Education Compliance and e-Learning in Australia: A Distance Education Model in Exercise and Sport Science

Ian Heazlewood
School of Psychological and Clinical Sciences
Charles Darwin University
Australia
Ian.heazlewood@cdu.edu.au

Abstract:

Universities attempt to develop innovative and current degrees in Australia, especially in exercise and sports science, which is one of the fastest growing disciplines within Australian universities. Compliance in higher education and relevant to exercise and sports science degrees is now an expected part of the higher education playing field in the Australian university sector as Australian and State governments through SCOTESE control the developing curricula by enforcing compliance with national agendas, such as the Australian Qualifications Framework that controls the expected level of academic achievement and skill levels within the 10 levels of educations from certificate level to doctoral level qualifications. In addition, professional associations also have input and demand compliance in terms of micromanaging the subjects and the curriculum in exercise and sports science, as well as university driven models of compliance in terms of developing and moderating degree programs. The development of e-learning fully and on-line degrees is now the norm not the exception with some universities delivering to 70% of their students using this education mode and Open Universities Australia a 100%. However, such development is controlled via compliance and how this compliance influences the product universities sell to their customers, the students.

Keywords – higher education, compliance, e-learning, exercise, sport science

I. INTRODUCTION

Currently in Australia many universities provide access to study by full on-line course delivery and assessment and in some cases over 70% of students access university higher education in this learning and assessment mode. Open Universities Australia, a consortium of Australian universities who deliver on-line distance education are at 100%. In the Australian higher education sector, such as Australian universities, is driven by Australian Government compliance. The majority of Australian universities are dependent heavily for funding from the Australian government as a consequence the Australian Government and State Governments have assumed a stronger role in university compliance. Currently, the higher education and university sector is driven by The Council of Australian Governments (COAG) Standing Council on Tertiary Education, Skills and Employment (SCOTESE) has been established. The purpose of SCOTESE, as set out in its Terms of Reference, is to ensure that Australia’s current and future workforce needs are met through increased participation, educational attainment, skills development and skills use to achieve greater productivity [1, 2]. SCOTESE held its inaugural meeting on 25 November 2011. The two of the Council’s five priority issues are:
1. Ensuring high quality tertiary education and training through strengthening of regulatory arrangements in the tertiary education and training sectors and promoting a sustainable and competitive international education sector.

2. Establishing a national tertiary education and training framework that drives responsiveness to labour market needs.

In the context of priority one, ensuring high quality tertiary education and training through strengthening of regulatory arrangements in the tertiary education and training sectors, demands are placed on universities to meet compliance guidelines. An example that impacts directly on the design and development of university curricula and subjects at all levels is the Australian Qualifications Framework [3, 4]. I will emphasise the Australian Qualifications Framework (AQF) is the national policy for regulated qualifications in Australian education and training. It incorporates the qualifications from each education and training sector into a single comprehensive national qualifications framework [3, 4]. The AQF was originally introduced in 1995 to underpin the national system of qualifications in Australia encompassing higher education, vocational education and training and has undergone a number of revisions, especially the modification of the original twelve levels of qualification from certificate 1 to doctoral level to the current ten levels of qualification. Of the seven objectives of the AQF objective 2, contributes to national economic performance by supporting contemporary, relevant and nationally consistent qualification outcomes which build confidence in qualifications and objective 5, underpins national regulatory and quality assurance arrangements for education and training are related to compliance issues.

II. AUSTRALIAN EDUCATION THE FLOW ON EFFECTS

Now that the education context has been provided for Australian education the flow on effects will now be addressed and will focus on the influence of undergraduate bachelor level degrees. In terms of purpose the bachelor degree qualifies individuals who apply a broad and coherent body of knowledge in a range of contexts to undertake professional work and as a pathway for further study. Normally, generic bachelor level degrees such as in science and arts are three years in duration full-time study whereas professional degrees, such as engineering, pharmacy and law are four years. In terms of expected skill outcomes graduates will have the cognitive skills to review critically, analyse consolidate and synthesise knowledge; cognitive and technical skills to demonstrate a broad understanding of knowledge with depth in some areas; cognitive and creative skills to exercise critical thinking and judgment in identifying and solving problems with intellectual independence; and communication skills to present a clear, coherent and independent exposition of knowledge and idea. These skills are presented in very generalised terms as are the following in the context of application and knowledge.

In terms of application of knowledge and skills graduates will demonstrate initiative and judgment in planning, problem solving and decision making in professional practice and/or scholarship; adapt knowledge and skills in diverse contexts and displaying responsibility and accountability for own learning and professional practice and in collaboration with others within a broad parameters. In terms of responsibility for accreditation and development, “accrediting authorities and those developing qualifications for accreditation must adhere to the AQF specifications for this qualification type and any government accrediting standards for higher education when accrediting a Bachelor Degree qualification,” (in this context universities) [3, p. 46].
III DEVELOPING A FULLY ON-LINE DELIVERED EXERCISE AND SPORT SCIENCE BACHELORS DEGREE

The next step is to integrate diverse inputs in terms of developing a fully on-line delivered exercise and sport science bachelor level degree that displays AQF compliance, e-learning technology, applications pedagogy-andragogy principles, that covers the subject area that represent exercise and sport science and all delivered via distance/external mode using e-learning management systems. In other words, blending learning theories and approaches for e-learning, e-learning pedagogical strategies, e-learning tactics and e-learning technology.

The first step is the developing the curriculum is the set of courses, and their content, offered at university [5]. Within Australia professional organizations are attempting to make universities compliant in terms of the curriculum offered by Australian universities in the domain of exercise and sport science such as the Exercise and Sport Science Australian (ESSA). In essence that believe they have authority in accrediting exercise and sport science programs and provided universities who wish to gain this accreditation a 126 page document with specific compliance criteria including specific micromanaged content in the subjects and subject sequence undertaken [6].

The units/subjects of study at Charles Darwin University CDU follow this model as shown in Table 1.

<table>
<thead>
<tr>
<th>Unit type</th>
<th>Credit Points</th>
<th>Specific requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common (2 units)</td>
<td>20</td>
<td>Units totalling 20 credit points selected from Common Units available at the time (e.g. units coded with a prefix of “CUC”). Students must complete the compulsory regional and indigenous issues unit (CUC107) and one of the two academic skills units. CUC107 Cultural Intelligence and Capability (compulsory) Plus either: CUC100 Academic Literacies or CUC106 Design and Innovation: Communicating Technology</td>
</tr>
<tr>
<td>Core (16 units)</td>
<td>160</td>
<td>Compulsory Core units totalling 160 credit points from units detailed below: ENG429 Biomedical Engineering PSY140 Introduction to Psychology 1 SBI171 Anatomy &amp; Physiology 1 SBI172 Anatomy and Physiology 2 SBI209 Design and Analysis of Biological Studies SBI261 Functional Anatomy SPE205 Biomechanics 1 SPE206 Exercise &amp; Sport Psychology 1 SPE209 Exercise Physiology 1 SPE210 Motor Control &amp; Development SPE305 Exercise &amp; Sport Science Practicum SPE309 Exercise Physiology 2 SPE310 Advanced Study in Exercise and Sport Science x 2 (repeatable unit)</td>
</tr>
</tbody>
</table>
The progression is 100 level units are year 1, 200 level units are year 2 and 300 level units are year 3. The expectation in the AQF model is a progression in depth of knowledge and skills as the student progresses from level 100 to level 300 units.

IV LEARNING THEORY PARADIGM

For those students undertaking the fully on-line external/distance degree the application of learning theory/paradigm was based on andragogy or adult learning as the majority of students in the BESS program are adult learners females and in their early thirties who are returning to study. Using this paradigm [7] it is believed that:

1. Adults are motivated to learn from being in situations in which they see a need to learn. Consequently, adult learning settings should begin with topics that address the adult audience’s current learning needs. At this point in time Australia has a significant national health problem related to obese and overweight adults and children and many of the topics in nutrition, exercise and health, anatomy and physiology and exercise physiology, exercise and sport psychology relate specifically to evaluating, understanding and treating this condition via exercise sport psychology and nutrition. The topic has a significant and direct social and health context for students.

2. Adults are oriented to the broad range of affairs in life, not to narrow subjects. Thus, adult teaching should be multidisciplinary rather than subject-oriented. Third year units allow for overlap of unit content when they undertake the two 300 level units in year three Advanced Studies in Exercise and Sport Science which promotes a multidisciplinary approach.

3. Adults learn from their experience. Therefore, the most productive adult learning comes from the analysis of adult experience. Both cased based problem solving and problem based learning approaches are implemented especially in 200 level and 300 level units, which are built around student experiences in exercise and sport as well as the social context of nutrition and exercise in the Australian community.

4. Adults have a deep need to be self-directing. Therefore, teaching adults should be involved in setting the agenda for their learning. The learning agenda allows the students at second and third year to select topics that interest them from a large range of potential topics within such subjects as exercise and sport psychology, biomechanics, exercise physiology and advanced studies in exercise and sport science. An
opportunity to specific integrate unit information into their world view (some educational neoconstructivism [8])

5. Individual differences broaden and harden with age. Therefore, adult teaching should make allowance for differences in style, time, place, pace, focus, and method. Distance/external fully on-line delivered education enable learning to suit the students’ day to day needs as education can be more easily integrated into their unique priority of needs and time demands.

V E-LEARNING MANAGEMENT SYSTEM

The next step is the development and delivery of contents is via the e-learning management system hosted by Blackboard. The Blackboard Learning System is a virtual learning environment and course management system supported by CDU’s LearnLine experts to assist staff developing their LearnLine sites for the new distance-external mode of the BESS degree. Staff who taught or were to teach the subjects to be migrated to distance-external mode were allocated a subject content specialist, as well as with expertise in Blackboard functionality as the Blackboard web based system is the underpinning structure of CDU’s LearnLine system. The funding per subject is substantial in terms of funding for external experts, internal support and up skilling staff both with software technology, hardware technology and application of pedagogical/andragogical theory and principles.

Using this LearnLine system functionality permits the students to:

• View and download content. PowerPoint’s, MP3 podcasts, links to videos, websites, learning activities, student assessments and Blackboard tutorials for staff and student users.
• Interact with your lecturer and fellow students via a range of communication tools both real time and asynchronously. Blogs, Blackboard collaborate, emails, discussion board, wikis, journals and so on.
• Undertake assessments and upload assignments under Safeassign.
• Track student progress and grades. Staff facility for whole of class tracking and individual students to track own assessment performance.
• Join online classrooms with Online Classroom collaboration tools and discussion board.
• Listen to lecture podcasts. MP3, websites and videos.

VI INTERNAL LEARNLINE WEBSITE COMPLIANCE AND RESOURCE DEVELOPMENT

However, another compliance issue involves all unit/subject websites passing a CDU quality audit conducted by LearnLine staff prior to release to CDU students. Such compliance involves copyright, minimum information for students as unit information, introductions, some lecture materials and assessments; and checks using 2012 Blackboard Exemplary Course Rubric [9].

One of the major problems with fully on-line science based subjects is how to manage practical classes using instrumentation and conducting evaluations that are an integral component to many exercise and sport science units. One method is to supplement student learning by conducting on campus student practical intensives where the students’ actually attend practical, whereas another method is the virtual classroom where students view and respond to practical demonstrations in terms of instrument set-up, data acquisition and data
analysis. A useful resource which the university subscribes to is Online - Sports Medicine and Exercise Science in Video [10] and which provides 219 videos totaling 268 hours of viewing across the domains of functional anatomy, exercise physiology, sport psychology, biomechanics, advanced studies in exercise and sport science, and exercise and health. These can be inserted within learning materials and learning activities to provide a comprehensive virtual classroom for exercise and sport science. According to the publishers the resources are “the most extensive video collection ever assembled in the areas of fitness and health assessment, disease management, injury treatment, nutrition, medical fitness, sport science, work-site wellness and exercise adherence” [10].

In addition, to extensive video resources most exercise and sport science textbooks some complete with extensive supporting materials that can uploaded onto LearnLine sites and these include chapter by chapter lecture driven PowerPoint presentations, streaming MP3 podcasts via textbook authors and subject experts, learning activities and chapter quizzes both random and non-random question selection, usually loaded via the software Respondus [11].

VII DEVELOPING AND STRUCTURING ASSESSMENTS

Structuring assessments to be compliant with AQF expectations within a bachelor’s degree and within year of learning within the degree is delivered in the following format in the core units in the BESS degree.

Year 1 involves some small assignments, some centralized examinations and dependence on multiple choice questions (MCQ’s).

Year 2 extended assignments (self-selected topics), simulations producing a research report (none self-selected), group based activities (self-selected) via a PowerPoint presentation and centralized examination (none self-selected) and all downloaded and uploaded via LearnLine.

Year 3 extended assignments (self-selected topics), simulations producing a research report (none self-selected), group based activities (self-selected and using on-line student collaborations) via a PowerPoint presentation and centralized examination (none self-selected) and all downloaded and uploaded via LearnLine. The emphases in these assessments are based on more complex models and which include multivariate statistical applications to exercise and sport science content.

In this context compliance with AQF and 2012 Blackboard Exemplary Course Rubrics re monitored internally via CDU internal quality assurance moderations. The newer learning and teaching 2013 quality assurance model for unit assessment items and LearnLine delivered units will require moderation via external peer review, although this process occurs already with moderation and quality assurance of CDU whole of undergraduate degrees.

VIII CONCLUSION

You are now finally at the step where the academic is able to deliver the fully on-line exercise and sport science degree built around the units/subjects utilizing fully on-line e-learning; supported by an extensive and comprehensive software (LearnLine built upon Blackboard), hardware (assuming the student has a functional computer system to support LearnLine), the academic who has been trained to use LearnLine, trained in pedagogy-andragogy to use e-learning, has uploaded relevant resources for learning and teaching that are
supported by the university e-learning management system. However all this is embedded in Australian and State Government compliance, the Australian AQF system; professional association compliance, ESSA; and internally driven university compliance, internal quality assurance of programs, units/subjects and LearnLine audits. The last and most important question is did the students learn anything and did compliance actually improve the product for our clients?

REFERENCES


   http://www.aqf.edu.au/


A survey of Data mining in the context of E-learning

Yousef A. ALMazroui
al_mazroui@hotmail.com

Abstract

In the recent past, every domain has benefited from the growth of Information and Communication Technologies (ICT). Education is one such sector with the rapid deployments of e-learning systems. Teaching and learning in e-learning domain is a serious business and there is a need for information processing in order to enrich the learning experience as well as to tap the potential of the growing e-learning business. In this context, the article aims to investigate the need for data mining, identify the problems within e-learning that data mining can solve as well as present the existing approaches, currently existing open research challenges and the future directions in this area.

1. INTRODUCTION

Over the last few decades ICT has evolved rapidly and brought forth a plethora of new services. However, the theme of data delivery using the client-server models has not changed. The most popular means of data delivery, apart from e-mail has been the World Wide Web (WWW). With the arrival of Web 2.0, content presentation has become richer and brought about the genesis of a large range of tools that enable content generation, adaptation as well as delivery [9]. Disseminating educational material for purposes of learning was achieved prior to Web 2.0. However, it is only subsequent to Web 2.0, the online learning space became participative and interactive [13].

Several tools, both general purpose (databases, blogs, Wikis, etc.) as well as specific (Moodle, Blackboard, etc.), are used in the context of e-learning. Together, as infrastructure, they provide the overall functionality of e-learning [19]. Each of these tools generates a log of user activities as a part of their operation. These logs are cumulative and it is possible to co-relate activities of a single user across these functional blocks to assess the utilization or extract resource access trends [9]. Beyond this, the possibility of extracting information to provide a feedback to the overall educational process to make it dynamic and adaptive is exciting. It is in this context of learning that two new but similar streams of research have emerged in the recent past, Educational Data Mining (EDM) [30] and learning analytics [31]. This article aims to investigate the role played by data mining techniques in e-learning particularly focussing on Educational Data Mining methods. Section 2, presents a brief background and the relation between E-learning and Educational Data Mining, section 3 elaborates on the various EDM methods and techniques while section 4 presents the state of the art and related work. Section 5 presents some of the open research challenges currently faced by the research community applying data mining techniques to e-learning and section 6 concludes the article with comments on future trends.

III. 2. E-LEARNING AND EDUCATIONAL DATA MINING

Clark and Mayer in [9] define e-learning as training delivered on a digital device such as a smart phone or a laptop that is designed to support individual learning or achieve organizational goals. They refer to self-study and similar supplementary systems as asynchronous e-learning and those with instructor-led learning...
(virtual classrooms, webinars, etc.) as synchronous e-learning. They very succinctly mention that the “e” in e-learning denotes how the course is digitized, stored and delivered whereas the “learning” refers to the content and the purpose of providing the content (the end user objective).

E-education is the term used by Bedi and Milic and Stedul in [6] to denote ICT methods and resources that supplement the traditional teaching process and resources. They emphasize that the term e-learning denotes a two-way interactive learning process involving a teacher and student using digital media. The teaching process is student-oriented leading to pro-active learning by the student and requiring quality online cooperation between the student and the teacher.

Bowles in [8] has clearly captured the multi-disciplinary aspect of e-learning. Figure 1 illustrates how the various fields of study interact to constitute e-learning.

![Figure 1: E-Learning - merging various fields of study (Bowles, M., 2004)](image)

It is important to observe that Knowledge Management (KM) is a key aspect in our context. It is from the requirement for KM that the need for information processing originates and evolves into the need for Data Mining.

The various functional elements of an e-learning infrastructure such as the generic and specific tools record all the user and system activity in the form of logs. These logs are detailed enough to track the utilization of resources, resource delivery throughputs as well as user activity and user requirement trends [24]. In addition to these inputs, the tools and software can be instrumented further to provide detailed inputs with respect to the user, including the mood of the user and/or her/his reaction to the usage of the system by deploying sensors that can track the user’s eye movement and facial expressions. All these inputs can be fed back to the learning system as well as provide valuable feedback to the teacher to enable a certain degree of adaptation, either system-wide or specifically to a student as demonstrated in [7].

Educational Data Mining (EDM) extends from the field of Data Mining (DM). DM attempted to fulfil many requirements that were not fulfilled by traditional methods such as Statistics, Artificial Intelligence, Pattern Recognition, Machine Learning, and so on [24]. It was also referred to as Knowledge Discovery in Databases (KDD) and involved processing large amounts of data to look for useful information [28]. EDM is DM applied in the context of e-learning with the objectives focused to the context.

Figure 2, summarizes the various stakeholders in the EDM domain and each of the above players can benefit in different ways from EDM. For instance, the learners can receive advice and customised
recommendations about resources and tasks that are the most suitable w.r.t. their current knowledge and learning objectives, educators can detect how effective their learning materials are, student performance etc, a study adviser can identify risk groups among the students and the directors of education can see how the students actually study and what the bottlenecks are in the current curriculum. In either case it is expected that the mined knowledge can give a better insight, facilitate and enhance the educational processes and the learning as a whole [10].

**Figure 2: Educational Data Mining in a Nutshell (Calders and Pechenizkiy, 2012) [10]**

EDM is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in [30]. EDM is concerned with developing, researching, and applying computerized methods to detect patterns in large collections of educational data – patterns that would otherwise be hard or impossible to analyze due to the enormous volume of data they exist within [22]. The data refers to both structured (logs, student registration information, student usage profiles, administrative information, etc.) and unstructured data (interaction with teachers via chat, collaboration with other students via chat, etc.).

**IV. 3. VARIOUS DATA MINING METHODS USED IN E-LEARNING**

EDM methods, like those in DM, are drawn from a variety of areas such as Statistics, Data Modelling, Information Visualisation, Machine Learning and Psychometrics [2]. Initially, EDM focused on Statistics and Visualisation [24]. Baker’s work on taxonomy of EDM methods in [3] provides an alternate view of the overall objectives and brings in two additional categories – Distillation of data for Human Judgement and Discovery with Models which are not in the traditional DM category. The use of models for information discovery has become increasingly popular as a technique in EDM research and is used to support sophisticated analyses [2].

In terms of the traditional DM tasks that are used to meet these objectives, Romero and Ventura in [24] report that regression, clustering, classification, and association-rule mining are the most frequently used and they correspond to the usage of decision trees, neural networks, and Bayesian networks, as techniques used for these tasks.

Scheuer & McLaren in [25] extend the taxonomy of EDM methods mentioned by Baker et.al in [3].
They modify and regroup techniques into the following methods:

- **Supervised model induction** includes machine learning techniques that infer prediction models from training instances where the value of the target attribute is known [25]. The goal of prediction is to develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of data (predictor variables) [4]. Classification, Regression and Density estimation are techniques used for prediction.

- **Unsupervised model induction** includes techniques similar to those above but when the value of the target attribute is not known [25]. Given a set of data, we need to identify classes or clusters of data and correlate them to find similarities. The similarity is measured by a “distance” factor. Clustering is the technique that is extensively used in this context.

- **Parameter Estimation** comprises means of estimating parameters of probabilistic models from the available data [25]. These models are used to predict the probability of specific events. The predicted variable is typically a probability density function such as a Gaussian function [4].

- **Relationship Mining** involves discovery of relationships between variables which may be associative, correlative, sequential, or casual [25]. If-Then rules are extensively used in relationship mining. These rules are then graded using “Support-Confidence” thresholds and those that are above both the thresholds are used for learning.

- **Distillation of data for human judgment** involves visualization of data. Data is presented to the teachers in the form of 3D graphs and similar visuals and a facility is provided to interact with this visualization to alter the forms and/or interpretation, as necessary [25].

- **Discovery with models** involves developing a model of a phenomenon via prediction, clustering or knowledge engineering. This model is then used as a component in another analysis, such as relationship mining or prediction [4].

In addition to these methods, Barahate in [4] lists a few other EDM methods:

- **Outlier detection** discovers data points which are significantly different from other data in the data set [16]. These could be data points that are either very high or very low in value. Outlier detection can be used to identify students with learning problems [4].

- **Text mining** is viewed as an extension of data mining to text data [16] and includes web content mining. It is used to extract information from semi-structured and unstructured data such as emails, web pages and text documents [13].

- **Social Network Analysis (SNA)** is used to measure relationships in the context of networked entities. They are typically used to assess online interactions between students. They can be effectively used to understand the group dynamics (structure and content) of educational communities [20].
These EDM methods are deployed effectively to improve student models, improving models of the knowledge structure of the domain, study the pedagogical support provided by learning software and discovery of trends for learning and learners [4].

4. STATE OF THE ART AND RELATED WORK

Data-mining that has been pre-dominantly used in e-commerce and for business applications is considered as a suitable candidate to fit into the domain of e-learning for one main reason i.e, e-learning similar to e-commerce is a large and growing business. Data mining techniques can potentially help identify the patterns that may be profitable to the e-learning business [15].

However, there are some critical differences in applying data mining to the e-commerce and e-learning domains.

1. Guiding students in their learning is the main purpose of e-learning unlike in e-commerce where the purpose is to guide the clients to purchase goods [21].

2. While increasing profits that can be quantified is the objective in e-commerce, in e-learning, improving the learning experience of students is the objective which is subjective [21].

It can be stated that the primary goal of applying data mining techniques in e-learning is to deal with the exponentially growing, unstructured information that the e-learning systems are generating. It was conceived that data mining techniques would enable data analysis to transform the huge volumes of data to relevant contextual knowledge. The broad objective was to understand the patterns of usage by the various players involved in the e-learning framework, mainly the teachers and students [11].

The current mainstream EDM research is primarily focused on mining logs generated by the e-learning systems [14]. However, Data mining in e-learning context is expected to address problems associated with different phases in the leaning process. The learning processes could be viewed in the following four lines namely; formal (e.g. tests) or informal (e.g. educational games), intentional (e.g. tutoring) or unexpected (e.g. using the social media) [7].

Although it is fairly a new stream of research, scientists have applied data-mining methods in e-learning context in some unique and interesting ways. The most common is the application of knowledge extraction techniques to formatively evaluate the e-learning systems with the purpose of continually improving the learning experience as presented by Arruabarrena et.al in [1].

Baker in [3] suggests applying EDM for improving student models, improving domain models, studying the pedagogical support provided by learning software, and scientific research into learning and learners. Five additional areas are added by Castro et al in [11].

They are applications dealing with the assessment of the student’s learning performance, course adaptation and learning recommendations based on the student’s learning behavior, evaluation of learning material and web-based courses, providing feedback to both teacher and students in e-learning courses, and detection of students’ learning behaviors. Romero and Ventura in [24] mention additional applications such as Student Grouping, Social Network Analysis, Developing Concept Maps, Planning & Scheduling and constructing Courseware.
Gong et al. in [12] have presented how educational data mining (EDM) methods are applied to look for empirical evidence to gain deeper understanding of the key factors impacting learning.

Beck and Mostow in [5] have studied learning decomposition that fits exponential learning curves to performance data relating a student’s later success to the amount of pedagogical support the student has received up to that point.

The work done by Zaiane in [29] centers around on how EDM methods can be used to support the development of more sensitive and effective e-learning systems. A specific example of such application is the work done by Tang et al. in [26]. Tang has proposed an evolving e-learning system using data mining techniques. Unlike a traditional e-learning system, in an evolving e-learning system, learning materials are automatically found on the web and integrated into the system based on users’ interactions with the system. Therefore, although users do not have direct interaction with the Web, new or different learning materials from the internet enrich their learning experiences through personalized paper recommendations.

Beck and Woolf in [32] have demonstrated how educational data mining prediction methods can be used to develop student models. They have presented a variety of variables to predict if the student will answer a particular question correctly. This work has inspired many related work and student modelling is a key theme in educational data mining.

Macfayden and Dawson in [18] have conducted a study that confirms that pedagogically meaningful information that is extracted from e-learning systems can be used to develop a customizable dashboard-like reporting tool for educators that will extract and visualize real-time data on student engagement and likelihood of success.

![Figure 1: Tutoring time (T_t) and waiting time (T_w) per session](image)

![Figure 2: Waiting factor (W) per session (in logarithmic scale)](image)

---

**Figure 3: Results on effective of teaching sessions (Source [34])**

Figure 3, shows results from the work done in [34] which studies the effect of student waiting times in a tutoring session and the efficiency of the tutoring session. The first part of the graph shows tutoring time is...
mostly higher than waiting time implying the higher efficiency of the session, the send half of the graph emphasizes that result, which represents the waiting factor in a log scale.

Figure 4: Illustration of an observation tool that encourages but does not require active observation.

The work presented in [35] aims to capture the student attention, an important parameter to measure in studying the effectiveness of learning methods. The method adopted by the authors is to redesign the learning interface so that some amount of additional interaction is encouraged and/or required for the reading. This means rethinking observation processes, transforming them from relatively passive activities to explicitly active processes.

5. RESEARCH GOALS AND CHALLENGES IN EDUCATIONAL DATA MINING

One of the critical research challenges for Educational Data mining is that most data mining techniques are built for business data and extract business value. There is need for building data mining methods that are specific to educational data [2].

An important and unique feature of educational data is that they are hierarchical. Data at the keystroke level, the answer level, the session level, the student level, the classroom level, the teacher level, and the school level are nested inside one another [2][24]. Three important features or parameters, that needs to be considered while capturing educational data are time, sequence, and context.

i) Time is an important parameter, such as the length of practice sessions or time to learn.

ii) Sequence represents how concepts build on one another and guide on how practice and tutoring should be ordered.

iii) Context is important for explaining results and knowing where a model may or may not work.

Methods for hierarchical data mining and longitudinal data modelling have been important developments in mining educational data and are considered to be critical research areas [7].

The research goals of educational data mining can be summarized as follows [7]:

1. Predicting students’ learning behaviour by creating student models that takes into account detailed information about the students. Information such as student’s knowledge about a subject matter, motivation and meta cognition which refers to higher order of thinking.

2. Discovering or improving domain models that characterize the content to be learned and optimal instructional sequences.
3. Studying the effects of different kinds of pedagogical (learning) support that can be provided by the e-learning software.

4. Advancing scientific knowledge about learning and learners through building computational models that incorporate the above three models of the student, the domain, and the software’s pedagogy.

Visual data analysis blends highly advanced computational methods with sophisticated graphics engines to tap the ability of humans to see patterns and structure in complex visual presentations [17]. It is essential to conduct research on the usability and impact of alternative ways of presenting fine-grained learning data to instructors, students, and parents.

Similarly, researchers need develop decision supports and recommendation engines that minimize the extent to which instructors need to actively analyse data [7].

Data is crucial for EDM or for learning analytics and many outstanding concerns in EDM research centres around data. Foundational issues are of data quality, ethics of use, scope of analytics activity, data standards, and integration data sets [7].

Potential barriers to adopting educational data mining are technical challenges, institutional capacity, and ethical issues. Successful application of educational data mining shall not come without effort, cost, and a change in educational culture to more frequent use of data to make decisions [27].

Technical challenges are mainly the high-end technical infrastructure that is required to process the vast amount of data. However, the actual technical challenge lies in ensuring the right data set are analysed to answer the e-learning questions that needs to be answered, as well as ensuring the data sets align with the questions [7].

Lack of data inter-operability is a major research challenge. Since most institutions have an increased decentralization of the source of education data, such as student information systems, teachers’ online grade books, homework submission systems, and publishers’ online assignments, homework help, and assessments. Systems provided by commercial vendors may not give the data they collect back to the teacher, district, or parents, so understanding who will own the data is an important consideration [7].

Educational data mining makes predictions and recommends actions based on increased visibility into student actions, and these give rise to a number of social and ethical concerns. Privacy and ethical issues are considered to be a major research challenge [7]. Personalized interaction and user modelling have significant privacy implications because personal information about users needs to be collected to customize software to individuals [27]. At the same time, privacy versus personalization is not a simple trade-off. Researchers need to develop techniques for anonymity of data and tools for data aggregation and disaggregation that protect individual privacy yet ensure advancements in the use of educational data.
Figure 5: Number of papers published until 2009 categorized based on the focus area of research. (Source: [24] Romero, C.; Ventura, S.,(2010))

Figure 5 depicts the focus of the educational data mining research community. Note that until 2009, there was a lot of work on visualization analysis as highlighted in [17], providing effective feedback to the learners, adaptive and efficient learning recommendations and predicting student behaviour and performance were other areas of research. It can be noted that researchers had barely begun exploring social network analysis as an educational data mining technique in 2009. However, the work in [33] demonstrates that the trend is changing and social network analysis is being used on educational data to predict learning behaviour, and assist in student evaluation and student support.

6. CONCLUSION

Data mining applications within e-learning systems is a rapidly growing phenomenon. Educational data mining and learning analytics are two streams of research focussed in solving e-learning problems using data mining techniques. Among them predictive modelling has particularly interesting applications on predicting student performance and the usefulness of e-learning systems. There are currently many open research challenges that exist in this domain, particularly centred around data privacy and ethics. Deployment of data mining techniques within commercial e-learning systems requires a joint effort by the ICT specialists, educationists and the learners.

7. REFERENCES


Assessing the impact of an e-Learning activity on the skills development of accounting students

Prof. Beverley Jackling, Riccardo Natoli, Nick Sculli and Salina Siddique

Financial Education Research Unit
School of Accounting & Finance
Faculty of Business and Law
Victoria University
Melbourne Australia

Abstract

The aim of this study is to evaluate an assessment activity (group assignment) that utilised e-Learning. The paper seeks to identify whether the use of the e-Learning tool enhanced collaborative group work processes (interaction and involvement). Using quantitative methods the results indicate that students from collectivist environments consistently gained more from the use of blogs than individualist cultural groups. As technology increasingly encompasses student learning, the results of this study demonstrate the need to have a greater degree of integration of e-learning tools such as blogs across the accounting curriculum. The study concludes by providing direction for educators in incorporating e-Learning activities when developing group tasks among culturally diverse groups and a broad based approach to the integration of technology to enhance the development of graduate capabilities.

Keywords: e-Learning, cultural diversity, group work, blogs

Note: not for publication or quotation without the express permission of the first named author.

Introduction

Empirical research evidence suggests that, despite the rising popularity of e-Learning, its innovative potential is yet to be systematically exploited by many universities (Zemsky and Massy, 2004 and Aoki and Ozkul, 2007). However, one such e-Learning tool – blogs, has attracted academic attention and has become increasingly employed by many higher education institutions since it provides the opportunity for self-reflection and collaborative learning in higher education contexts (Dippold, 2009; Williams and Jacobs, 2004).

The aim of this study is to evaluate an assessment activity (group assignment) that utilised e-Learning, with a particular focus on cultural diversity of the student cohorts. Specifically, the first objective of this paper is to identify whether the use of the e-Learning tool enhanced group work processes (interaction and involvement). A second objective is to examine whether the introduction of e-Learning as part of a group learning activity is perceived more positively by collectivist cultural groups than individualistic cultural groups.

The paper is structured as follows: Section 2 reviews the literature on the use of e-Learning with particular reference to its use as a reflective tool, pedagogical issues arising from its use particularly with culturally diverse student groups. Section 3 explains the research methods employed, while the results of the study are discussed in section 4. The final section provides direction for educators in terms of enhancements to e-Learning for accounting education purposes and possible direction for future research.
Literature review

E-Learning

The role of e-Learning includes, but is not limited to, information transfer including cross-pollination of ideas among students, helping students with understanding and developing concepts, replacing part of the teacher’s responsibilities and developing student understanding through engagement with online resources (Gonzalez, 2010; Lameras et al., 2008; Ellis et al., 2006, Brown, 2000). Furthermore, the use of e-Learning is also highly recognised in responding to the needs of increasingly heterogeneous groups of students (O’Neill et al., 2004). Of the many e-Learning tools, it is blogs that has garnered significant use by many higher education institutions (Hourigan and Murray, 2010).

Blogs

A blog can be defined as a continuing, chronological publication of personal thoughts of users in their own words, online (Yang, 2009). Such publication can be aggregated over time, purposefully, in the form of an ePortfolio, to include ideas, evidence, reflections and feedback, etc. with a view to demonstrate an individual’s learning and capability to a selected group of audience (Housego and Parker, 2009; Dippold, 2009; Williams and Jacobs, 2004). Creating, presenting, managing and sharing ePortfolios requires e-Learning tools such as PebblePad, which was used for the group assignment, to allow its users to upload a wide range of online documents anywhere, any time.

Through PebblePad users can record their thoughts and reflections on learning experiences and achievements via files, photos, audios and videos in the form of “assets”.

Reflective learning and blogs

Previous studies demonstrate that blogs are effective in providing a suitable social networking interface for demonstrating effective collaboration and evaluation. For instance, Munday (2010) showed that blogs assist students to take responsibility for reflecting and documenting their own learning and to create the kind of impact they wanted to happen and to showcase it for various viewers including employers. In this way blogs have addressed concerns about the lack of collaborative learning in groups. Clark and Hardham (2010) used blogs to design a series of self-directed reflections on clinical practice which enabled many students to use reflection to continually develop and improve their own strategy to complete their assignment. Another study undertaken by Palmer et al. (2008) investigated students’ perception of appropriateness of an online journal in achieving unit learning outcomes. The results showed that blogs provided the majority of the students with the ability to compare their understanding of the course material with other students.

Pedagogical and student issues arising from blogs

While the pedagogical advantages claimed for blogs include interactivity, reflectivity and collaborative learning, it is argued that such advantages are not necessarily inherent. A UK based study (Wilson, 2010) highlighted the pedagogical issues related to applying blogs as a medium for online submission and assessment to two different cohorts of new teaching practitioners as part of their personal development planning. The study reported that: (i) designing the assessment task appropriate for online submission, (ii) providing clear instructions and links, (iii) managing the assessment in terms of starting, (iv) keeping track and shutting down
the ability to change after submission date, and (v) provision of early support to the students and the assessors are the main issues in a successful application of a blog.

**Cultural Diversity**

The influx of international undergraduate students studying accounting in Australia includes many students from a Chinese background, typified by a Confucian Heritage Culture (CHC). The CHC is linked to the dimension of individualism-collectivism which has been used in multicultural studies to explain differences between cultural groups (Hofstede, 1980). Typically non-Western countries have been characterised as being collectivist in nature, placing more emphasis on the group rather than the individual good. In an educational context this has been reflected in studies that show Hong Kong students, for example, prefer a more collaborative learning environment which they view as promoting deep learning strategies (Chan and Watkins, 1994; Lee et al., 2003; Yin, 2009). Similarly, Tang (1996) found that CHC students spontaneously collaborated in study outside the tertiary classroom more than Western students.

This study examines whether accounting students felt that the use of blogs enhanced their group work experience and particularly collaborative skills. As mentioned previously, given that a large proportion of students enrolled in accounting programs at Australian universities are from an international background (Hannon and D’Netto, 2007) this study also seeks to identify differences in perceptions of blogs between culturally diverse students, with a focus on international students from a collectivist cultural background. Consequently, the paper examines two main research questions:

**RQ 1.** Does the use of e-Learning (blogs) enhance collaborative learning in groups?

**RQ 2.** Are there any differences between students from a collectivist cultural background when compared with students from an individualistic cultural background, using blogs as a means of promoting collaborative learning?

Following the examination of these research questions, the paper assesses how the use of blogs could be enhanced by promoting the importance of the online activity. In addition, suggestions will be made to improve the operational function of blogs in developing assignment tasks with specific reference to accounting education.

The significance of this study is that it provides an insight into how technology can be incorporated to complement an assessment activity to enhance the quality of the group work experience. In addition, the collaborative design can be used as an exemplar for future planning and designing of learning technologies, instructional design and generic courseware to account for language background differences in student cohorts.

**Research approach**

**Details of the group assignment**

The assessment task that features in this study was designed for use in a second year corporate accounting unit of study in degree course in accounting, undertaken at a Victorian university. The cohort of students studying the undergraduate degree typically consists of approximately an even mix of local and international students. The assignment represented 20 per cent of the total assessment in the unit of study and
was based on the topic of asset impairment using various newspaper articles related to Australian listed companies as examples of the application of accounting standards related to impairment.

**Group formation**

The group size is a consideration to be examined when considering group work. The general consensus from the literature is that group size should remain small (e.g., Gillies, 2003). In this instance the group size of two ensured that each member of the group was committed to making a contribution via the blog entries and thus, it was anticipated the smallness of the groups would minimize the risk of free riders.

The blog formed part of the assessment regime being worth 5 per cent of the total assessment in the unit of study. Blog entries were required to be completed every two weeks for 8 weeks via group consultation. An activity sheet provided students with a structured approach to the use of the blog over the 8 week period. To introduce students to the technology, a number of resources were made available, including an introduction to the use of the blogs followed by an actual demonstration by the Web based Learning coordinator in the first lecture of the semester. The Coordinator also provided a range of resources such as an online video about the use of blogs via an online help movie titled “Using blogs”. The blog afforded students the opportunity to engage in collaborative activities within the group as well as reflect on their activities. In particular students were encouraged to develop a mind map to graphically represent elements of the problem, relationships between elements and information provided in the assignment including various media releases related to the corporation used as a case study for the assignment.

**Research design and data collection**

To evaluate the implementation of an e-Learning activity as a means of enhancing collaborative learning, a questionnaire was distributed independently from the lecturer in charge of the accounting unit, to students at the completion of the assignment task. The questionnaire was completed by 121 students undertaking the unit of study.

Students were asked via the questionnaire to provide their views on the blogs. This section of the questionnaire consisted of 7 items related to students’ group learning experience with blogs and was based on the work of Ballantine and McCourt Larres (2007). Further, the questionnaire also collected background information relating to gender; language spoken at home and country of permanent residence.

The questionnaire was completed the week the assignment was due in class (week 8 of a 12 week semester). The timing of the completion meant that the assignment was ‘fresh’ in the mind of students, and was therefore, likely that responses would accurately reflect experiences.

**Descriptive statistics**

Of the 121 responses received, 112 were usable, which represented a response rate of approximately 35.3 per cent based on the number of students who sat the final exam. In terms of the sample size, 51 students (45.5%) were male and 61 students (54.5%) were female. Slightly over half (50.9%) of the participants were international students while the remainder (49.1%) were domestic. Of the international students completing the questionnaire, over 73 per cent were of Chinese background and were identified as the sample of students from a collectivist cultural background. All other students were categorized as being from an individualistic cultural
background based on the classification of Hofstede (1980). The majority of students spoke a language other than English at home (58%) compared to 42% who spoke English at home. Table 1 below, provides details of the characteristics of the respondents.

(Insert Table 1 here)

**Results and discussion**

**Group learning experience - Blogs**

Students’ perception of their group learning experience using blogs was measured using a 5 point Likert scale where rankings ranged from ‘strongly disagree’ to ‘strongly agree’. The items are located in question 4 of the survey (items 1-7). Table 2 below reflects the perception of all students towards their group learning experience.

(Insert Table 2 here)

In addressing research question 1, from an overall perspective, the results in Table 2 show that students did not have a positive learning experience in terms of the enhancement of collaborative learning using blogs. The most successful outcome related to the blogs ability to allow students to reflect on their own involvement in the group learning process (mean =3.34) and to reflect on their contribution to the group assignment (3.33).

In both the present study and that of Ballantine and McCourt Larres (2007), marks were allocated for the group activity. However the low ranking of responses might reflect that no marks were awarded for the ongoing online dialogue in maintaining the blog. As Ballantine and McCourt Larres indicate, it is unlikely that students value educational approaches that are not subsequently assessed by the instructor. Furthermore, given the novel characteristics of the blog, there may be a case to be made for ongoing monitoring of the blogs by academics, rather than reviewing the work only when the assessment is submitted for marking. A further factor that may explain the lack of engagement with the blogs was that the percentage mark awarded for the group assessment component of the assignment (5 per cent) may have been perceived as being too low in terms of the overall unit assessment, to justify a high level of commitment to the blog.

Despite the explanations provided above, the accompanying high standard deviation figures (1.057 and 1.123 respectively) shown in Table 2 for the responses suggest that there were some major variation. In fact, large standard deviation figures (greater than 1) occurred for all seven items. These variations warrant further testing. Hence, differences in the perception of the two cultural groups were tested using the Mann-Whitney test. The Mann-Whitney test is a non-parametric test that is applicable for tests of independence between two independent sample means. It is specifically suited to nominally scaled data as collected in this study. The results of the test are shown in Table 3 below.

(Insert Table 3 here)

The results showed that the two cultural groups of students held significantly different attitudes regarding their perceived group learning experience obtained from the use of blogs. Across all 7 items Chinese students from a collectivist cultural background attained higher mean scores than all other students completing the questionnaire. The Mann-Whitney test results demonstrated that significant differences exist across five of the seven items between the attitudes of the two cultural groups of students regarding their perceived group
learning experience obtained from the use of the blog. Specifically, significant differences occurred with Chinese students perceiving that the blog was ‘useful for reflecting on group interaction’ (U=1138.000, z=-1.903, p=0.057), ‘valuable part of group assessment’ (U=1094.00, z=-2.155, p=0.031), to ‘better reflect on their involvement in the group learning process’ (U=1028.00, z=-2.598, p=0.009), to ‘better reflect on their own contribution to the group assignment’ (U=1070.500, z=-2.324, p=0.020), and the blog was ‘more able to enhance their responsibility for their own learning’ compared with students from individualistic cultural backgrounds (U=962.000, z=-3.021, p=0.003). The findings support the view that students from collectivist cultural backgrounds are more likely than students from an individualistic cultural background to have positive perceptions about group work. In this instance it would appear that engagement with e-Learning via blogs was more readily perceived as contributing to collaborative learning by Chinese students.

Although there was no significant difference regarding blogs being ‘an effective deterrent against free riders’ (U=1281.000, z=-.980, p=0.327), the Chinese student mean score was greater indicating that this cohort felt that the blog was more effective in combating the free rider issue. More importantly this result shows that the other students, being mainly domestic students, - who typically are most aggrieved with the free rider issue (see: Aggarwal and O’Brien, 2008; Pieterse and Thompson, 2010) - did not believe that the use of blogs addressed this major ongoing group work concern.

**Conclusion**

This study explored accounting students’ attitudes towards a carefully designed assessment activity (group assignment) that utilised an e-Learning tool to identify whether the use of blogs enhanced group work activity amongst culturally diverse groups. The results demonstrated that students from a collectivist cultural background consistently gained more from the use of blogs as a useful communication tool and as a basis for reflection. They also believed it enhanced their own contribution by sharing ideas and information which comprise a valuable part of the group assessment.

The findings from this study have implications for accounting educators facing a similar mix of domestic and international students. In designing group tasks, there is a need to consider not only group size, self-selection or structured formation of the groups, but also the cultural mix. Educators wanting to advance the collaborative aspects of group work may well give consideration to blending students from collectivist cultures with students from other cultural backgrounds as a means of enhancing collaborative processes.

**References**


**TABLE1** Descriptive Statistics

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>51</td>
<td>45.5%</td>
</tr>
<tr>
<td>Female</td>
<td>61</td>
<td>54.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residence</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian</td>
<td>55</td>
<td>49.1%</td>
</tr>
</tbody>
</table>
International 57 50.9%

*Country of residence* (international)

<table>
<thead>
<tr>
<th>Country</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>42</td>
<td>80.8%</td>
</tr>
<tr>
<td>Vietnam</td>
<td>4</td>
<td>7.7%</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>2</td>
<td>3.8%</td>
</tr>
<tr>
<td>India</td>
<td>1</td>
<td>1.9%</td>
</tr>
<tr>
<td>Other *</td>
<td>3</td>
<td>5.8%</td>
</tr>
</tbody>
</table>

*Main language*

<table>
<thead>
<tr>
<th>Language</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>47</td>
<td>42.0%</td>
</tr>
<tr>
<td>Non-English</td>
<td>65</td>
<td>58.0%</td>
</tr>
</tbody>
</table>

Note: n = 112; *5 students did not identify their residence

<table>
<thead>
<tr>
<th>Statement</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The blog was a useful communication device between the members of the group</td>
<td>107</td>
<td>3.00</td>
<td>1.036</td>
</tr>
<tr>
<td>The blog was useful as a means of reflecting on the interaction of the group members</td>
<td>107</td>
<td>3.23</td>
<td>1.095</td>
</tr>
<tr>
<td>The blog is an effective deterrent against free riders (i.e. people who do not make a fair contribution to the task when completing group work)</td>
<td>107</td>
<td>3.16</td>
<td>1.100</td>
</tr>
<tr>
<td>The blog is a valuable part of the group assessment</td>
<td>107</td>
<td>3.05</td>
<td>1.147</td>
</tr>
<tr>
<td>The blog enabled me to reflect on my involvement in the group learning process</td>
<td>107</td>
<td>3.34</td>
<td>1.057</td>
</tr>
<tr>
<td>The blog enabled me to reflect on my contribution to the group assignment</td>
<td>107</td>
<td>3.33</td>
<td>1.123</td>
</tr>
<tr>
<td>The blog enhanced my responsibility for my own learning</td>
<td>107</td>
<td>3.30</td>
<td>1.157</td>
</tr>
</tbody>
</table>

*Note:* Nine students did not identify themselves on the demographic section in the questionnaire (questions 7-11). Their responses were excluded from further analysis.
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>**</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>The blog is a valuable part of the group assessment</td>
<td>3.05</td>
<td>1.147</td>
<td>Individual</td>
<td>70</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>Collectivist</td>
<td>41</td>
<td>3.34**</td>
<td>0.938</td>
<td></td>
</tr>
<tr>
<td>The blog enabled me to reflect on my involvement in the group learning process</td>
<td>3.34</td>
<td>1.057</td>
<td>Individual</td>
<td>70</td>
<td>3.14</td>
</tr>
<tr>
<td></td>
<td>Collectivist</td>
<td>41</td>
<td>3.68**</td>
<td>0.934</td>
<td></td>
</tr>
<tr>
<td>The blog enabled me to reflect on my contribution to the group assignment</td>
<td>3.33</td>
<td>1.123</td>
<td>Individual</td>
<td>70</td>
<td>3.14</td>
</tr>
<tr>
<td></td>
<td>Collectivist</td>
<td>41</td>
<td>3.66**</td>
<td>0.990</td>
<td></td>
</tr>
<tr>
<td>The blog enhanced my responsibility for my own learning</td>
<td>3.30</td>
<td>1.157</td>
<td>Individual</td>
<td>70</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>Collectivist</td>
<td>41</td>
<td>3.71**</td>
<td>1.055</td>
<td></td>
</tr>
</tbody>
</table>

**Differences between groups significant at 1% level (p < 0.01)
*Differences between groups significant at 5% level (p < 0.05)

1 Nine students did not identify themselves on the demographic section in the questionnaire (questions 7-11). Their responses were excluded from further analysis.
Instruction Scheduling For Low Power Using Genetic Algorithm

Bui Ngoc Hai, Nguyen Ngoc Binh
Faculty of Information Technology
University of Engineering and Technology (UET)
Vietnam National University, Hanoi (VNU)
Hanoi, Vietnam
{haibn.mi10, nnbinh}@vnu.edu.vn

Abstract:

Optimizing power consumption is an important topic in embedded system engineering, especially for embedded systems that use battery power source. The optimized power consumption helps prolong the lifetime of the system. Instruction scheduling is an effective method for reducing power cost of processor(s). In this paper, we apply the genetic algorithm to low power instruction scheduling. The genetic algorithm is a flexible algorithm that can be applied in many different fields; it is an effective algorithm for problems which have large search spaces such as the one in scheduling problems. When designing the genetic algorithm for the scheduling problem, we use the method introduced by C. Moon et al, this is the chromosome encoding method that is suitable for the traveling salesman problem with precedence constraints. In the experiment section, we use two open source simulation tools that are SimpleScalar Tool Set and SimplePower, the algorithm is applied to assembly programs of SimpleScalar Instruction Set, these programs are compiled and then have their power consumptions measured by SimplePower. The experimental results have shown the effectiveness of our proposed method and algorithm.

Keywords – embedded systems, instruction scheduling, genetic algorithm, SimpleScalar, SimplePower

I. INTRODUCTION

In embedded system engineering, optimization is an important problem. Because embedded systems always have limited resources such as the size of memory, the speed of the processor, power supply… Optimization will make the system work more efficiently with allowed resources. In embedded system, optimizing power consumption is an important issue, especially for embedded systems using battery power source. The optimized power consumption helps prolong the life time of the system. Recently there have been many researches on optimizing the power consumption of the processor, especially the optimizing power consumption in Instruction Level – Instruction Scheduling for Low Power - This work is reordering the assembly instructions of a program so that power consumption is reduced without the program changing the semantics.

Instruction Scheduling for low power is first presented by V. Tiwari [1, 2, 3] as the instruction level power model and the idea of reordering assembly instructions. Instruction scheduling for low power aims to reduce the circuit state overhead, which is the energy dissipated due to switching from execution of one instruction to another. Instructions can be reordered to have less amount of circuit state overhead. Some scheduling algorithms for low power are presented and evaluated in [8]. In [7] C.-L. Su et al proposed a cold scheduling algorithm to reduce the switching activity of the processor. P. Dongale [6] in his master thesis,
proposed the algorithm force-directed scheduling for low power (FP-ISLP), this is the application of the classic force-directed scheduling algorithm to low power problem. This algorithm has the advantage of avoiding the local optimum. In [4], the authors presented another method by formulating the instruction scheduling problem as a traveling salesman problem (TSP), they used minimum spanning tree and simulated annealing technique for optimization. Some other scheduling methods were introduced in [11, 12, 13].

In this paper, we use a scheduling method using the genetic algorithm. Traditional genetic algorithm is not suitable for the scheduling problems, because it will be stuck to the constraints of the data flow graph, i.e. when we create a new order of instructions, we are not sure whether it satisfies the data flow graph or not. Here, we use the genetic algorithm with a chromosome encoding that solves the data dependency problems better. This method was introduced in [5], the authors proposed this method and used it to solve TSP; we use their idea to apply to the scheduling problem to reduce energy consumption.

II. PROBLEM DESCRIPTION

Our scheduling problem involves the following steps:

Divide an assembly program code into basic blocks

Construct Data Flow Graph for each basic block

Apply Scheduling algorithm for each basic block

A Data Flow Graph (DFG) is a graph which presents the data dependency of the instructions in a basic block, it is a directed acyclic graph, each instruction of the basic block is presented by a vertex, each arc present the dependency of an instruction pair. Fig. 1 shows an example of DFG.

Scheduling is similar to Topological sorting problem, from the Data Flow Graph, we have to choose an order that satisfies the constraints of the graph. Our scheduling problem is finding a topological order so that the total cost through all vertices is the smallest or smaller the original’s one, the costs between two vertices in a row here is the overhead cost between two vertices. Here, we cannot measure the overhead cost between each pair of instructions, but we can measure the energy consumption of each pair of instructions. This power includes base energy cost of each instruction and overhead cost between them. Measure the power consumption in pairs as above, we build a Power Dissipation Table (PDT), which is a matrix, where each element \((i, j)\) of the table represents the power dissipation when instruction \(i\) is followed by instruction \(j\), this table will be used instead of the overhead cost table. The scheduling algorithm takes as input a DFG for a basic block and the PDT. To create the PDT table, we do the following: each element \((i, j)\) of the table is measured by giving the instruction \(j\) after instruction \(i\), followed by the instruction \(nop\), repeated 20,000 times to avoid loop overheads. This work is described in Fig.2. Two open source simulation tools - SimpleScalar Tool Set [9] and SimplePower [10] are used. We use SimpleScalar ISA to create assembly programs and to create the table PDT, the compiler ssbig-na-sstrix-gcc of SimpleScalar are used to compile assembly programs. SimplePower is a power simulator for SimpleScalar ISA, it is used to measure the elements of the PDT and to measure the power consumption of assembly programs.
III. GENETIC ALGORITHM FOR LOW POWER SCHEDULING

In this section, we introduce the genetic algorithm (GA) approach based on topological sorting for our problem. Genetic algorithm is a search algorithm that is very effective in problems which have large search space; this is a flexible algorithm and can be applied in many different areas. The main problem when designing GA for our problem is initialization of the initial population, construction of the cross over operator, mutation operator, building fitness function. The parameters such as population size pop_size, number of generation max_gen, the probability of cross over pc and the probability of mutation pm also need to be chosen.

A. Topological Sort

The topological sort is an ordering of vertices in a directed graph, such that if there is a path from vertex vi to vertex vj, then vj appears after vi in the ordering. In the topological sorting procedure, in each step, select any vertex without incoming edges and then store the vertex and its position. Then, the vertex and all the arcs from this vertex are removed from the graph. But more than single sequence of vertices can be derived from a directed graph using this topological sorting procedure. To overcome this issue, and to obtain a feasible complete path from a directed graph, an ordering technique using the topological sort and random assignment of priority is used. To derive a unique sequence from a directed graph, a random priority assignment technique to randomly assign a different priority to each vertex in the graph is used. Therefore, a string of priorities can represent a feasible path. The topological sorting procedure with priorities assignment is shown in Fig.3.
B. Representation of chromosome

According to the topological sorting algorithm above, instead of using a sequence of vertices of the graph, we can use a string of priorities to represent chromosome, each string of priorities will represent one individual in the population. Fig.4 shows an example of chromosome representation.

C. Cross Over Operator

We use the operator called Moon Cross Over introduced in [5]. This operator generate a new chromosome from two old chromosomes. Fig.5 shows an example of Moon Cross Over operator.

D. Mutation Operator

In the mutation operation, any two genes will be randomly selected and swapped. Fig.6 shows an example of Mutation operator.

E. Fitness Function

Our fitness function is the total costs from the first vertex to the last vertex of the sequence of vertices which have been sorted. In our problem, the path between each pair of vertices is the corresponding PDT value when switching between two instructions that correspond to these vertices. Our goal is to select the sequence that has the smallest fitness function value.

<table>
<thead>
<tr>
<th>Vertex</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_3 )</th>
<th>( v_4 )</th>
<th>( v_5 )</th>
<th>( v_6 )</th>
<th>( v_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>priority</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 4. Chromosome representation

<table>
<thead>
<tr>
<th>Vertex</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_3 )</th>
<th>( v_4 )</th>
<th>( v_5 )</th>
<th>( v_6 )</th>
<th>( v_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>priority</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

Chromosome2

<table>
<thead>
<tr>
<th>Vertex</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_3 )</th>
<th>( v_4 )</th>
<th>( v_5 )</th>
<th>( v_6 )</th>
<th>( v_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>priority</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

New chromosome

<table>
<thead>
<tr>
<th>Vertex</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_3 )</th>
<th>( v_4 )</th>
<th>( v_5 )</th>
<th>( v_6 )</th>
<th>( v_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>priority</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 5. Moon cross over operator

<table>
<thead>
<tr>
<th>Vertex</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_3 )</th>
<th>( v_4 )</th>
<th>( v_5 )</th>
<th>( v_6 )</th>
<th>( v_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>priority</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vertex</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_3 )</th>
<th>( v_4 )</th>
<th>( v_5 )</th>
<th>( v_6 )</th>
<th>( v_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>priority</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 6. Mutation operator
F. Genetic Algorithm for low power scheduling

Overall Algorithm is described in Fig.7 this algorithm is performed on a DFG. Firstly, for initiation, we choose the \( \text{pop\_size}, \text{max\_gen}, \text{pc} \) and \( \text{pm} \), we create the initial population by creating the number \( \text{pop\_size} \) of chromosomes. Each chromosome is created randomly by assigning a priority number between 1 and N for each vertex, where N is the number of vertices, so that different vertices have different priorities. Then, we select the best solution by using topological sorting and fitness function for each chromosome. Secondly, for each generation, we create new chromosomes by applying \textit{moon cross over} and \textit{mutation} operators, update the best solution and select the next population. The algorithm stops when the number of generations is equal to \( \text{max\_gen} \).

\[
\text{Initiation} \\
\quad t = 0; \\
\quad \text{Set } \text{pop\_size}, \text{max\_gen}, \text{pc}, \text{pm}; \\
\quad \text{Create} \text{ randomly the initial population } P(t); \\
\quad \text{Select} \text{ the best solution by using topological sorting and fitness function for each chromosome}; \\
\text{While} (t < \text{max\_gen}) \text{ do} \begin{align*}
\quad & \text{Generate } C(t) \text{ from } P(t) \text{ by applying } \text{Moon Cross Over and Mutation}; \\
\quad & \text{Update} \text{ the best solution from } P(t) \text{ and } C(t); \\
\quad & \text{Select } P(t+1) \text{ from } P(t) \text{ and } C(t); \\
\quad & t = t+1;
\end{align*}
\text{end}\text{.}
\]

\[\text{Figure 7. Genetic algorithm for low power scheduling}\]

IV. EXPERIMENT RESULTS

In the experimental section, we use the compiler \textit{ssbig-na-sstrix-gcc} to compile an assembly program into benchmarks. From initial assembly programs, we create benchmarks and execute them on \textit{SimplePower} to measure power consumption. Then, we apply our scheduling algorithm to these programs, create benchmarks and measure with \textit{SimplePower} for comparison. We created four assembly programs manually and used a number of sample programs available of \textit{SimplePower} to test. The programs created manually is named from \textit{example1} to \textit{example4}, they have a feature that their instructions less interdependent than the available sample programs. We selected \( \text{pop\_size} = 100, \text{max\_gen} = 200, \text{pc} = 0.8, \text{pm} = 0.5 \). The experimental results are shown in Table 1, where nF is nanofarad. We note that applying our approach made a lower power consumption in all cases, up to 20% of power consumption reduction have been obtained. From the results we also can see that when the instructions are less interdependent, the algorithm perform more effectively.

V. CONCLUSION

In this paper, we applied genetic algorithm for the problem of scheduling for low power, an effective chromosome encoding has been used. Assembly programs were scheduled and measured power consumption on the power simulator \textit{SimplePower} for visual observation. The experimental results have shown that the genetic algorithm is a good approach for the problem of scheduling for low power, with large a search space and a
difficulty to find the optimal solution. However, in our paper there are some issues to be considered further. e.g. 
the algorithm moon cross over sometimes gave a new chromosome identical to the parent chromosome; another 
issue is the selection of the algorithm parameters and it also needs to be considered thoroughly. These issues are 
our future work to have a more effective and efficient method in solving a low power problem by software 
approach.

<table>
<thead>
<tr>
<th>No.</th>
<th>Benchmark</th>
<th>Power Unscheduled (mF)</th>
<th>Power Scheduled (mF)</th>
<th>Power Reduction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>example1</td>
<td>18387.7622</td>
<td>14609.6939</td>
<td>20.55</td>
</tr>
<tr>
<td>2</td>
<td>example2</td>
<td>19304.6189</td>
<td>16837.0278</td>
<td>12.78</td>
</tr>
<tr>
<td>3</td>
<td>example3</td>
<td>13951.9957</td>
<td>11574.0013</td>
<td>17.04</td>
</tr>
<tr>
<td>4</td>
<td>example4</td>
<td>15083.6704</td>
<td>11987.8005</td>
<td>20.52</td>
</tr>
<tr>
<td>5</td>
<td>Quick sort</td>
<td>2077771.2516</td>
<td>1926899.1609</td>
<td>7.26</td>
</tr>
<tr>
<td>6</td>
<td>Bubble sort</td>
<td>12365628.8339</td>
<td>11335254.7681</td>
<td>8.33</td>
</tr>
<tr>
<td>7</td>
<td>Binary search</td>
<td>11500.0162</td>
<td>11179.6200</td>
<td>2.78</td>
</tr>
<tr>
<td>8</td>
<td>Hanoi</td>
<td>6341659.6373</td>
<td>6063484.9539</td>
<td>4.38</td>
</tr>
<tr>
<td>9</td>
<td>Heap sort</td>
<td>3598279.6097</td>
<td>3427785.7324</td>
<td>4.73</td>
</tr>
<tr>
<td>10</td>
<td>Permutation</td>
<td>24001185.1045</td>
<td>22654096.0146</td>
<td>5.61</td>
</tr>
<tr>
<td>11</td>
<td>Matrix Multiply</td>
<td>110309.0015</td>
<td>103655.2787</td>
<td>6.03</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENTS

This work is partly supported by VNU Research project QGTĐ.12.20. The authors would like to thank the Vietnam’s NAFOSTED fund for their grant.

REFERENCES


Effective way of learning abstract lexical items: focusing on color in vocabulary learning

Haruko Miyakoda
Department of English
Tsuda College
Tokyo, Japan
miyakoda@tsuda.ac.jp

Kei-ichi Kaneko
Department of Computer Science
Tokyo University of Agriculture and Technology
Tokyo, Japan
k1kaneko@cc.tuat.ac.jp

Aya Sasaki
Department of English
Tsuda College
Tokyo, Japan

Abstract:

The use of ICT in the classroom has lead to a revival of interest in vocabulary learning. Since computers and mobile devices allow users to employ not only text data but also visual and sound data as well, learners are at an advantage in being able to use ICT in vocabulary learning.

There are many studies in the literature that claim, in particular, the effectiveness of visual data in enhancing vocabulary learning ([1], [2]). However, although visual data may facilitate the learning of concrete lexical items, we encounter problems when dealing with non-concrete, abstract items because it is not easy to link these items with visual data. Even if visual data do enhance vocabulary learning, it is difficult to take this factor into consideration when studying abstract words. We need to find alternative ways that may facilitate attainment of abstract words without relying too much on visual data.

Some studies have claimed the effectiveness of employing colors in enhancing memorization [3]. Based on this finding, we conducted an experiment to test the effect of colors on vocabulary learning. The result of this experiment indicated that among the three colors black, blue and red, the retention rate of words printed in red letters was highest.

Since color does seem to play a positive role in vocabulary learning, we developed an online vocabulary system that creates learning materials based on colored letters. The result of our preliminary evaluation test suggests that vocabulary learning can be enhanced further if learners are given a chance to select the color of their preference when creating the learning entities.

Keywords-abstract words: color, memory enhancement, u-learning, vocabulary
I. INTRODUCTION

With the development of computers and computer-assisted teaching programs, there has been a revival of interest in vocabulary teaching in recent years. Educational tools for studying vocabulary have been presented in the literature ([4], [5]). In addition, many researches with respect to learning materials can also be found in the literature, mostly supporting the effectiveness of visual data in facilitating the vocabulary learning process (cf. [1], [2]). However, as can easily be expected, although visual data may be effective, not all lexical items can be expressed visually. Particularly, we encounter problems with abstract lexical items, for which visual images are difficult to find and create to begin with. Therefore, no matter how effective visual data may be in vocabulary learning, we still need to pursue alternative ways of attaining vocabulary effectively without relying too much on visual data.

In this paper, we will focus on employing colors as an alternative to visual data. There are deeply rooted beliefs or superstitions in every culture concerning the effect of employing colors in studying. For example, blue is considered by some people to be the best choice for studying because this color is claimed to help reduce blood pressure, and learners can relax and better concentrate on their study. For this reason, some preparatory schools in Japan instruct their students to use blue pen when memorizing English words [6]. On the other hand, some instructors insist on using red in studying because it is claimed to have the effect of maximizing the learner’s motivation towards study, thus leading to efficient learning [7].

More “advanced” scientific research on the effectiveness of certain colors in enhancing memorization is also already widely in progress [3]. The aim of this paper is to apply the findings in the literature in developing a u-learning vocabulary system.

II. EXPERIMENT 1: TESTING THE ROLE THAT COLOR PLAYS IN VOCABULARY LEARNING

In order to test the so-called “common beliefs” related to vocabulary learning, we conducted an experiment to see whether color does indeed play a role in enhancing vocabulary attainment. Six subjects participated in the experiment. The procedures are as follows:

1) Prior to the experiment, a vocabulary test was conducted on all participants in order to distinguish the lexical items that they were familiar with from those that were unfamiliar. The items that the participants chose as “not familiar” were considered to be candidates for the experiment.

2) Based on the result in 1), the following 30 items were chosen for the experiment:

   acclaim, aggravate, barren, broil, collide, complexion, compulsory, courtesy, delinquent, deputy, dismal, dwelling, erupt, excursion, famine, folly, glare, inclination, loom, luncheon, menace, municipal, nuisance, outskirts, resentment, subside, subsidy, tedious, vicious, warrior

3) A set of flash cards were prepared for the 30 items in the following way:

   a. Stacked in the following order: 10 words printed in black letters followed by 10 in red, then 10 in blue
   b. Stacked in the following order: 10 words printed in red letters, followed by 10 in blue, then 10 in black
   c. Stacked in the following order: 10 words printed in blue letters, followed by 10 in black, then 10 in red
d. 30 words appearing repeatedly in the following order: letters printed in:

black→red→blue→black→red→blue, …etc.

e. 30 words appearing repeatedly in the following order: letters printed in:

red→blue→black→red→blue→black,…etc.

f. 30 words appearing repeatedly in the following order: letters printed in:

blue→black→red→blue→black→red,…etc.

The above mentioned six sets of flash cards were prepared in order to eliminate any effect of ordering.

4) Each subject participated in four tests in total: Pre-test 1, Post-test 1, Pre-test 2, Post-test 2. Pre-test 1 was conducted prior to the first learning session; Post-test 1 was conducted right after; Pre-test 2 was conducted a week later from the first learning session, and Post-test 2 was conducted right after the second learning session.

This can be schematized as follows:

```
Pre-test 1 ─────── Learning session 1 ─────── Post-test 1
         ↓<One week later>
Pre-test 2 ─────── Learning session 2 ─────── Post-test 2
```

Figure 1: Procedure of Experiment 1

The result of the experiment is summarized below in Figure 2.

As indicated in Figure 2, if we calculate the retention rate by comparing the results of Post-1 test with that of Pre-2 test, we find that red had the highest retention rate (52%) followed by blue (44%), and then black (32%). According to the forgetting curve by Ebbinghaus, memory retention rate after one week is expected to be 24% [8], and our test results indicate that red as well as blue drew out far better results than the predictions made by the forgetting curve. However, if we turn our attention to the result of Post-2 test, we find that no significant difference can be observed in the final result among the three colors.

Although the result of the experiment seemed to suggest that color may be effective in learning vocabulary, no significant difference was observed among the three colors in the end if we observe the result of the post-2 test. In order to get a better understanding of the role that color plays in vocabulary learning, we thought it might be worthwhile to test the learning effect of using colors that the learners themselves have chosen.
In other words, many learners have a certain preference for colors, and this preference may lead to different results. For some learners, red may lead to efficient learning, but for others, blue, black, or any other color may be the color that works best.

In order to allow learners to easily choose and employ their favorite color in vocabulary learning, a system for creating learning materials based on colored letters was implemented. The details of the system will be depicted in the next section.

III. ONLINE VOCABULARY SYSTEM “CITRON”

In this section, we report on the “Citron” system that we have designed and developed. This system supports users in creating the vocabulary learning materials for u-learning. It allows learners to easily choose and employ their favorite color in vocabulary learning. It is implemented by using Microsoft Visual Basic 2008 Express Edition. Figure 3 depicts the screen for inputting the vocabulary data, i.e., the spelling of the foreign word (1), the corresponding meaning (2), and the color used for the letters (3).

![Figure 3: The screen for inputting vocabulary data](image)

If the learner, for example, inputs the English word *broil* in ☐, then chooses the color “red” in ☐, the system automatically creates a learning entity as indicated in Figure 4:

![Figure 4: Example of learning entity](image)
If the learner prefers to display all the lexical items in black, he/she clicks the B/W button, and instantly all the colored letters change to black. The system can be terminated by clicking the Quit button.

In the following section, we report on the preliminary evaluation experiment for Citron.

IV. EXPERIMENT 2: EVALUATING “CITRON”

In order to evaluate the effect of employing colors that the learners themselves have chosen, an evaluation experiment was conducted on nine university students attending a university in Tokyo.

Three students were allotted to three groups, Groups A, B, and C, respectively. In addition, three different types of learning materials were prepared for the experiment:

1. Materials using only black letters
2. Materials using pre-fixed colored letters (i.e., the colors were chosen by the authors beforehand)
3. Materials using colored letters whose colors were chosen by the participants themselves

In this experiment, the three Groups A, B, and C memorized 30 words in total. Here again, in order to eliminate any effects of ordering, the above mentioned three types of learning materials were presented to each group in different order. That is to say, the first 10 words in the word list for Group A were black letters (Type (1) mentioned above), the next 10 words were displayed using pre-fixed colored letters (Type (2)), and finally the last 10 words were created by the participants themselves by choosing their color of preference prior to the experiment (Type (3)). The order of the materials for each group is summarized in Table 1.

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type (1)</td>
<td>Type (3)</td>
<td>Type (2)</td>
</tr>
<tr>
<td>Type (2)</td>
<td>Type (1)</td>
<td>Type (2)</td>
</tr>
<tr>
<td>Type (3)</td>
<td>Type (2)</td>
<td>Type (1)</td>
</tr>
</tbody>
</table>

Each participant was instructed to memorize the 10 words for each type in five minutes (e.g., the participants in Group A memorized the first 10 words of Type (1) in five minutes, the next 10 words of Type (2) in five minutes, and the last 10 words of Type (3) in 5 minutes, 15 minutes in total). After one week, a test was carried out to check the memory retention rates of the participants. Since no significant difference was observed among the three groups, we treat the results together.

The test results after one week are shown in Table 2, where for seven subjects, Type (3) material (i.e., learners chose the colors themselves) drew out the best memory retention rate (47%), followed by the pre-fixed Type (2) material (33%), then Type (1) in black letters (24%). The retention rate for Type (1), as you may recall, matches the prediction made by Ebbinghaus’ forgetting curve (i.e., 24%), while the one for Type (3) far exceeds it.
This result implies that employing colors in vocabulary learning has a positive effect on the memory retention rate. Furthermore, the memorization process may be facilitated if the learners are allowed to choose their favorite color for each vocabulary item.

V. CONCLUSION

In this study, we attempted to make clear the role that color plays in vocabulary learning. There are the so-called “common beliefs” related to vocabulary learning as well as research findings on the effectiveness of colors in enhancing memorization. We conducted an experiment to see whether color does indeed enhance word attainment. The result of our experiment indicated that among the three colors black, blue and red, the retention rate of words printed in red letters was highest. This result implies the positive role that color plays in vocabulary learning, however, no significant difference could be observed in the final result among the three colors.

In order to evaluate the effect of employing colors that the learners themselves have chosen, we first developed Citron, a system that supports its users in creating vocabulary materials for u-learning. One of the characteristics of this system is that it facilitates the creation process of vocabulary materials for u-learning by allowing learners to easily choose and employ their favorite color in vocabulary learning. From the result of the evaluation experiment we had conducted, we found that the memory retention rate for the materials in which the learners themselves chose the color was the highest (47%), followed by pre-fixed colored letters (33%), then black (24%). The result implies that the employing color does indeed enhance vocabulary learning, and furthermore, if the learners themselves are allowed to choose their favorite color in creating the material, this leads to better results in memory retention.

It has been claimed in the literature that visual images help enhance vocabulary learning. However, the problem of employing such multimedia materials is that the data size for each becomes huge, and in addition, heavy burden is imposed on users in creating the learning materials, since there is the need to find visual images
for each material. In order to overcome these problems, we need to consider methods that facilitate the material creation process and at the same time do not rely too much on multimedia data that may lead to complication.

One of the advantages of employing colors in creating learning entities in terms of data management is that the data size for storing the materials can be drastically reduced. Furthermore, the learners are freed from the burden of having to find visual data for each lexical item. In this sense, the vocabulary system Citron that we have developed has the potentiality to help learners attain prominent achievement in vocabulary learning.

Although the results of our preliminary experiment suggest that memorizing the meaning of words using colored letters enhances the memory retention rates, we were not able to obtain statistically significant results due to the small number of participants. There is the need to carry out evaluation experiments on a larger scale. This we leave for future research.

ACKNOWLEDGMENT

The authors are grateful to Yuichi Yuzuhara for his contribution to this study.

REFERENCES


Optimization of the Hybrid Cloud Operation by Dynamic Placement Method with MIP

Hironori Emaru\(^1,2\) and Yoshiaki Takai\(^3\)

hironori.emaru@hds.com, takai@iic.hokudai.ac.jp

\(^1\)Research & Development Division, Hitachi America, Ltd. Santa Clara, CA, United States
\(^2\)Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido, Japan
\(^3\)Information Initiative Center, Hokkaido University, Sapporo, Hokkaido, Japan

Abstract:

As economic pressure builds globally, enterprises, municipalities and universities are starting to look at hybrid cloud build by combining public cloud and private cloud as a potential choice to reduce a total cost of ownership for IT. To optimize the hybrid cloud, it is necessary to minimize operational cost. We proposed the dynamic placement method which determines deployment and migration plan of applications to minimize the operational cost of the hybrid cloud. We showed that the total cost of this method was 43.9% lower than the conventional method in 5 years even if timing and frequency of the deployment and the characteristic of applications are unknown in advance. However, generally speaking, the time variability of compute utilization varies more widely than the storage utilization. As a result, dynamic placement method has a challenge that compute utilization is lower than storage utilization in the private cloud. In this paper, we propose a dynamic placement method with MIP which combines our dynamic placement method with conventional mixed integer program method. This method migrates applications among clouds using MIP when the difference between storage utilization and compute utilization expands in the private cloud. We show the effectiveness of the proposed method by evaluating in a simulation environment.

Keywords- hybrid cloud; dynamic placement method with MIP; optimization

I. INTRODUCTION

Cloud computing is one of the today's most promising technologies. Cloud computing is a model for enabling on-demand network access to a shared computing resource pool that can be rapidly provisioned and unprovisioned [1]. In addition to public cloud and private cloud, hybrid cloud has been built recently by combining public clouds and private clouds. The deployment of multiple applications on hybrid cloud is expected to become common in the future.

For these dynamic environments, there is no particular method to use private cloud and public cloud properly. Therefore, adaptivity for the dynamic behavior of applications is required to deploy applications on the hybrid cloud efficiently.

We proposed the dynamic placement method (hereinafter referred to as “DPM”) which determines deployment and migration plan of applications to minimize the operational cost of the hybrid cloud. We showed that total cost of this method was 43.9% lower than the conventional method in 5 years even if timing and frequency of the deployment and the characteristic of applications are unknown in advance. However, generally speaking, time variability of compute utilization varies more widely than the storage utilization. As a result, the challenge for DPM is that compute utilization is lower than storage utilization in the private cloud. In this paper, we propose a dynamic placement method with mixed integer program (hereinafter referred to as “DPM with
MIP”) which combines DPM with conventional mixed integer program method. This method migrates applications among clouds using MIP when the difference between storage utilization and compute utilization become expanding in the private cloud. We show the effectiveness of the proposed method by evaluating in a simulation environment.

II. BACKGROUND

A. Related Work

There are several related works on application deployment in the hybrid cloud. Strebel and Stage [2] developed an economic decision model that compared costs in the private cloud and the public cloud. They formulated this problem as a mixed integer program (MIP) and evaluated it using a linear regression model. Mazhelis, et al. [3] developed a cost model of the hybrid cloud. Their model assumes that private cloud consumes maintenance cost with or without use whereas public cloud is a pay-as-you-go model, where unit price of the public cloud is expensive than the private cloud, and workload of the cloud is volatile. Based on these assumptions, they advocated that hybrid cloud was less costly than private-only or public-only cloud.

MIP method was formulated as follows according to [2]: Maximum compute and storage utilization of the day for each application is supposed to be known at the beginning of the day. Applications are reallocated every day to utilize the resources of the private cloud close to 100% by solving MIP. This method is expected to bring the cost without migration to the lowest.

B. Dynamic Placement Method

The objective of the MIP method is to reach 100% resource utilization of the private cloud. This method supposes that future workload and deployment timing of applications is known, and/or migrates applications every day based on the workload of applications. However, the cost of the MIP method may not be the lowest because migration consumes compute cost and network cost in addition to the operational cost. Therefore, we proposed the DPM which has following two policies:

As the first policy to minimize the operational cost of the hybrid cloud, we consider keeping the utilization of the private cloud at the predefined range. This range is expressed by the target ratio the $T_c$ (%), the lower limit threshold $T_b$ (%) and upper limit threshold $T_u$ (%). Here, the target ratio $T_c$ satisfies $T_b < T_c < T_u$. Whereas public cloud is charged on a pay-as-you-go basis, private cloud requires CAPEX in addition to OPEX regardless of use. Generally, OPEX of the private cloud is less than the amount of billing of the public cloud for the same amount of resources as described in [3]. Therefore, the higher the utilization of the private cloud, the lower the total operational cost of the hybrid cloud. On the other hand, if the target utilization of the private cloud is 100%, shortage of resources in the private cloud may result. It is necessary to avoid this situation because it causes failures such as unexpected stop of the applications or deployment failure of the new application.

As the second policy, we consider to combine deployment with migration. Some methods of selecting the target of deployment have been proposed in the past. However, there is no method which utilizes migration to minimize the total operational cost. Because existing methods such as MIP method assume that the amount of communication and workload is known in advance, there is no adaptivity for the dynamic behavior of applications.
Based on the above policies, an application is deployed on the private cloud if utilization of the private cloud is less than $T_u$, and an application is deployed on the public cloud if utilization of the private cloud is more than $T_c$. When the utilization of the private cloud exceeds $T_u$, migration from private cloud to public cloud is executed. An evaluation value is calculated for each application in private cloud, then selects applications with higher evaluation value as a target of migration. On the other hand, when the utilization of the private cloud falls below the lower limit threshold $T_b$, migration from public cloud to private cloud is executed. An evaluation value is calculated for each application in public cloud, then selects applications with higher evaluation value as a target of migration.

III. DYNAMIC PLACEMENT METHOD WITH MIP

DPM tries migration from public cloud to private cloud when the storage or compute utilization falls below the lower limit threshold $T_b$. However, sufficient migration can not be executed if the other utilization is nearly $T_u$. For example, applications can not be migrated when the storage utilization exceeds the upper threshold $T_u$, even if the compute utilization falls below the lower limit threshold $T_b$. Because the DPM does not execute aggressive migrations, when the above situation occurs once, the resource utilization keeps low for a while.

We consider the combination of DPM with MIP method to solve the above problem. That is, applications are reallocated using MIP method when the storage or compute utilization falls below the lower limit threshold $T_b$ and the other utilization exceeds $T_u$. MIP method used here aims to reach upper limit threshold $T_u$ as opposed to the conventional MIP method that aims reach to 100% resource utilization of private cloud. This prevents the shortage of the resources in private cloud. Here, how to determine the lower limit threshold $T_b$ is important. If this value is too low, almost none of the migration is performed. On the other hand, if this value is too high, frequent migration causes cost increase. We will examine the suitable value of $T_b$ by simulation.

IV. EXPERIMENTS AND CONSIDERATIONS

A. Condition of the Simulation

We prepare 10 patterns of application traces based on [4], [5] and [6]. The amount of compute and storage capacity of the private cloud is set at the time of a start of simulation and is not expanded during the simulation. These parameters of the public clouds are set to unlimited. The simulation period is set to 5 years.

The hybrid cloud is supposed to be charged based on maximum compute and storage usage for each application during a single day. That means the minimum billing period is one day. When we start the simulation, the compute utilization of the private cloud was set to 70% and all public clouds are unused.
B. Experiment 1

In this section, we compare MIP, DPM and proposed method. We prepare 5 patterns of the proposed method by the difference of the lower limit threshold $T_b$.

Figure Error! Main Document Only. shows the 5 year cost of each method. Upper part of the stacked bar chart shows the cost of the migration and the lower part shows the other cost. The numbers of the DPM with MIP (40% - 80%) shows the lower limit threshold $T_b$. This graph shows that cost of the migration of the MIP is especially high. The cost of the proposal method is lower than the DPM when the lower limit threshold $T_b$ is set to 50%-70%. Among them, 50% is the lowest cost. In this case, migration cost is 41.2% and total cost is 4.5% lower than the DPM.

Figure Error! Main Document Only. shows the analysis of the migration. Stacked bar chart (left vertical axis) shows the total data amount of migration in 5 years. The upper part of the stacked bar chart shows the amount of migration from private cloud to public cloud and the lower part shows the opposite. The line chart (right vertical axis) shows the number of MIPs. The solid line shows the number of MIPs which is executed when
the storage utilization falls below the lower limit threshold $T_b$ and the compute utilization exceeds the upper limit threshold $T_u$. The dotted line shows the number of MIPs which is executed when the compute utilization falls below the lower limit threshold $T_b$ and the storage utilization exceeds the upper limit threshold $T_u$. MIP method executes migration every day; therefore, the migration is executed 1825 times in 5 years although this is not shown in Figure Error! Main Document Only.. The higher the lower limit threshold $T_b$, the more frequent the number of MIP. However, there are no correlation between number of MIP and amount of migration. That is, the number of MIP is the lowest when the $T_b$ is 40% whereas amount of migration is the least when the $T_b$ is 50%. If the lower limit threshold $T_b$ is too low, it is not efficient because amount of migration of each MIP is huge. On the other hand, if the lower limit threshold $T_b$ is too high, it is not efficient because too much migration is executed. However, even if the lower limit threshold $T_b$ is set to 80%, amount of migration is only 17.3% of the MIP method as shown in Figure Error! Main Document Only..

Figure Error! Main Document Only. shows the efficiency of the resource utilization in the private cloud. In this simulation, nearly 3,000 applications are deployed in 5 years. 100% stacked bar chart (left vertical axis) shows the deployment ratio of applications. That is, this graph shows which cloud is selected when each application is deployed. The upper part of the chart shows the ratio of applications which are deployed on the public cloud, and the lower part shows the opposite. By using DPM, sometimes applications cannot be deployed on the private cloud even if the private cloud has unused resources. That is, if storage or compute utilization exceed the upper limit threshold $T_u$, applications cannot be deployed even if the other resource has enough capacity. Storage utilization and compute utilization become equalized by using MIP together with DPM. The higher the lower limit threshold $T_b$, the more frequently the number of equalizations. Therefore, more applications are deployed on the private cloud. The line chart (right vertical axis) shows the resource utilization. The solid line shows the storage utilization and dotted line shows the compute utilization. The higher the lower limit threshold $T_b$, the smaller the difference between storage utilization and compute utilization, but the values themselves are 75%-78%. These values are smaller than 95.8% of the MIP method.

C. Experiment 2

In this section, a different deployment pattern of applications from experiment 1 is applied. That is, the ratio of the storage’s heavy workload is increased compared with experiment 1.

Figure Error! Main Document Only. shows the 5 year cost of each method. Structure of this graph is same as Figure Error! Main Document Only.. The total cost is the lowest when the lower limit threshold $T_b$ is set to 60% of DPM with MIP method. This threshold is different from experiment 1. In this case, migration cost is 62.4% and total cost is 13.7% lower than the DPM.

Figure Error! Main Document Only. shows the analysis of the migration. Structure of this graph is same Figure Error! Main Document Only.. In this experiment, demand for the storage resource is larger than experiment 1. Therefore, we predict that storage utilization becomes higher than compute utilization and compute utilization falls below the lower limit threshold $T_b$ frequently. In practice, storage utilization falls below the lower limit threshold $T_b$ frequently as we supposed.
Figure Error! Main Document Only. shows the efficiency of the resource utilization in the private cloud. Structure of this graph is same as Figure Error! Main Document Only. The higher the lower limit threshold $T_b$, the smaller the difference between storage utilization and compute utilization. However, resource utilization is almost same even if the lower limit threshold $T_b$ is set to more than 50%. Therefore, we think optimal value for $T_b$ is between 50% and 60% for most cases.

Figure Error! Main Document Only. Comparison of the cost for storage heavy workload.

Figure Error! Main Document Only. Efficiency of the resource utilization in the private cloud for storage heavy workload.

Figure Error! Main Document Only. Amount of migration and number of MIP for storage heavy workload.

Figure Error! Main Document Only. shows the resource utilization of the private cloud and cumulative cost in case of lower limit threshold $T_b$ is set to 60%. The solid line and the dotted line (left vertical axis) show storage utilization ratio and compute utilization ratio in private cloud. The chain line and the two-dot chain line (right vertical axis) show cumulative total cost and migration cost. Compute utilization falls below 60% temporarily. But average utilization ratio of the compute utilization is 75.5%. DPM with MIP method improves resource utilization and reduces the total cost.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a DPM with MIP method which combines our dynamic placement method and conventional mixed integer program method. This method determines the deployment and migration plan of applications based on the DPM on a constant basis, and migrates applications among clouds using MIP when the
difference between storage utilization and compute utilization expands in the private cloud. We show the effectiveness of the proposed method based on our evaluations in a simulation environment. As a result, total cost of the DPM with MIP method is 4.5%-13.7% lower than DPM.

We have two challenges: First one is how to determine the lower limit threshold $T_b$. In this paper, we select the value of the lower limit threshold $T_b$ heuristically. However, the setting of the lower limit threshold $T_b$ depends on the characteristics or deployment pattern of the applications. Second one is improvement of MIP method. In this paper, MIP method does not use historical information. By utilizing historical information, we can calculate a variance of compute amount or a growth rate of storage capacity, which may result in lower cost. We will implement proposed method in the actual environment when we resolve these challenges.

REFERENCES


Secure Hybrid MDC Thin Client Concept in Capgemini BPO

a novel security concept within BPO environment

Gary Wu
IT Innovation CoE
Capgemini BPO
Nanhai, China
Gary.wu@capgemini.com

Abstract:

Key challenges in Business Process Outsourcing IT are the minimization of the risks of customer data loss and to provide a technical basis for a global, highly reliable, continuous and flexible BPO service delivery. In order to cope with those challenges, the IT innovation CoE of Capgemini BPO has developed the novel “Model Delivery Center” (MDC) thin client concept and already successfully rolled it out as a new IT standard to a significant part of its employees in BPO service delivery globally (China, India, Europe and Latin America). The MDC thin client concept is a “hybrid” thin client approach (meaning application functionality from various sources can be consolidated on a single use desktop) which allows that data of the various stakeholders (Capgemini BPO as a service provider, customers etc) stay completely separated. Data especially from Capgemini BPO’s customers in never persistently stored on Capgemini BPO end user IT equipment, minimizing the risk for data loss. The concept meets the ISO 27001 Information Security Management System requirements. From the technology perspective, combines cost efficient “commercial off the shelf” (COTS) hardware, open source software for the lightweight software load, SaaS for the Capgemini internal application functionality and state of the art remote application access technologies for the applications of Capgemini BPO’s customers. The concept also helps to reduce the complexity of Capgemini IT architecture by globally centralizing security monitoring and alerting; reducing effort on IT audits by standardizing set up in all delivery locations; Improved IT recovery time objective in a disaster recovery situation by no local data; and able to continue service delivery from alternative locations for BCP purpose. This article focuses on the aspect security of the MDC thin client concept.

Keywords--Security; Hybrid; Thin Client Concept; Google Apps; Cloud

Introduction

When conventional fat client PCs with locally (or conventional network share) stored data (can comprise information from communication or document processing applications etc.) from Capgemini and its customers
are used for BPO service delivery, certain risks for data loss do exist, even when professional measures are taken to prevent it.

Furthermore, when applying regionally different approaches for IT user equipment in BPO service delivery (Capgemini BPO services are provided from North and South America, Europe, Asia and Australia), the efforts for reaching certifications like ISO/IEC 27001:2005 (received in October 2008) are higher compared to a situation with global standards.

In this context, Capgemini BPO IT Innovation Center of Excellence (CoE) has developed and started to deploy its Model Delivery Center (MDC) thin client concept as a global IT user equipment standard for its more than 10,000 employees. Cloud based services are integral components of the concept.

The “hybrid” (application functionality from various sources can be consolidated on a single use desktop) MDC thin client concept addresses the before mentioned aspects while increasing the flexibility for Business Continuity Planning (BCP) by enabling the capability for securely “working from alternate location” (WFAL). Despite the significant improvements, the MDC thin client concept does not increase the IT cost.

This paper focuses on how the MDC thin client concept minimizes the risk for loss of customer and Capgemini BPO data by cloud solutions.

Reducing risks for customer data

The approach to further minimize the risks for customer data is separating customer data from Capgemini BPO internal data and ensure that neither Capgemini BPO IT user devices nor delivery center network storage comprises customer data persistently. Customer data remains on the customer side and under customer management and is securely accessed by Capgemini BPO employees through the customer remote business applications but not copied or moved. In this way, also efforts on the customer side to prepare and conduct audits are significantly reduced as the BPO service provider does not intensively need to be involved or checked for that.

Fig 1 describes the how customer and internal application functionality is accessed in the case of the MDC thin client concept being applied. It is worth noting that the amount of applications locally installed on the thin client is reduced to a minimum, as it can be expected. Client software for accessing remote applications (web browser, Citrix receiver, VMView client, RDP client) is present besides connectivity related software.
Customer applications (ERP, Email, MS Office, Instant Message, Internet Browser, etc.) are securely accessed through technologies like Citrix XenApp, VMView, RDP or web services. The published applications (we say customer cloud) allow the Capgemini employees to deliver their services without creating copies of customer data persistently outside the regime of the customer. This approach allows the customer to maintain and apply his own controls on the data and continue to apply his standards on managing his data in customer cloud.

A. Secure connectivity between Capgemini BPO data centers and customer data centers

Connectivity between Capgemini BPO data centers and customer data centers is using Multiprotocol Label Switching (MPLS) or leased line to make sure data is secured and transferred between Capgemini BPO and Clients. For the customer’s data hosted in Capgemini BPO data centers, it can be referred to Fig #2 to see an overview of the network structure of a sample Capgemini BPO data center. Within the scope of this network structure, customer’s data are safely stored on Capgemini Data Center and customer owns their data.

B. Summary of key hardening measures in the thin client concept’s context

1. Application software installed on the thin client devices reduced to the minimum, business application functionality to be received through the respective software clients.
2. No persistent storage of customer and internal data on Capgemini BPO user equipment and the delivery centers
3. Mounting of external storage media at the thin client devices technically avoided
4. Customer application and data are in the cloud

This leads to the following advantages:
1. Less complex network configuration for both Capgemini and clients
2. Risk of export of data through external storage devices minimized.
3. Prevention of data losses due to damage or stolen hardware
4. Business continuity and data recovery plans in place to ensure that service can be maintained in case of a disaster or an emergency and that any data loss will be recovered

C. Robust Software load

The software load/image component of the MDC thin client concept is based on the Ubuntu Linux technology with the above mentioned security hardening measures implemented. Ubuntu Linux is well known for its stability, good performance, network friendliness, flexibility, compatibility, fast and easy installation, multitasking, security and more reliability. Compared to other open source Linux, Ubuntu has an enterprise grade support.

D. Flexible BCP and WFAL

The approach of thin client concept uses COTS hardware (current standard is Dell Latitude 2120 Netbook) which has great mobility. Working together with Capgemini VPN solution--Cisco ASA with IPSEC VPN technology applied. Capgemini employees can take the netbook with them and access Capgemini network through this VPN solution, which make the users working like in the office. This approach can increase the BCP (Business continuity planning) flexibility and option. And also provide a working from alternate location (WFAL) solution. Since no persistent data will be saved in thin client PC and end users will work in virtual desktop/application, no matter the netbook broken or stolen, customer’s data will not loss and remind confidential. At this point, security hardening is applied.

Reducing risks for Capgemini BPO data

Many companies like Capgemini BPO want to protect data and prevent data from data corruption and unauthorized access. The second part of this paper describes how this hybrid thin client concept helps Capgemini BPO to secure data and improve the audibility of the data for security purpose by cloud services.

A. Data Security on thin Client image component

- In order to prevent data leak from USB ports, MDC thin client image component locks the USB ports by software control. When users connect USB storage devices to MDC thin client, the USB storage devices will not be recognized. Except USB storage devices, others USB devices are functional.
- In order to prevent data leakage from broken or lost hardware and prevent virus for Capgemini daily working data, MDC thin client concept forces user to save data on Google Cloud or Capgemini file servers. With low
privilege accounts, MDC image users are able to store data & change system setting only for session time. After users log out, all locally saved data is erased. Therefore, users will have to save their data or edit their data in cloud web services (Google Cloud/Apps).

- MDC thin client concept uses a cloud basis centralized monitoring tool--landscape. 1. Keep the MDC thin client’s security patches up to date. The software center and the patches are controlled by Capgemini BPO by pointing the patch sources to Capgemini repository instead of public sources. Administrators will update the latest security patches to MDC PCs on a regular basis. 2. System Administrators will receive emails if any of these monitored devices have abnormal actions including like additional software installed, virus, etc. 3. Easier to support the end user.

B. Data Security on Capgemini emails & Documents

To prevent accidental or malicious disclosure, modification, or destruction of records, data sets, and program segments is an important part of security function in every company. MDC thin client concept cloud component uses Google Cloud/Apps to achieve this purpose. Google Apps includes dozens of critical security features specifically designed to keep data safe, secure and in Capgemini control. Capgemini data belongs to Capgemini users and Apps tools enable users to control it, including the way users are sharing with other defined. Google data center network provides exceptional security and guarantees reliable access to the data (24x7x365.25). Capgemini users login Capgemini enterprise Google email by their Corp username/password in Capgemini delivery centers through SSO technology. While they are not in the delivery centers, users have to use a token generator to login Capgemini enterprise email. The token generator will be provided to users after the security manager’s approval.

Google Cloud/Apps has applied the following security strengthening measures:

1. Works are always backed up: while users are working, all critical data is automatically backed up on Google servers.

2. Users own and control their data, which can be managed and keep track of the usage and data via Google dashboards.

3. Google increase security and reliability by building applications and not including unnecessary hardware or software. In this way it can guarantee 99.9% uptime and build-in robust disaster recovery.

4. Strong encryption and authentication: Google Apps offers an extra layer of security with two factor authentication, which greatly reduces the risk of hackers stealing usernames and passwords. Google Apps also automatically encrypt browser sessions with SSL for Apps users without the need for VPNs or other costly, cumbersome infrastructure.

Here are several Google Apps that are important for Capgemini BPO.
● Google Email: The fact that Google Email has 25GB of storage means users never have to delete anything, which can keep the data safely stored. Auto-backup feature will keep the email in the draft if computer or browser crashes. Features like two-step authentication, attachment viewing in the browser, encrypted connections to Google’s servers, simultaneous replicated storage for your email, built-in disaster recovery, spam filtering and combined with Capgemini SSO authentication are working in for a great security way.

● Google Documents: Google Docs is a suite of products that enable Capgemini users to create different kinds of online documents, which can replace Microsoft office. It also has the auto-backup feature which can back up the documents users are working on the moment user lost connection for any reason. It also keeps revision records, which allow users to view at a glance all changes made to a document by each collaborator or himself. While it may not work exactly like a track changes tool, Google Docs revision history enables user to view and revert to earlier versions of his doc, and this can help users to reduce mistakes.

● Google Vault: Google Apps Vault is a solution for managing information critical to business and preserving important data. It allows personnel involvement in investigations and audits by searching, preserving, exporting and auditing company email/data. Vault also enhances management and information administrative capabilities by proactively archiving, retaining and preserving emails and on-the-record chats. Vault gives management, IT, legal and compliance users a systemized, repeatable and defensible platform that reduces the risks of doing business.

Other advantage of MDC Thin Client Concept

A. Easy to support

With this Hybrid thin client image component, IT supports can easily support end users. Not only IT supports can keep off installing/supporting customer’s business applications, but also benefited from simply support on MDC thin client software load. With the advantage of no local data is stored and all applications are in the cloud, whenever the end user has any issue(software or hardware), IT support just need to re-image their PC or replace with another one. This can reduce the IT work load and increase the efficiency.

B. Reduce the Auditing effort

All customer data is in customer cloud which can reduce the involvement of Capgemini during customer’s audit. While Capgemini Data is in Capgemini BPO cloud (or Google cloud, which data is fully controled by Capgemini), auditing becomes easier and simpler.

Summary of Capgemini BPO hybrid thin client

Unlike traditional desktop PC solution, hybrid MDC thin client concept, which makes application functionality from various sources can be consolidated on a single use desktop, is global standard IT solution of Capgemini BPO. This approach reduces data loss by totally separating Capgemini data from customer data by using customer cloud; uses reliable, robust, Capgemini customized, open source software image to reduce the data leak for customer and Capgemini data; protects Capgemini daily working data by using web services (Google cloud); uses enterprise level COTS mobile hardware to increase the flexibility of BCP and WFAL; provides easy IT support and cost efficiency solution (without increasing IT cost). For these reasons, the hybrid...
MDC thin client concept becomes Capgemini BPO standard IT Solution and is deployed across the delivery centre network.

References

5. https://support.google.com

An online Peer-Tutoring Platform for Programming Languages based on Learning Achievement and Teaching Skill

Yu-Chen Kuo
Department of Computer Science and Information Management
Soochow University
Taipei, Taiwan
yckuo@csim.scu.edu.tw

Zhe-Yu Wu
Department of Computer Science and Information Management
Soochow University
Taipei, Taiwan
zheyu.w@gmail.com

Abstract:

The research is to establish an online Peer-Tutoring Platform for Programming Languages and delve into the one-to-one peer-tutoring activities, while we conduct the learning strategy about the peer education. Students with better learning ability serve as tutors, instructing those with worse learning ability, called tutees. Tutors proceed with the peer-tutoring activities of the online programming language with respectively different teaching skills. The aim of that is to enable the students to learn the programming language and ameliorate their program designing ability. Furthermore, we delve into how their peers are affected in learning, when the students as tutors utilize on different teaching skills to carry the learning activities out.

Keywords- peer-tutoring; programming language; learning achievement; teaching skill

I. INTRODUCTION
With the rapid development of technology and the universal internet, both computer-assisted instruction and web-based collaborative learning have been commonly applied to the school education[1]. Recently, several scholars replaced the traditional classroom learning activities with the web-based learning strategy. The internet is used to process the learning activities, which enables teachers to efficiently manage students’ learning behaviors and learning portfolios[2, 3]. Besides, although the web-based collaborative learning can increase students’ motivation for learning, some scholars point out that the learning activity process is usually more complex in the process[4]. The reason for that is students with inferior learning ability usually lengthen their collaborative time[5, 6]. Hence, the issue that we need to understand is how to draw on the effective learning strategy for assisting the on-line learning activities like that.

The peer-tutoring is a structural learning strategy. It mainly encourages children’s interactive learning. Some scholars propose that the peer-tutoring on the internet can prompt students to learn the mother tongue, modern languages, and the attitude toward modern languages[7]. In addition, it can enhance students’ ability for designing the programming language[8].

In college, programming language is an important core capability for students whose major is information-related. However, it is a difficult subject for many students. Kelleher and Pausch pointed out that students usually face complicated languages and instruction codes in thinking about how to solve the programming language problems and understand the executive process of the program[9]. These abstract structures cause students to have the idea that learning the programming language is a hard and complex task. Many studies show that at school, students who select the subject of programming language are diminishing and more and more students have less passion and interest in programming language, which cause them to give up on it[10]. In traditional teaching about programming language, teachers always lecture on the platform. Therefore, there is only one-way communication between teachers and students. Students do not know how to ask and teachers can not clearly find out students’ problems. This one-way teaching reduces students’ learning desire. With a view to this, more and more scholars offer research related to computer-assisted instruction. It aims to help students learn programming language. Since 1980, the systems for computer-assisted instruction have been gradually developed, such as WebToTeach[11]. In 1995, Intelligent Tutoring Systems (ITS) which possess the strength of web, such as ELM-ART, started to appear[12, 13]. The aforementioned systems are mainly to aid teachers. They help teachers with course arrangement, exams and assist students individually with virtual teaching assistants. There are limited effects about communication, flexibility, and learning results in the face of virtual TAs[14, 15]. In 2008, scholars proposed a programming language teaching platform which supports peer-tutoring[16]. Students do not learn by virtual TAs anymore; instead, it is the peers that give instruction as TA. Nonetheless, because a TA faces more than one student, the learning results are rather limited.

The research is to establish an online Peer-Tutoring Platform for Programming Languages. The learning strategy for peer-tutoring is used to process one-to-one peer-tutoring activities. Students with superior learning ability act as tutors. They instruct students, tutees, with inferior learning ability. They are provided with pre-activity training. By this, tutors can enhance their programming ability. Then we delve into how their peers are affected in learning, when tutors utilize on different teaching skills to carry the learning activities out.

II. LITERATURE REVIEW
A. Programming Language Learning Systems

The computer-assisted programming language learning systems have been discussed by many scholars. Much research is conducted on automatic virtual TAs. These systems only provide students with a good learning environment and strengthen the learning effects of students with formalized feedback. Arnow and Barshay proposed a programming language learning system, WebToTeach[11]. WebToTeach, a web-based teaching environment, supports many programming languages and diverse training question patterns in order to allow students to acquire more information. Brusilovsky proposed LISP of ELM-ART, which provides virtual TAs, so that students can obtain more help by virtual TAs[12]. Besides, the system can arrange the difficulties of chapters for letting students fit in quickly, based on students’ capability. However, the feedback these systems can provide is limited can the feedback that students can get is still the formalized answer which teachers devised beforehand. Students are only able to get relatively limited help. Chien proposed the programming language teaching platform which supports peer learning[16]. This reverses the traditional views about learning programming languages. Students serve as TAs through the online learning performance. In literature, TAs need to encounter several students. Nonetheless, they have not had any pre-learning training, which resulted in insufficient teaching capability.

B. Peer-Tutoring

Peers have been playing a significant role in students’ career. Students can have a mutual learning and encouragement by learning with peers at their age and with the similar background. This behavior can not only increase tutees’ motivation and confidence for learning, but decrease their learning pressure and fortify the learning effect. Moreover, the tutor’s social skill and mind development will be improved. Peer-Tutoring is a teaching system where peers help and discuss with each other and share their own experience[17]. It mainly highlights the tutee learn by accepting the education of the tutor at his age and with the similar thinking pattern. On the other hand, the tutor can learn in two ways. One is the preparation and training before teaching; the other, the feedback of teaching tutees. Frick indicated that the peer-tutoring is a teaching strategy that students teach others of similar age and with inferior learning ability one on one[18]. This strategy applies to every age group, capability level and learning field. Fantuzzo, Riggio, Connelly and Dimeff thought that peers’ instructing and cooperating with each other can help them solve confusion on studies and have a shared experience[19].

From the above review, we see that most scholars tend to explore on the impacts on students who use learning systems, or to introduce the effectiveness of peer-tutoring. Not many scholars discuss about whether the tutor is capable on teaching or not during peer-tutoring. Moreover, whether there is any discrepant between the teaching technique of the tutor and the effects of peer-tutoring activity is still to be discussed.

III. RESEARCH METHODS

A. System Structure

This research is to set up an online Peer-Tutoring Platform for Programming Languages. By applying the strategy of peer-tutoring, we divide the students into tutor and tutee according to their capabilities, and adopt one-to-one peer-tutoring activity. During the allowed time for answering, the tutee writes program codes and they are sent synchronously to the tutor through the internet. When the tutee faces certain problems, he may ask...
help from the tutor through the online chat-room. On the other hand, the tutor notices about the mistakes of the tutee immediately through the screen. The tutor then gives guidance to the tutee, and he also may provide some help through the learning platform.

B. Design of experiments

Before the experiment begins, we first conduct a pre-test and then divide the students into groups according to their capabilities. Following the strategy of peer-tutoring, we put student with higher capability and student with lower capability into one group. Therefore, this study will group the students into Experimental Group A and Experimental Group B. For Experimental Group A, the student who acts as the tutor is going to perform Peer-tutoring activity, through conducting an online program language by using the teaching skill of demonstration method (Fig. 2). And for Experimental Group B, the tutor is going to perform peer-tutoring activity, through conducting an online program language by using the teaching technique of the flow chart method (Fig. 4). The procedure of the experiment is: students will first spend 30 minutes on pre-test. The aim of such pre-test is to know whether the familiarity level on programming language of each student varies greatly or not. The pre-test is also used to distinguish tutors and tutees. Before conducting the online peer-tutoring, the tutors in each group will receive pre-training classes for twice, 30 minutes each time. This is to reinforce the programming and tracing code capability of the tutors. Next, after classes every week, Experimental Group A and Experimental Group B use the online Peer-Tutoring Platform for Programming Languages, and two problems on programming languages will be asked for peer-tutoring activity, lasting for 5 weeks. After the designed peer-tutoring activity, there will be a 30-minute post-test. The aim of the post-test is to know that after conducting the online peer-tutoring activity, whether the students’ capability of programming language has improved or not. Lastly, in order to find out the satisfaction of the online system, 15 minutes will be spent on conducting a related survey. Then we can analyze the learning activity situations and effects of the students.

![Figure 1: experimental group A_tutee](image1)

![Figure 2: experimental group A_tutor](image2)
C. Experimental Subjects

The experimental subjects of this study are the sophomores of Class A and Class B studying in the Department of Information Management. For convenience, classes will be seen as units, and are divided into Experimental Group A and Experimental Group B. The problems used for testing the students are designed by experts; aim to test on the common programming language topics, especially on the main concept of recursive programming language. To most students, recursive programming language is a lot more difficult than others like if-else, do-while and the For Loop. Therefore through this study, we aim to strengthen students’ concept and familiarity towards recursive program.

IV. CONCLUSION AND FUTURE WORK

The subject of programming language includes complicated program syntax and logic concepts, and these characteristics will definitely reduce the students’ willingness to learn. Therefore, this study changes traditional learning methods. In after-school practices, students now can practice and learn through peer-tutoring, rather than practicing alone. When students face problems, their peers can help them to solve problems. The learning result and willingness will be increased by using this method.

During the development process of the platform, we will invite experts to carry out tests on the accuracy, availability and feasibility of the system. We estimate that the learning results and learning willingness of the students will be improved, through the peer-tutoring activity conducted on the online Peer-Tutoring Platform for Programming Languages. This aims to make students learn and teach at the same time.

The platform developed in this study is now used in the Data Structure courses in the university. Experimental subjects includes totally 106 students from two classes, and they are divided into Experimental Group A and Experimental Group B according to their classes. The experimental questions in this study mainly come from the concept of recursive language in the programming language. Pre-test are conducted before the experiment to ensure the two groups both have a considerable degree of knowledge. Before the learning activity, the tutor will undergo some pre-training in order to reinforce their ability of programming and tracing code. In Experimental Group A, the teaching skill used in peer-tutoring by the tutor is demonstration; and for
Experimental Group B, flow charts are used during teaching. After the learning activities, post-test and survey will be carried out.

REFERENCES


An Open MPI-based Cloud Computing Service Architecture

WEI-MIN JENG and HSIEH-CHE TSAI

Department of Computer Science Information Management
Soochow University
Taipei, Taiwan
{wjeng, 00356001}@csim.scu.edu.tw

Abstract:

Cloud computing is generally accepted as the workable solution for computation intensive problems such as big data analysis. Among the major building blocks for the underlying infrastructure, MPI (Message Passing Interface) is one of the most effective solutions. This study aims to propose a Java-based service architecture using Open MPI for cloud computing operations. Web-based interfaces are provided for further inquiries of the operation results. Additional interfaces are supported for computational needs as well for users who are not familiar with the language details. Simple protocol is available for Java developers to transfer their job for further processing. Professional MPI library function users can integrate their codes with Java programs and therefore get to improve the way they manage their MPI programs more flexibly.

Keywords- Open MPI; JAVA; Spring Framework; cloud computing


I. INTRODUCTION

With the rapidly growing Internet and Information Technology, cloud computing [1, 2] became a reality for a variety of applications. It refers to the technology of using remote high-capacity computing resources via the ubiquitous Internet services intended for high-performance utility type of computing. With or without the use of cloud computing, the popular use of Internet and its applications have already become part of our lives. The amount of information computers handle increase tremendously in an enormous way, issues such as big data analytics [3] have been brought to our attentions. Local computers in general are not designed for effective processing, therefore, there is an immediate need for a cloud computing architecture in order to provide available computing resources.

Parallel and distributed computing refers to the computational technique by dividing the computational problem into smaller problems whereas each of the tasks may be performed on separate computing resources. The intermediate results then are put together and return as the final solution to the calling application. Both MPI and Hadoop [4] fall into this computational technique category and have been heavily used recently. MPI works with major languages to perform the computation and communication. The major drawback is that it has to be customized by experienced developers in order to arrange all the collective operations. The original MPI library version does not facilitate either Internet computing or Web programming interfaces and thus open source Java-based Open MPI [5] implementation paves the way for the development of many diversified applications. Hadoop has better support from Java; however, its application coding styles are limited due to the limitations from MapReduce operation.

This study proposes an Open MPI-based cloud computing service architecture, using Java to develop the web-based application. Java-based application programming interfaces can be harnessed to save the user from using the script or manual instructions. The submission of user-defined MPI programs can be conducted over Web interfaces or server end directly. Network socket mechanisms are exploited to implement the required communication services for independent service modules. Multiple-user services are provided by using concurrent threads, thread pool of resources are used in the server for better management. In order to monitor the execution status of MPI programs, Java-based Web interface implementation is provided to retrieve the information after the identification is validated. JSP and Servlet are the two major implementations for Java Web application development, whereas JSP codes are transformed into Servlet ones immediately after its first invocation. Program framework is frequently used for developing Java programs, Spring Framework [6,7] is employed in this study as the Web page development framework. Lightweight Container is used in its kernel to realize IoC (Inversion of Control) to complete the setting in the configuration file rather than using the in-line setting. Program components with loose coupling are therefore testable with this container approach. Application programs done this way further are benefited from great deal simplification in both component and maintenance.

Section one briefs the research motivation and goals and introduces the programming framework of this study. Section two compares different proposed design approaches, specifically MPI and other cloud computing implementations. Section three addresses the modes of use along with the protocols categorized by the types of prospective users. Section four introduces the hardware platform used by this study. Section five
II. LITERATURE SURVEY

A. Open MPI

Open MPI is an open source library implementation complying with the MPI-2 standard. It is composed of three major parts [8] as shown in Figure 1. MCA (MPI Component Architecture), a major part of MPI, provides the fundamental operation mechanism for other parts of the MPI architecture. Component Framework is responsible for the management of all modules in which functional programs and connectors are contained. MAC relays the requests to the underlying modules upon receiving the operation needs. Figure 2 illustrates the way how an Open MPI code works with a Java program Ring.java [9] to conduct the operation in a Ring topological architecture.

![Figure 1. Open MPI Component](image)

Open MPI programs check the availability of all the participating computing nodes of the clustered architecture for its operational status. Problematic node, if any, is displayed in the case of connection failure. The developed programs were interpreted into class files first and the command line instruction is entered as `mpiexec -host master, slave1, slave2 java Ring`, where master, slave1, and slave2 are the computing nodes used in this study and will be detailed explained in Section four.
B. Runtime.exec()

Runtime.exec() is a built-in method of the java.lang.Runtime class used for invoking an external program. A simple example of this method use is shown in Figure 3 for the Linux Runtime.exec() operation. The above program runs javac command by using Runtime.exec() method. BufferStream is used to store either the output or the erroneous information and provides the execution status after its completion. This study makes use of the Java method as the means to invoke the MPI programs.

C. Thread Pool
Thread is not reusable and its creation is a resource demanding process. The performance of a socket server can be degraded by a great deal with repetitive thread creation requests. Thread pool [10] pre-creates the thread and supplies the thread upon request. This study utilizes the design of thread pool to manage the system resources for MPI program operations.

D. Mybatis Framework

The pitfall of embedding the SQL codes in the programs to access to and from the database is the recompilation problem when the codes have to be changed. The Mybatis Framework[6] approach takes on the problem by separating the SQL codes from the host languages. XML (Extensible Markup Language) is used to write the SQL codes and mapping is done to the host language in the later phase. The database programming is done in this fashion to facilitate the access of user profile information, MPI function execution history, and web page information.

E. Eucalyptus and Amazon EC2

Frequently used cloud computing platforms are a remotely accessible personal computer node. One of the most famous examples, Amazon EC2, offers a fee-based remote virtual host service using VPS (Virtual Private Server) in which MPI works as its underlying computational architectural building blocks. The management of the cloud computing underlying architecture becomes an important issue with different settings. Adapt to work with a Linux environment, Eucalyptus [11] is designed to provide the cloud computing platform management with Amazon EC2 interfaces. Important components include Cloud Controller, Cluster Controller, and Node Controller to organize its architecture with adequate management web interfaces. Different from the computational platforms such as Amazon EC2 and Eucalyptus, this study aims to propose a cloud computing service platform with great deals of flexibility of use. Via the help of Java method invocation, customized MPI library function services can be supported for both professional MPI developers and less skilled users.

III. DEFINITION OF DIFFERENT USERS

In addition to the system administrator, three types of users are classified according to their MPI proficiency levels, they are general users, advanced users, and MPI developer users. The brief definitions are as follows.

A. General users

General users are the people who directly use the available MPI codes with the help of Web interfaces. They are able to upload the data for computation and conduct the MPI code invocation.

B. Advanced users

The general program developers who make use of the proposed MPI Connect method invocation wrapped using Java JAR and retrieve the execution status on demand. This approach alleviate the local computing node from the heavy use of local resources by submitting the data for remote processing.

C. MPI developer users
The users who are capable of developing MPI codes and further accommodate their codes into the existing MPI services.

**D. System administrator**

The administration staffs who are responsible for manage the platform users, execution history, and system configuration setup information.

**IV. HARDWARE ARCHITECTURE**

The experimental setup consists of the cluster of three computers running Ubuntu Linux operating system with Open MPI installed used for the underlying cloud computing hardware component. Figure 4 illustrate the architectural view of the setup. The Master computer node is equipped with two network cards for their corresponding inbound and outbound purposes. It takes the MPI invocation requests and invokes the MPI codes by Open MPI interfaces. The associated database software MySQL and the Web server software are both installed on the server. Slave1 and Slave2 are excluded from the outside world and they work as the computing nodes for providing services.

![Architectural view of the setup](image)

**V. SOFTWARE SYSTEM COMPONENTS**

Figure 5 shows the conceptual view of the Software system components. The system is composed of three major components, Web component, SQL component, and MPI component. Each of the three components can be running on a separate machine if necessary. The three are installed and run on the same machine in our experiment. The fourth component, MPI Connect, runs on different machine to provide the connection service for advanced users and the details are explained in Section three.
A. Web Component

This major Web component consists of two Web services, Spring Framework and Mybatis Framework-based programming component, and independently installed MPI Client component to establish the connection with the MPI Server.

1) Web USER

Different user/administration interfaces are provided according to the service targets. The users are able to use the four functions and services after the successful log-in process. The functions are MPI code invocation, uploading the data for MPI code execution, execution results and status inquiry, and History of use inquiry.

2) Web MANAGER

The functions and services dedicated for the system administrator after successful logon to the administrator interface include User profile management, MPI code execution history management, and MPI Server configuration setting.

3) Spring and Mybatis

Web page programming is done collectively by Controller, Manager, and Dao of Spring Framework. Controller handles the data flow in between the Web interfaces and invokes the corresponding programs of the Manager. MPI client will be activated if the input is a request for MPI code invocation. MPI Server configuration setting is done in the same way. SQL codes benefit from the integration of Dao design and Mybatis Framework.

4) MPI Client

Specific method, designed by this study, is used by Spring’s Manager and sets up the connection session with MPI Server using Socket Client design. The time to terminate its session is determined by the return values.
B. SQL Component

MySQL is chosen as the database management software and the archived information include user profile (user account, password, privilege) and MPI code execution history.

C. MPI Component

It consists of MPI Server, MPI Dispatch Module, Open MPI, and Record System.

1) MPI Server

MPI Server connects with MPI Client by using the proposed Socket Server mechanism. The thread design underlying the Socket Server is based on the Thread Pool notion. The Thread Pool parameter setting is stored independently in XML format for easy of modification. When the system administrator requests a change of setting by Web interface, the corresponding configuration setting child program within the MPI Server is invoked to make the change. Both program execution and data file requests are done in the same manner.

2) MPI Dispatch Module

In order to work with the characteristic of the system for a String-only data transmission, Data Conversion Engine is provided for the format transformation. Figure 6 displays the step-by-step conversion engine for getting the resulting String format. The input data is formatted first by converting into the Generic Object type, Object to Byte conversion is performed as the second operation, and a Byte-to-String conversion is conducted as the last operation. Runtime.exec() can then be used for data transmission. The invoked MPI codes reverses the process and the original data format can be recovered for computation. The same MPI Dispatch Module is used again once the MPI code completes its operation.

![Figure 6. Data Conversion Engine.](image)

3) Open MPI

This constitutes the core of our study, it includes all the Open MPI computation by means of the MPI Dispatch Module invocation. With the help of Ring program, the available computing resources will be compiled into a list for user’s selection. All the MPI code execution status will be recorded and therefore the associated costs can be estimated.

4) Record System

The status, output message, and execution results will be stored in the archive for later retrieval. The archived message is purged periodically in order to keep the history data size manageable.
VI. CONCLUSION AND FUTURE WORK

MPI has been used extensively in high performance computing applications which may involve huge amount of data in nature. To complement its language features, host languages such as Java is commonly used to provide a comprehensive solution for many applications. Lacking of the language constructs for different kinds of users, it has not been successful to find a turn-key solution to combine the two. This study proposes a comprehensive cloud computing architecture by using the improved Java-Open MPI invocation mechanism to different users. Web interface services are offered to enable the execution of MPI codes over the Internet. Due to the network transmission speed limitation, the communication time may be too long and the chances for possible network failure may result. The fault tolerance and disaster recovery issues can be investigated and hopefully provide more stable services in the near future. However, the initial result of our proposed architecture sheds the light for a viable cloud computing architecture.

REFERENCES


Text Hidden in Picture Using Steganography: Algorithms and Implications for Phase Embedding and Extraction Time

Anwar H. Ibrahim, Waleed M. Ibrahim
Department of Electrical and Electronics Engineering
School of Engineering & Technology Infrastructure
Infrastructure University Kuala Lumpur
anwar@iukl.edu.my

Abstract:

Companies, institutions and military often have a necessity to communicate a highly sensitive message and always carry the risk of capture text by unauthorized parties. Utilization of text message hiding in the picture using stegabography appears to be an alternative of predicting the limited text and capacity during embedding and extraction. For comparative analysis this paper proposed amendments to the existing techniques structure, using Robet, Canny, Sobel, Perwitt and Log algorithms. It exhaustively observes these techniques by source coding using Matlab for small text messages and makes recommendations of such features. After analyzing the results Canny method shows the better phase embedding and extraction in all the text given. It also proposed security enhancements and presents a novel form of stegangraphy.

Keywords-component; Steganography, Robet, Canny, Sobel, Perwitt and Log algorithms, text message

I. INTRODUCTION

Steganography is the procedure of hiding private information within any picture or media for incidence of communication (Thiyagarajan et al. 2012). Steganography is frequently confused with cryptography since the two are related in the way that both are used to keep confidential material (Esra & Hakan 2012). The difference between Steganography and cryptography is in the appearance in the managed output; the output of steganography operation is not speciously visible but in cryptography the output is scrambled so that it can draw attention (Johnson et al. 2001). Steganalysis is processed to detect of the presence of steganography. In this method the secret message is embedded into the image. The three parameters of the hidden process of the test namely phase embedding time, phase extraction time and offset from the original text are varied to represent encoded secret text messages. All steganographic methods concept is to achieve the minimal amount of time for text embedding and extraction for the execution to reduce the probability of error. Consequently, for the image was initially saved in the JPEG format, the progress of message embedding will not exclude the characteristic procedure produced by the JPEG compression and it can simply determine whether a known image has been stored as JPEG in the phase embedding. Essentially, the images used for testing is too small; and it can consistently recover the values of the JPEG execution time and capacity available from the image. However, the image will not be changed and look like the original after embedding, in the sense that it should be potential to prove the particular technique for coefficient determination.

II. STEGANOGRAPHY TECHNIQUES

There have been several approaches is essential and assessing the security of a steganographic scheme. Zollner et al (1998) is one of the first author to introduce the security aspect using steganographical systems.
Steganography technique is being developed to secure the protected text from the third party without any doubt is called steganography. Generally, the word steganography is formed from the Greek words which mean “Covered Writing” (Mehdi et al 2004). Consequently, the people also have the hidden information from a combination of approaches and variations. Hiding data in image can be prepared by a number of techniques such as Robert, Log, Sobel, Canny, and Prewitt. The most common procedure between all these algorithms is maintained low bit embedding, when the appropriately describes steganography as the performance of encoding data into the least significant bit in the image (Martin & Ira 2006).

III. INFORMATION HIDING IN JPEG IMAGES

JPEG images are commonly used for the text steganography in most of the current research. The JPEG image format customized a discrete cosine transform (DCT) to transform successive 8×8-pixel blocks of the image into 64 DCT coefficients each. The leastsignificant bits of the quantized DCT coefficients are used as redundant bits into which the hidden message is embedded. The modification of a single DCT coefficient affects all 64 image pixels (Andreas & Pfitzmann 1999). Recently, many different steganography techniques proposed a method to hide text in JPEG image. A number of techniques are classified steganographic system is the best way to secure short message in the image. Some of the methods were categorized them according to the kind of covers used for concealing secret communication between sender and receiver. The classification of the steganography in JPEG image identified according to the cover adjustment applied at the edge embedding development. Images are the most popular cover objects used for steganography. In the field of digital images several different image file formats are existing, and most of them used for specific applications. In this research a JPEG image file format, is used for steganographic algorithms.

IV. PHASE EMBEDDING

Originally it was believed that text message embedding using steganography would not be possible to use with JPEG images, meanwhile they use lossy compression which results in parts of the image data being changed. One of the main characteristics of steganography is the fact that information is hidden in the redundant bits of an object and since redundant bits are left out when using JPEG it was feared that the hidden message would be destroyed. Even if one could somehow keep the message intact it would be difficult to embed the message without the changes being noticeable because of the harsh compression applied. A method of phase embedding based on the difficulty of message data with different techniques is presented. The theory behind this method is analyzed was shown that the techniques can be used to hide text in images. The details of the procedures used for carrier proposed were detailed in the result. However, the properties of the compression algorithm have been exploited in order to develop a steganographic algorithm for JPEGs. Steganographic technique based on lina image is proposed to apply in this system. The secret message is hidden in the lina image as shown in Figure 1.

![Figure 1. Embedded text in image](image-url)
V. **PHASE EXTRACTION**

One of these properties of JPEG is exploited to make the changes to the image invisible to the human eye. During the DCT transformation phase of the compression algorithm, rounding errors occur in the coefficient data that are not noticeable (Johnson, N.F. & Jajodia 1998). Although this property is what classifies the algorithm as being demolished, this property can also be used to hide messages. On the receiver side a sequential change detection technique is depending on the selection algorithm which called Cumulative Sum. The stega-image of sequential change by selecting techniques from the algorithm options and also the change detection setback by choosing an image encoded in the stega-text of press load image button in the extraction phase. Selective algorithms are developed to observe the changes in the parameter of concern. After the training of extraction phase was considered, the whole system based on Stega extraction techniques and a classification step to be chosen for certain results. This feature summarization is extremely dependent on the image itself, so it depends on image format and hence pre-embedding giving out and investigate sightings of a technique should be carefully described. The high dependence between stega-techniques by the sender and images used in testing results can be explained by the following considerations. The user can be able to save the extracted text in notepad format. Furthermore, the smaller amount of text is embedded into the image, the smaller the prospect of introducing obvious object by the embedding and extracting process. An additional important factor is the removal timer of the text from the image. The sender should avoid using high capacity images that would be easy to analyze for incidence of secret messages. Figure 2 shows the original image of Lina after the extracting procedure.

![Figure 2. The original image after extracting the text](image)

VI. **RESULTS**

The embedding phase has been designed with all these considerations in mind. More specifically, all techniques have different ways to match the execution process for data hidden. Table 1 shows the comparative study between six presented algorithms, in terms of the total capacity characters and time-consuming to precede the embedding messages.

**TABLE 1 COMPARATIVE STUDY OF STEGA ALGORITHMS USING DIFFERENT IMAGE FORMAT**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Total Capacity char (TIFF)</th>
<th>Total Capacity char (PENG)</th>
<th>Total Capacity char (JPEG)</th>
<th>Embedding Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canny</td>
<td>1226</td>
<td>974</td>
<td>1200</td>
<td>1.3878</td>
</tr>
<tr>
<td>Sobel</td>
<td>599</td>
<td>575</td>
<td>663</td>
<td>0.5768</td>
</tr>
</tbody>
</table>
Prewitt 591 592 349 0.5883
Robert 560 683 506 0.5165
Log 1016 956 895 0.8262

The result in table 1 found that, Canny technique is the best method to obtain the text embedding due to more embedding characters (1226) in the image, comparing with others. The cases of the stega-analysis of the same techniques used by the sender for phase embedding and capacity embedded for phase embedding. It was found that Canny technique is the best algorithm in terms of capacity for phase embedding. However, the most time-consuming involved in this technique was a more than the other techniques.

Techniques and methods are provided for embedding English text data into an image was considered into details by using different methods in the embedding and extraction procedure. A distributed techniques feature of the extracted text in a particular time is calculated and compared with a set of predefined techniques corresponding to the Figure and tables to be encoded. From Figure 2 it was found that; the best technique in terms of time-consuming for phase embedding was Demirel technique when using Lina image (JPEG) format.

Figure 2. Embedding time condition using phase embedding techniques for JPEG format

Table 2 shows the cases of the stega-analysis of the various techniques used by the sender for phase embedding and different techniques used by the receiver for phase extraction.

<table>
<thead>
<tr>
<th>Extraction case</th>
<th>Receiver</th>
<th>Total Capacity char</th>
<th>Extraction Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>Canny</td>
<td>1226</td>
<td>1.9711</td>
</tr>
<tr>
<td>Case 2</td>
<td>Sobel</td>
<td>663</td>
<td>1.2612</td>
</tr>
<tr>
<td>Case 3</td>
<td>Prewitt</td>
<td>595</td>
<td>1.2955</td>
</tr>
<tr>
<td>Case 4</td>
<td>Robert</td>
<td>536</td>
<td>1.3121</td>
</tr>
<tr>
<td>Case 5</td>
<td>Log</td>
<td>895</td>
<td>1.4222</td>
</tr>
<tr>
<td>Case 6</td>
<td>Demirel</td>
<td>500</td>
<td>1.0231</td>
</tr>
</tbody>
</table>

Figure 3 shows the detail analysis in different cases when the extraction time was considered phase extraction in JPEG format.
VII. CONCLUSION

Briefly, the system exposes an embedding algorithm to various types of the method processing such as total capacity characteristic, image format compression, Embedding time, extraction time and other procedure that a robust text embedding technique would be required to hold the process. Figures and tables indicated in the results we conclude that under the stability and reasonable time constraint for embedding and extraction, the Canny estimation algorithm performs well in JPEG format. However the time spending to proceed the embedding and extraction phase is long compared with Sobel, Perwitt and Log. Furthermore, the JPEG image format allowed the system to embed more data in the edge comparing with TIFF and PENG. Although only some of the main image steganographic techniques were discussed in this paper, one can see that there exists a large selection of approaches to hiding information in images by using canny. All the major image file formats have different methods of hiding messages, with different strong and weak points respectively. Least significant bit (LSB) in JPEG format makes up for this, but all approaches result in suspicious files that increase the probability of detection when the presence of a warden.

REFERENCES
