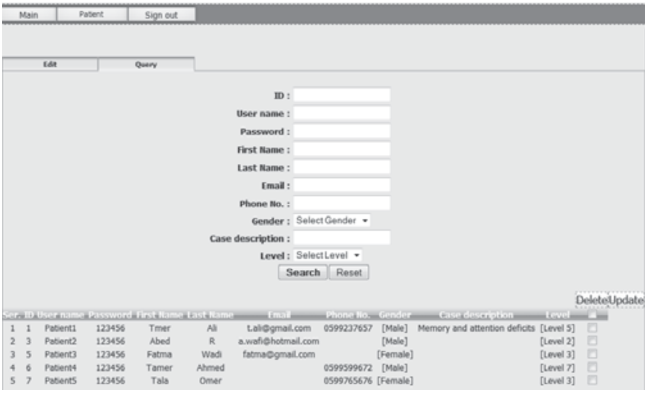


Fig. 10. RGS patient editor.

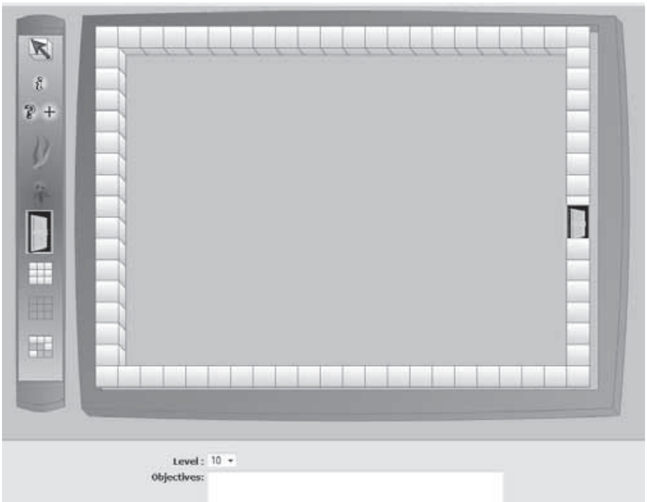


Fig. 11. The game design environment.

At the bottom part of the game design environment, there is a text box and a "save" button. Therapists can enter texts in the text box describing the objectives of the game, and the instructions on how to play the game.

When patients login to RGS to play the assigned game, these texts will be launched as the game introduction. In the end, after the therapist completes the design stage, he/she can simply click on the button marked “save” and the final result will be saved.

The RGS questions editor, shown in Fig. 12, enables therapists to create customized questions with different difficulty levels. Patients will have to answer a number of questions. The questions can be about patient’s background, such as “How many siblings do you have?” and/or general knowledge questions such as “How many days are there in a week?” and/or mathematical equations such as “5-2=?” Each question has three possible answers. After clicking on one of the answers, the game will provide feedback to the patient about his/her answer. The complexity and the number of questions for patients depend on the game level designed by the therapist. The patient’s answers will allow therapists to analyze his/her memory and/or cognitive progressions.

With patient assessment, it is possible to “prescribe” training that targets specific cognitive functions. The therapists, based on their patient’s abilities, limitations, and preferences, will create this game training. As shown in Fig. 13, therapists can access the RGS design environment and start the design process using the tailoring tools, draw the Maze pathway, add the game objects, identify the behavior of the opponents, such as how they move or react during game play, and edit and add questions to the playing field.

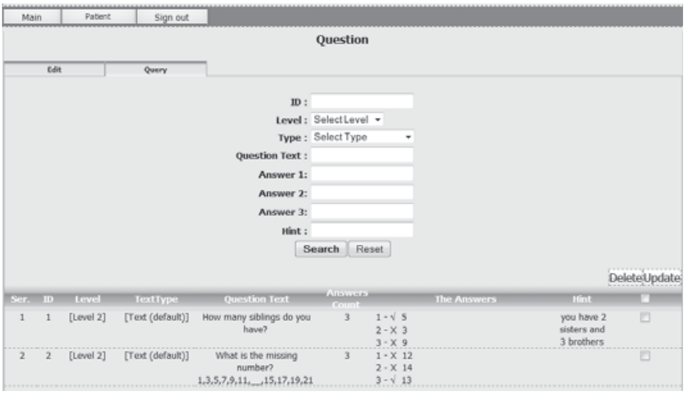


Fig. 12. RGS questions editor.

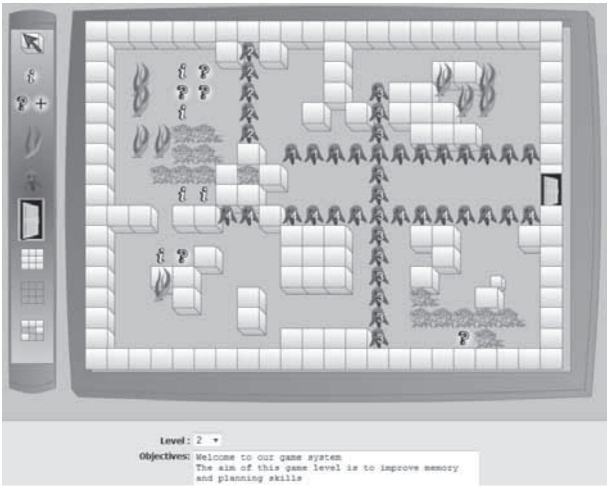


Fig. 13. Game level design by therapist.

Once a patient's profile is created and the game training is tailored and assigned by the therapist, the patient will then have access to a personalized set of game tasks. Patients will have to login to the system first. Either the patients or caregivers (for those who have interaction difficulties) can do this simple task. Then, a short introductory screen will appear to explain the game's purpose and/or how it is played as shown in Fig. 14. Patients can then start the game, as shown in Fig. 15, by clicking on the "green arrow" button at the bottom of the screen.

Fig. 14. Introductory screen.

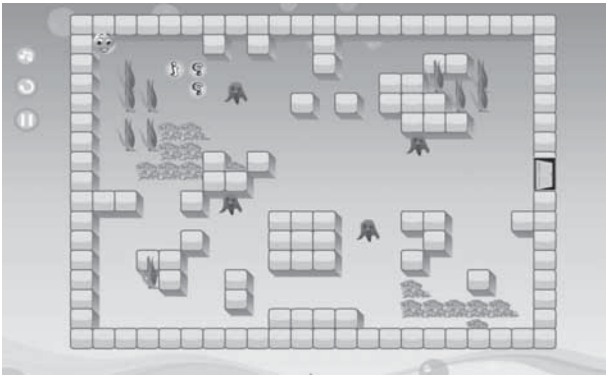


Fig. 15. The patient’s gameplay interface.

Patients control the yellow “smiley face” character using the keyboard's arrow keys and steer it from the upper left corner of the playing field to the Maze destination (i.e. the opened door). Unfortunately, there are opponents blocking the way, which are continuously moving. Players need to avoid collision with them to reach the destination. These restrictions force patients to watch out while walking and they must plan their pathway before moving to the Maze’s destination. Whenever the “smiley face” hits one opponent while moving, it will fly, and then automatically move back to the playing field. The length of time taken to reach the destination is recorded. The quicker the player can reach the destination, the better the results he/she can achieve.

Moreover, while patients go through the Maze toward the destination, he/she will be asked to answer some customized questions. There are a number of question marks on the playing field. Once the smiley face hits a question mark, it will disappear and a question will be presented via a pop-up screen as shown in Fig. 16. The patient will have three choices of answers: when he/she thinks that he/she has found the answer, he/she can click on it and he/she will be given a feedback telling him/her if he/she has chosen correctly. Then, the patient can press the space bar or simply click the “arrow button” to continue playing. In addition, help icons labeled “i” are available on the playing field; patients can use them to answer the questions as shown in Fig. 17.

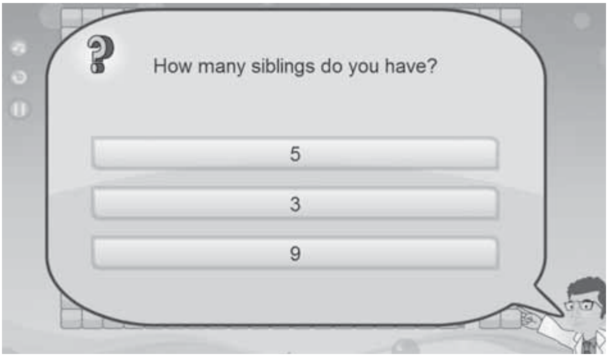


Fig. 16. Game's question interface.

Fig. 17. Helping aids interface.

Outside the Maze, in the upper left corner, three buttons are provided for the patient. The first one allows patients to replay the game level. The second button can pause the game and the last button can mute the game's music. Patients continue until they reach the Maze destination (i.e. when the "smiley face" touches the opened door). The score will be calculated based on the total time taken to complete the game level and the number of questions the patient had answered correctly. The result will be presented and automatically saved. The patient will then be given the opportunity to advance to the next game level.

6. Conclusion

In this chapter, a conceptual framework for designing effective games for cognitive rehabilitation of patients with acquired-brain-injury has been discussed. Each component of this framework plays a vital role in ensuring the effectiveness of the therapeutic game that is produced. This framework can be a useful guide for game developers and other practitioners in designing effective therapeutic games that can significantly affect patients, therapists, and health care systems. The proposed framework simulates conventional rehabilitation procedures. In conventional rehabilitation, there is a significant coordination between the patient and the therapist. The therapist formulates a treatment plan based on the patient's assessment, and then selects rehabilitation exercises and techniques for him/her. The patient depends on the therapist's tailored plan to reduce his/her deficits and increase his/her participation, while the therapist depends on the patient to cooperate with the established rehabilitation program. Therefore, the therapist has to be considered together with any potential technology. As a consequence, in the proposed framework, the therapist remains an integral part in the planning (designing) of the game-based rehabilitation intervention. Therefore, a user-friendly, therapist-oriented game environment with "tailoring tools" is important in therapeutic game development, producing feasible games that can precisely cater to the varied abilities and needs of brain-damaged patients. To demonstrate the implementation of the framework, a Rehabilitation Gaming System (RGS) is developed. RGS benefits the rehabilitation process and fulfills its real needs such as increasing patient's motivation by providing an individualized rehabilitation game experience while simultaneously reducing the development costs associated with it; allowing therapists to track patient activities and to assess their progress. Furthermore, it is likely to open new opportunities for home-based and unsupervised rehabilitation. At the end of it all, a question arises on whether or not such interventions will be accepted by the target group (i.e. patient and therapist). Hence, in the near future, an evaluation of the developed rehabilitation gaming system that targets therapists will be conducted to determine their acceptance and satisfaction level towards the system.

References

Alankus, G., Lazar, A., May, M. and Kelleher, C. 2010. Towards customizable games for stroke rehabilitation. *The 28th international conference on Human factors in computing systems*, hlm. 2113-2122.

Bartle, D., Rossoff, S., Whittaker, D., Gooch, B., Kerns, K. and MacSween, J. 2010. Cognitive games as therapy for children with FAS. *ACM SIGGRAPH 2010 Posters*, hlm. 48.

BELCHIOR, P. D. C. 2007. Cognitive training with video games to improve driving skills and driving safety among older adults. University of Florida.

Bizzocchi, J. and Paras, B. 2005. Game, motivation, and effective learning: An integrated model for educational game design. *International DiGRA Conference*, hlm.

Broeren, J., Bellner, A., Fogelberg, M., Göransson, O., Goude, D., Johansson, B., Larsson, P., Pettersson, K. and Rydmark, R. 2008. Exploration of computer games in rehabilitation for brain damage. *Proceedings of the 7th International Conference on Disability, Virtual Reality and Associated Technologies. September*, hlm. 75-80.

Burdea, G. 2003. Virtual rehabilitation-benefits and challenges. *Methods of Information in Medicine-Methodik der Information in der Medizin* 42(5): 519-523.

Burke, W., McNeill, M., Charles, D., Morrow, P., Crosbie, J. and McDonough, S. 2010. Designing engaging, playable games for rehabilitation. *Proc. 8th Intl Conf. on Disability, Virtual Reality and Assoc. Technologies, Viña del Mar/Valparaíso, Chile, August*, hlm. 195-201.

Carmien, S., DePaula, R., Gorman, A. and Kintsch, A. 2003. Increasing workplace independence for people with cognitive disabilities by leveraging distributed cognition among caregivers and clients. *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, hlm. 95-104.

Cruz, V. T., Pais, J., Bento, V., Mateus, C., Colunas, M., Alves, I., Coutinho, P. and Rocha, N. P. 2013. A Rehabilitation Tool Designed for Intensive Web-Based Cognitive Training: Description and Usability Study. *JMIR Research Protocols* 2(2):

Doidge, N. 2007. *The brain that changes itself: Stories of personal triumph from the frontiers of brain science*. Ed. Penguin. com.

Elaklouk, A. M. and Zin, N. A. M. 2012a. Games for cognitive rehabilitation. *5th Annual International Conference on Computer Games Multimedia and Allied Technology*, hlm.

Elaklouk, A. M. and Zin, N. A. M. 2012b. Requirements for game-based cognitive intervention system for acquired brain Injury *The GSTF Journal on Computing (JoC)*

ÉML., P. 2009. Role of neuropsychological assessment in the differential diagnosis of Alzheimer's disease and vascular dementia. *Dementia & Neuropsychologia*

Fok, S. 2009. Internet-enabled exercises and prosthesis for home-based cognitive rehabilitation. *International Journal of Biomedical Engineering and Technology* 2(1): 29-43.

Garris, R., Ahlers, R. and Driskell, J. E. 2002. Games, motivation, and learning: A research and practice model. *Simulation & gaming* 33(4): 441-467.

Hays, R. T. 2005. The effectiveness of instructional games: A literature review and discussion. DTIC Document.

Jaeggi, S. M., Buschkuehl, M., Jonides, J. and Shah, P. 2011. Short-and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences* 108(25): 10081-10086.

Kearney, P. and Pivec, M. 2007. Recursive Loops of Game-Based Learning: a Conceptual model. *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, hlm. 2546-2553.

Koenig, S. 2012. Individualized Virtual Reality Rehabilitation after Brain Injuries. College of Engineering, University of Canterbury

Kolb, D. A. 1984. *Experiential learning: Experience as the source of learning and development*. Ed. Prentice-Hall Englewood Cliffs, NJ.

LOPresti, E. F., Bodine, C. and Lewis, C. 2008. Assistive technology for cognition [Understanding the Needs of Persons with Disabilities]. *Engineering in Medicine and Biology Magazine, IEEE* 27(2): 29-39.

Maclean, N., Pound, P., Wolfe, C. and Rudd, A. 2002. The concept of patient motivation a qualitative analysis of stroke professionals' attitudes. *Stroke* 33(2): 444-448.

Mattheiss, E., Kickmeier-Rust, M., Steiner, C. and Albert, D. 2009. Motivation in game-based learning: It's more than 'flow'. *Lernen im Digitalen Zeitalter, Potsdam* 77-84.

Oates, J., Weston, W. W. and Jordan, J. 2000. The impact of patient-centered care on outcomes. *Fam Pract* 49(796-804.

Panic, A. S. 2010a. Addressing patient motivation in virtual reality based neurocognitive rehabilitation. BA Thesis, Hogere Technische School Arnheim, Arnheim.

Panic, A. S. 2010b. Research Project: Virtual Reality Based Neurocognitive Rehabilitation. Delft University of Technology.

Perry, J. C., Andureu, J., Cavallaro, F. I., Veneman, J., Carmien, S. and Keller, T. 2010. Effective game use in neurorehabilitation: user-centered perspectives. *Handbook of Research on Improving Learning and Motivation through Educational Games, IGI Global*

Prensky, M. 2001. The digital game-based learning revolution. *Digital game-based learning*

Rego, P., Moreira, P. and Reis, L. 2011. Natural user interfaces in serious games for rehabilitation. *Information Systems and Technologies (CISTI), 2011 6th Iberian Conference on*, hlm. 1-4.

Rego, P., Moreira, P. M. and Reis, L. P. 2010. Serious games for rehabilitation: A survey and a classification towards a taxonomy. hlm. 1-6.

Rosenbaum, P. and Stewart, D. 2004. The World Health Organization International Classification of Functioning, Disability, and Health: a model to guide clinical thinking, practice and research in the field of cerebral palsy. *Seminars in Pediatric Neurology*, hlm. 5-10.

Scherer, M., Jutai, J., Fuhrer, M., Demers, L. and Deruyter, F. 2007. A framework for modelling the selection of assistive technology devices (ATDs). *Disability & Rehabilitation: Assistive Technology* 2(1): 1-8.

Scherer, M. J. and Glueckauf, R. 2005. Assessing the Benefits of Assistive Technologies for Activities and Participation. *Rehabilitation Psychology* 50(2): 132.

Steiner, W. A., Ryser, L., Huber, E., Uebelhart, D., Aeschlimann, A. and Stucki, G. 2002. Use of the ICF model as a clinical problem-solving tool in physical therapy and rehabilitation medicine. *Physical therapy* 82(11): 1098-1107.

Tagliaferri, F., Compagnone, C., Korsic, M., Servadei, F. and Kraus, J. 2006. A systematic review of brain injury epidemiology in Europe. *Acta neurochirurgica* 148(3): 255-268.

Tam, S. 2010. An Environment for Stroke Therapy Game Authoring. Washington University.

Torrente, J., del Blanco, Á., Moreno-Ger, P. and Fernández-Manjón, B. 2012. Designing serious games for adult students with cognitive disabilities. *Neural Information Processing*, hlm. 603-610.

Torrente, J., Moreno-Ger, P., Fernández-Manjón, B. and Sierra, J. L. 2008. Instructor-oriented authoring tools for educational videogames. *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on*, hlm. 516-518.

Tsaousides, T. and Gordon, W. A. 2009. Cognitive rehabilitation following traumatic brain injury: assessment to treatment. *Mount Sinai*

Journal of Medicine: A Journal of Translational and Personalized Medicine 76(2): 173-181.

Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., Orvis, K. L. and Conkey, C. 2009. Relationships Between Game Attributes and Learning Outcomes Review and Research Proposals. *Simulation & gaming* 40(2): 217-266.

CONTEXT-AWARE E-LEARNING INFRASTRUCTURE

STANIMIR STOYNOV, HUSSEIN ZEDAN, EMIL DOYCHEV, VESELINA VALKANOVA, ASYA STOYANOVA-DOYCHEVA AND VLADIMIR VALKANOV

Abstract

Within this chapter is presented a service- and agent-oriented infrastructure created to support the delivery of context-aware education services and teaching content provision, known as Distributed eLearning Centre (DeLC). The current state of DeLC and its ongoing transformation into a Virtual Education Space are described in detail. Furthermore, various formal tools for support of context-aware behaviour of DeLC and VES are discussed. Calculus of Context Aware Ambients (CCA) will be used for modeling and verification of DeLC and VES infrastructure. Security and Context-Aware Flow (SC-Flow) will be used for the specification and implementation of scenario-based management in both infrastructures. Tempura will be applied for processing temporal dependencies in these scenarios.

Keywords: service- and agent-oriented architectures, e-learning, education services, intelligent agents, context-aware ambients, scenario-based management, context-aware workflow.

1. Introduction

The Distributed eLearning Centre project (DeLC) aims at the development of an infrastructure for context-aware delivery of electronic education services and teaching content, personalized and customized for each individual user (Stoyanov et al., 2005; Stoyanov et al., 2008) In this chapter, the current state of DeLC and its ongoing transformation into a virtual education space are presented.

In conformity with (Dey, 2000), a context is any information that can be used to characterize the situation of an entity where an entity may be a student, a teacher, a place, or an object (a service, an agent, a structure of teaching material, a digital library, an event) that is considered relevant to the interaction between a user and an application, including the user and the application themselves. Context-awareness is the ability to identify changes in the environment and to personalize and adapt education services and teaching content in accordance, i.e. the personalization and adaptation are very important attributes of context-aware systems. To satisfy this requirement new types of software architectures are needed, which are enabled to sense aspects of the environment and use this information to adapt their behaviour in response to the changing situation. The DeLC infrastructure is service-oriented and agent-based providing wireless and fixed access to services and content.

The development of context-aware architectures can be benefited from some ideas and approaches of pervasive computing, internet of things and semantic web. Pervasive computing is a new paradigm for the next generation distributed systems where computers disappear in the background of the users' everyday activities. In such a paradigm computation is performed on a multitude of small devices interconnected through a wireless network. Fundamental to pervasive computing is that any component (including user, hardware and software) can be mobile and that computations are context-aware. As a result, mobility and context-awareness are important features of any design framework for pervasive computing applications. Context-awareness requires applications to be able to sense aspects of the environment and use this information to adapt their behaviours in response to changing situations.
Furthermore, we present various formal tools to support the context-aware behaviour of DeLC and VES. Calculus of Context Aware Ambients (CCA in short) will be used for modeling and verification of the DeLC and VES infrastructure that is context-aware. This process calculus is built upon the calculus of mobile ambients and introduces new constructs to enable ambients and processes to be aware of the environment in which they are being executed. This results in a powerful calculus where both mobility and context-awareness are first-class citizens. We present the syntax and a formal semantics of the calculus. We also demonstrate a new theory of equivalence of processes which allows the identification of systems that

have the same context-aware behaviours. We prove that CCA encodes the Pi-calculus which is known to be a universal model of computation. We have used our CCA to specify DeLC in its entirety, hence achieving its correctness. Such a dynamic system must enforce complex policies to cope with security, mobility and context-awareness. We show how these policies can be formalised and verified using CCA. In particular, an important liveness property of the mLearning system is proved using the reduction semantics of CCA. SC-Flow will be used for the specification and implementation of scenario-based management in both infrastructures. Tempura will be used for processing temporal dependencies in the scenarios.

2. DeLC overview

2.1 Infrastructure of DeLC

The distributed eLearning Center (DeLC) is a reference architecture, aiming to support context-aware provision of education services and electronic content. The DeLC architecture is modeled as a graph which consists of separate nodes. DeLC can be used to model real units (laboratories, departments, faculties, colleges, and universities) which offer a complete or partial education cycle.

There are various types of nodes:

- eLearning Nodes (eLNs) – these operate as autonomous hosts for delivery of electronic services and teaching content;
- Development Nodes (DNs) – they host various specialized tools for support of the education process, for example teaching material editors;
- Supporting Nodes (SNs) – nodes for assistance of eLearning Nodes transparent to users;
- Interface Nodes (INs) – nodes liaising with external (legacy) systems transparent to users.

The configuration of the network edges is such as to enable the access, incorporation, use and integration of electronic services located on the different eLNs. The eLNs can operate as isolated nodes or integrated in more complex virtual structures, called clusters. Remote eService activation and integration is possible only within a cluster. In the network model we can easily create new clusters, reorganize or remove existing clusters (the reorganization is done on a virtual level, and it does not affect the real organization). For example, the reorganization of an existing cluster can be made not by removing a node but by denying the access to the offered one by its services. The reorganization does not disturb the function of the other nodes (as nodes are autonomous self-sufficient education units providing one or more integral education services).

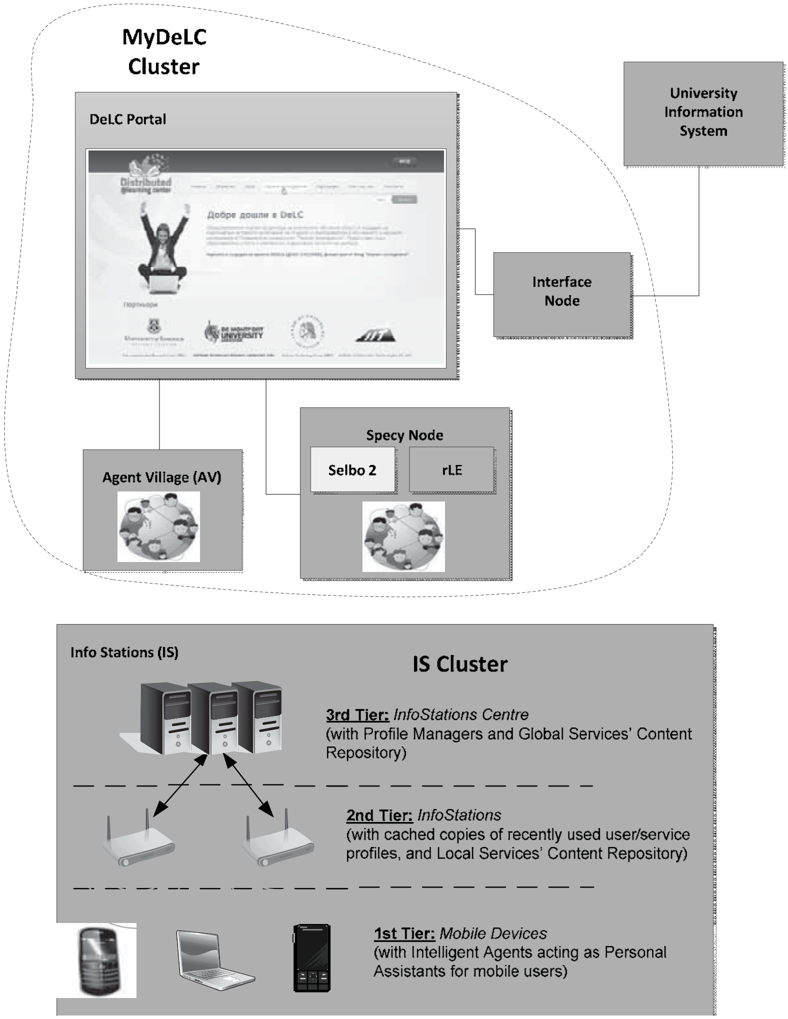


Figure 2.1 DeLC Infrastructure

An important feature of the eLearning Nodes is the access to supported services and electronic content. In relation to the access there are two kinds of nodes:

- Mobile eLearning Node, and
- Fixed eLearning Node.

For both nodes individual reference architectures are proposed within DeLC.

The current DeLC infrastructure (Fig. 2.1) consists of two separate education clusters described in more details in this chapter. The first one, known as MyDeLC, delivers education services and teaching content through an education portal. The second one provides mobile access to services and content over an extended local network known as InfoStation network.

2.2 Context-Awareness in DeLC

As we mentioned earlier, the context plays an important part in the operational behaviour of DeLC. One of the many challenges is to design the infrastructure to operate in pervasive environments. This requires the provision for a computational model in which the context is made a first class citizen and each of these contexts has its own policies that constrain its behaviours – from access control to privacy and confidentiality provisions. Such a computation model has to treat both constraints in a uniform and integrated fashion.

Contexts can take a variety of forms: different education services, digital libraries, teaching material, events, hand-held devices, etc. A context is presented as (what we call) a *context frame*, which is a set of variables (or attributes) of interests. For example, attributes of interests of a context such as a PDA could be its processor speed, memory size, and battery life time. On the other hand, attributes such as age, qualification and work experience will be of interest in the case of a human context; yet body temperature, blood pressure and kidney functions are attributes more appropriate in the case of a hospital ward context. In this sense, a system is context-aware, if it uses context frames for the delivery of desired information and services to the user, where usability is dependent on the the user's action. Usually, the values of the attributes in the context frames are supplied from the environment. The purpose of adaptation is to provide seamless, transparent and adequate enforcement of consumer requests for services and/or electronic content. The purpose of the personalization is to be able to adapt depending on the users' individual characteristics, both taking into account the corresponding context frame. So users should have the feeling that the environment is designed to meet their needs.

In DeLC the adaptation and personalization (A&P) is supported by a general model which will be presented here briefly. In respect of A&P everything in DeLC, which can be identified, accessed or activated is considered as an *information object* (IO). There are various types of IOs, such as electronic services, teaching material (an integral knowledge piece, lecture, lesson, module, and lecture course), digital libraries, users, and events.

Each IO type is presented by an appropriate classification scheme and each IO has an individual context frame. Classification is a preferred method of organizing information resources in virtual spaces and it usually uses one dimensional method. In our case, we use a multi-dimensional classification applying some ideas of the Resource Space Model (RSM) (Zhuge, 2008), for the following two reasons:

- Various information objects have to be analyzed from different points of view;
- Increasing or decreasing the dimensions is an effective way of generalization or specialization of knowledge, i.e. a powerful means to support A&P.

In this way an IO can be uniquely characterized by the location in the classification scheme and the individual profile. Multidimensional classification schemes are presented as the so-called *characteristic spaces* where the dimensions present various attributes of IOs and can be assessed and classified by different methods. Characteristic spaces and profiles build a meta-level in DeLC. A segment of the characteristic space of MyDeLC users is presented in Fig. 2.2. A characteristic space of electronic education services can be seen in (Ganchev et al., 2009).
Moreover, the A&P is supported at various levels and must take account of the environment. In our case we distinguish two basic levels:

- *Architectural level* – various kinds of active components with sufficient flexibility, power and efficiency are supported in the infrastructure;
- *Content level* – appropriate structuring of content and passive components where the SCORM 2004 standard is the preferred presentation.

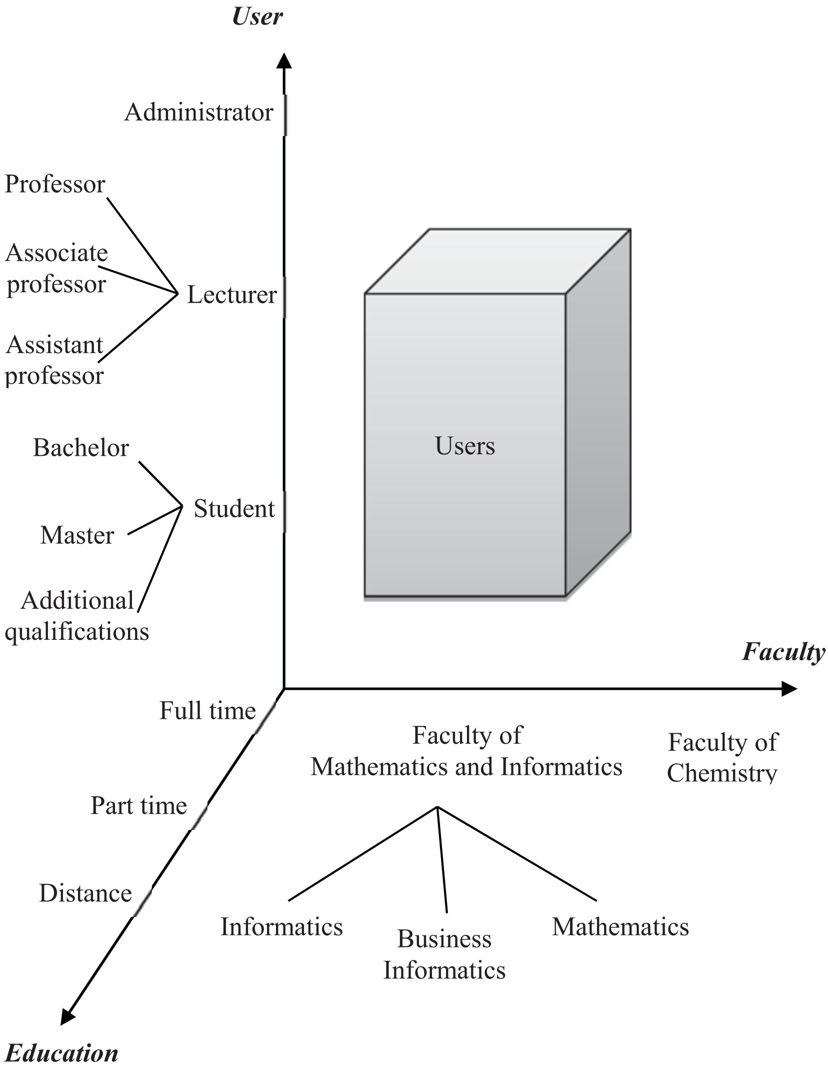


Figure 2.2 User Character Space

A&P can be done in two ways:

- *Statically* – during the initialization by configuration parameters or by extracting information from external nodes;

- *Dynamically* – depending on the progress of the education process.

3. MyDeLC Cluster

MyDeLC cluster is composed of the following four nodes:
- MyDeLC portal – the kernel of the cluster supports the planning, organizing, and conducting the education process;
- Agent Village – it provides intelligent help to the portal;
- Specy Node – this is a specialized node hosting various e-learning tools;
- Interface Node – it implements an interface to the University Information System.

In this section, the first three nodes will be presented in more detail.

3.1 Education Portal MyDeLC

The education portal MyDeLC is a front-end system for direct interaction with lecturers, students and administrators. We try to combine the advantages of various e-learning standards in order to ensure favorable conditions for context-aware delivery of education services and teaching content:
- SCORM 2004 (ADL, 2009) – SCORM was developed to support the creation and portable delivery of reusable teaching content for self-spaced computer-based training;
- Common Cartridge (CC) – IMS GLC describes CC as ‘‘standards for organization, publishing, distribution, delivery, search, and authorization of a wide variety of collections of digital learning content, applications, and associated online discussion forums used as the basis for or in support of online learning of any type” (IMS GLC, 2009).

The portal architecture is service-oriented and multi-layered consisting of four layers (Fig. 3.1): user interface, e-services, engines, and digital libraries.

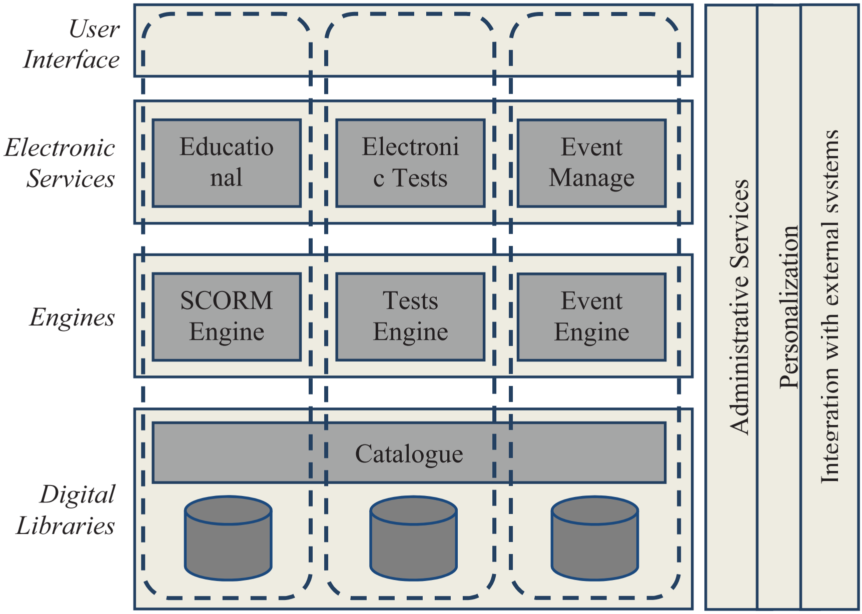


Figure 3.1 MyDeLC Portal Architecture

The user interface supports the connection between users and the portal. Through it users can register in the system and create their own personalized education environment. The user interface visualizes and provides access for the user to services, depending on the role assigned to them during the registration.

In accordance with the service-oriented architecture of the portal its functionality, intended for operation by users, is offered as separate electronic services. For flexible and efficient management we have developed a classification of the provided services. The services are divided into three major groups by their functionality:

- Services for the preparation, organization and planning of the teaching process – this group consists of services for providing information about the education process and the creation of curricula and curriculum schedules;
- Services, directly supporting the learning process – this group consists of services, such as e-lectures, e-test, and electronic consultation;

- Services for the recording and documentation of the teaching process – these services support the automatic generation of documents required for the recording of the learning process, e.g. examination protocols, an electronic student book, a teacher's electronic diary, and archives.

According to their purpose, services are divided into three categories:

- Services for students – they are intended for training and supervision of trainees. These include services for accessing and reading pre-prepared electronic lectures (based on an intelligent search module), for sitting for tests (mandatory or optional) in studied disciplines, for consultations with teachers, reviews of a student's book as well as some additional services not directly related to teaching such as an intelligent personal calendar;
- Services for teachers – these are intended for the teaching staff and allow the addition of lectures in the digital library, creating a library of test questions, constructing, reviewing and evaluating tests, and an intelligent personal calendar;
- Services for administrators – they are intended for the management, configuration, and planning of the teaching process and for related users, rights and services. These include services for the creation, activation, and deactivation of users and roles, the addition and removal of services and rights of access, curriculum management, an academic calendar, early warning systems, and a calendar of events (e.g. upcoming exams, sessions, and important dates).

The *engines* operate as system services transparent for the users and their basic purpose is to process the *characteristic spaces* and *context frames* so as to ensure context-aware support of the education services. In MyDeLC architecture, the following engines are implemented:

- SCORM Engine – it is implemented in the portal architecture for delivery of an interpreter of the electronic content, developed in accordance with the SCORM 2004 standard;
- Test Engine – it assists the electronic testing;
- Events Engine – it provides a model for the management of user-oriented events.

The SCORM 2004 engine (Fig.3.2) architecture includes the following components:

- *SCORM Repository* – this is the storage place for the SCORM content which consists of a database and file resources;

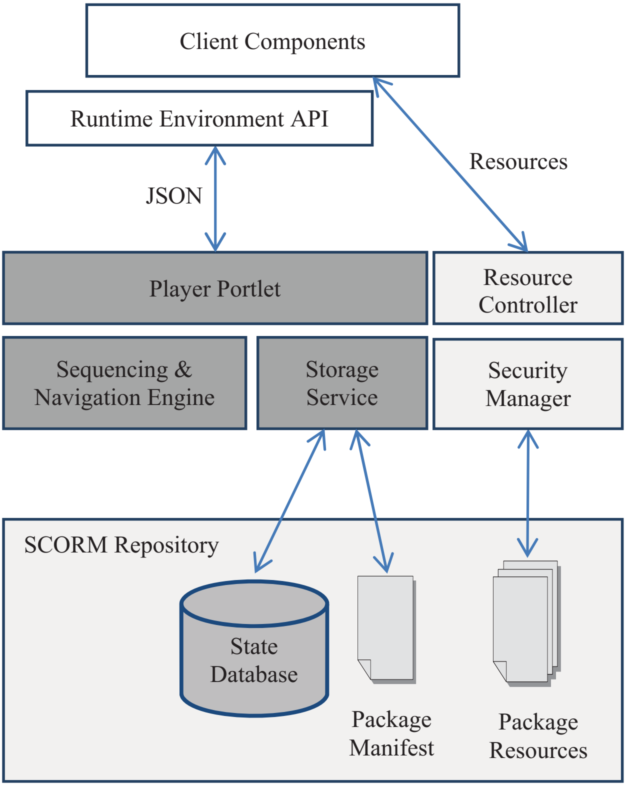


Figure 3.2 SCORM Engine Architecture

- *Storage Service* – it provides operations for storage and retrieving of the SCORM content;
- *Security Manager* – it guarantees that only the allowed content will be delivered to the learner;
- *Sequencing & Navigation Engine* – this is the core of the SCORM Engine. Here the rules defined by the content developer are processed to resolve the next SCO/Asset for delivery;
- *Player Portlet* – it implements the user interface and its connection with the Sequencing & Navigation Model;
- *Resource Controller* – it delivers the device resources (usually a browser), requested by the client, if they are allowed by the Security Manager;

- *Runtime Environment API* – ECMAScript (Javascript implementation) containing the defined in SCORM API functions, which are used by the SCORM packages in runtime;
- *Client Components* – these are user interface components that are working in the client browser.

The Test Engine is implemented for preparation, execution and evaluation of electronic tests. The Test Engine architecture (Fig. 3.3), implemented as interacting portlets, consists of the following components:

- *Library Portlets* for managing the content library. They are intended to be used by the lecturers (content creators);
- *Assessment Portlets* for the execution of an examination. These portlets are intended for the lecturers as well;
- *Player Portlets* for generation of a pseudo-unique test per learner and delivering it to him/her;

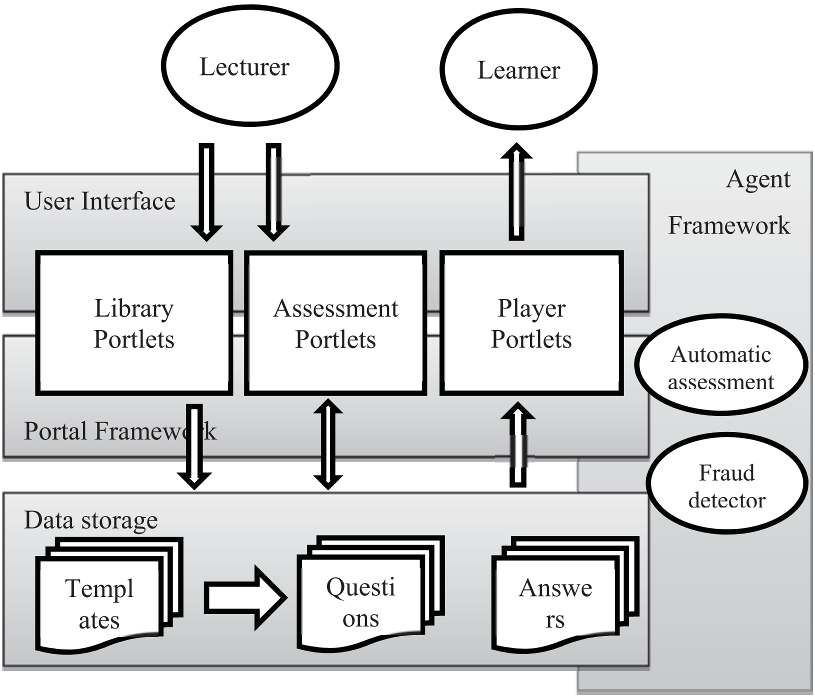


Figure 3.3 Test Engine Architecture

The Test Engine provides two editors to the lecturers, namely Question Editor and Test Editor. The question editor enables the lecturers to create test questions for various subject disciplines stored in a specialized digital library. The test editor can be used for the creation of test templates or ordinary tests. The test templates are constructed when a lecturer would like to implement personalized testing. In this case the Test Engine automatically generates individual instances for each student. The lecturers can use a fixed set of questions for each learner. In this case the test editor is used for creating ordinary tests. The Test Engine evaluates multiple choice questions automatically. The automatic evaluation of essay-type questions is discussed in the next section. The electronic testing is completed with the generation of exam protocols and statistics.

The Test Engine deals with two basic types of questions namely multiple choice questions (learners are supposed to choose the right answer among several predefined options) and essay-type questions (learners write their answers as a long-answer text). The context frames of both types consist of a specific part and a common part known as an *abstract type question* (Fig. 3.4).

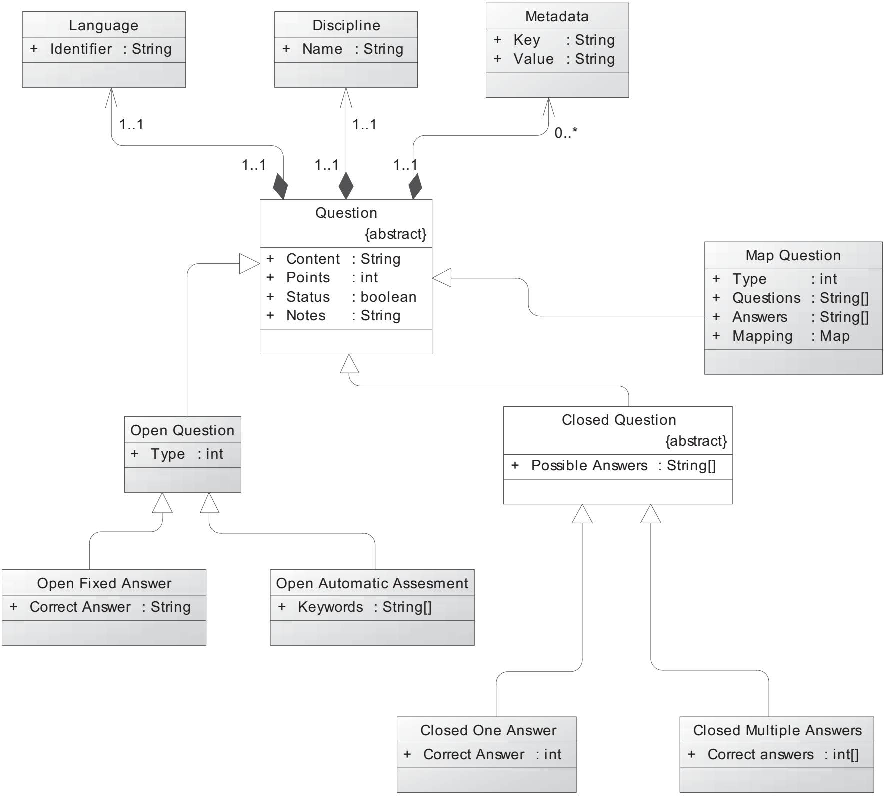


Figure 3.4 Question Class Diagram

The template is the key element which allows personalized generation of individual tests. The context frame of a template (Fig.3.5) is specified as a set of criteria applied to the question library. The template generation can be presented as the function **template** ({*criteria*}, *result_count*) □□ *question_set*, where *result_count* is the number of needed questions and *question_set* are questions selected randomly from the library that comply with the criteria.

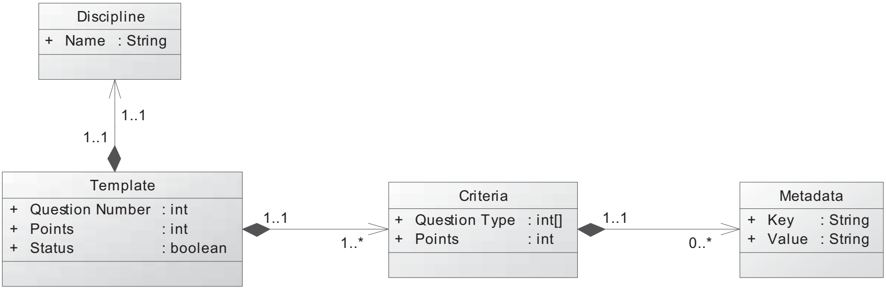


Fig.3.5 Template Structure

Furthermore, a template context frame can include parameters extracted from the users' requests such as:

- The number of questions from subject X;
- The number of questions from type T;
- The number of questions that are valid only for Y education form and are from type F.

The Test Engine generates individual instances of an intended test template (Fig. 3.6) during an exam.

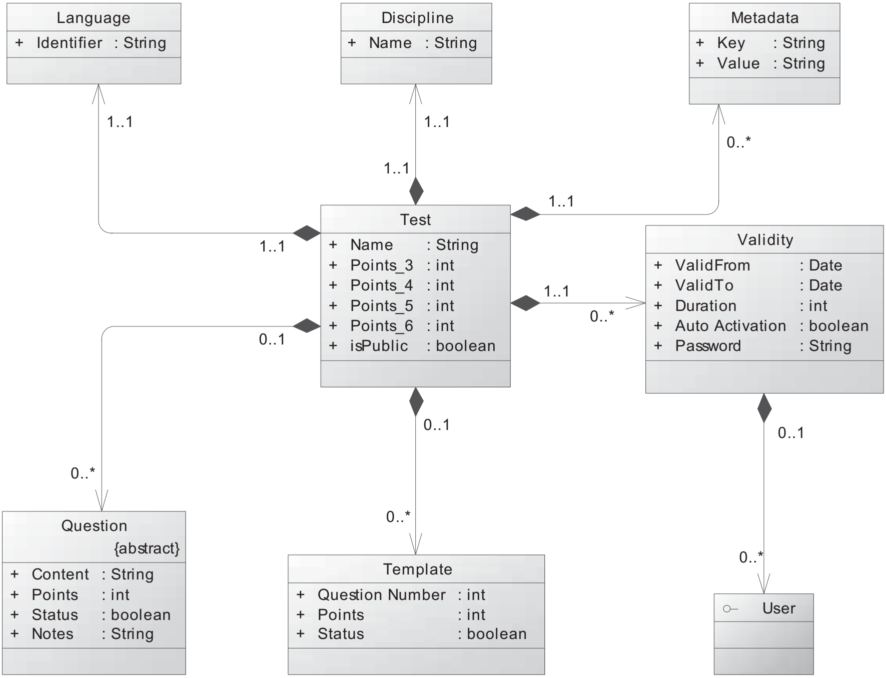


Figure 3.6 Test Instance Diagram

The Event Engine supports a model for event management, enabling the users to see and create events and also be notified for them in advance. The events in the system reflect important occasions for the users, such as a lecture, examination, test, national holiday, birthday, etc. The event context frame includes attributes such as a name, start and end date and time, details, and information if it is a recurring event, as well as rules for its recurrence. The Event Engine supports yearly, monthly, and weekly repetition. Thus, various types of timetables, schedules and individual calendars (Fig.3.7) are supported in the portal.

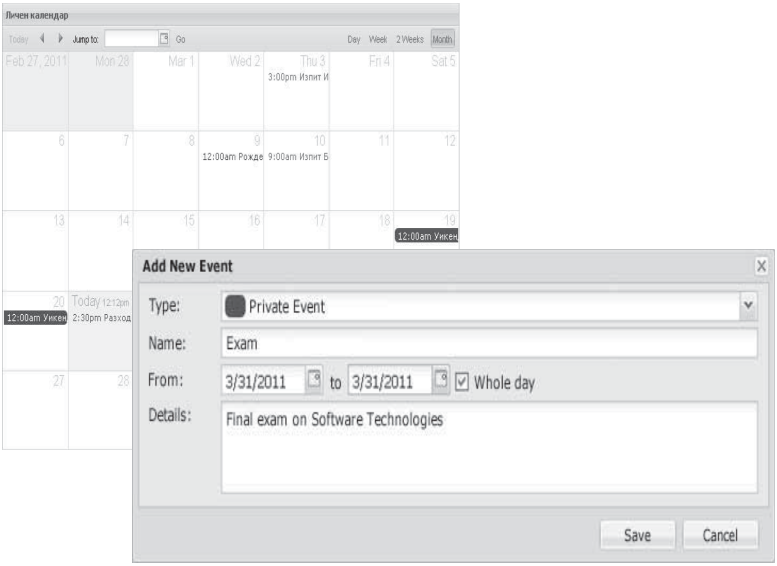


Figure 3.7 Individual Calendar

The third layer contains electronic content in the form of repositories, known as digital libraries. In the current version are supported lecture courses, a digital library, question library, test templates library, course projects library and diploma theses library. The supported portal services work directly with the digital libraries. The digital libraries content can be navigated with the help of a generalized catalogue.

3.2 Agent Village

The Agent Village (AV) is a SN in MyDeLC implemented as an agent-oriented server (Fig. 3.8), the task of which is to host assistants for support of the services located on the education portal. Interaction between this node and the portal is done by a specialized interface, the behaviour of which depends on the direction of the required assistance: from the portal to AV it is *reactive* and in the opposite direction it is *proactive*. In the reactive case the interaction between the two nodes is initiated by the portal. The reactive behaviour is necessary when the services need "expert" assistance. In this case, the services address the corresponding agents located in the AV. The problem is that in their nature the services are passive and static software modules intended mainly for the convenient implementation and integration of some business functionality. Therefore they have to "transfer" the responsibility for the activation and support of the connection to an active component of the architecture, as agents do. In this situation, the services send the needed data to the AV which reacts by interpreting this data. Depending on the identified need of assistance a suitable agent is activated by the AV. The reactive behaviour of the architecture can be implemented by using:

- Synchronous model – this model is analogous to calling subroutines in programming languages. In this model the service sends data to the AV and waits for the result from the corresponding agent before continuing its execution;
- Asynchronous model – in the asynchronous model the interaction is accomplished through a messaging mechanism.

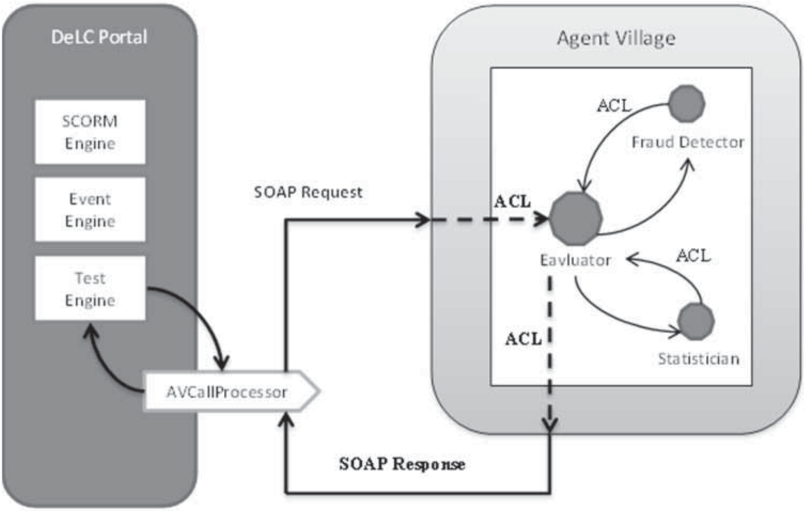


Figure 3.8 Agent Village

In the proactive behaviour (agents work "on behalf of the user"), an agent from the AV can determine that in its environment "something is happening" that would be interesting for the user assisted by that agent. The agent activates itself for performing certain actions to satisfy the preferences (wishes) of the user who will be informed by the education portal. As the portal is designed for reaction to the user's requests, the pro-activity can be managed only asynchronously and for this purpose a specialized service periodically checks a "mailbox" for incoming messages from the AV.

According to our architecture, the reactivity and the pro-activity are possible if the environment of the agents (the Agent Village node) doesn't remain *passive* (Fig. 3.8). The agents need a wrapper which behaves as a web service in order to be identified by the portal and in this way a connection between the two nodes to be made. Thus the portal can transfer the request for assistance to the AV masked as a service. In its turn, the AV transforms the request into an ACL message understandable for the agents. In a similar manner the active environment transforms ACL messages into SOAP responses, which can be processed from the portal services.

Currently, the following three assistants reside in the AV node:

- Evaluator Assistant (EA);
- FraudDetector;
- Statistician.

The Evaluator Assistant (EA) provides expert assistance to the teacher in the assessment of electronic tests. The Test Engine of the portal evaluates multiple choice questions automatically. Usually essay-type questions are assessed by the teacher and the scores are entered manually in the service to complete the final assessment of the test. In the cluster, the Test Engine calls the EA in order to evaluate essay-type questions. Therefore the received SOAP Request messages are transformed into ACL messages by the AV understandable for the EA. A transformed ACL message consists of the following parts:

- An answer of an essay-type question to be evaluated;
- Parameters for the used estimation method;
- The maximum number of points for this answer.

The EA plans the processing of the request where two evaluation methods can be applied:

- Word Matching (WM) method – it counts the "exact hits" of the keywords in the answer. The minimum threshold of percentage match (i.e. a keyword to be considered as "guessed"), which has been laid in the experiments, is between 70% and 80%. Intentionally, the method does not look for 100% match in order to give a chance to words with some minor typos to be recognized as well. To calculate the points a coefficient is formed in the following way: the number of hits is divided by the number of keywords. The actual number of points for an answer is calculated as the maximum number of points is multiplied by this coefficient;

- Optimistic Percentage (OP) method – it makes an optimistic estimation of the points for an answer. In essence, the method iterates over the keywords list and summarizes their percentage matches. Thus, the calculated amount of rates for each keyword, divided by the maximum possible match (in %), gives the reduction coefficient. The actual number of points for an answer is calculated by multiplying the maximum number of points by the coefficient of the reduction. This method is more "tolerant" to allowing spelling mistakes in the answers because low percentage matches are not ignored (unlike the first method) and are included in the formation of the final amount of points.

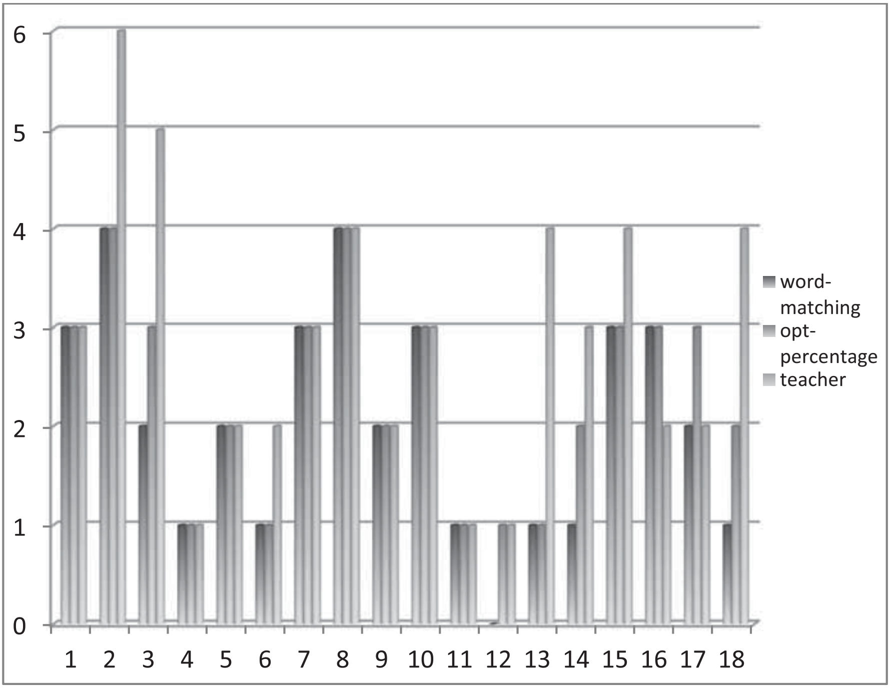


Figure 3.9 Comparison of the points given by the algorithms and the teacher (IDB)

When the calculations finish, the EA generates an answer in the form of an ACL message which is then transformed by the AV into a SOAP Response message (a result of a web service call). In the answer, there is a parameter representing the calculated amount of points that can be extracted afterwards by the Test Engine in the portal. A comparison of the scores, given by the two methods and by the teacher, is presented in Fig. 3.9. In the diagram, 18 electronic tests in the subject “Introduction to Databases” are summarized and the results show a trend that the scores given by the teacher are the “most generous”, the optimistic method (OP) is in the second place, and the most exacting method is the one for matching the words (WM).

The FraudDetector tries to recognize any attempts to cheat in an answer given by a student. Such attempts would be to guess the keywords or copy/paste results from Internet search engines. This assistant cooperates with the Evaluator agent and if its receptors detect a probability of a cheating attempt, it informs the Evaluator agent which, for its part, informs the assessing teacher that this answer requires special attention, because it is a suspicious one.

The Statistician stores information about all processed answers with a full history of the details from all calculating methods used by the Evaluator agent. This assistant needs feedback about how many points are finally given by the teacher for each answer. Thus it accumulates a knowledge base for each teacher and is able to decide which of the methods best suits the assessment style of the current assessing teacher. Upon returning the results of the Evaluator assistant, the information by this agent determines which results from each method will be submitted to the teacher as a main result, and the results of the other methods will be presented as an alternative. Another feature of this agent will be also to provide actual statistics on the performance of each of the calculating methods, as the "weakest" of them goes out of service until new and better performing methods are added to the Evaluator agent. This monitoring of the methods' behaviour becomes really significant when the so-called genetic algorithms are added, which we are still working on – as it is known, they can be "trained" and thus their effectiveness can change. In this process a knowledge base is developing for each specific subject, which supports the methods in their work.

3.3 Specy Node

This node hosts various tools to support e-learning and is not related to the services offered by the portal directly. Currently, the node is implemented as a service- and agent-oriented server where the following two tools reside:

- rLE – a specialized tool for software engineering education;
- Selbo 2 – a development tool for creating electronic teaching content.

3.3.1 Refactoring Learning Environment

rLE is an agent-oriented environment assisting e-learning in software engineering, and in software refactoring in particular. Although the refactoring process could be realized by hand, the possibility of applying automatic tools is of great importance. The rLE architecture consists of two components:

- *Front-end (FE)* – the environment, which is used by the students for editing, compilation, and testing of the source code.
- *Back-end (BE)* – an intelligent agent, known as Refactoring Agent (RA), assisting the students during the code improving.

The RA continuously analyses and assesses the source code in FE and it accesses the complete source code via the sensors. This implies not only the files being edited, but also the completed ones that have not been opened in the FE for editing. In this way the agent could make a profound analysis and give an adequate assessment of the required changes on the basis of all the code rather than the part that is currently being modified. The sensors also provide some basic metric information to the agent, which is used for initial filtering of the possible refactoring methods that can be further evaluated. The possible metrics are LOC (Line of Code) per class/method, number of methods/attributes per class, and so on. The role of the effectors is to trigger different events that assist the students during the accomplishment of their tasks in FE where they are working. Such events could be:

- Underlying particular parts of the code by highlighting them with a selected colour;
- Displaying messages in dialogue windows, balloon messages, etc.;
- Emitting sound-signals, vocal messages;
- "Incarnating" the agent in the form of an animation to exalt the effect.

The collaboration of the sensors and effectors is coordinated by the Local Control (LC) of the agent. For this purpose, the LC uses the information from the sensors and the refactoring rules stored in the Refactoring Knowledge Base (RKB) of the agent. The analysis of the source code is done by individual steps. In the first step, the RParser parses the source code and transforms it into a tree-like structure. In the next step, the RAnalyzer analyses this structure and, in accordance with the results, it takes the first filtering of the suitable refactoring methods ("the short list"). The RKB is a set of rules, each one of which describes the conditions that allow a particular refactoring method to be entered into the "short list". Depending on the refactoring method to be applied, the agent can react in three different ways:

- *Automatic Refactoring* – to apply the method automatically after receiving a confirmation from the user.
- *Refactoring Proposal* – to display detailed instructions, explaining to the user where and how the particular refactoring method should be applied.
- *Refactoring Questionnaire* – to ask the user additional questions in order to clarify the conditions and define the appropriate refactoring method.

3.3.2 Selbo 2

Selbo 2 is a development environment and an intelligent workplace for teachers which aids them in creating interactive learning material (Mitev, 2009). The Selbo 2 environment is not monolithic, but rather a set of editors interconnected into a single user interface (Fig. 3.10). The editors, in turn, can be dynamically added, removed and replaced without modifying the rest of the system (plug-in system). Editors themselves are components of the GUI. They are used for the actual editing of the various electronic resources. In general, the editors can be attributed to two groups. The first group contains the editors of the standard formats, independent of a specific domain, such as the text editor and HTML editor. The second group contains specialized editors relevant to a specific domain, for example to "Software Engineering" UML editor, editor of the source code, and others. An important specialized editor is for creating standard electronic content compatible with the SCORM 2004. SCORM electronic packages can be provided for further processing (update, modification, integration with other components present in the environment, etc.) in the "workplace" of the teacher and integrated in the education portal. During the import teachers can create annotations of the courses and determine rights of access to the educational content. Annotations are an essential part of the mechanism for finding lectures, which are appropriate for or match students' personal preferences and interests. Annotations are visible to all students with access to the relevant course. The lecturers themselves define the access rights, which may be user roles, individual users, user groups, organizations, and communities. Moreover, teachers can set rules governing the period of access to the educational material.

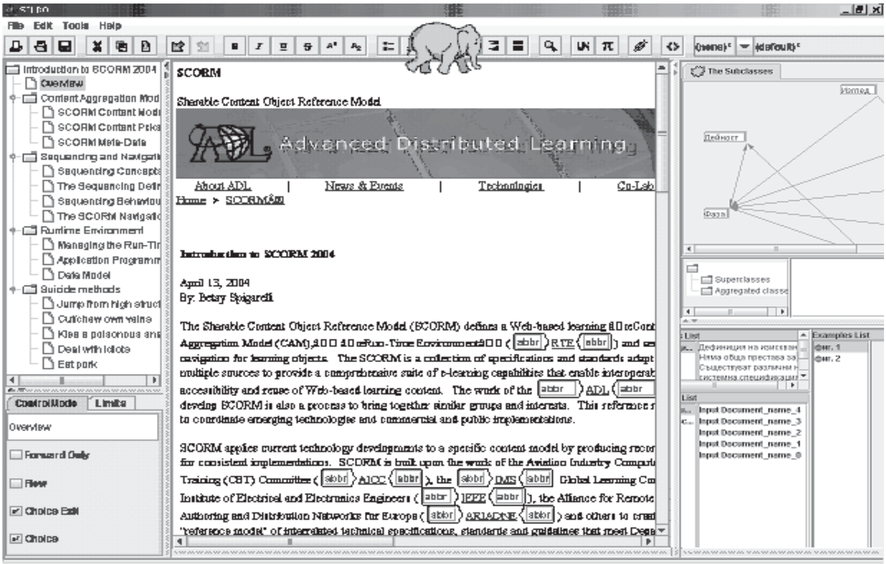


Figure 3.10 The Selbo 2 environment

Student access to the electronic content is determined according to the students' rights and roles. This service supports an intelligent search of lectures corresponding to the personal interests and preferences of students. Information about this search is derived from the users' profiles. These profiles are continually updated with each attended lecture or sitting for a test in a given discipline. Visualization of the electronic content is performed by the integrated SCORM engine.

4. InfoStation Cluster

A distinguishable feature of contemporary mobile e-learning (m-learning) systems is the anywhere-anytime-anyhow aspect of delivery of electronic content, which is personalised and customised to suit a particular mobile user (Barker, 2000; Maurer, 2001). In addition, mobile service content is expected to be delivered to users always in the best possible way through the most appropriate connection type according to the always best connected and best served communication paradigm (O'Droma, 2007; Passas,2006). In the light of these trends, the goal is to develop an intelligent mobile eLearning node which uses an InfoStation-based communication environment with distributed control (Frenkiel, 1996; Ganchev, 2007). The InfoStation paradigm is an extension of the wireless Internet, where mobile clients interact directly with Web service providers (i.e. InfoStations). By their mobile devices users request services from the nearest InfoStation utilizing Bluetooth or WiFi wireless communication.

4.1 InfoStation-Based Network

The continuing evolution in the capabilities and resources available within modern mobile devices has precipitated an evolution in the realm of e-learning. The architecture presented here attempts to harness the communicative potential of these devices in order to present learners with a more pervasive learning experience, which can be dynamically altered and tailored to suit them. The following network architecture enables mobile users to access various m-learning services via a set of intelligent wireless access points, or InfoStations, deployed in key points across the University Campus.

The InfoStation-based network consists of three tiers. The first tier encompasses the user mobile devices (cell phones, laptops, PDAs), equipped with intelligent agents acting as Personal Assistants to users. The Personal Assistant gathers information about the operating environment onboard the mobile device, as well as soliciting information about the user. Supplied with this information, the InfoStation can make better decisions on applicable services and content to deliver to the Personal Assistant. The second tier consists of InfoStations, satisfying the users' requests for services through Bluetooth and/or WiFi wireless mobile connections. The InfoStations maintain connections with mobile devices, create and manage user sessions, provide interface to global services offered by the InfoStation Centre, and host local services. The implementation of these local services is an important aspect of this system. By implementing particular services within specific localised regions throughout the University campus, we can enrich the service users' experience within these localities. The division of global and local services allows for a reduction of the workload placed on the InfoStation Centre. In the original InfoStation architecture, the InfoStations operated only as mediators between the user mobile devices and a centre, on which a variety of electronic services were deployed and executed. The InfoStations within this architecture do not only occupy the role of mediators, they also act as the primary service-providing nodes. The third tier is the InfoStation Centre which is concerned with controlling the InfoStations and overall updating and synchronisation of information across the system. The InfoStation Centre also acts as the host for global services.

In order to ensure a context-aware service provision we propose that an application is built as an integration of two components (Stoyanov, 2011):

- A standardized **middleware**, which is able to detect the dynamic changes in the environment during the processing of user requests for services *(context-awareness*) and correspondingly to ensure their efficient and non-problematic execution (*adaptability*);
- A set of **electronic services** realizing the functionality of the application area (education), which can be activated and controlled by the middleware.

4.2 Context-Aware Provision of Mobile Services

Some important attributes of interest of mobile services are the following:

- The mobile device location (*device mobility*) – in some cases this mobility leads to changing the serving InfoStation. This is especially important due to the inherent mobility within the system as users move throughout the University campus. This information has a bearing on the local services deployed within a particular area i.e. within the University Library;
- User device (*user mobility*) – this mobility offers different options for the delivery of the service request results back to the user. What is important here is to know the capabilities of the new device activated by the user so as to adapt the service content accordingly;
- Communication type – depending on the current prevailing wireless network conditions/constraints, the user may avail of different communication possibilities (e.g. Bluetooth or WiFi);

- User preferences – service personalisation may be needed so as to reflect the changes made by users in their preferences, e.g. the way the service content is visualised to them, etc.;
- Goal-driven sequencing of tasks engaged in by the user;
- Environmental context issues such as classmates and/or learner/educator interactions.

To ensure adequate support for user mobility and device mobility (the first two attributes of interest), the following four scenarios are supported by the middleware (Ganchev, 2008):

- *'No change'* – an mLearning service is provided within the range of the same InfoStation and without changing the user mobile device;
- *'Change of user mobile device'* – due to the inherent mobility, it is entirely possible that during an mLearning service session, the user may shift to another mobile device, e.g. with greater capabilities, in order to experience a much richer service environment and utilize a wider range of resources;
- *'Change of InfoStation'* – within the InfoStation paradigm, the connection between the InfoStations themselves and the user mobile devices is by definition geographically intermittent. With a number of InfoStations positioned around a University campus, the users may pass through a number of InfoStation serving areas during the service session. This transition between InfoStation areas must be completely transparent to the user, ensuring that he/she has continuous access to the service;
- *'Change of InfoStation and user mobile device'* – the most complicated scenario whereby the user may change the device simultaneously with the change of the InfoStation.

To support the third aspect of the context change (a different communication type) the development of an intelligent component (agent) is envisaged which is working within the communication layer. This component operates with the capability to define and choose the optimal mode of communication, depending on the current prevailing access network conditions (e.g. congestion level, number of active users, average data rate available to each active user, etc.). The user identification and corresponding service personalisation is subject to a middleware adaptation for use in the particular application area. In the case of eLearning, the architecture is extended to support the three fundamental eLearning models – the education domain model, the user/learner model, and the pedagogical model.

4.3 Agent-oriented middleware

The main implementation challenges within this system are related to the support of distributed control, as the system should be capable of detecting all relevant changes in the environment (context-awareness) and, according to these changes, facilitate the service offerings in the most flexible and efficient manner (adaptability). The system architecture presented in the previous section is implemented as a set of cooperating intelligent agents. An agent-oriented approach has been adopted in the development of this architecture in order to:

- Model the real distributed infrastructure adequately;
- Allow for realisation of distributed models of control;
- Ensure pro-active middleware behaviour which is quite beneficial in many situations;
- Use more efficiently the information resources spread over different InfoStations.

Moreover, the agent-oriented architecture can easily be extended with new agents (where required) that cooperate with the existing ones and communicate by means of a standardized protocol, in this case the FIPA Agent Communication Language (ACL) (FIPA, 2002). Indeed, the InfoStations and the InfoStation Centre exist as networks of interoperating agents and services, with the agents fulfilling various essential roles necessary for system management. Within each of these platforms agents take responsibility for selecting and establishing a client-server cross-platform connection, conveyance of context information and the delivery of adapted and personalised service content. This multi-agent approach differs from the classic multi-tier architectures in which the relationships between the components at a particular tier are much stronger.

Conceptually, we define different layers in the system architecture in order to present the functionality of the middleware that is being developed in a more systematic fashion. Implementation-wise, the middleware architecture is considered as a set of interacting intelligent agents. Communication between the user mobile devices and the serving InfoStations can be realized in two ways:

- An agent operating within the InfoStation discovers all new devices entering the range and subsequently initiates communication with them; or

- Personal Assistant agents on the user mobile devices are the active part in the communication and initiate the connection with the InfoStation.

In the current implementation of the prototype architecture, the former approach is used for Bluetooth communication, whereas the latter applies for WiFi communication. The system architecture of the cluster ensures continuity to the service provision, i.e. support for the continuous provision of services and user sessions in the case of scenario change or resource deficiency (Stoyanov, 2012). The agents which handle the connection and session establishment perform different actions, such as:

- Searching for and finding mobile devices within the range of an InfoStation;
- Creating a list of services required by mobile devices;
- Initiation of a wireless connection with mobile devices;
- Data transfer to- and from mobile devices.

Also, illustrated in Fig. 4.1, are the components which serve to facilitate a level of context sensitivity and personalisation to the presented services. A short description of the various agents (for Bluetooth communication) within the architecture is presented below.

The first step in the delivery of the services involves the Scanner agent, which continuously searches for mobile devices (Personal Assistants) within the service area of the InfoStation. In addition, this agent retrieves a list of services required by users (registered on their mobile devices upon installation of the client part of the application), as well as the profile information, detailing the context (i.e. device capability and user preference information). The Scanner agent receives this information in the form of an XML file, which itself is extracted from the content of an ACL message. The contents of this XML file are then passed on via the Connection Advisor agent, to the Profile Processor agent, which parses the received profile and extracts meaningful information. This information can in turn be utilized to perform the requisite alterations to services and service content.

The information is also very important in relation to the tasks undertaken by the Scenario Manager agent. The role of this agent is to monitor and respond to changes in the operating environment, within which the services are operating (i.e. change of mobile device). In the event of a significant change of service environment, this agent gathers the new capability and preference information (CPI) via the Scanner agent. Then, in conjunction with the Query Manager agent and the Content Adaptation agent, it facilitates the dynamic adaptation of the service content to meet the new service context.

The main duty of the Connection Adviser agent is to filter the list (received from the Scanner agent) of mobile devices as well as requested services. The filtration is carried out with respect to a given (usually heuristic) criterion. Information needed for the filtration is stored in a local database. The Connection Adviser agent sends the filtered list to the Connection Initiator agent, who takes on the task of initiating a connection with the Personal Assistant onboard the mobile device. This agent generates the so-called Connection Object, through which communication with the mobile device is established via Bluetooth connection. Once this connection has been established, the Connection Initiator generates an agent to which it hands over the control of the connection, called a Connection agent.

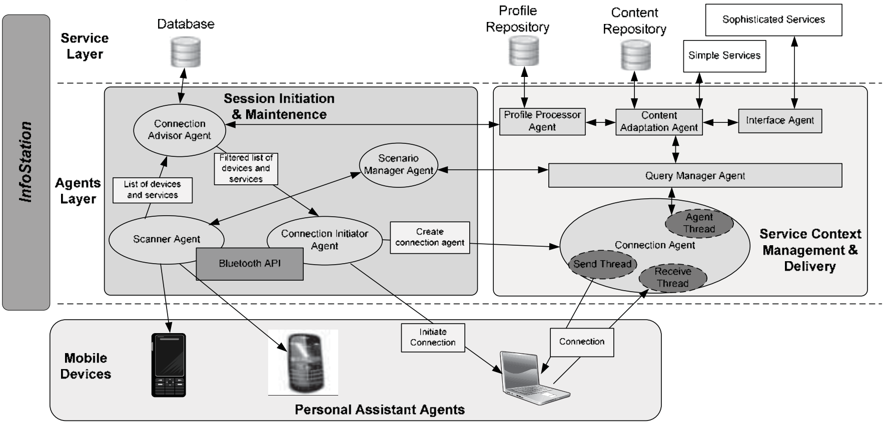


Figure 4.1 The Agent-Oriented Middleware Architecture

From this point on, all communications between the InfoStation and the Personal Assistant are directed by the Connection agent. The internal architecture of the Connection agent contains three threads: an agent thread used for communication with the Query Manager agent, and a Send thread and Receive thread, which look after each direction of the wireless communication with the mobile device.

The Query Manager performs one of the most crucial tasks within the InfoStation architecture. It determines where information, received from the mobile device, is to be directed, e.g. straight to simple services, or via Interface agents to sophisticated services. It also transforms messages coming from the Connection agent into messages of the correct protocols to be understood by the relevant services, i.e. for simple services – UDDI or SOAP, or for increasingly sophisticated services by using more complicated, semantic-oriented protocols, e.g. OWL-S (OWL-S, 2004). The Query Manager agent also interacts with the Content Adaptation agent in order to facilitate the Personal Assistant with increasingly contextualised service content. This Content Adaptation agent, operating under the remit of the Query Manager agent, essentially performs the role of an adaptation engine, which takes in the profile information provided by the Profile Processor agent and executes the requisite adaptation operations on the service content (e.g. file compression, image resizing, etc.).

The Query Manager agent receives user service requests via the Connection agent, and may communicate with various services. Once it has passed the request on to the services, all service content is passed back to the Query Manager via the Content Adaptation agent. The Profile Processor agent parses and validates received profiles (XML files) and creates a Document Object Model (DOM) tree (W3C, 2009). Using this DOM tree the XML information may be operated on, to discern the information most pertinent to the adaptation of service content. The Content Adaptation agent receives requests-responses from the services, queries the Profile Processor agent regarding the required context, and then either selects a pre-packaged service content package which closely meets the requirements of the mobile device, or applies a full transformation to the service content to meet the constraints of the operating environment of the device.

The tasks undertaken by the Content Adaptation agent, the Scenario Manager agent, and the Profile Processor agent enable the system to dynamically adapt to changing service environments, even during a particular service session. Once the connection to a particular service has been initialized and the service content adapted to the requisite format, the Connection agent facilitates the transfer of the information to the user mobile device.

5. Virtual Education Space

Virtual Education Space (VES) is a further step in building DeLC which takes into account the current trend in the development of Internet technologies. Currently, the integration of the virtual world into the physical one of the real education process is not satisfactory. E-learning systems stand more like a "patch" in the real learning process rather than as their organic continuation and extension. Recently, the Internet has been changing from a network of computers to an Internet of things, i.e. more closely integrated in the physical world. We note also the new opportunities offered by the emerging Semantic web (Berners-Lee et al., 2001).

A virtual education space may introduce new approaches and scenarios to solve complex problems in the field of e-learning. A significant trend in e-learning is that it is based on the integrated nature of the high-tech world in which people live and learn. By integrating different technologies for VES development we expect:

- Increased commitment and interest of students to teaching in ways that were impossible before;
- Creating new opportunities for learning and teaching;
- Improving and extending the interaction with local and global communities.

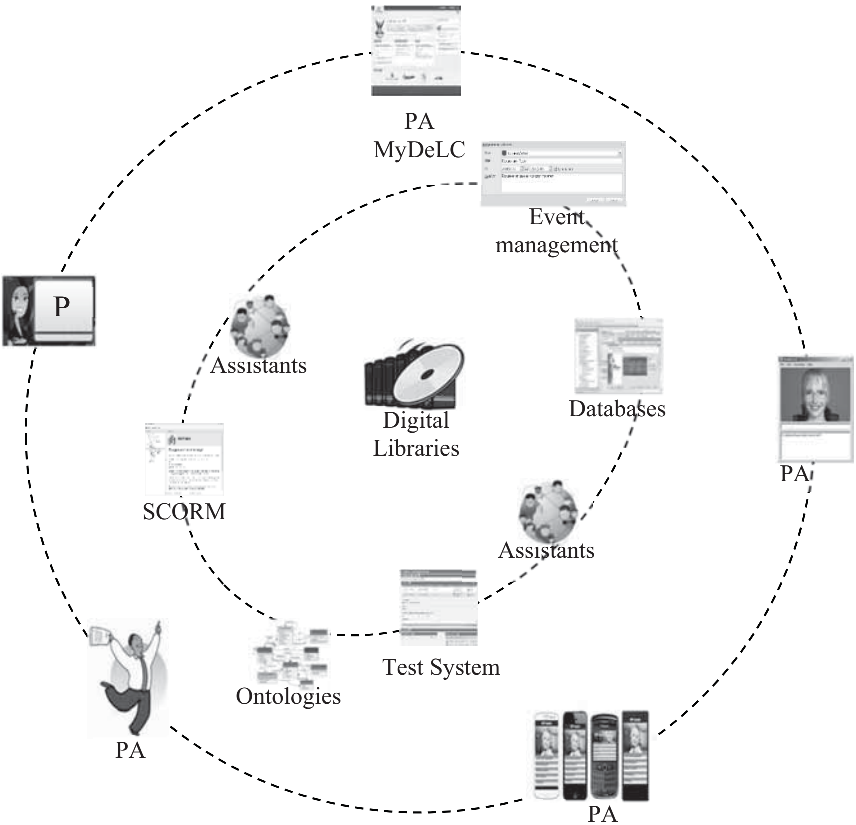


Figure 5.1 VES Infrastructure

VES (Fig. 5.1) is developing through a transformation of both university clusters. During the transformation we are going to break the "well-ordered" layered architecture of MyDeLC and replace it with an infrastructure consisting of autonomous intelligent components. At the same time we intend to keep the components existing in DeLC and in addition to enhance their interactive, reactive and proactive behaviour. For this purpose there are intelligent agents under construction known as *assistants* to cater for all passive components in the space, such as digital libraries, databases, ontologies, and services.

Unlike the web satisfying AAA slogan (Anyone can say Anything about Any topic), VES is a controlled space specialized for use in e-learning. The internal infrastructure of the space is transparent to users. Access to the services and content can be achieved only by pre-defined "entry points". An entry point will be a “thin” version of the DeLC portal where services, machines, and digital libraries are exported as separate components in the space. Other entry points will be Personal Assistants (PAs) located on different consumer devices such as mobile devices, tablets, and laptops. In the execution of user requests a PA can interact with one or more appropriate assistant(s).

6. Calculus of Context-Aware Ambients

Context-awareness requires applications to be able to adapt themselves to the environment in which they are being used such as user, location, nearby people and devices, and users' social situations. In this section we use small examples to illustrate the ability of CCA to model DeLC and VES that are context-aware.

6.1. Syntax of Processes and Capabilities

This section introduces the syntax of the language of CCA. Like in the π-calculus (Milner, 1999; Sangiorgi, 2001), the simplest entities of the calculus are *names*. These are used to name for example ambients, locations, resources, and sensors data. We assume a countably-infinite set of names, elements of which are written in lower-case letters, e.g. n, x, and y. We let $\tilde{y}$ denote a list of names and $|\tilde{y}|$ the arity of such a list. We sometimes use $\tilde{y}$ as a set of names where it is appropriate. We distinguish three main syntactic categories: processes P, capabilities M, and context expressions κ.

The syntax of processes and capabilities is given in Table 6.1, where P, Q and R stand for processes, and M for capabilities. The first five process primitives (inactivity, parallel composition, name restriction, ambient and replication) are inherited from MA (Cardelli, 2000). The process 0 does nothing and terminates immediately. The process $P \mid Q$ denotes the process P and the process Q running in parallel. The process $(\upsilon n)\ P$ states that the scope of the name n is limited to the process P. The replication $!P$ denotes a process which can always create a new copy of P. Replication was first introduced by Milner in the π-calculus (Milner, 1999). The process $n[P]$ denotes an ambient named n whose behaviours are described by the process *P*. The pair of square brackets `[' and `]' outlines the boundary of that ambient. This is the textual representation of an ambient. The graphical representation of that ambient is:

The graphical representation highlights the nested structure of ambients.
CCA departs from MA and other processes calculi such as (Zimmer, 2005; Bucur, 2008; Bugliesi, 2004) with the notion of *context-guarded capabilities*, whereby a capability is guarded by a context-expression which specifies the condition that must be met by the environment of the executing process. A process prefixed with a context-guarded capability is called a *context-guarded prefix* and it has the form κ? M.P. Such a process waits until the environment satisfies the context expression κ, then performs the capability *M* and continues like the process *P*. The process learns about its context (i.e. its environment) by evaluating the guard. The use of context-guarded capabilities is one of the two main mechanisms for context acquisition in CCA (the second mechanism for context acquisition is the call to a process abstraction as discussed below). The syntax and the semantics of context expressions are given below. We let *M.P* denote the process True?*M.P*, where True is a context expression satisfied by all contexts.

Table 6.1 Syntax of CCA processes and capabilities

P, Q, R::=	**Process**
0	inactivity
P|Q	parallel composition
(υn) P	name restriction
n[P]	ambient
! P	replication
κ! M. P	context-guarder action
x ⊳ ($\tilde{y}$).P	process abstraction
α ::=	**Locations**
↑	any parent
n ↑	parent n
↓	any child
n ↓	child n
::	any sibling
n ::	sibling n
ϵ	locally

M::=	**Capabilities**
del n	delete n
in n	move in n
out	move out
$\alpha\, x\langle\tilde{z}\rangle$	process call
$\alpha\, (\tilde{y})$	input
$\alpha\, \langle\tilde{y}\,\rangle$	output

A process abstraction $x \rhd (\tilde{y}).\mathbf{P}$ denotes the linking of the name *x* to the process *P* where $\tilde{y}$ is a list of *formal parameters*. This linking is local to the ambient where the process abstraction is defined. So a name *x* can be linked to a process *P* in one ambient and to a different process *Q* in another ambient. A call to a process abstraction named *x* is done by a capability of the form $\alpha\, x\langle\tilde{z}\rangle$ where †specifies the location where the process abstraction is defined and $\tilde{z}$ is the list of *actual parameters*. There must be as many actual parameters as there are formal parameters to the process abstraction being called. The location α can be $'\uparrow'$ for any parent, $'n\uparrow'$ for a specific parent *n*, $'\downarrow'$ for any child, $'n\downarrow'$ for a specific child *n*, $'::'$ for any sibling, 'n ::' for a specific sibling *n*, or ϵ (empty string) for the calling ambient itself. A process call $\alpha\, x\langle\tilde{z}\rangle$ behaves like the process linked to *x* at location α, in which each actual parameter in $\tilde{z}$ is substituted for each occurrence of the corresponding formal parameter. A process call can only take place if the corresponding process abstraction is available at the specified location.

In CCA, an ambient provides context by (re)defining process abstractions to account for its specific functionality. Ambients can interact with each other by making process calls. Because ambients are mobile, the same process call, e.g. $\uparrow x\langle\tilde{z}\rangle$, may lead to different behaviours depending on the location of the calling ambient. So process abstraction is used as a mechanism for context provision while the process call is a mechanism for context acquisition.

Ambients exchange messages using the capability $\alpha\,\langle\tilde{z}\rangle$ to send a list of names $\tilde{z}$ to a location α, and the capability $\alpha\,(\tilde{y})$ to receive a list of names from a location α. Similarly to a process call, an ambient can send a message to any parent, i.e. $\uparrow\langle\tilde{z}\rangle$, a specific parent n, i.e. $n\uparrow\langle\tilde{z}\rangle$; any child, i.e. $\downarrow\langle\tilde{z}\rangle$; a specific child n, i.e. $n\downarrow\langle\tilde{z}\rangle$; any sibling, i.e. $::\langle\tilde{z}\rangle$; a specific sibling n, i.e. $n::\langle\tilde{z}\rangle$; or itself, i.e. $\langle\tilde{z}\rangle$.

An *input prefix* is a process of the form $\alpha(\tilde{y})$.P, where $\tilde{y}$ is a list of variable symbols and P is a continuation process. It receives a list of names $\tilde{z}$ from the location α and continues like the process $P\{\tilde{y} \leftarrow \tilde{z}\}$, where $P\{\tilde{y} \leftarrow \tilde{z}\}$ is the process P in which each name in the list $\tilde{z}$ is substituted for each occurrence of the corresponding variable symbol in the list $\tilde{y}$.

The mobility capabilities in and out are defined as in MA (Cardelli, 2000) with the exception that the capability out has no explicit parameter in CCA, the implicit parameter being the current parent (if any) of the ambient performing the action. An ambient that performs the capability in n moves into the sibling ambient n. The capability out moves the ambient that performs it out of that ambient parent. Obviously, a root ambient, i.e. an ambient with no parents, cannot perform the capability out. The capability del n deletes an ambient of the form n[0] situated at the same level as that capability, i.e. the process del n. P|n[0] reduces to P. The capability del acts as a garbage collector that deletes ambients which have completed their computations. It is a constrained version of the capability open used in MA to unleash the content of an ambient. As mentioned in (Bugliesi, 2004), the open capability brings about serious security concerns in distributed applications, e.g. it might open an ambient that contains a malicious code. Unlike the capability open, the capability del is secure because it only opens ambients that are empty, so there is no risk of opening a virus or a malicious ambient.

6.2. Context Model

In CCA the notion of an ambient, inherited from MA, is the basic structure used to model entities of a context-aware system such as: a user, a location, a computing device, a software agent, or a sensor. As described in Table 1, an ambient has a name, a boundary, a collection of local processes and can contain other ambients. Meanwhile, an ambient can move from one location to another by performing the mobility capabilities in and out. So the structure of a CCA process, at any time, is a hierarchy of nested ambients. This hierarchical structure changes as the process executes. In such a structure, the context of a sub-process is obtained by replacing that sub-process in the structure by a placeholder ′⊙′. For example, suppose a system is modelled by the process P|n[Q|m[R|S]]. So, the context of the process ◌̨ in that system is P|n[Q|m[⊙|S]], and that of ambient m is P|n[Q|⊙]. Following are examples of contexts in the smart phone system described in Sect. 5.3. The following context is the context of the smart phone carried by Bob when Bob is inside the conference room with Alice:

$$e_1 \triangleq conf\,[\,P\,|bob\,[\odot]\,|\,alice\,[Q]\,],$$

where P models the remaining part of the internal context of the conference room and Q the internal context of the ambient alice. We assume that there is only one ambient named *alice* in the conference room.

If Bob is inside the conference room while Alice is outside that room, the context of the smart phone carried by Bob can be described as follows:

$$e_2 \triangleq alice\,[\,Q\,]\,|\,conf\,[\,P\,|\,bob\,[\odot]\,\,],.$$

Bob might carry with him another device, a PDA say, while inside the conference room. In this case the context of the smart phone can be modelled as:

$$e_3 \triangleq conf\,[\,P'\,|\,bob\,[\odot]\,|\,pda[\,R\,]\,],$$

where P′ models the remaining part of the internal context of the conference room, pda is the name of the ambient modelling the PDA device and R specifies the functionality of the PDA.

Our context model is depicted by the grammar in Table 6.2, where the symbol *E* stands for context (environment), n ranges over names and P ranges over processes (as defined in Table 1). The context 0 is the empty context, also called the *nil* context. It contains no context information. The position of a process in that process' context is denoted by the symbol ⊙. This is a special context called the *hole context*. The context (υn) E means that the scope of the name n is limited to the context vE. The context n[E] means that the internal environment of the ambient n is described by the context E. The context E|P says that the process P runs in parallel with the context E, and so E is part of the process P's context.

Ground context. A ground context is a context containing no holes.
Note that a context contains zero or one hole; and that a ground context is a process. We do not allow multi-hole contexts because they are not suitable to our purpose.

Table 6.2 Syntax contexts

E ::=	Context
0	nil
$\odot$	hole
n[E]	location
(υn) E	restriction

$$E_1(E_2) = \begin{cases} E_1 & \textit{if } E_1 \textit{ is a ground context} \\ E_1\{\odot \leftarrow E_2\} & \textit{otherwise.} \end{cases}$$

Context evaluation. Let E_1 and E_2 be contexts. The evaluation of the context E_1 at the context E_2, denoted by E_1 (E_2), is the context obtained by replacing the hole in E_1 (if any) by E_2, viz

where $E_1\{\odot \leftarrow E_2\}$ is the substitution of E_2 for $\odot$ in E_1.

The hole $\odot$ plays an important role in our context model. In fact, a context E containing a single hole represents the environment of a process P in the process E(P). A process modelling Bob using a smart phone in the conference room with Alice can be specified as:

$$e_1(\text{ phone [S]}) \triangleq \text{conf [P | bob [phone [S]] | alice [Q]]},$$

where e_1 is the context specified in Example 5.2 and S is the specification of the smart phone. A process modelling Bob using a PDA in the conference room can be specified as:

$$e_3(0) \triangleq \text{conf [P' | bob [pda [R]]]},$$

where e_3 is the context specified in Example 5.2. The syntax of CEs is given in Table 6.3 where κ ranges over CEs, n ranges over names and x is a variable symbol which also ranges over names.

Table 6.3 Syntax context expressions

κ ::=		Context Expressions
	True	true
	n=m	name match
	•	hole
	¬κ	negation
	k1 | k2	parallel composition

k1 ∧k2	conjunction
n[κ]	location
new(n, k)	relevation
□κ	spatial next modality
✧ κ	somewhere modality
∃x. κ	existential quantification

6.3. Modelling of InfoStation-Cluster

As we have mentioned earlier that e-learning is becoming an authentic possible alternative educational approach as the technologies regarding that area are developing so fast, and there is a recognisable growth of a great variety of wide-band telecommunication delivery technologies. This section introduces at a glance each of the mobile services provided by the InfoStation cluster:

- *AAA*: in order for any user to use any mService in the system, the user device should be registered. The AAA service (Authentication, Authorisation, and Accounting) allows the users to register their devices with the system to gain the ability of using the mLearning services offered by the system.
- *mLecture*: this service allows the users to gain access to the lecture material through their mobile devices. The users can request a specific lecture, which is adapted according to the capabilities of the user devices and then delivered to their mobile devices.
- *mTest*: this service is crucial to the learning process. The mTest service allows the users to gain access to test materials that provide means of an evaluating process. A user can request, like the mLecture service, a specific test, which is also adapted to the capabilities of the user device then delivered to the user mobile device. The mTest service may only run individually on a user device and unaccompanied with any other service whatsoever.
- *mTutorial*: this service allows the users to gain access to a self-assessment test. It is a combination between the mLecture and the mTest services. A user can request a self-assessment test in a similar way as requesting an mLecture. After the user submits their answers, he/she receives a feedback on his/her performance and the correct answers to the questions he/she got wrong.
- *Intelligent Message Notification (IMN in short)*: this service allows the users to communicate with each other by exchanging messages via their mobile devices.

- *VoIP*: this service allows the users to communicate with each other via phone calls throughout the infostation-based mLearning system.

The InfoStationCentre (ISC) provides the User Authentication, Authorisation and Accounting (AAA) service which identifies each mobile user and provides him/her with a list of services the user is authorised to access. This service is regulated by the following policies:

- When a user is within the range of an IS, the intelligent agent (PA) of the user's device and the IS mutually discover each other. The PA sends a request to the IS for user Authorisation, Authentication and Accounting (AAA). This request also includes a description of the mobile device currently being used and any updates of the user profile and the user service profile.
- The IS forwards this AAA request to the ISC along with the profile updates. If the user is successfully authenticated and authorised to utilise the services by the AAA module within the ISC, a new account record is created for the user and a positive acknowledgement is sent back to the IS. Then the IS compiles a list of applicable services and sends this to the PA along with the acknowledgement. The PA displays the information regarding these services to the user who then makes a request for the service he/she wishes to use.

If the user chooses the mLecture service, then the following policies of the mLecture service apply:

- The PA forwards the mLecture service request to the InfoStation, which instantiates the service. If the IS is unable to satisfy fully the user service request it is forwarded to the ISC, which is better equipped to deal with it. In either case, the lecture is adapted and customised to suit the capabilities of the user devices and the user own preferences, and then delivered to their mobile devices.
- During the execution of the service, the user is free to move into a different infostation, to switch between devices or to do both.
- A user cannot use the mLecture and mTest services simultaneously. The mTest service should operate unaccompanied at all occasions.

This section presents the formalisation of the policies of the InfoStation cluster. We first introduce some naming conventions which are used in the specification of the system. Then we give the specification of two mobile services. The following naming conventions are used to differentiate between variables' names and constants. A variable name begins with a lowercase letter while a constant begins with a number or an uppercase letter. The list of the constant names that are used in the formalisation process is given in Table 6.4. and the list of variable names is given in Table 6.5. The system consists mainly of one central ISC, multiple ISs and multiple user devices. Each component of the system is modelled as an ambient. That is, the ISC, each IS and each user device is modelled as an ambient. In particular, a device, *PC* say, being used by a user, *303* say, is modelled by an ambient named *PC303*. The ISC ambient runs in parallel with the IS ambients, and all the user devices within the range of an IS are child ambients of that IS ambient.

Table 6.4 Constants

Notations	Descriptions
ISC	the InfoStation Centre
IS_i	the i-th InfoStation
$Phone$	a phone
PDA	a PDA
PC	a PC
$SLIST$	list of mServices
ACK	an acknowledgement
NULL	empty message
DENIED	request denied

Table 6.5 Variables

uid 301, 302, 303	a user's ID
dtype Phone, PDA, PC	a user's device type
aname Phone301, PC303	an ambient's name
lect Lect001	a lecture's ID
reply OK, DENIED, content	a reply to a request
content CONTENT	lecture's content
slist SLIST	list of services
ack ACK	acknowledgement

This is textually represented by the following process (Eq.1):

$$\begin{aligned} ISC[P_{ISC}] \;\;|\;\; & IS1[PDA303[P_{PDA303}] \mid PC401[P_{PC401}] \\ & \mid P_{IS1}] \\ | \;\; & IS2[PDA301[P_{PDA301}] \\ & \mid Phone402[P_{Phone402}] \mid P_{IS2}] \\ | \;\; & IS3[Phone300[P_{Phone300}] \mid P_{IS3}] \\ & \mid IS4[Phone403[P_{Phone403}] \mid PC302[P_{PC302}] \\ & \mid P_{IS4}] \end{aligned}$$

where each P_x is a process modelling the behaviour of the corresponding ambient x.

Now we give the formal specification of the ISC and the ISs below. An abstract model of an infostation IS_i (for some integer i) has the following main components: the AAA request ambient $AAAreq_i$, the lecture ambient $Lectreq_i$, and the cache ambient $Cache_i$. The InfoStation is a parent to the inside ambients which are siblings to each other. The specification of each of these ambients is as follows: $AAAreq_i$ ambient is responsible for handling AAA requests sent by user devices willing to register with the InfoStation IS_i. The $AAAreq_i$ ambient receives a request from a device and, immediately, forwards it to the InfoStation, then receives a reply from the InfoStation and again, forwards it to the user's device. This behaviour is modelled by the following process:

$$PA_i \triangleq\; ! :: (uid, dtype, aname).IS_i \uparrow \langle uid, dtype, aname \rangle..IS_i \uparrow (ack, aname, slist).aname :: \langle aname, slist \rangle.\ 0$$

where *uid* is the user ID, *dtype* is the device type and *aname* is the name of the ambient sending the request.

The InfoStation accordingly receives a request from the $AAAreq_i$ ambient, forwards it to the InfoStation Centre, and after receiving the reply from the InfoStation Centre it forwards it to the $AAAreq_i$ ambient. This behaviour is modelled as (Eq.2):

$$\left(\begin{array}{l} !AAAreq_i\downarrow(uid, dtype, aname).ISC :: \langle AAAreq, uid, \\ dtype, aname, IS_i\rangle.\mathbf{0} \mid \\ !ISC :: (ack, aname, slist).(\mathsf{has}(aname))? \\ AAAreqi\downarrow\langle ack, aname, slist\rangle.\mathbf{0} \end{array} \right)$$

Lectreq_i ambient handles all the mobile education service requests sent by the user devices. It receives a lecture request from a user device and forwards it to the infostation IS_i i.e. (Eq.3)

$$! :: (\ lectid, uid, dtype, aname\ ).IS_i\uparrow \langle\ lectid, uid, dtype, aname\ \rangle.0$$

Then it gets the reply from that infostation and forwards it to the user device which initiated the request, i.e. (Eq.4)

$$! :: (\ lectid, reply, aname\ ).aname :: \langle\ lectid, reply\ \rangle.0$$

So the whole behaviour of the Lectreq_i ambient is (Eq.5)

$$PL_i \qquad \triangleq \text{Eq.(3)} \mid \text{Eq.(4)}$$

We show how the InfoStation handles a request from the Lectreq_i ambient after we have specified the Cache_i ambient. This is the ambient where the InfoStation stores copies of requested lectures for future rapid access. It models a cache memory. A lecture is stored as an ambient (named after that lecture's id) which contains three persistent memory cells, each containing a version of the lecture suitable to a specific type of device (phone, PDA or PC). When an InfoStation receives an mLecture service request from a device, it checks for the requested material in its cache first rather than getting it from the InfoStation Centre directly. The process of checking the availability of a lecture inside the cache is done by sending a request to the Cache_i ambient which then checks whether it has the ambient of the requested lecture or not. If the requested lecture is available the cache ambient retrieves it and sends it back to the InfoStation, otherwise, it replies immediately to the InfoStation that this lecture does not exist. The behaviour of the Cache_i ambient is modelled by the following process (Eq.6):

$$P_{c_i} \ \widehat{=}\ !\uparrow(lectid, uid, dtype, aname). \left(\begin{array}{l} \mathtt{has}(lectid)?lectid\downarrow\langle dtype, aname\rangle. \\ lectid\downarrow(reply, aname). \\ \uparrow\langle lectid, uid, dtype, reply, aname\rangle.\mathbf{0} \mid \\ \neg\mathtt{has}(lectid)?\uparrow\langle lectid, uid, dtype, \\ NULL, aname\rangle.\uparrow(lectid, content, dtype). \\ lectid\downarrow\langle content, dtype\rangle.lectid\downarrow(ack).\uparrow\langle ack\rangle.\mathbf{0} \end{array}\right)$$

The behaviour of each lecture ambient (named after the lecture's id *lectid*) in the cache is modelled by the following process (Eq.7):

$$P_{lectid} \mathrel{\widehat{=}} !\uparrow(dtype, aname).\left(\begin{array}{l} \mathtt{has}(dtype)?dtype\downarrow\langle\rangle.\ dtype\downarrow(reply).\\ \uparrow\langle reply, aname\rangle.\ \mathbf{0}\ |\\ \neg\mathtt{has}(dtype)?\uparrow\langle NULL, aname\rangle.\\ \uparrow(content, dtype).dtype\downarrow\langle content\rangle.dtype\downarrow().\\ \uparrow\langle ACK\rangle.\mathbf{0} \end{array}\right)$$

The InfoStation will act as follows. First, it receives a request from $Lectreq_i$, then it checks the availability of the lecture in its cache by sending a request to the $Cache_i$ ambient, i.e. (Eq.8)

$$!Lectreq_i\downarrow(lectid, uid, dtype, aname).Cache_i\downarrow\langle lectid, uid, dtype, aname\rangle.0$$

If the cache replies with the content of the lecture, it will send a request to the InfoStation Centre with a flag set to 1 (meaning that the requested lecture exists in its cache) asking whether the user is currently taking an mTest. If the user is taking an mTest, then the mLecture service request must be denied. If the cache did reply with NULL as a lecture's content, then the infostation will send a request to the InfoStation Centre with the flag set to 0 (meaning that the lecture does not exist in its cache) asking for both the requested lecture and to check whether the user is taking a Test. This behaviour of the IS is modelled as (Eq.9)

where C and N are defined as follows:

$$!Cache_i\downarrow(lectid, uid, dtype, creply, aname).\left(\begin{array}{l} \neg(creply = NULL)?\ ISC :: \langle lectid, uid, dtype,\\ aname, 1\rangle.\mathbf{C}\ |\\ (creply = NULL)?\ ISC :: \langle lectid, uid, dtype,\\ aname, 0\rangle.\mathbf{N} \end{array}\right)$$

and

$$\mathbf{C} \ \widehat{=}\ ISC :: (lectid, reply, aname).\left(\begin{array}{l}(reply = OK \wedge \mathtt{has}(aname))?\\ Lectreq_i\!\downarrow\!\langle lectid, creply, aname\rangle.\mathbf{0} \mid \\ (reply = OK \wedge \neg\mathtt{has}(aname))?\\ ISC :: \langle lectid, uid, dtype, aname, 0\rangle.\mathbf{0} \mid \\ (reply = DENIED \wedge \mathtt{has}(aname))?\\ Lectreq_i\!\downarrow\!\langle lectid, reply, aname\rangle.\mathbf{0}\end{array}\right)$$

$$\mathbf{N} \ \widehat{=}\ ISC :: (lectid, aname, reply).\left(\begin{array}{l}(\neg(reply = DENIED) \wedge \mathtt{has}(aname))?\\ Cache_i\!\downarrow\!\langle lectid, reply, dtype\rangle.Lectreq_i\!\downarrow\!\langle lectid,\\ reply, aname\rangle.\ Cache_i\!\downarrow\!(ack).\mathbf{0} \mid \\ (\neg(reply = DENIED) \wedge \neg\mathtt{has}(aname))?\\ Cache_i\!\downarrow\!\langle lectid, reply, dtype\rangle.ISC :: \langle lectid, uid,\\ dtype, aname, 0\rangle.\ Cache_i\!\downarrow\!(ack).\mathbf{0} \mid \\ (reply = DENIED \wedge \mathtt{has}(aname))?\\ Lectreq_i\!\downarrow\!\langle lectid, reply, aname\rangle.\mathbf{0}\end{array}\right)$$

Thus, the whole behaviour of an infostation IS_i is modelled as (Eq.10):

$$PIS_i \triangleq \text{Eq.(2)} \mid \text{Eq.(8)} \mid \text{Eq.(9)}$$

A model of the ISC encompasses ambients modelling users' accounts and named after the users' IDs; an ambient named *Lectures* that contains all the lecture ambients, each named after the corresponding lecture ID. Each lecture ambient contains three persistent memory cells named *Phone*, *PDA,* and *PC*, each storing the lecture's version suitable for the corresponding type of device. The mTest service is not represented as we are only dealing with the mLecture service in this chapter. These components of the ISC are formalised below.

Users' accounts

An ambient modelling a user's account contains two ambients named *Loc* and *Utest*. Each of these two ambients models a persistent memory cell which stores, at any time, the current location of that user (for the former) or a Boolean indicating whether that user is taking an mTest or not (for the latter). We understand by location of a user the IS the user is registered with. The behaviour of the *Loc* ambient and the *Utest* ambient are specified exactly with appropriate initial values. The ISC requests the value of any of these cells by sending the name *Loc* or *Utest* to the user's account ambient (see Eq. (14)) which then can *get* (i.e. read) the value of the corresponding child ambient as follows, where the parameter *x* is the corresponding child ambient name (Eq. 11):

$$!\,ISC \uparrow (x).x \downarrow \langle\rangle.\, x \downarrow (\,y).\, ISC \uparrow \langle y\rangle.\, 0$$

The user's account ambient can also put (i.e. write) a value in any of its child ambients as follows, where the parameter x is the corresponding child ambient name and the parameter n is that value (Eq.12):

$$!\,ISC \uparrow (x, n).\, x \downarrow \langle n\rangle.\, x \downarrow (\,).\, ISC \uparrow \langle x, ACK\rangle.\, 0$$

So the whole behaviour of a user's account ambient named uid is specified as:

$$P_{uid} \triangleq Eq.\ (11)\ |Eq.\ (12)$$

Lectures

As mentioned above, this ambient contains all the lectures that are available in the mLecture service. Each lecture has a unique ID and the corresponding ambient is named after that ID. The behaviour of a lecture ambient is specified in Eq.(7). The *Lectures* ambient behaves exactly as the cache ambient specified in Eq.(6).

Infostation centre

We now formalise the behaviour of the ISC when it receives a request from an IS. We are interested in two types of request in this chapter: an AAA request and a lecture request.
When the ISC receives an AAA request from an IS, it updates the user's account with its new location and then replies to that IS with an acknowledgement along with a list of available services. For the sake of simplicity, the service list is represented by the name *'SLIST'*. This is modelled by the following process (Eq. 13):

$$!\,IS_i{::}\,(\,aaareq, uid, dtype, aname\,).uid \downarrow \langle IS_i\,\rangle.uid \downarrow (\,ack\,).$$

$$IS_i{::}\,\langle\,ack, aname, SLIST\rangle.\, 0$$

After receiving a lecture request, the ISC checks whether the user requesting the service is currently taking an mTest. This is done by it sending a message to the corresponding user's account ambient. That user's account ambient replies with 0 for ‘No’ and 1 for ‘Yes’. If the reply is ‘Yes’ then the ISC fetches the current location of the user and forwards a ‘DENIED’ message to that location. If the reply is ‘No’ and the flag is set to 1, an ‘OK’ message is forwarded to the current user location; otherwise (i.e. reply is ‘No’ and the flag is set to 0), the ISC fetches the appropriate lecture version for the user device and sends it to the current user location. This behaviour is represented by the following process (Eq.14):

$$!\ IS_i :: (\, lectid, uid, dtype, aname, flag\,).\ uid \downarrow \langle\, Utest \rangle.\ uid \downarrow (\, y).\ (PISC_1 \mid PISC_2)$$

where

$$PISC_1 \triangleq (y = 1)?\ Uid \downarrow \langle Loc \rangle.uid \downarrow (z).z :: \langle lectid, DENIED, aname \rangle.\ 0$$

and

$$P_{ISC2} \,\widehat{=}\, \left(\begin{array}{l} (y = 0 \wedge flag = 1)?uid{\downarrow}\langle Loc \rangle.uid{\downarrow}(z). \\ z :: \langle lectid, OK, aname \rangle.\mathbf{0} \mid \\ (y = 0 \wedge flag = 0)?Lectures{\downarrow}\langle lectid, uid, \\ dtype, aname \rangle.Lectures{\downarrow}(lectid, uid, \\ dtype, reply, aname).uid{\downarrow}\langle Loc \rangle.uid{\downarrow}(z). \\ z :: \langle lectid, reply, aname \rangle.\mathbf{0} \end{array} \right)$$

So the whole behaviour of the ISC is modelled by the following process:

$$P_{ISC} \triangleq \text{Eq.}(13) \mid \text{Eq.}(14)$$

Now that a formal model of the infostation-based mLearning system has been presented, we show how this model can be used to validate the properties of the mLearning system. Lamport proposed two main classes of system properties: *safety properties*, which state that ‘nothing bad will happen’; and *liveness properties*, which assert that ‘something good will happen, eventually’. In the light of this classification, we wish to establish the liveness property that every lecture request from a user is eventually replied to by the system, provided that the user does not become infinitely unavailable after that request has been made.

Theorem 7.1 *Every user's lecture request will eventually get a reply, provided that the user stays long enough in the system.*

The proof of this theorem is based on the reduction semantics of a given by a congruence relation ‘≡’ and a reduction relation ‘→’. In this proof we will assume, without loss of generality, that the user is using a laptop

(PC) to access the system from an infostation IS_i. Given that the user is mobile, the following cases must be considered:

- the user sends the request and waits for the reply in the same infostation IS_i (i.e. the user may move around within the range of the infostation);
- the user sends the request and moves into a different infostation IS_j , $j \neq i$.

In Case 1, the user behaviours can be modelled by the following ambient (Eq.15):

$$PC303[\ Lectreq_i :: \langle Lect001, 303, PC\ , PC303\rangle.\ Lectrrq_i :: (lectid, reply).\ 0]$$

This ambient sends a lecture request to the $Lectreq_i$ ambient and waits for a reply from that ambient, and then terminates. The lecture request contains the following information: (i) the lecture ID, *Lect001*; (ii) the user ID, *303*; (iii) the device type, *PC*; and (iv) the name of the ambient to reply to, *PC303*.

The behaviour of the $Lectreq_i$ ambient is modelled by the process PL_i in Eq.(5), which basically receives a lecture request from a sibling ambient (e.g. a user device), forwards the request to the infostation IS_i, gets the reply from that infostation and forwards it to the very ambient which initiated the request. How the infostation IS_i interacts with the $Lectreq_i$ ambient is specified in Eq.(8). These interactions between a user device, the $Lectreq_i$ ambient and the infostation IS_i can be expressed as a sequence of derivations using the reduction relation '→'. Because of the space limit, we cannot give the full sequence of derivations in this chapter. For illustration, the following sequence of derivations describes how a lecture request sent by the user gets to the infostation to be processed:

$$
\begin{array}{ll}
 & \text{Eq. (8)} \mid \text{Eq. (15)} \mid \mathit{Lectreq}_i[P_{L_i}] \\
\rightarrow & \{\text{Rule (Red Com S2) in Table 10; the request is sent to the} \\
 & \mathit{Lectreq}_i \text{ ambient}\} \\
 & \text{Eq. (8)} \mid \mathit{PC303}[\mathit{Lectreq}_i :: (\mathit{lectid}, \mathit{reply}).\mathbf{0}] \mid \\
 & \mathit{Lectreq}_i \left[\begin{array}{l} IS_i \uparrow \langle \mathit{Lect001}, 303, \mathit{PC}, \mathit{PC303}\rangle.\mathbf{0} \mid \\ ! :: (\mathit{lectid}, \mathit{uid}, \mathit{dtype}, \mathit{aname}). \\ IS_i \uparrow \langle \mathit{lectid}, \mathit{uid}, \mathit{dtype}, \mathit{aname}\rangle.\mathbf{0} \mid \\ !IS_i \uparrow (\mathit{lectid}, \mathit{reply}, \mathit{aname}). \\ \mathit{aname} :: \langle \mathit{lectid}, \mathit{reply}\rangle.\mathbf{0} \end{array} \right]
\end{array}
$$

$$\rightarrow \; \{\text{Rule (Red Com R6) in Table 10; the } \mathit{Lectreq}_i \text{ ambient forwards the request to the infostation}\}$$

$$\begin{array}{l} \text{Eq. (8)} \mid \mathit{Cache}_i \downarrow \langle \mathit{Lect001}, \mathit{303}, \mathit{PC}, \mathit{PC303} \rangle.\mathbf{0} \\ \mid \mathit{PC303}[\mathit{Lectreq}_i :: (\mathit{lectid}, \mathit{reply}).\mathbf{0}] \mid \\ \mathit{Lectreq}_i \left[\begin{array}{l} ! :: (\mathit{lectid}, \mathit{uid}, \mathit{dtype}, \mathit{aname}). \\ IS_i \uparrow \langle \mathit{lectid}, \mathit{uid}, \mathit{dtype}, \mathit{aname} \rangle.\mathbf{0} \mid \\ !IS_i \uparrow (\mathit{lectid}, \mathit{reply}, \mathit{aname}). \\ \mathit{aname} :: \langle \mathit{lectid}, \mathit{reply} \rangle.\mathbf{0} \end{array} \right] \end{array}$$

At this stage, the infostation IS_i has received a lecture request from the $Lectreq_i$ ambient and is willing to check with the $Cache_i$ ambient whether it has the requested lecture for the specified type of device. The behaviour of the $Cache_i$ ambient is specified by the process PC_i in Eq.(6). If the requested lecture *Lect001* exists in the cache for the specified type of device, then the $Cache_i$ ambient gets a copy of the lecture for the specified type of device by interacting with the child ambient named *Lect001* whose behaviour is specified by the process $P_{Lect001}$ in Eq.(7). It can also be seen from Eq.(6) and Eq.(7) that if the requested lecture *Lect001* does not exist in the cache for the specified type of device, then a reply message 'NULL' is forwarded to the infostation IS_i. So in either situation, the infostation IS_i receives a reply from the cache. Once a reply is received from the $Cache_i$ ambient, the infostation IS_i contacts the infostation centre *ISC* as specified by the process in Eq.(9). How the *ISC* reacts is modelled by Eq.(14); it replies with a 'DENIED' message if the user requesting the lecture is currently using an mTest service, otherwise it replies with an 'OK' message and possibly a copy of the requested lecture if it is not available locally in the IS_i 's cache. How each of these types of reply is handled by the IS_i is modelled by the components *C* and *N* in Eq.(9). One can see that for every case where the user is still in the range of the infostation IS_i (i.e. the context expression 'has(*aname*)' holds), the infostation IS_i sends a reply to the $Lectrq_i$ ambient which subsequently forwards the reply to the user device as specified in Eq. (4). This completes the proof of Case 1.

$$\mathit{PC303} \left[\begin{array}{l} \mathit{Lectreq}_i :: \langle \mathit{Lect001}, 303, \mathit{PC}, \mathit{PC303} \rangle.\texttt{out}. \\ \texttt{in } IS_j.\mathit{AAAreq}_j :: \langle 303, \mathit{PC}, \mathit{PC303} \rangle.\mathbf{0} \mid \\ \texttt{at}(IS_j)?\mathit{AAAreq}_j :: (\mathit{ack}, \mathit{slist}). \\ \mathit{Lectreq}_j :: (\mathit{lectid}, \mathit{reply}).\mathbf{0} \end{array} \right]$$

The proof of Case 2 can be done in a similar manner as in Case 1, with the user behaviours specified as in Eq.(16), where $i \neq j$, i.e. the request is

sent from one infostation and the reply to that request is received after the user has moved to a different infostation.
This ambient sends a lecture request from the infostation IS_i, moves to a different infostation IS_j, registers with this infostation by sending an AAA request and then waits for the acknowledgement of its registration. Once its registration has been confirmed, it prompts to receive the reply to the lecture request and then terminates.

7. CS-Flow

The process management in the virtual education space is a significant problem that we are trying to solve. We believe a scenario-oriented management is appropriate where various education scenarios can be activated depending on the specific situation in the space. At the same time we are looking for opportunities for a formal specification of these scenarios. Context Secure Flow (short CS-Flow) is a formal interpretable language for presenting context-aware workflow. Presently, we are going to examine the language use in VES. In this section, we introduce CS-Flow briefly; a detailed description can be found in (Zedan, 2012).

Fundamental to the design philosophy is that the computational model has a context as a first class citizen which has a name to identify (model) its location and a frame to identify all observable attributes of interest. Another important consideration is the ability to express security (e.g. access control) and context requirements in a unified fashion using the same mechanism and constructs. In the computational model, the following three distinct structures are supported: *context*, *activity,* and *guard*.

As we mentioned earlier *context* plays an important part in the operational behaviour(s) of modern workflows. Existing modelling techniques assume a centralised execution infrastructure. Therefore, one of the many challenges is to design a workflow system to be executed in pervasive environments. This requires the provision for a computational model in which the context is made a first class citizen and that each of these contexts has its own security policies that constrain its behaviours – from access control to privacy and confidentiality provisions. Such a computation model has to treat both constraints (security and context) in a uniform and integrated fashion.

Contexts can take a variety of forms: different platforms and operating systems, hand-held devices, web-services, users, teaching materials, etc. A context is characterised by what we call a context frame, which is a set of variables (or attributes) of interests. For example, attributes of interests of a context such as a PDA could be its processor speed, memory size, and battery life time. On the other hand, attributes such as age, qualification, and work experience will be of interest in the case of a human context; yet body temperature, blood pressure and kidney functions are attributes more appropriate in the case of a hospital ward context.

In the model, these attributes are predicated upon to form a context guard – in the case of a context, so as a decision may be taken to execute an activity or choose a different but more suitable context, etc. In addition, they are also important as mechanisms to express security policies and for the design of a variety of enforcement mechanisms of these policies that, for example, controls the access to sensitive data/information. As we shall see later, these guards play an important role in the specification and design of context properties and security policy constraints.

An *activity* in the model does not exist in isolation. Indeed, it requires a context to house it. Activities within a workflow move into a context to be executed but may choose to move out to another context in order to complete its functionality. In this way, a context can be nested in a larger context in a compositional fashion.

Our unit of computation in a workflow system is an activity which describes a piece of work that contributes toward the accomplishment of a given (functional and business) goal. Hence, an activity:

- has a goal;
- has an input;
- has an output;
- performs in a particular order – e.g. in sequence/parallel/alternate with others;
- is associated with a particular context – contexts can be an organisation, a device, a service (e.g. web-based service) or a computational environment;
- uses resources/information;
- may affect more than one organisation unit;
- creates some value for users;
- properly terminates – in the same or in a different context.

Central to an activity is the *activity frame*. The frame is a set of variables, known as context variables, which are allowed to change during the execution of the activity. The context variables characterise the attributes of interest about the context, e.g., location, time, temperature, etc. Their changes are first only observed and then acted upon. Normally, an activity starts, executes and then properly terminates within the same context it has started with. An activity starts in one context but may terminate in a different context. This means that an activity has the ability to be mobile and to move from one context to another. But as an activity in our model is tightly associated with a context, mobility occurs at a context level, i.e. an activity moves with its context. This is achieved by the execution of the context operation to α. (i.e. moving out of the current context and/or moving into a new context, α).

In a language, activities may be composed concurrently to produce a new activity which terminates if and only if all of its components terminate, i.e. we adopt the distributed termination convention. Furthermore, without loss of generality, we assume a single clock for an instant of a workflow. Activities are also composed in alteration and in a non-deterministic fashion. An activity can also be conditionally executed after the passability of its condition or guard.

Each activity/context is governed by a set of context and/or security policies/constraints which are continually changing due to either the occurrence of an event and/or the passage of time. Within workflow systems, access *control policies* play a fundamental role. Traditionally, an access control policy is expressed in terms of:

- subjects – such as a human, activities, and platforms. These need to be authenticated before being allowed to access a resource (an object) – an object cannot be modified unless an access right is granted;
- object – this is a resource which is there to be used. It has a state which a subject can alter once it is granted permission to do so;
- action – this is an activity where once the access is granted, it can be executed.

When designing a policy we inevitably think of the system in a specific context and define the rules that restrict the system in this context. It is natural to define policies with respect to their context and then compose the policies to yield the overall system policy. Security policies are expressed within the body of the activity and their enforcement can be part of the infrastructure of the activity itself or its context.

The presented computational model is supported by a design language, known as Context Secure-Flow (CS-Flow), with which we are able to design and analyse workflows. Fig. 7.1 depicts the syntax of CS-Flow based on two syntactic categories: activities (denoted by P or Q) and guards (denoted by G_1 or G_2). We assume a countable-infinite set of names which are written in lower-case letters, e.g. *x* and *y*. Let *any* denote an unspecified value. This is used in communicating signals amongst activities and/or context. We let *b* denote Boolean variables and $\tilde{y}$ denote a list of names. As it is customary we use ˜ to denote a set, i.e. $\tilde{y}$ designates a set of names where it is appropriate. We reserve t to represent time, such that $t \in N$ and t_α and t_β denote the start and termination time of an activity, such that $t_\alpha \leq t_\beta$, i.e. the termination time of an activity will be at or after its release time. We also use the names α, β, γ, ... to identify contexts.

$$
\begin{aligned}
P, Q \; ::= \; & \text{skip} \mid \text{abort} \mid x := v \mid \text{delay}(t) \mid [t_1 \ldots\; t_n]\; P \mid c!v \mid c?x \\
& \mid \alpha <x> \; : \{P\} \mid \text{to}(\alpha) \mid \text{var } \tilde{x} \text{ in } P\ \{Q\} \mid \text{chan } \tilde{c} \text{ in } P\ \{Q\} \\
& \mid \text{in } \alpha \,.\, P(\tilde{x}) \mid P\, ;\, Q \mid P \parallel Q \mid P \rhd.\, {}^{G}{}_{t}\, Q \mid \text{while } G.\ \text{do } P \text{ od} \\
& \mid [p_1] : G_1 \rightarrow P \; \square \; [p2] : G2 \rightarrow Q \\
\\
G \; ::= \; & \text{true} \mid b \mid \text{not } G \mid G1 \text{ and } G2 \mid \text{somewhere}(a)\,.\; G
\end{aligned}
$$

Figure 7.1 Syntax of CS-Flow

A segment of the context "teacher" which models the behaviour of a teacher while conducting lessons is shown (Fig. 7.2) in order to demonstrate the intended application of the language.

$\mathbf{P}_{\textbf{t-conduct-lesson}} \ \widehat{=}\ $ chan $\text{chan}_{t.pat}$ in

{ while true

$\text{chan}_{t.pat}$! any;

to(class-room);

$\text{chan}_{t.pat}$! ready-req;

$\text{chan}_{t.pat}$? ans_{rr};

ans_{rr} = ok → BlendedLearning(C_i);

□

not (ans_{rr} = ok) → TraditionalLearning(C_i);

$\text{chan}_{t.pat}$! finish-req;

to(teacher-office);

}

Figure 7.2 Context “Teacher” (segment)

8. Tempura

The most important characteristic of the scenarios is that they are time-dependent; therefore we need a proper formalism to describe them and an interpreting mechanism to correspond with our system. A good approach for developing the required interpreting mechanism is to look for a formal notation which on the one hand allows to specify the scenarios with their temporal characteristics and on the other hand permits the existence of a possibility for a program interpretation of those specifications. Usually formal logic-based models are suitable in such situations. Apparently, in this case a formal notation using first order logic is inapplicable according to time-critical specifics of the system. Thus, the formal notation Interval Temporal Logic (ITL) was chosen for the following reasons (Moszkowski, 1986):

- ITL is a time-dependent logic allowing the presentation and control of linear and parallel processes. A basic characteristic of this logic is considering time as a discrete sequence of points in time called intervals. Losing the concept of time as an endless notion we can describe different processes in a way similar to the operation of modern computers and software systems;

- For ITL there is an interpreting mechanism and its program realisation called Tempura.

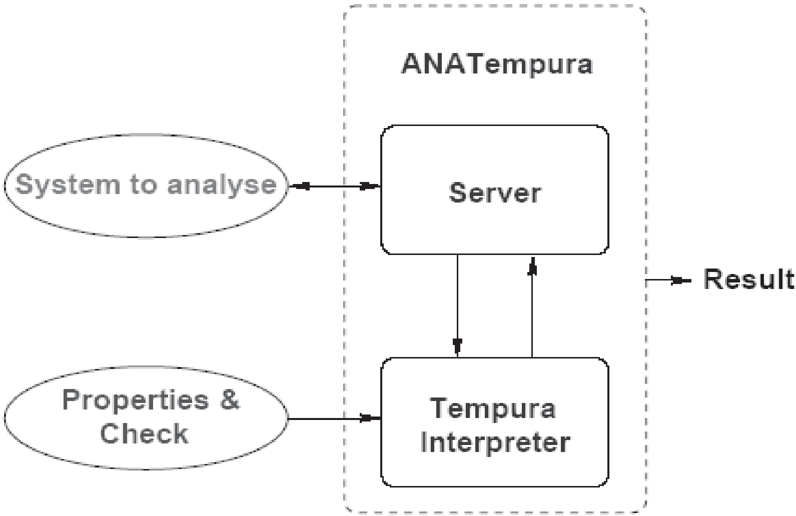


Figure 8.1 Architecture of Ana-Tempura

In practice Tempura is not used by itself but in a special environment called Ana-Tempura (Fig. 8.1). This environment is used as an instrument for verification of time-dependent processes using the ITL mechanisms. Verification is based on special assertion points in the code which Ana-Tempura checks for consistency with criteria for time and security which are presented with ITL notations. During the execution of a given program, those code points generate a sequence of values for certain variables which are matched with pre-defined ITL statements to prove their validity.

In the case of the mobile access we could define a set of variables for identification of an InfoStation (idIS) and a mobile device (idMD). In a particular moment in time, the combination of these variables would indicate where the user is and what kind of device is in use. Following and analysing the alternation of these variables will allow us to recognize which of the scenarios has happened. In fact, this is the basic functionality of Tempura and its surrounding environment Ana-Tempura.

According to the architecture of a specific InfoStation network we could define an ITL statement with idIS and idMD variables which will describe a model of user "behaviour" (i.e. a mobile device type and an IS in range). During the runtime we will use Tempura to match their values alternation to our pre-defined statement and catch the change of the scenarios. Of course to keep our system more homogeneous, we will propose some changes in the original Tempura – Ana-Tempura system:

- The Tempura interpreter will be reengineered to receive an agent-oriented Tempura. The process will be divided into two steps, first the interpreter will be rewritten into Java to receive an object-oriented version and then it will be transformed into a system of agents. The basic reason for this is to keep the homogeneous architecture of DeLC and VES;
- Ana-Tempura is like a surrounding environment of the ITL interpreter Tempura, in our system we will replace Ana-Tempura with specific intelligent agents. Their role will be to supply Tempura with information about changing values of our identification variables.

9. Conclusion

This chapter is structured into two parts. The first one presents the current state of the project Distributed eLearning Center (DeLC) which aims to build a modern e-learning infrastructure. DeLC consists of two separate clusters known as MyDeLC cluster and IS cluster. The delivery of electronic education services and teaching content in the first cluster is achieved by a portal and in the second cluster by a specialized mobile communication network known as InfoStation network. Furthermore, a Virtual Education Space (VES) as an extension of DeLC is presented in the first part.

The second part deals with the application of formal tools for modelling and management of DeLC and VES. In this part their formal system is presented in more detail. Calculus of Context Aware Ambients will be used for modeling and verification of the DeLC and VES infrastructures that are context-aware. SC-Flow will be used for the specification and implementation of scenarios for process management in VES. The new version of Tempura will be used for processing temporal dependencies in the scenarios.

The proposed DeLC infrastructure is implemented by an integrated development environment based on Eclipse including the Java Agent DEvelopment (JADE) framework (Bellifemine et al., 2007) with inter-agent interactions facilitated through Bluetooth communication (in conformance with JSR-82). The JADE framework simplifies the multi-agent systems implementation through the provision of a predefined set of services and management tools in addition to the runtime library and agent programming library. The portal is implemented by the Liferay (Liferay, 2013) framework.

Acknowledgments

The authors wish to acknowledge the partly support of the NPD - Plovdiv University under Grant No. NI-FMI-004 “Innovative basic and applied research in computer science, mathematics and pedagogy training”, 2015-2016 and Grant No. IT15-HF-001 “Improving the access to educational services in the context of Lifelong Learning”, 2015-2016.

References

ADL. (2009). SCORM 2004, 4th edition documentation. Advanced distributed learning, Available at: <http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/2004/>.

Barker, P. (2001). Designing Teaching Webs: Advantages, Problems and Pitfalls, In: *Proc. of ED-MEDIA 2001 World Conference on Educational Multimedia, Hypermedia Telecommunication*, Charlottesville, VA, 54-59.

Bellifemine, F. et al. (2007). *Developing Multi-Agent Systems with JADE*, John Wiley & Sons Ltd.

Bucur, D., Nielsen, M. (2008). Secure Data Flow in a Calculus for Context Awareness, In: *Concurrency, Graphs and Models*, Springer.

Bugliesi, M. et al. (2004). Access Control for Mobile Agents: The Calculus of Boxed Ambients, In: *ACM Trans. on Programming Languages and Systems*, 26 (1), 57-124.

Cardelli, L. Gordon, A. Mobile Ambients, In: *Theoretical Computer Science*, 240, 177-213.

Burners-Lee, T. et al. (2001). The Semantic Web, *Scientific American,* May (2001).

Dey, A.K., Abowd, G.D. (2000). Towards a Better Understanding of Context and Context-awareness, In: *Proceedings of the Workshop on the What, Who, Where, When and How of Context-Awareness*, New York, ACM Press.

FIPA. (2002). ACL Message Structure Specification, Foundation for Intelligent Physical Agents. Available at: http://www.fipa.org/repository/aclspecs.html.

Frenkiel, R., Imielinski, T. (1996). Infostations: The Joy of `Many-time, Many-where' Communications, In: *WINLAB Technical Report*.

Ganchev, I. et al. (2009). Development of InfoStation-based and Context-aware mLearning System Architectures, In: Hijón-Neira, R. (ed.) *Advanced Learning*, In-The, Vucovar, Croatia.

—. (2007). An InfoStation-Based Multi-Agent System Supporting Intelligent Mobile Services Across a University Campus, *Journal of Computers*, 2(3), 21-33.

—. (2008). InfoStation-Based Adaptable Provision of m-Learning Services: Main Scenarios, *International Journal Information Technologies and Knowledge* 2, 475-482.

IMS GLC. (2009). Consortium Current Activities. IMS Global Learning Consortium. Available at: <http://www.imsglobal.org/activities.html>

Liferay. (2013). Liferay. Available at: http://www.liferay.com/.

Maurer, H., Sapper, M. (2001). E-Learning Has to be Seen as Part of General Knowledge Management, In: *Proc. of ED-MEDIA 2001 World Conference on Educational Multimedia, Hypermedia Telecommunications*, Chalottesville VA, 1249-1253.

Milner, R. (1999). *Communication and Mobile Systems: The π-Calculus*, Cambridge University Press, 1999.

Mitev, D. et al. (2009). Selbo2 – An Environment for Creating Electronic Content in Software Engineering, *Cybernetics and Information Technologies*, 9(3), 96-105.

Moszkowski, B. (1986). *Executing Temporal Logic Programs*, Cambridge University Press, Cambridge.

O'Droma, M., Ganchev, I. (2007). Toward a Ubiquitous Consumer Wireless World, *IEEE Wireless Communications*, 14, 52-63.

OWL-S. (2004). OWL-S: Semantic Markup for Web Services. Available at: http://www.w3.org/Submission/OWL-S/.

Passas et al. (2006). Enabling Technologies for the 'Always Best Connected' Concept: Research Articles, *Wirel. Commun. Mob. Comput.*, 6, 523-540.

Sangiorgi, D., Walker, D. *The π-calculus: A Theory of Mobile Processes*, Cambridge University Press.

Stoyanov, S. et al. (2005). From CBT to e-Learning, *Journal Information Technologies and Control*, 3(4), 2-10.

—. (2008). An Approach for the Development of InfoStation-Based eLearning Architectures, *Comptes Rendus de l'Academie Bulgare des Sciences*, 61(9), 1189-1198.

—. (2011). Semantic Multi-Agent mLearning System, In: Elci, A. et al. (eds.) *Semantic Agent Systems: Foundations and Applications*, Springer Verlag

—. (2012). Intelligent Distributed eLearning Architecture, In: Koleshko, V. M. (ed.), *Intelligent Systems*, InTech.

W3C. (2009). W3C, Document Object Model (DOM). Available at: http://www.w3.org/DOM/.

Zhuge, H. (2008). *The Web Resource Space Model,* Springer Verlag.

Zedan, H. (2012). *CS-Flow. Computational Model – Lunguistic Support*, Internal Report, STRL, De Montfort University.

Zimmer, P. (2005). A Calculus for Context-awareness, *Technical Report BRICS*

Automating Helpdesk in the ICT Age

Nithya Kannan and Edmond Prakash

1. Introduction

The rise in Information and Communication Technologies (ICT), have transformed human lives. Organisations are shifting towards thin and cost-effective technology. Almost all transactions are performed electronically nowadays, which in turn eliminates a large amount of paperwork. The number of users of ICT systems within an organisation is also increasing rapidly. On the other hand, the necessity to provide more high quality support services automatically increases the cost of infrastructure and resources. An intelligent way of providing support service is becoming increasingly important to bring satisfaction to users and to enable timely resolution of all ICT related issues within budget.

One of the most significant developments of ICT deployment is in the banking industry. Currently, banking services have been automated by ATMs with instant access to applications like online account checking and money transfer. This significant progress with the deployment of advanced ICT technology has also reduced the need for customers to visit the bank in person and the need for face-to-face consultation with bank employees. Similar to ATM and online bank facilities that empower their clients, service-desks tools are improving and service-desk technologies are empowering customers of day-to-day IT services. By automating the service-desks, the technicians and their managers can devote their attention to resolving issues rather than directly interacting with irate customers.

Use of ICT in a university environment is prevalent in every activity of education with multiple stakeholders including students, teaching, administration and support staff. In fact the utilization of ICT at universities is gaining acceptance at a rapid pace when compared with other organisations. Higher education institutions have deployed ICT infrastructure facilities in all departments. The most important information systems are the student registration system, online library system, online examination system, blackboards, etc. The systems used in a university environment are also under constant evolution. This poses a heavy burden on the service staff, as the number of staff does not increase in tandem with the rising demands for support services and software. The demand and availability expected of these services are also increasing rapidly. As an example, if during the conduct of an online examination, a problem (incident) occurs with the online examination systems; this results in major consequences, if it is not resolved on time. During this critical period, communication is the most important part. An effective solution would be to have an integrated single point of contact, one-stop solution where all stakeholders can record their incident/request as soon as possible.

Most help-desks still operate in a reactive mode, they respond to issues only after the user calls and reports the incident. During peak hours, help desks are inundated with calls, most of them from customers requesting the status of the service request/incidents or logging new calls. These calls divert help-desk technicians from resolving the calls and consequently making the customers unhappy. This leads to a critical situation that affects the smooth running of the business. Secondly, in most of the cases help-desk staff members are not able to assign the call to appropriate IT technicians and are unable to prioritize the calls based on their impact.

To eliminate many of these issues, and to make sure the progress and advancement in ICT sector reaches a wider customer base, it is time to make rapid progress and expansion of the help-desk technology. The expected outcome is a help-desk that will enhance customer satisfaction. The overall aim is to achieve a smooth transition by providing access to the users, instant access to solutions which includes: self-service web-based tools to track their service request/incident without interfering with the day-to-day work of the help-desk personnel.

2. Help Desk

Help Desk Mobilisation: Innovation in mobile technology has raised the service-desk to the next level. Mobilisation is not the right tool for every job. For example, if an employee’s responsibility is to attend the calls at the service-desk then it is not a suitable tool to use. On the other hand, when the technician or administrator is continuously engaged in resolving the issues at different locations then mobilisation will be a more effective and efficient tool to deal with the issues in a timely manner. This enables the technician to solve the incident from any location. Today mobilisation of service-desk is still in the emerging phase in enterprises. When moving to a mobilised service-desk some key areas are focused; For example, when a support team needs quick action for certain requests such as performance of quantity of data transfer or stable connectivity. Point focused application methodology is an area of progress in mobilisation of service-desk [1].

Heat System: Baylor University uses a 'heat system' web tool to manage calls effectively. Any call that cannot be resolved by service desk staff is assigned with a ticket number and then routed to the support team. Earlier versions of the system had limited functionality and were only accessed by service-desk staff. Recently, using content management system the service-desk web tool is modified to accommodate other stakeholders including students, faculty and staff. For service desks in Baylor University, Schaaf and Brenner [2] present a model driven approach with a set of twelve different models based on the OSI architecture for IT management for integrating the IT service tool with web service [3]. Gupta [4] introduced a business driven approach (BDIM model) to utilize web service for integrating management IT tool to support parts of IT service management that can be automated.

Case Based Reasoning: Case based reasoning (CBR) is widely used in intelligent systems for decision support. It contains information of both problem description and resolution description. However when classifying the call type, only description of request is taken into account, additional information is ignored. Recently, various CBR systems have been proposed but the application domain of incident needs to be taken into account. On the other hand using ontological description, the system identifies the most important conceptualisation by describing the domain specific concept and their relationships [5].

2.1 ICT: Classification of Help Desk Technology

Interaction of service providers with customers has been transitioned due to the innovation in communication technologies. For example, banks have changed the interaction with their customers. There are five different methods by which technology relates the provider and clients.

Type A: Technology –free: In the service counter (e.g. consulting services), technology is not involved as shown in Figure 1.

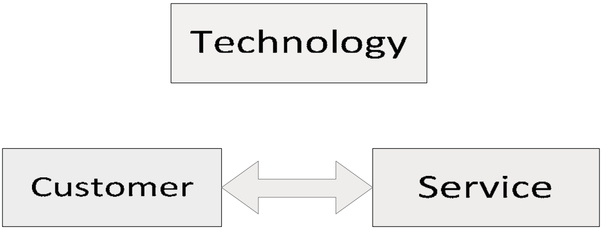


Figure 1: Technology-free

Type B: Technology-assisted: Service provider has access to the technology in the service counter (e.g. A bank personnel uses a terminal to check the customer's transactions) as shown in Figure 2.

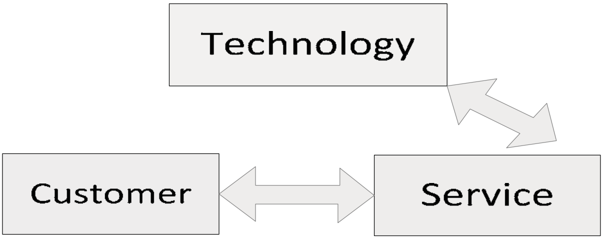


Figure 2: Technology-assisted

Type C: Technology-facilitated: Both the service provider and customer have access to the same technology in the service interaction as shown in Figure 3.

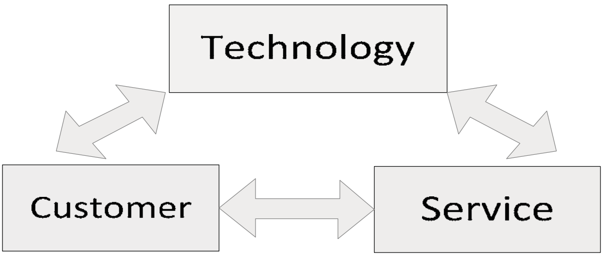


Figure 3: Technology-facilitated

Type D: Technology-mediated: Service provider and customer are not in the same location in the service transaction as shown in Figure 4. Communication may be achieved via phone or email (e.g. a customer who receives technical support from a service desk).

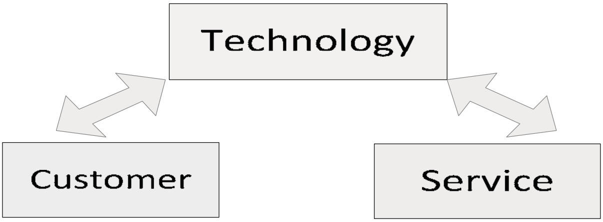


Figure 4: Technology-mediated

Type E: Technology-generated: Service provider completely supported by technology, generally called as self-service (e.g. ATM terminals and online banking) as shown in Figure 5.

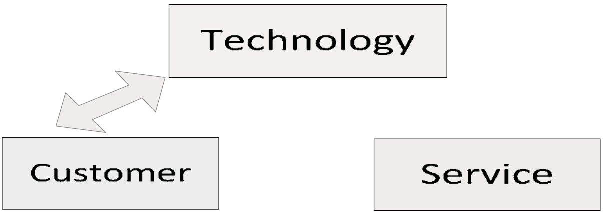


Figure 5: Technology-generated

Service transactions are considered in respect to customers. Customers may be technical or non-technical. Type B or C is effective when the incident is complex and unclear than preferring Type E. When incidents are routine, for example, password change, Type E may be effective. Other types are of secondary consideration. Type D may have language and time-zone implications.

3. ICT: Technology Mediated Intelligent Help-desk

Simple and routine incidents are easily resolved using automation when all the attributes are identical. Promoting online knowledge-base management with search and navigation facilities is effective to resolve such routine incidents. Some of the automated solutions in service management are self-service terminals, online technical support and software applications. Users are provided with self-service tool choice as the first line of support to resolve the significant routine incidents. It reduces the number of incidents that needs to be managed by staff.

The proposed system follows a conventional incident management strategy. The feature of this model is automation of incident classification. The service request is assigned a group, based on the keyword class or token dictionary. Once the incident has been logged, the assigned group accepts the ticket and works on the ticket. If the incident has been resolved then the resolution notes will be updated on the ticket and it will be closed. If it is not resolved, then it is escalated to higher workgroups. The resolution procedure will be continued until the ticket is resolved. The following subsection describes the taxonomy applied for automating the ticket.

3.1 Service Automation

Automation has a massive impact on functioning of service assets such as people, process, management and information. Progress in artificial intelligence and rich media technologies has raised the performance level of software-based agents to manage different tasks. The process of automation helps to achieve the quality of service.

With some level of automation, it is feasible to manage routine service requests/incidents. This needs the study of business activity that exists with each consumer. When going for automation, the following principles should be applied: simplify the process before automating; make clear about the flow of activities, allocation of jobs, and information. The simplification of tasks is made easier through messages, interactive terminals and websites.

3.2 Intelligent System Database

The configuration items (CI) of hardware and software are stored in the configuration database. Relationships of the items are also stored in the database. For instance, the Oracle database is maintained as schema, table, stored procedures and also has explicit and implicit relationships of the objects. Usually search is performed based on the keyword extracted from incident description on the ticket. The extracted keyword is then compared with all configuration items stored in the database. The CI attribute value matches with those keywords are returned as keyword search results.

3.3 Taxonomies

For identifying the incident type, appropriate work groups for the incident different nomenclatures are used. They are incident class, keyword class, and token dictionary. Each type is discussed below:

A. Incident Class Taxonomy

A user calls or logs a call to the service desk to report an incident or request a service. Automation of incident classification is used to support initial incident resolution and facilitate escalation of the incident/request to the appropriate group. In order to automate the classification, "is-a" relationship is used to describe the entities and their relationships. The incident class taxonomy is shown in Figure 6.

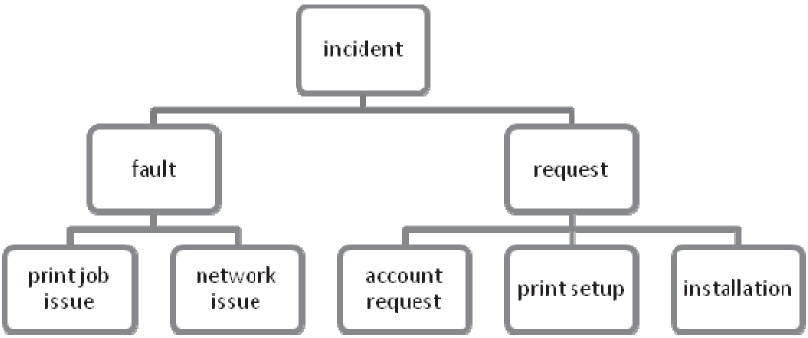


Figure 6: Incident class taxonomy

B. Keyword Class Taxonomy

Keyword-class taxonomy follows "is-a" relationship as shown in Figure 7. The main objective is to enhance the possibility of capturing the keywords using their relationships.

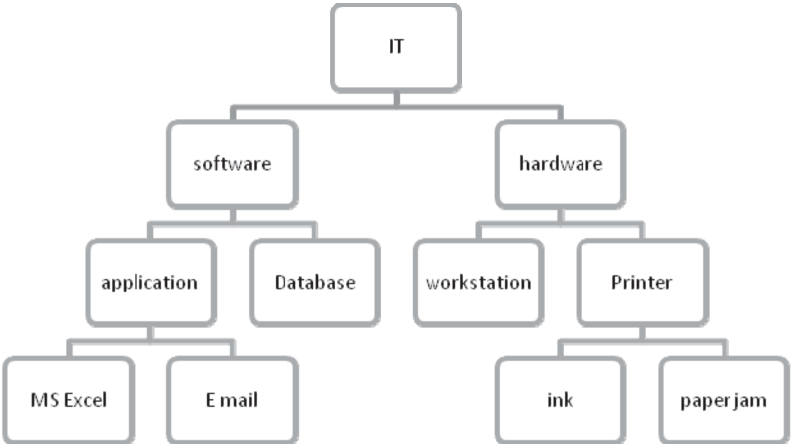


Figure 7: Keyword class taxonomy

C. Token Dictionary Taxonomy

When incident description is entered, tokens are separated by removing the verbal words or stop-words (‘a’, ’is’, ’the’, punctuation, etc.). From the token (data) dictionary stored in the database (see Figure 8), the token is matched and automatically assigned the ticket to appropriate group to resolve the incident. The schema of *token dictionary* consists of tuples of general-tokens, alias terms and keyword class.

General-tokens	Alias terms	Keyword-class
Print	Printer, printing error, print, etc.	printer
web	www, internet, w3, etc.	internet

Figure 8: Token dictionary taxonomy

3.4 Help-Desk Flow

For example, a ticket with incident description: “printer is faulty”

Incident classification: Based on incident classification this is an incident (fault) not the request

Tokens: printer

Keyword-classes: Printer

The proposed system incident management is processed as follows:

Detection: Customer logs a call via online web tool system that automatically generates the ticket with incident number and the ticket contains information ‘customer perception’ in the description. The short and detailed description is known as incident description.

Classification: Based on the incident classification (incident or request) the system will provide the initial resolution and support to escalate the incident to workgroup.

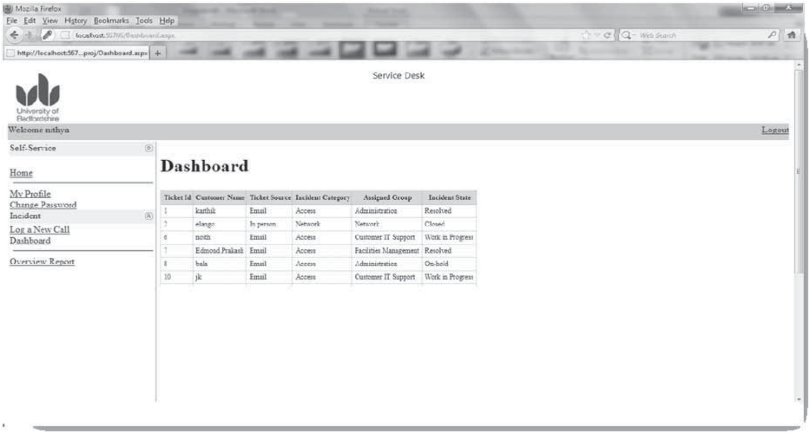
Routing: With the help of the incident description, the classification on the ticket is used to extract the keyword from token dictionary; the incident will then be automatically routed to the appropriate work group.

Resolution: The assigned group follows the resolution procedure or else the ticket will be escalated to the next higher group. This process is continued until the incident is resolved.

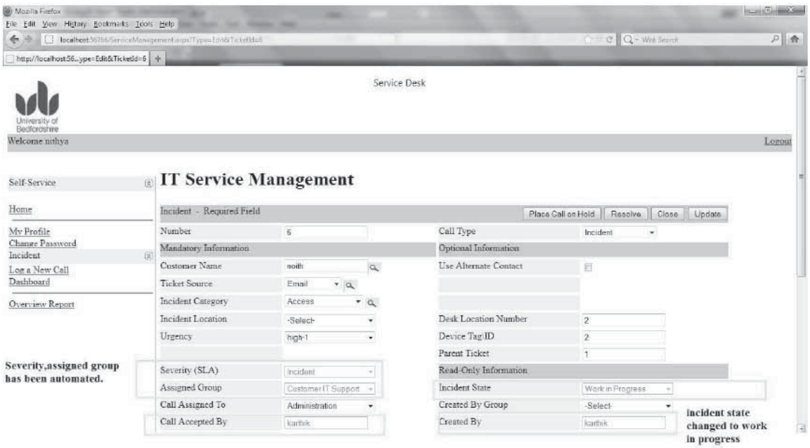
The proposed technology mediated help-desk system filters the incident type, description and classification from the incoming incident. The keyword obtained from the description is compared with token dictionary stored in the database and improves the process of automation of help desk. This technology/system promotes to deliver better quality of service and increase the customer satisfaction by resolving the incoming incident as quickly as possible compared to a typical service desk process. The unresolved ticket will be escalated to higher workgroup and the resolution procedure continued until it is resolved.

3.5 Prototype HelpDesk

We have developed a prototype Incident Response system as shown in Figure 9. The system has been developed using Microsoft asp.net, SQL server management studio and C# programming language. The web forms were simple and easy to comprehend. Though, there were many challenges to run the interface without any error with the database, debugging tools like SQL profiler and Microsoft visual studio IDE made the testing easier. In future, the plan is to extend the work to implement the knowledge-base portal. Based on the similarity evaluation of incident description, similar incidents can be resolved. This knowledge-base management will help the users to deal with the incidents on their own. Combining the incident management and knowledge-base management helps to reduce the number of calls to service desk.



(a) After successful login, the system directs the logger to a dashboard (click on the ticket-id to view the status of the incident/request).



(b) Once the incident/request is accepted by the technician the incident state automatically changes to 'work in progress.'

Figure 9: Snapshot of the working system

4. Summary

We have presented an overview of an ICT enabled Help Desk. We have implemented a prototype of this system to automate the helpdesk. There are many ways this system can be tailored to any organisation. In addition, the system can enhance the performance through meta-data in the incident report, can narrow down and channel the incidents to a specific work group in an organisation. Another direction of improvement will be to add semantic information extracted from the incidents. It is also possible to learn from human experts and this knowledge can help to improve the performance of the help-desk further.

References

Pansa, I.; Walter, P.; Abeck, S.; Scheibenberger, K., "Model-based integration of tools supporting automatable IT Service Management processes," *Network Operations and Management Symposium Workshops (NOMS Wksps), 2010 IEEE/IFIP* , pp.99,102, 19-23 April 2010 doi: 10.1109/NOMSW.2010.5486595

Brenner, M; Garschhammer, M; Sailer, M; and Schaaf, T. CMDB – Yet Another MIB? On Reusing Management Model Concepts in ITIL Configuration Management, Large Scale Management of Distributed Systems, Lecture Notes in Computer Science Volume 4269, 2006, pp 269-280

Nelson, K; Kramolis, T; and Kramolis, T. 2005. Open your eyes...speak your mind...help desk communicating. In *Proceedings of the 33rd annual ACM SIGUCCS conference on User services* (SIGUCCS '05). ACM, New York, NY, USA, 253-256.
DOI=10.1145/1099435.1099494
http://doi.acm.org/10.1145/1099435.1099494

Gupta, R.; Prasad, K.H.; Mohania, M., "Information integration techniques to automate incident management," *Network Operations and Management Symposium, 2008. NOMS 2008. IEEE* , vol., no., pp.979,982, 7-11 April 2008
doi: 10.1109/NOMS.2008.4575262

Jantti, M.; Tanskanen, K.; Kaukola, J., "Knowledge Management Challenges in Customer Support: A Case Study," *Information, Process, and Knowledge Management, 2009. eKNOW '09. International Conference on*, pp.78,83, 1-7 Feb. 2009
doi: 10.1109/eKNOW.2009.7

Kang, YB; Zaslavsky, A; Krishnaswamy, S and Bartolini, C. A knowledge-rich similarity measure for improving IT incident resolution process", in Proceedings of the 2010 ACM Symposium on Applied Computing, ACM, 2010, pp. 1781-1788.