PROCEEDINGS

DMS 2007

The 13th International Conference on Distributed Multimedia Systems

Sponsored by

Knowledge Systems Institute Graduate School, USA

Technical Program

September 6 - 8, 2007
Hotel Sofitel, San Francisco Bay, California, USA

Organized by

Knowledge Systems Institute Graduate School

ISBN 1-891706-21-7 (paper)

Additional Copies can be ordered from:
Knowledge Systems Institute Graduate School
3420 Main Street
Skokie, IL 60076, USA
Tel:+1-847-679-3135
Fax:+1-847-679-3166
Email:office@ksi.edu
http://www.ksi.edu

Proceedings preparation, editing and printing are sponsored by
Knowledge Systems Institute Graduate School and the University of Salerno, Italy

Printed by Knowledge Systems Institute Graduate School

Foreword

It is with great pleasure that we welcome you to the 13th International Conference on Distributed Multimedia Systems DMS 2007.

Multimedia Computing is permeating an increasing number of application fields. While in the beginning it would draw the attention of few specific fields, such as entertainment and arts, the need to process multimedia information has arisen in more and more types of computer systems and applications, such as ambient intelligence, security, environmental control, distance education, and so on, yielding many new research issues. Thus, in the last two decades, research in multimedia computing has considerably enlarged its scope, facing many new problems than merely those related to the encoding of multimedia information. In particular, the research community has enlarged its attention on issues related to the efficient distribution of multimedia information over networks, the management of multimedia objects through DBMSs extended with capabilities to enable indexing and content based retrieval, and on software engineering issues in the development of multimedia information systems. These activities have grown in parallel with the internet boom, which has itself boosted the exploitation of distributed multimedia computing, contributing to further research developments in this area. This event seeks to promote research exchanges that will benefit both the academic and industrial communities around the world. To support latest trends, in the current edition we aimed to bring together scientists and students working in main of the above mentioned research fields and applications of Multimedia Computing to share the latest advances in these areas. We have received 70 submissions from 16 countries, regarding most of the above mentioned application fields and research areas, witnessing an increased interest for applications of multimedia in ambient intelligence and environment management, for which we organized a dedicated session.

We are extremely grateful to the many colleagues who helped in organising the conference and supported the review process. In particular, we would like to thank the sponsors of this conference, Knowledge Systems Institute, USA and the University of Salerno, Italy.

Taieb Znati
DMS'07 PC Chair

Giuseppe Polese
DMS' 07 PC Co-chair

Mohan Kankanhalli
DMS' 07 PC Co-chair

The 13th International Conference on Distributed Multimedia Systems (DMS 2007)

September 6-8, 2007
Hotel Sofitel, San Frnacisco Bay, California, USA

Organizers & Committees

Conference Co-Chairs

Shi-Kuo Chang, *University of Pittsburgh, USA*
W. Grosky, *University of Michigan, USA*

Program Chair

Taieb Znati, *University of Pittsburgh, USA*

Program Co-Chairs

Mohan Kankanhalli, *National University of Singapore, Singapore*
Giuseppe Polese, *University of Salerno, Italy*

Program Committee

Frederic Andres, *National Institute of Informatics, Japan*
Elisa Bertino, *Purdue University, USA*
Ing-Ray Chen, *Virginia Tech (VPI&SU), USA*
Shu-Ching Chen, *Florida International University, USA*
Chabane Djeraba, *University of Sciences and Technologies of Lille, France*
David H. C. Du, *Univ. of Minnesota, USA*
Jean-Luc Dugelay, *Institut EURECOM, France*
Larbi Esmahi, *National Research Council of Canada, Canada*
Farshad Fotouhi, *Wayne State University, USA*

Moncef Gabbouj, *Tampere University of Technology, Finland*
Forouzan Golshani, *Wright State University, USA*
Masahito Hirakawa, *Shimane University, Japan*
Joemon Jose, *University of Glasgow, UK*
Michael Lew, *Leiden University, Netherlands*
Fuhua Lin, *Athabasca University, Canada*
Jonathan Liu, *University of Florida, USA*
Chien-Tsai Liu, *Taipei Medical College, Taiwan*
Brahim Medjahed, *University of Michigan, USA*
Vincent Oria, *New Jersey Institute of Technology, USA*
Sethuraman Panchanathan, *Arizona State Univ., USA*
Syed M. Rahman, *Minnesota State University, USA*
Ishwar Sethi, *Oakland University, USA*
Timothy K. Shih, *Tamkang University, Taiwan*
Peter Stanchev, *Kettering University, USA*
Hari Sundaram, *Arizona State University, USA*
Genny Tortora, *University of Salerno, Italy*
Joseph E. Urban, *Arizona State University, USA*
Athina Vakali
Son T. Vuong, *Univ. of British Columbia, Canada*
Ellen Walker, *Hiram College, USA*
Kang Zhang, *The University of Texas at Dallas, USA*

Proceedings Cover Design

Gabriel Smith, *Knowledge Systems Institute Graduate School, USA*

Conference Secretariat

Judy Pan, *Chair, Knowledge Systems Institute Graduate School, USA*
Beverly Crockett, *Knowledge Systems Institute Graduate School, USA*
David Cohen, *Knowledge Systems Institute Graduate School, USA*
C. C. Huang, *Knowledge Systems Institute Graduate School, USA*
Daniel Li, *Knowledge Systems Institute Graduate School, USA*

International Workshop on Distance Education Technologies (DET 2007)

September 6-8, 2007
Hotel Sofitel, San Frnacisco Bay, California, USA

Organizers & Committees

Program Chairs

Paolo Maresca, *University Federico II, Napoli, Italy*

Program Co-Chairs

Yasir Khalifa, *University of Pittsburgh, USA*
Xin Li, *Google Inc., USA*

Program Committee

Saud Aldarmi, *Kuwait University, Kuwait*
Alfio Andronico, *University of Siena, Italy*
Giovanni Adorni, *University of Genova, Italy*
Aziz Barber, *American University College of Science & Technology, Lebanon*
Jinwei Cao, *University of Delaware, USA*
Heng-Shuen Chen, *National Taiwan University, Taiwan*
Luigi Colazzo, *University di Trento, Italy*
Rita Francese, *University of Salerno, Italy*
Kwei-Liang Ho, *Taipei Municipal University of Education, Taiwan*
Kao-Shing Hwang, *National Chung Cheng University, Taiwan*
Pedro Isaias, *Portuguese Open University, Portugal*
Kusai Kayali, *University of Aleppo, Syria*
Souhail Khawatmi, *University of Aleppo, Syria*
Ammar Kheirbek, *Damascus University, Syria*
Gabriele Meiselwitz, *Towson University, USA*
Atif Memon, *University of Maryland, USA*
Giuseppe Scanniello, *University of Salerno, Italy*

Georgi Stojanov, *American University in Paris, France*
Yu-Huei Su, *National Hsin Chu University of Education, Taiwan*
Mahbubur R. Syed, *Minnesota State University, USA*
Gang Tan, *Boston College, USA*
Goran Trajkovski, *Towson University, U.S.A*
Huahui Wu, *Google Inc., USA*
Kui Xie, *Mississippi State University, USA*
Ling-Ling Yang, *Taipei Medical University, Taiwan*
Changhe Yuan, *Mississippi State University, USA*

International Workshop on Visual Languages and Computing (VLC 2007)

September 6-8, 2007
Hotel Sofitel, San Frnacisco Bay, California, USA

Organizers & Committees

Workshop Co-Chairs

Gennaro Costagliola, *University of Salerno, Italy*
Giuliana Vitiello, *University of Salerno, Italy*
Monica Sebillo, *University of Salerno, Italy*

Program Chair

Kang Zhang, *University of Texas at Dallas, USA*

Program Co-Chair

Paul Fishwick, *University of Florida, USA*

Program Committee

Alfonso F. Cardenas, *University of California, USA*
Maria Francesca Costabile, *University of Bari, Italy*
Philip Cox, *Dalhousie University, Canada*
Stephan Diehl, *University of Trier, Germany*
Jing Dong, *The University of Texas at Dallas, USA*
Filomena Ferruci, *University of Salerno, Italy*
Maolin Huang, *University of Technology, Sydney, Autralia*
Jun Kong, *North Dokota State University, USA*
Zenon Kulpa, *Institute of Fundamental Technological Research, Poland*
Robert Laurini, *University of Lyon, France*
Kim Marriott, *Monash University, Australia*

Rym Mili, *University of Texas at Dallas, USA*
Piero Mussio, *University of Milan, Italy*
Marc Najork, *Microsoft, USA*
Luca Paolino, *University of Salerno, Italy*
Minho Park, *Stephen F. Austin State University, Texas, USA*
Joseph J. Pfeiffer, *New Mexico State University, USA*
Hyunju Shim, *Samsung Corporation, Seoul, South Korea*
Nenad Stankovic, *University of Aizu, Japan*
David Stotts, *University of North Carolina, USA*
Noam Tractinsky, *Ben-Gurion University of the Negev in Beer-Sheva, Israel*
Thanos Vasilakos, *University of Western Macedonia, Greece*

Table of Contents

DMS-3 Distributed Multimedia Systems/Computing I

DMS-4 Multimedia Software Engineering

DMS-5 Multimedia for Ambient and Environment Management

DMS-6 Distributed Multimedia Computing II

DET Workshop

DET-1 Eclipse & Teaching: Case studies and best practice

DET-2 Educational Technologies for Distance Learning

DET-3 Intelligent System for Education Management

VLC Workshop

VLC-1 Diagram Analysis and Usability

VLC-2 Application and Interface Design

VLC-3 Image Processing and Visual Environments

Note: (S) means short paper.

Keynote
Future Technical Challenges in Distributed Multimedia Systems

Eric W. Hendricks, PhD
Deputy Chief Technology Officer and
Head of Science and Technology
SPAWAR Systems Center, San Diego

Plenary Talk
The Future of Visual and Iconic Language Applications and Inscription Technology: A New Way to Chat

LorRaine Tauchi Duffy

Senior Scientist
SPAWAR Systems Center, San Diego

Chat and instant messaging have become a way of life for the armed forces, just as it has for business executives, professors, teenagers, and "soccer-moms." It is particularly efficient on frail, discontinuous networks in hostile environments. The US Navy's Tactical Situation Assessment (SA) Technology (TSAT) project will attempt to develop a more proficient language for chat-based SA updates, incorporating icons, symbols and gestures that can be transmitted electronically via a gesture-recognition data glove that serves as a next-generation keyboard. The research focuses on leveraging Web 2.0 phenomenon, such as instant messaging via visual symbols ('leet speak), embedded within a context that evolves, dependent upon interfaces that are mobile and evolving (a desktop is no longer the only option). The specific intent is to improve the ability of Chem-bio personnel, as well as fire fighters, lunar astronauts, divers, and other teams to more quickly and more efficiently communicate their situational understanding across a net-centric environment, despite typical obstacles that include heavy chem-bio resistant gear, intense noise, need for covertness, little ambient light, and with jitter in the network links. TSAT focuses on three areas to improve the ability to communicate SA context and improve distributed understanding gained through communication. We intend to improve what team members say, how they say it, and the way they convey their message in a rapidly evolving environment. Discussion will also cover the intense multi-disciplinary nature of research and development projects mandated by revolutionary conceptual development.

Plenary Talk
Abstract Art vs. Information Visualization

Kang Zhang

Department of Computer Science
The University of Texas at Dallas
Richardson, Texas 75083-0688

In this talk, we will study some of the theories and practices in abstract art and their existing and potential applications to information visualization. We discuss the three dimensions of painting, i.e. form, color, and texture, various visual cognition principles, and finally aesthetic compositions used in abstract painting but applicable to information visualization. Our objective is to bridge visual arts with information visualization, so that the latter could learn from the former in creating more aesthetic visualizations and thus making the viewer's visualizing process a pleasant experience.

Proceedings
International Conference on
Distributed Multimedia Systems
(DMS 2007)

Co-Editors

Taieb Znati, *University of Pittsburgh, USA*
Mohan Kankanhalli, *National University of Singapore, Singapore*
Giuseppe Polese, *University of Salerno, Italy*

Assessing and Comparing Blended and Pure Learning Approaches: a Controlled Experiment

Andrea De Lucia[1], Carmine Gravino[1], Giuseppe Scanniello[2], Genoveffa Tortora[1]
[1] *Dipartimento di Matematica e Informatica, University of Salerno,*
Via Ponte Don Melillo, 84084 Fisciano (SA), ITALY
E-mail: {adelucia, gravino, tortora}@unisa.it

[2] *Dipartimento di Matematica e Informatica, University of Basilicata, Viale Dell'Ateneo,*
Macchia Romana, 85100, Potenza, ITALY
E-mail: giuseppe.scanniello@unibas.it

Abstract

Demand for e-learning is being driven by a need of flexibility according to the knowledge improvement and learning performance. To identify the learning paradigm that better address these issues, systematic and quantitative evaluations in terms of surveys, controlled experiments, and case studies should be performed. In this paper we present the results of a controlled experiment aimed at assessing and comparing blended and pure learning approaches. The data analysis of the experiment revealed that the pure approach does not significantly affect the knowledge acquisition of the subjects.

1. Introduction

The widely adopted learning paradigms are purely computer based, or also a combination of traditional and computer based learning, i.e., blended learning. In the pure learning no face-to-face tutor contact is furnished. The benefit of using one approach rather than the other is not straightforward and can depend on the context. Thus, systematic and quantitative evaluations (i.e., survey, case study, and controlled experiment) should be adopted to assess and compare blended and pure learning approaches in terms of acquired knowledge and learning performance. Among the evaluation techniques, controlled experiments are recognized to be an important mean as they represent the classical scientific method for identifying cause-effect relationships.

In this paper we present a controlled experiment aimed at assessing and comparing pure and blended learning approaches. The context of the experiment was constituted of students of the Bachelor program in Computer Science at the University of Salerno. The didactic contents used in this experiment were adaptive and have been defined using the visual based tool proposed in [3]. We considered adaptive didactic contents to fill the gap between traditional learning and e-learning according to the students' diversities. In fact, in the e-learning field the diversities among students and virtual classrooms are only marginally supported from the methodological and technological point of view, thus contradicting the third good teaching requirement proposed by Leblanc [9] (i.e., good teaching is about listening, questioning, being responsive, and remembering that each student and class is different). The defined adaptive didactic contents have been deployed in E-World platform [5]. The experiment was performed on line in a research laboratory at University of Salerno.

The remainder of the paper is organized as follows: Section 2 presents related work, while the tool and the platform used to define and enjoy adaptive didactic contents are described in Section 3. The experiment design and the achieved results are reported in Section 4, while final remarks conclude the paper.

2. Related work

Some case studies have been proposed in the literature [4][13][16] to assess and compare pure and blended learning approaches. For instance, Yeh *et al*. [16] adopted the multiple linear regression analysis and the ANOVA test in blended learning approach to analyze the relationship between learning contents and

acquired knowledge. The authors considered four independent variables: scores of the online tests, the reading time, the total number of login times, and the total number of online discussions, while the dependent variable represented the acquired knowledge, assessed using the tests performed during two month and the final exam. The context of this study was represented by high school students. The data analysis revealed that: the score of the online tests was highly correlated with acquired knowledge, while the remaining independent variables were not highly correlated with the acquired knowledge. Differently, Chen [4] verified in a pure learning context that the reading time, the total number of login times, and the total number of online discussions were highly correlated with the independent variable acquired knowledge. The difference between the two contexts was motivated by the fact that in the blended approach the learning content is presented both in the classroom and online, thus reducing the impact of the online presentation. An empirical analysis has been proposed in [13] to assess the group learning in blended and pure learning contexts. The assessment was based on four groups composed of four students. Three groups of students conducted the experiment on the pure learning approach, while the remaining group experimented the blended learning approach. In order to compare the acquired knowledge of the involved students, three variables were considered: the number of logins, the groups' task performance, and the coherence of the group discourse. The study revealed that knowledge improvement in the group learning does not solely depend on the communication setting. An interaction between the communication setting and the learning approach was also observed. In particular, subjects in the blended learning condition who worked together face-to-face led a much more coherent discourse than students in the pure learning conditions.

3. The e-learning platform

The adaptive didactic contents used in the presented controlled experiment have been defined using the Adaptive Self Consistent Learning Object SET (ASCLO-S) language [3], which is a special case of flow diagrams where an instructional designer defines classes of students, represented by stereotypes, and then specifies for each class the more appropriate learning process. The knowledge of a student is represented as a learning profile, which is updated during the fruition of the didactic contents. Students' profiles and stereotypes are compared to adapt the didactic content to the actual competences of the students.

The language proposes three different granularity levels of knowledge: *ASCLO*, *Knowledge Fragment*, and *Raw Contents*. In particular, a sentence of the ASCLO-S language is an aggregation of Adaptive Self Consistent Learning Objects (ASCLO), while an ASCLO is a set of didactic contents that are presented to the student considering him/her knowledge. An ASCLO is considered as a logical collection of Knowledge Fragments and a self-assessment test (used to update the student profile). A Knowledge Fragment is composed of Raw Contents, which are textual or multimedia objects presented in linear way. To improve the expressiveness and the readability the ASCLOs can be grouped in swimlanes.

The visual language has been then implemented in a Java prototype, which also enables instructional designers to deliver the defined adaptive contents in the E-World platform [5]. E-World has been designed to manage the adaptivity of the didactic contents and to trace the times that the students spend to enjoy the Knowledge Fragments. This platform has been implemented using J2EE technologies. In particular, Apache and Tomcat were used as Web Server and Web container, respectively. Instead, the main technologies used to dynamically deliver contents are Java Server Pages (JSP) and servlet. Finally, to integrate the component to trace the times that the students spend to enjoy the adaptive didactic contents we used a Web Service implemented using AXIS SOAP Engine [1]. Further details can be found in [5].

4. The controlled experiment

In this section we describe the proposed controlled experiment following the template suggested in [14], a well-known text book on software engineering experimentation.

4.1 Experiment definition and context

The purpose of the experiment is to evaluate the use of adaptive contents as defined in [3] in blended and pure learning scenarios. The subjects of the experiment were fifteen students attending the Web Development Technologies course of the Bachelor program in Computer Science at the University of Salerno. They

were volunteers with comparable background as the questionnaire filled in at the end of the controlled experiment revealed. The experiment has been performed online at the University of Salerno in the Web Development Technologies research laboratory. The PCs involved in the experiment had similar hardware and software configuration.

We performed the experiment in two laboratory sessions. In particular, the subjects employed in the experimentation of the blended learning approach accomplished the laboratory session first. They were nine and attended a classroom lesson and after an online lesson on the same knowledge contents. The subjects experimented the pure learning approach were six and enjoyed only the online lesson. It is worth noting that the subjects were involved only in one of the laboratory sessions.

In order to compare the two learning approaches two learning tasks to be accomplished within two hours have been defined:

- T_1: enjoying a course on XML (eXtended Mark-up Language) and performing a self assessment test evaluate the acquired knowledge;
- T_2: enjoying a course on XSLT (eXtensible Stylesheet Language Transformations) and performing a self assessment test evaluate the acquired knowledge.

The subjects selected to experiment the blended learning approach performed the first task, while the remaining subjects accomplished the second task. The self-assessment tests of both the learning tasks were composed of ten multiple choice questions.

4.2 Hypothesis formulation

To compare the blended and pure learning approaches the following null hypotheses have been formulated:

- **H_{n1}:** There is not a significant difference between the knowledge that the subjects acquired performing the learning tasks in the blended and pure contexts.
- **H_{n2}:** There is not a significant difference between the times to accomplish the learning tasks in the blended and pure contexts.

The alternative hypotheses are:

- **H_{a1}:** There is a significant difference between the knowledge that the subjects acquired performing the learning tasks in the blended and pure contexts.
- **H_{a2}:** There is a significant difference between the times to accomplish the learning tasks in the blended and pure contexts.

4.3 Selected variables and experiment design

In order to properly design the experiment and analyze the results, we considered the following independent variable:

- **Method:** this variable indicates the factor on which the study is focused, i.e. blended and pure.

The following two dependent variables have been identified:

- **Time:** the time required to accomplish the task (including the time to answer the self assessment test).
- **Score:** the score of the self assessment test used to evaluate the acquired knowledge.

4.4 Preparation

We have used as server a PC, where the E-World platform and the required technical infrastructure were installed. On the other hand, the subjects used PCs equipped with the web browser Internet Explore 6.0. A training session of fifteen minutes has been carried out to give subjects an equal prior knowledge of the experiment. We also presented E-World to let subjects get confidence with it. The training session was concluded presenting detailed instructions on the tasks to be performed. The following subsection provides more details on the steps composing the controlled experiment.

4.5 Material and execution

Each laboratory session was divided in four steps and was carried out without invoking any kind of tutor support and time limit. After the training session we asked the subjects to explore the platform to familiarize with it. Finally, depending on the method the subjects performed the learning tasks described in Section 3.1.

Once the subjects accomplished the learning tasks the supervisors collected the information traced by E-world, namely the time spent to enjoy each knowledge fragment and the score of the self assessment test. To carry out the controlled experiment each subject was also provided with a folder containing a pencil, some white sheets, and the E-world user manual.

4.6 Experiment results

The descriptive statistics of the dependent variables *Score* and *Time* are reported in Table 1, while the distributions of these variables are summarized by the boxplots in Figure 1. The boxplots of the *Score* variable are quite similar, while the boxplots of the variables *Time* present some differences. In particular, the median of *Time* of the Pure Method is larger than the median of *Time* of the Blended Method. It is worth noting that the boxplots of the variables *Time* of the Pure Method do not present outlier.

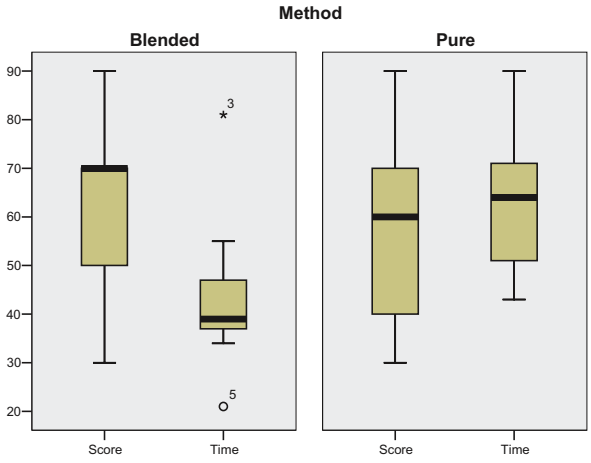


Figure 1. Boxplots of the dependent variables

To verify the considered null hypotheses we decided to use the one-way ANOVA test. In particular, we first considered the basic output of the ANOVA test, i.e., the source table (the F test), the means and the standard deviations, and a plot of means, and then in case there is a statistically significant difference between the observations of the two methods and dependent variables a further investigation is required to investigate the cause [6].

To apply one-way ANOVA the following four assumptions have to be verified: the observations are independent; the scale of measurement for the observations is interval; the distribution of the observations in each group is normal; and the variance of the observations in the groups is the same.

The first and the second assumption were easily verified since the two groups of subjects are composed of separate individuals and the scale of measurement for *Score* and *Time* is continuous. To verify the normality of the distributions we have performed both the Kolmogorov-Smironov and Shapiro-Wilk tests [11] (see Table 2). The results of these tests revealed that the normality is verified for the considered variables. To verify the last assumption, we adopted the Levene statistic test (see Table 3). The null hypothesis of the test cannot be rejected, thus the variance of the variables *Score* and *Time* is homogenous. Concluding the one-way ANOVA can be applied on the observations of the variables *Score* and *Time*.

Table 4 shows the results of the one-way ANOVA on the dependent variable *Score*. The results suggest that we cannot reject the null hypothesis H_{n1}, i.e., there is not a significant difference between the knowledge augmentation of the subjects who conducted the experiment in the blended and pure learning contexts.

Table 1. Descriptive statistics

Method	Independent Variables	Obs	Mean	Std. Dev.	Min.	Max.
Blended	Score	9	60	18.708	30	90
	Time	9	43.778	16.776	21	81
Pure	Score	6	58.333	22.286	30	90
	Time	6	63.833	16.449	43	90

Table 2. Tests for normality of the distributions

		Kolmogorov-Smironov			Shapiro-Wilk		
	Method	Statistic	df	p-val	Statistic	df	p-val
Score	Pure	0.200	6	0.200	0.958	6	0.801
	Blended	0.259	9	0.083	0.926	9	0.442
Time	Pure	0.165	6	0.200	0.977	6	0.938
	Blended	0.202	9	0.200	0.887	9	0.185

Table 3. Test of Homogeneity of Variance

		Levene Statistic	df1	df2	p-val
Score	Based on Mean	0.332	1	13.000	0.575
	Based on Median	0.307	1	13.000	0.589
	Based on Median and With adjusted df	0.307	1	110.511	0.590
	Based on trimmed mean	0.332	1	13.000	0.575
Time	Based on Mean	0.014	1	13.000	0.906
	Based on Median	0.048	1	13.000	0.830
	Based on Median and With adjusted df	0.048	1	120.046	0.830
	Based on trimmed mean	0.026	1	13.000	0.873

Table 4. ANOVA with Score as dependent variable (R Squared = 0.002 and Adjusted R Squared = -0.075)

Source	Type III Sum of Squares	df	Mean Square	F	p-val
Method	100.000	1	100.000	0.025	0.878
Error	52830.333	13	4060.410		
Total	581000.000	15			

Table 5. ANOVA with Time as dependent variable (R Squared = 0.287 and Adjusted R Squared = 0.232)

Source	Type III Sum of Squares	Df	Mean Square	F	p-val
Group	14480.011	1	14480.011	50.223	0.040
Error	36040.389	13	2770.261		
Total	453010.000	15			

The one-way ANOVA performed with *Time* as dependent variable (see Table 5), revealed that there is a significant difference between the subjects of the Blended and Pure Methods. Hence, the null hypothesis H_{n2} can be rejected. In order to accept the alternative hypothesis H_{a2} a further investigation is required. To this end, the k Independent-Samples t-test was adopted. The results confirmed a significant difference (*p-value* = 0.040) between the times required to accomplish the learning tasks. In particular, we observed that to accomplish the learning task the subjects of the Blended Method spent significantly less time than the subjects of the Pure Method. This result is further confirmed by the Error Bar graphs (see Figure 2), which shows that the mean times to accomplish the learning task T_2 is not contained in the 95% Confidence Interval of the times to perform the learning tasks T_1. We expected this result as the subjects of Blended Method already attended one or more classroom lessons on the same topic.

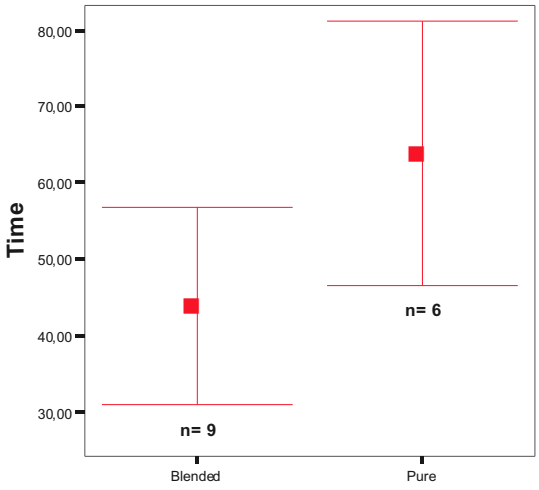


Figure 2. Error bar graph

4.7 Threats to validity

The validity of conclusions is generally influenced by factors that can be originated from within and outside an empirical study. In particular, the threats to validity that could affect our experiment are: internal, construct, external, and conclusion validity threats.

Generally, the internal validity is only relevant in studies that try to establish a causal relationship. Thus, the internal validity threats are relevant for our study since we aimed at comparing the use of two different learning approaches within adaptive learning processes. The key question in internal validity was whether observed changes can be attributed to the effect of the considered learning approaches and not to other possible causes. Concerning our experiment the internal validity threats are mitigated by the experiment design. In particular, to avoid frustration and disappointment during the fruition of the adaptive contents, the subjects were not informed of the traceability performed by E-World during the fruition of the adaptive contents. Furthermore, the subjects did not know exactly the hypothesis of the experiment. Finally, the subjects were not evaluated on their performance within the laboratory sessions.

Like external validity, construct validity is related to the generalization of the achieved results. The construct validity threats that could be present in this experiment, i.e. the interactions between different treatments, were mitigated by a proper design of the experiment. Indeed, the recruited subjects experimented either blended or pure learning approach. Moreover, depending on the treatment, the measurement of the dependent variable was performed either analyzing the log files and the database of the adopted e-learning platform.

External validity refers to the approximate truth of conclusions involving generalizations within different contexts. In our case the external validity threats could be present as we performed the controlled experiment within a specific teaching context and employing special conceived didactic contents. Thus, replications in different learning contexts using different didactic contents should be performed with different subjects in order to confirm or contradict the achieved results.

Conclusion validity concerns the issues that affect the ability to draw a correct conclusion. The conclusion validity threats are mitigated by the experiment design and by the properly selection of the population we would like to generalize the considered learning approach. In particular, we can assert that the selected sample was very representative of the population. In fact, they were students with the habit to study and learn.

5. Final remarks

Controlled experiments have been widely employed in several fields, such as Software Engineering [2][10], Information and Communication Technology [17], Biology [7][8], and Medicine [12][16] as they represent an important mean for identifying cause-effect relationships.

In this paper we have presented the results of a controlled experiment which has been performed to assess and compare blended and pure learning approaches. The subjects involved in this experiment were students attending the Web Development Technologies course of the Bachelor program in Computer Science at the University of Salerno. The didactic contents used to compare the learning approaches were adaptive to bridge the diversities among students and have been developed using the visual based tool proposed in [3]. The adaptive learning processes defined by employing this tool are deployed and traced by the E-World platform [5].

The data analysis of the controlled experiment revealed that the pure approach does not significantly affect the knowledge acquisition of the subjects. As expected, the experiment also revealed that the pure learning approach significantly affects the time to enjoy an adaptive didactic content.

Future work will be devoted to replicate the controlled experiment in several educative contexts to understand and improve the external validity of the achieved results, as well as to identify important environmental factors that could affect these results. Furthermore, we will replicate the experiment on didactic contents implemented using different adaptive learning models.

References

[1] Apache "Axis Project". http://ws.apache.org/axis/, 2004

[2] V. R. Basili, F. Shull, and F. Lanubile, "Building Knowledge through Families of Experiments", *IEEE Transactions on Software Engineering*, vol. 25, no. 4, 1999, pp. 456-473.

[3] G. Casella, G. Costagliola, F. Ferrucci, G. Polese, G. Scanniello. "A Visual System for Designing and Realizing Adaptive Distance Courses". *In Proceedings of 1st International Conference Technology Enhanced Learning*, November 20-21, 2003, Milan, Italy.

[4] M. Chen, "The development and effectiveness of a web-based learning system for in-service teachers," *In Proceedings of the Working Conference of Research and Development for On-the-Job Training of Teachers*, 2000, pp. 1-12.

[5] G. Costagliola, A. De Lucia, F. Ferrucci, G. Scanniello "E-World: A Platform for Managing and Tracing Adaptive E-learning Processes" *In Proceedings of the 12th International Conference on Distributed Multimedia Systems*, G. Canyon, USA, August 30 – September 1, 2006, Knowledge Systems Institute Press, pp.: 258-263.

[6] J.L. Devore, N. Farnum, *Applied Statistics for Engineers and Scientists*, Duxbury, 1999.

[7] G. Giaever, A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, et al., "Functional profiling of the Saccharomyces cerevisiae genome", *Nature* 2002, pp. 387-391.

[8] M. Krallinger, A. Valencia, "Text-mining and information-retrieval services for molecular biology", *Genome Biol* 2005, vol. 6, pp. 224-.

[9] R. Leblanc "*Good Teaching: The Top Ten Requirements*", The Teaching Professor, vol. 12, no. 6, 1998.

[10] D. Perry, A. Porter, L. Votta, "Empirical Studies of Software Engineering: A Roadmap", *The Future of Software Engineering*, Ed: Anthony Finkelstein, ACM Press, 2000, pp. 345-355.

[11] P. Roston, "An extension of Shapiro and Wilk's test for normality to large samples", In *Applied Statistics*, vol. 31, no.2, 1982, pp. 115-124.

[12] H.S. Sacks, J. Berrier, D. Reitman, V.A. Ancona-Berk, T.C Chalmers, "Meta-Analyses of Randomized Controlled Trials", *The New England Journal Medicine*, 316(8), Feb. 1987, pp. 312-455.

[13] K. Schweizer, M. Paechter, B. Weidenmann, "Blended Learning as a Strategy to Improve Collaborative Task Performance", *In Journal of Educational Media*, vol. 28, no 2-3, 2003, pp. 211 – 224.

[14] C. Wholin, P. Runeson, M. Host, M. C. Ohlsson, B. Regnell and A. Wesslen, Experimentation in Software Engineering – An Introduction, Kluwer, 2000.

[15] J.M. Yancey, "Ten Rules for Reading Clinical Research Reports", *American Journal Orthodontics and Dentofacial Orthopedics*, vol. 109, no. 5, May 1996, pp. 558-564.

[16] D. Yeh, C-H Lee, P-C Sun, "The Analysis of Learning Records and Learning Effect in Blended e-Learning", in *Journal of Information Science and Engineering,* vol. 21, 2005, pp. 973-984.

[17] M.V. Zelkowitz, and D. R. Wallace, "Experimental models for validating technology", *Computer*, 1998, pp. 23-31.

A Scalable Architecture for Latency Sensitive Massively Multiplayer Online Games

Kamran S. Malik and **Son T. Vuong**
Computer Science Department, University of British Columbia
Vancouver, BC Canada V6T 1Z4
{kmalik, vuong}@cs.ubc.ca

Abstract

P2P overlays are a natural architecture for scalably supporting Massively Multiplayer Online Games. However, computing consistent game state in a distributed fashion without compromising responsiveness places very hard to meet bandwidth, latency and scalability requirements on the architecture. We present a system that utilizes the limited Area of Interest of game entities to arrange player contributed machines in an overlay that optimizes bandwidth consumption, reduces event propagation delays and scales dynamically with an increasing number of players. Network proximity of nodes is also taken into account during the overlay construction process. In this paper, we describe how to split game state management over unreliable player machines while meeting the fault tolerance, state persistency and consistency requirements of games. We anticipate our system, with its increased scalability and lower latency, opening the possibility of completely new genre of MMOGs with fast paced reflexive action at a fraction of a cost of the more traditional client server architecture.

1 Introduction

Massively Multiplayer Online Games (MMOGs) are distributed multimedia games played over the Internet that involve thousands of concurrent players who interact in real time with other players and entities that share the game world. MMOGs are distinguished from other multiplayer games by the large number of players and the long term persistent storage of the game state. Games like Lineage have recorded 180000 concurrent players and the number of its subscribers exceeds two million [1].

Most commercial MMOGs available in the market use the traditional client-server model that limits the scalability of the system and places an enormous financial burden on the publishing company to furnish enough resources to accommodate large number of players. Peer-to-peer architectures allow games to utilize the storage, bandwidth and computing power available on the player machines. However, a number of technical issues are impeding the commercial popularity of this approach.

In this paper we discuss these impediments and as a solution propose an architecture that allows player machines to contribute their resources for game state management by becoming a part of an overlay that dynamically distributes the load over peer machines. The architecture exploits area of interest management to maximize the utilization of the pooled resources. We also describe methods for maintaining the consistency of the game state spread over the peers and means for guaranteeing state persistence in the face of failure.

The rest of the paper is organized as follows: Section 2 describes the existing commercially used Client-Server architecture and its drawbacks; Section 3, makes a case for a P2P approach, and discusses some recently proposed attempts to adopt the approach and lists their shortcomings. In Section 4, we provide a list of goals that we attempted to achieve through our architecture described in Section 5. The resulting prototype is presented in Section 6. And we close with sections on possible extensions of the architecture to service discovery, Section 7 and conclusion, Section 8.

2. Existing system support

Traditionally, MMOGs have used a client server architecture where the server is responsible for storing and computing the state of the shared world based on the inputs gathered from all the participating players. The clients receive input from the user, convert them into commands and pass them on to the server. The server collects the commands from player machines on the network, runs a game world specific simulation algorithm on the gathered commands to generate the next state of the game world and sends updates to the clients based on their area of interest. Since the state of the world is derived from input gathered from all the clients at a single central location, the players can see the effect of the combination of each other's actions.

It is much simpler to maintain consistency and deal with delayed and lost messages from clients as only the server is involved in calculating the state of the entire world. The game company also retains control over player admission and game state. However, all the clients have to communicate with the server which can quickly become a computing and network bottleneck because of the potentially large number of players involved in MMOGs and large amount of game state.

2.1 Problems with client server

Although these architectures can be made to scale with the number of players by adding more centralized resources, it fundamentally lacks flexibility and the centralized server(s) must be over-provisioned to handle peak loads. Game publishers must estimate in advance the expected number of players. Finacncial constraints often dictate under-provisioned servers, limiting the growth of appealing games. Even if the company deploys the correct amount of resources, these resources remain under utilized while the game popularity rises to the predicted level. Similarly, when a game's popularity starts to wane the gaming company is unnecessarily forced to maintain the original level of resources.

Moreover, client server architectures limits the responsiveness since clients have to wait for commands to propagate to a potentially busy server and the resulting

updates to return to the client. The higher lag prevents certain genres of games, like first person shooter games, from becoming massively multiplayer.

3. Peer to Peer Architectures

Peer to peer architectures are naturally scalable as the available resources grow with the increasing number of participants and they can be designed to not consume any finite centralized resources to accommodate new members.

Early multiplayer games and even more recent research and commercial work rely on a simple P2P model. Games like MiMaze [14] and Age of Empires [15] implement a completely decentralized unstructured P2P system where every node receives moves from all the players in the game and decides the game state on its own. These games depend on broadcast or multicast support from the underlying network. All nodes are equal and model the entire world based on the messages received from all the other players. Since all nodes run the exact same algorithm on identical sets of inputs, all players see the same world on their screens. Other P2P systems such as AMaze [17] and Mercury [16] utilize multicast groups to send messages only within groups of nearby players to minimize the bandwidth usage. However, IP multicast is still not widely supported and architectures that try to mimic multicasting by the use of unicast links introduce too much latency.

To adapt P2P technology for games with large number of players spread all over the Internet, we can no longer assume that (i) messages can be easily broadcasted to all participants; and (ii) the messages are delivered to all recipients within guaranteed time constraints. The system should also persistently store the state for shared mutable environment object and player entities.

3.1 Related Work

Knutsson et al. [2] propose making all player machines members of a Pastry Overlay [11]. The game map is split into zones and all players within a zone form a group by joining the Scribe [12] multicast tree for the zone. Zone IDs are used to hash onto Pastry nodes to select zone coordinators, which also serve as roots for the multicast trees. Game state is updated by broadcasting the moves on the multicast tree using the coordinator's ID.

Kawahara et al [7] advocate a fully distributed scheme where all nodes are equal. A player node maintains a list of a fixed number of nearest neighbors and forms direct connections with the neighbors included in the list. Neighbors periodically exchange their lists to help each other detect new incoming nodes.

In Solipsis [8] nodes attempt to maintain links with a ring of nodes around themselves instead of just connecting to the N nearest nodes. The nodes in the ring act as sentinels and notify the node when an unknown node approaches the central node from any direction.

3.2 Problems with existing P2P solutions

Knutsson et al. describe the most complete system of the three architectures as Solipsis and Kawahara et al. [7] do not consider the issue of game state management. However, their approach still has some undesirable properties. The cell sizes do not correspond to the area of interest of the players. Therefore, players cannot see across artificial cell boundaries introduced by the game architecture otherwise non-existent in the game world. The authors recommend players with greater visibility requirements to join multiple zones. But this results in a node receiving unnecessary messages from locations beyond its area of interest. More seriously the scheme seems to go against the very reasons that make the peer approach appealing. Player updates are routed through the coordinator, which is not much different from a client sending updates to a server. Secondly, the location of players and proximity of neighboring virtual world zones are not used in deciding how to stitch nodes together while constructing the overlay. As a result, messages have to be routed through multiple nodes that may not even be interested in the contents of a message. There may be as many as 50 virtual hops [5] in a decent sized game. To make matters worse, each hop between nodes handling neighboring virtual world zones translates to multiple hops on the physical network. Moreover, as Pastry tries to randomize neighbors in the overlay to improve fault tolerance, the messages have to haphazardly bounce all over the Internet before it gets to the destination. All of this adds to the delivery time of updates, adversely affecting the responsive of the application.

The main shortcoming of Knutsson's approach is that it does not fully utilize the direct connections between peers. On the other hand, Kawahara's method maximizes the hop count efficiency by maintaining direct connections between players at the cost of increased network traffic. Moreover, maintaining a list of a fixed number of nearest nodes does not guarantee global connectivity. For instance, if all the N closest nodes crowd together on only one side of the node, the player is rendered completely oblivious to a player approaching from the other side.

Like Kawahara et al., Solipsis utilizes direct links between neighbors improving the event delivery delay. It also solves the problem of node blindness due to overcrowding on one side by ensuring that there is always a ring of neighbors directly connected to the node regardless of their distances. However, proper neighbor discovery is still not guaranteed when players move too fast. Secondly, to maintain global connectivity players have to maintain connections even if they are too far apart to be visible to each other.

4 Goals

While designing our system our objectives were to:

- Maximize the game responsiveness by reducing the message delivery time.
- Maintain consistency of the game across all clients.

- Minimize the bandwidth and computational requirements for a single client.
- Provide fault tolerance for unpredictable peer exits.
- Provide game state maintenance for player controlled avatars (PC), non-player controlled entities (NPC), immutable and mutable terrain objects.
- Provide distributed persistent storage of game state.
- Retain admission control with the game publisher.
- Avoid using network supported multicast.

5 Proposed System

Our design exploits the limited area of interest of a player. Each node is responsible for just the avatar of the machine's owner. Avatars have a limited area in which they can bring about changes. This is known as the *publish area*. An avatar also has a *subscription area* that conforms to the sensing abilities of the avatar. These areas can have different dimensions depending on the capability and sensing ability of the player's avatar. A player can see a player, B, only if its subscription area overlaps the publish area of player B. All players whose areas overlap are connected directly with each other. Every time a player moves it sends a move event to all the connected peers and checks if the areas still overlap. If they do not, the peers drop the connection and are no longer aware of each other.

Using publish/subscribe areas is the most efficient and flexible interest management technique. It ensures that only participants who are interested receive only the events that can possibly affect them.

5.1 Neighbor Discovery

Selected peers are elevated to the role of a zone coordinator. Zone coordinators are assigned fixed portions of the game world (called *zones)* to manage. The coordinators have knowledge of all the avatars residing in their zone. They keep a record of their location and publish/subscribe areas so that they can inform players when they come close enough to each other to interact. It is up to the players to decide when to stop communicating with a neighbor.

Zone coordinators form a structured peer-to-peer overlay. The entire game world is split up into non-overlapping rectangular regions/zones. Each coordinator maintains direct links with the coordinators of the zones bordering its own zone. Through these neighboring coordinators it has indirect links to coordinators covering the entire game space. These indirect links can be used to route a message to any point in the virtual world. In this manner, coordinators maintain global connectivity by only keeping a record of a small subset of neighboring zone coordinators.

Players utilize the overlay of coordinators to seamlessly move to any part of the virtual world and acquire knowledge about their immediate surroundings. As a player crosses a zone boundary the coordinator informs the player that it has crossed into another zone. At the same time it also provides the player with information to connect to the coordinator of this new zone. The player then registers its current location and publish/subscribe area with the new coordinator and begins sending updates of its location.

5.2 Game State

A complete MMOG middleware should maintain game state consistency and persistence. Game state can be categorized as Player controlled Character (PC), Non-Player Controlled Character (NPC), Immutable terrain objects and Mutable terrain objects. Interaction between different classes of objects require varying level of latency and consistency guarantees.

Terrain objects are always located at certain points on the map. The zone coordinator for that position of the map computes and stores the states of these objects. As an avatar moves through the world the coordinator informs its *playerMD* (the middleware component running on the player's machines) of the positions and states of the all the terrain objects in the avatar's surrounding. The playerMD caches these values and informs the relevant coordinator whenever its avatar interacts with a terrain object. The coordinator resolves any conflict associated with concurrent updates to a terrain object and refreshes the cached value of the shared environment object in all the players in the vicinity.

For more latency sensitive interactions like PC-to-PC interactions, playerMDs are allowed to optimistically synchronize with each other, involving the coordinator only to resolve inconsistencies. For two players to interact they must have overlapping publish/subscribe areas. Such players have direct connections which they use to inform nearby players of their actions. The playerMDs gather these events and periodically recalculate the new state of the world applying the Trailing State Synchronization algorithm [4]. The same algorithm is run on the coordinator but with a greater lag, which ensures more accurate state generation at the coordinator. Therefore, the coordinator's state is considered to be authoritative and is used to resolve differences between the state of interacting playerMDs.

5.3 Fault Tolerance

Every coordinator has a number of replicas, which are periodically updated. The exact number is determined by the fault tolerance requirements of the game. Only one of the replicas actively exchanges messages with the players and neighboring coordinators and computes the states of the objects located in the zone. At a fixed interval the active coordinator updates one of the replicas, which relays the updates to the remaining replicas. Since the replicas expect to be updated at regular intervals, if they do not receive an update for some time they conclude that a replica upstream from it has gone down and initiate a process to select new playerMD to replace the lost replica.

5.4 Persistence

To insure tolerance for catastrophic failure and allow a minimum level of persistence there is a lightweight background process, which backs up the state of the different zones of the game world onto the central server. The publishing company can keep a copy of the state of the entire world secure on its own machine. This copy may be out of date by several minutes but it provides a minimum level of protection against a situation where there is a mass exodus of players. In the event of such a catastrophic failure the game will lose the last few minutes of moves. Given the non-critical nature of the application, this is fairly acceptable. Only in very rare circumstances the game will have to be reset to a slightly outdated state saved at a backup server. Most failures will be sufficiently handled by the replicas.

5.5 Coordinator Overlay

The design of the overlay centers on a 2 dimensional Cartesian coordinate space. The entire coordinate space is dynamically partitioned among the coordinators. Each coordinator is assigned its own independent zone within the overall space. A coordinator discovers and tracks coordinators for zones bordering its own zone. Using this set of coordinators a node can route messages to any random point on the virtual coordinate space.

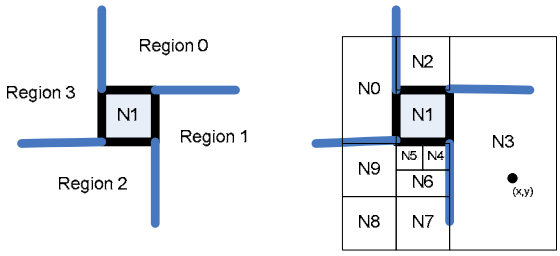
5.6 Overlay Construction

The overlay is constructed incrementally with the initial coordinator started by the central server on a game publisher's computer. The initial coordinator manages the entire game space and maintains a list of all player machines currently playing the game. As players join the game the initial coordinator can become overloaded. When that occurs the system dynamically selects one of the player nodes to share the coordinator's load. The overloaded coordinator splits its zone in half and transfers control of one half to the selected player machine while retaining responsibility for the other half. The zone split and handover is achieved through the following multi-step process:

1. A coordinator, *C*, detects an overload locally and sends a distress signal to the central server.
2. The central server maintains a list of all the players currently logged in. It chooses the best candidate to share the distressed coordinator's load and sends it a signal to spawn the coordinator process on its machine.
3. The spawned coordinator, *D*, informs the overloaded coordinator that it is ready to share the load.
4. The overloaded coordinator decides the orientation of the split and passes the dimensions and coordinates of the newly created zone that the new coordinator will own.
5. The original coordinator informs its neighbors of the split. The new coordinator is also passed information about the coordinators that will be neighbors for the zone it inherits. Based on this information new links are established between the new coordinator and the old neighbors of C.
6. C also passes on to D states of all the players and objects that lie within its zone. These transferred players are asked to establish connections with D.
7. As a last step, D asks the central server to pick player nodes to act as replicas and copies its acquired game object states to all the replicas.

5.7 Routing

Each node divides its neighbors in four sets as shown in figure 1. On receiving a message to route to a position (x,y) the node first decides which neighbor set or region it should search for the next hop.



Neighbor Registry at Node N1		
Region	**Registered Neighbors**	**Trespasser**
0	N2, N3	False
1	NULL	True
2	N9, N5, N4	False
3	N0	False

Figure 1: Neighbor Registry and Routing

Within a set the neighbors are sorted in increasing order of their origins - with the origin being the coordinates of top left corner of the zone. The node goes through this list looking for a coordinator that has the smallest Euclidean distance to the message's destination and forwards the message to the selected coordinator. To handle regions like N3, called Trespassers, the algorithm also considers zones that have their origin in one zone but overlap into another region.

5.8 Transparent Borders

The routing mechanism is used most extensively when a new player logs in. A new player must contact the central server to join the game. At this point the server can refuse the player's entry into the game based on the company's admission control policy. Each new player must have some initial position, which lies within some coordinator's zone. The server generates a register message for the player with its starting position as the destination of the message and injects the message into the overlay. The message ultimately ends up at the coordinator that oversees the starting position of the player.

The coordinator decides, based on the information in the

message, whether the player's publish/subscribe areas encroach into the zones of neighboring coordinators. It generates a list including all such coordinators and itself. This list is sent to the joining player asking it to register with all the coordinators in the list.

Players forward their actions to all the coordinators they are registered with so that their movements can be tracked. This makes the boundaries fairly transparent. In all other proposed zoning techniques perception of the players is limited by the boundaries of the zone in which its origin lies. In these schemes two nearby players separated by the artificial boundary are oblivious to each other's presence.

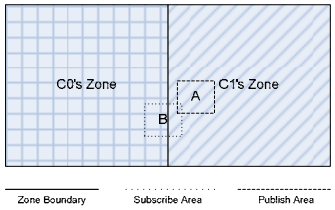


Figure 2: Transparent Boundaries

However, our scheme makes the boundaries completely transparent through a combination of the publish/subscribe areas and the requirement for a player to register with all coordinators whose space the player's publish/subscribe areas overlap. In figure 2, since B's AoI encroaches into C1's zone, B sends position updates to C1 as well, even though B itself resides in C0. Based on this information C1 can track B's location and inform A when B's subscribe area starts overlapping its own publish area.

5.9 Splitting & Merging

A zone is partitioned in half with the split occurring along a line running through the center either horizontally or vertically. When the height and width of a zone are equal it is split vertically resulting in two rectangular zones. If the height and width are unequal the zone is split horizontally. In this way, a coordinator that is repeatedly split alternates between rectangular and square zones. This results in more regularly shaped regions.

Merging is much more involved. The relationship between coordinators can be classified as parent-child. The *parent* being the coordinator that split and handed over half of its space to the *child* coordinator. Each coordinator stores the ID of its parent and a list of its children in the order in which they are 'created'. Whenever a coordinator becomes under utilized it sends a merge request to its parent if it does not have any children. Otherwise, it has to send a merge request to the last created child. A child can turn down the request under two circumstances (i) it is overloaded and merging will, in all likelihood, only result in an even more overloaded coordinator (ii) the child has other children. If this is the case, the parent has to wait for its child to become available. Only once a node has recombined with all of its children in reverse order of creation can it initiate a merge with its parent or accept a parent's invitation to merge.

The architecture is flexible enough to support more complicated splits. For example, instead of just splitting in half, the vertical or horizontal line could be drawn away from the center and the direction of the split could be altered to equally split the number of players. Alternatively, game semantics can also be used to coincide the zone boundaries along natural obstructions in the game like walls or mountain ranges.

5.10 Dynamic Area of Interest (AoI)

As the zone split and become progressively smaller to handle higher player densities it becomes increasingly likely that a player's AoI will overlap multiple regions. To prevent this from happening the AoI of a player are resized proportionally to the dimensions of the zone in which it lies. As a result, a player passing through small zones will have its AoI decreased. This scheme is an acceptable adaptation of the *cocktail party effect* [2] already well established in interest management systems. Even in real life a person in a crowded space tends to only concentrate on listening to few close by people can blank out the background noise. Since smaller zones mean higher number of players in the region, the algorithm simulates the same effect by simply reducing a player's AoI when it enters smaller regions.

5.11 Overload Detection

A coordinator is expected to perform its task against strict time constraints so that the latency and responsiveness demands of the application are adequately met. The overload detection algorithm uses this to decide when a coordinator needs to shed some of its workload. The coordinator times how long it takes to perform state computations and dissemination of the updates. If it sees that it is unable to process the state before the next state is due, it deduces that it is overloaded. The processing time also serves as a weak indicator of the load the node is putting on its network connection. As overloading the network forces the network to drop packets some of which have to be retransmitted by the coordinator resulting in higher processing time.

5.12 Topology Aware Overlay Construction

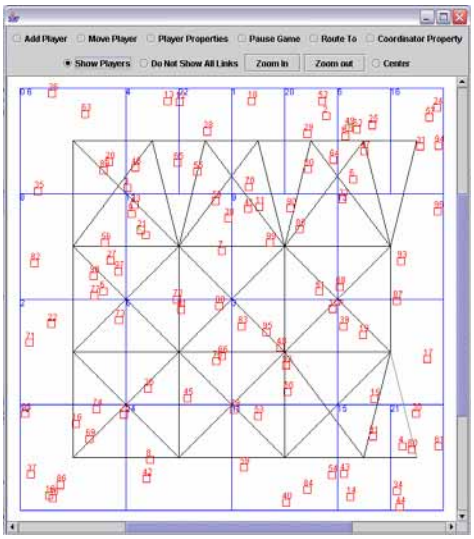
While choosing a child for an overloaded parent the central server considers the network proximity of the child and the parent. Choosing a child which is topologically closest to the parent results in lower message exchange latency, leading to better consistency and lower aggregate network consumption. Since neighboring coordinators need to exchange large amounts of data, this can have a considerable effect on performance.

The child selection algorithm uses network coordinates [10] as an index for network proximity. Each node is assigned network coordinates on the basis of ping times to certain landmarks on the network. Two nodes are considered to be close when the Euclidean distance between their network coordinates is small. By selecting children that are closest to

the coordinator being split we ensure that neighboring coordinators are optimally located on the network and hence the entire overlay is optimal.

6 Prototype and Experiments

We have implemented a prototype that allows the coordinators to dynamically split and merge as needed while maintaining a consistent overlay topology. We have also devised advanced overlay routing algorithms to route messages with only O(log N) hops compared to O(N) hops required by other similar overlay structures like CAN [13]. The prototype also implements protocols to handle player registration, player movement, AoI overlap notification and inter coordinator communication.



The GUI displays the Zone Boundaries, Player Locations and Inter-Coordinator and Inter-Player Connections. Also more detailed player and coordinator properties can be viewed (not shown in the diagram)

Figure 3: Screen Shot of Prototype Visualization

7 Future Work: Extensions to Service Discovery

We optimized the overlay topology by arranging the nodes according to their location in the virtual world. This also resulted in a simple direct mapping from an entity's location in the game space to the overlay's key space which allows more efficient range queries based on the location.

These properties of our overlay can be exploited for Service Discovery applications [18] which somewhat like games need to track the location of mobile devices and services. Recent research work on incrementally deployable distributed service discovery infrastructure propose using P2P overlays [19, 20] which do not use any location information in their construction. Moreover, these overlays are highly inefficient in performing range queries, which are very common in service discovery. Therefore, we feel that our architecture is much better suited for such an application. As future work, we plan to adapt the architecture for scalable global service discovery.

8 Conclusion

In this paper we have provided a general overview of MMOG, its requirements and some architectures that have been used to meet these requirements. We also proposed an architecture that utilizes the players' machines to allow gaming companies to scalably reach out to larger audiences without a huge investment in the infrastructure. The proposed ideas open the possibility of bringing MMOG to a completely new genre of games involving fast paced action that have very strict latency requirements. By adopting the publish-subscribe mechanism we guarantee near optimal area of interest management at the same time as partitioning the workload dynamically without introducing unnatural, in-transparent boundaries in the virtual world. We also show how the tried and tested Trailing State Consistency protocol nicely fits on to the architecture to provide persistent and fault tolerant management of players and terrain objects and propose an extension of the architecture for service discovery.

Bibliography

[1] Honghui Lu, Bjorn Knutsson, Margaret Delap, John Fiore, and Baohua Wu. The Design of Synchronization Mechanisms for Peer-to-Peer Massively Multiplayer Games.
[2] Bjorn Knutsson, Honghui Lu, Wei Xu, and Bryan Hopkins. Peer-to-Peer Support for Massively Multiplayer Games.
[3] Jouni Smed, Timo Kaukoranta, and Harri Hokonen. Aspects of Networking in Multiplayer Computer Games.
[4] Chris Dickey, Daniel Zappala, and Virginia Lo. A Distributed Architecture for Massively Multiplayer Online Games.
[5] Shun-Yu Hu, and Guan-Ming Liao. Scalable Peer to Peer Networked Virtual Environment.
[6] Eric Conan, Burton Filstrup, and Anthony Kurc. A Distributed Multiplayer Game Server System.
[7] Nobutaka Matsumoto, Yoshihiro Kawahara, Hiroyuki Morikawa, and Tomonori Aoyama. A Scalable and Low Delay Communication Scheme for Networked Virtual Environments.
[8] Joaquin Keller, and Gwendal Simon. Solipsis: A Massively Multi-Participant Virtual World.
[9]K.L. Morse. Interest management in large scale distributed simulations
[10] T. S. Eugene Ng, and Hui Zhang. Predicting Internet network distance with coordinates-based approaches.
[11] Anthony Rowstron and Peter Druschel. Pastry: scalable, decentralized object location and routing for large scale peer-to-peer systems.
[12] Miguel Castro, Michael B. Jones, Anne-Marrie Kermarrec, Anthony Rowstron, Marvin Theimer, Helen Wang, and Alec Wolman. An evaluation of scalable application-level multicast built using peer-to-peer overlays.
[13]Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, and Scott Shenker. A Scalable Content Addressable Network.
[14] C. Diot, and L. Gautier. A distributed architecture for multiplayer interactive applications on the internet.
[15] Paul Bettner, and Mark Terrano. GDC2001: 1500 Archers on a 28.8: Network Programming in Age of Empires and Beyond.
[16] Ashwin R. Bharambe, Sanjay Rao, and Srinivasan Seshan. Mercury: a scalable publish-subscribe system for internet games.
[17] E. J. Berglund and D. R. Cheriton. Amaze: A multiplayer computer game.
[18] Feng Zhu, Matt Mutka, and Lionel Ni. Classification of Service Discovery in Pervasive Computing Environments.
[19] Filipe Araujo, and Luis Rodrigues. GeoPeer: A Location Aware Peer to Peer System.
[20] Magdalena Balazinska, Hari Balakrishnan, and David Karger. INS/TWINE: A Scalable Peer-to-Peer Architecture for Intentional Resource Discovery.

E-EDUCATION TO REACH THE UNREACHED

Dr.Ramakrishna Avvaru
Associate Professor
Dept. of Education
Osmania University
Hyderabad, INDIA- 500 007
arknrd_rrrst1234@yahoo.co.in

Ms. Meghna Jhaveri
B.E. 4/4 (CSE) II Sem., Dept. of CSE
Chaitanya Bharathi Institute of Technology
Gandipet, Hyderabad, INDIA- 500 075
jhaveri_meghna@yahoo.co.in

Ms. Adilakshmi
B.E. 4/4 (CSE) II Sem., Dept. of CSE
Chaitanya Bharathi Institute of Technology
Gandipet, Hyderabad, INDIA- 500 075
adi.murty@gmail.com

Ms. Deepthi Anantha Uppala
B.E. 4/4 (CSE) II Sem., Dept. of CSE
Chaitanya Bharathi Institute of Technology
Gandipet, Hyderabad, INDIA- 500 075
deepthi.uppala@gmail.com

Ms.Ramadevi Narayana,
Associate Professor, Dept.of CSE,
Chaitanya Bharathi Institute of Technology
Gandipet, Hyderabad, INDIA – 500 075
nrdcbit@rediffmail.com

Abstract

The research paper presents an extremely effective system demonstrating the power of software, instead of relying on expensive hi-tech gadgets. Widely available wireless communication platforms are used to reach out to the masses. As a result, a novel, nifty, yet simple use of mobile phones and community radios have been developed, to provide last mile connectivity. E-Education has been taken up to reach the unreached. It consists of three main parts – a central system, a mandal system and last mile connectivity. The central system comprises of the talk to the teacher and a website component. The mandal system is the heart of e-education and includes eight modules such as P-SCAN, education jockey, idol-speaks, games, PSE, broadcast & feedback modules. The last mile connectivity is made through inexpensive technology of community radios. The paper examines the problems of the outreach system for robustness, inexpensiveness and social development. Finally, the authors conclude that e-education is not limited by the availability of computers and internet.

1. Introduction

Interoperable mobile communication systems and services are becoming truly global and mass market decreasing the digital divides. Essentially, due to their penetration and outreach, mobile gadgets and their connectivity impact not only our life management and communication habits but also shape our societies through basic and more advanced services they enable like community and e-governance applications.

Helping the less privileged children who cannot go to schools is the main purpose of this paper. The concept of **"**Schools to children, if not children to schools**"** is taken as the core issue in the national literacy policy of Government of India. It can be viewed as an end-to-end solution making use of the most innovative software that runs on existing hardware and infrastructure.

Computer hardware, even today, with big projects like OLPC and others, is still not affordable to most sections of the society. The educational-potential of technology, we feel, should not be restricted, due to the lack of affordable hardware [1].

The research paper presents an extremely effective system demonstrating the power of software, instead of relying on expensive hi-tech gadgets. Widely available wireless communication platforms are used to reach out to the masses. As a result, a novel, nifty, yet simple use of mobile phones and community radios have been developed, to provide last mile connectivity.

Table 1. Technologies and the corresponding modules

Technologies used	*Module name and other details*
. Net Framework 3.0	WPF, WCF (P2p N/W) And For Other Features Needed Through Out The System
Ms OCS 2007	The Speech Server 2007 (Shipped As Part Of OCS)For IVR System And The Windows Live API
Directx 10.0	For The Gaming Module
Microsoft Expression Studio	For Providing Users With Rich Media Content On Website (Ms Expression Web) As Well As Windows Applications (Ms Expression Blend)
Expression Media	With the need to quickly and easily import, organize, search, annotate, repurpose and archive digital files (audio). This is the way to go.
Share Point Portal Server	Single Integrated Platform For Connectivity (Central Site).
Microsoft Vista	For Deployment
SQL Server	For Storing All The Content At Central System And The Reporting And Analysis Services For Data Analysis.
Multi Point Sensor API	For Gaming And AI Modules
Microsoft Encarta	For Content.

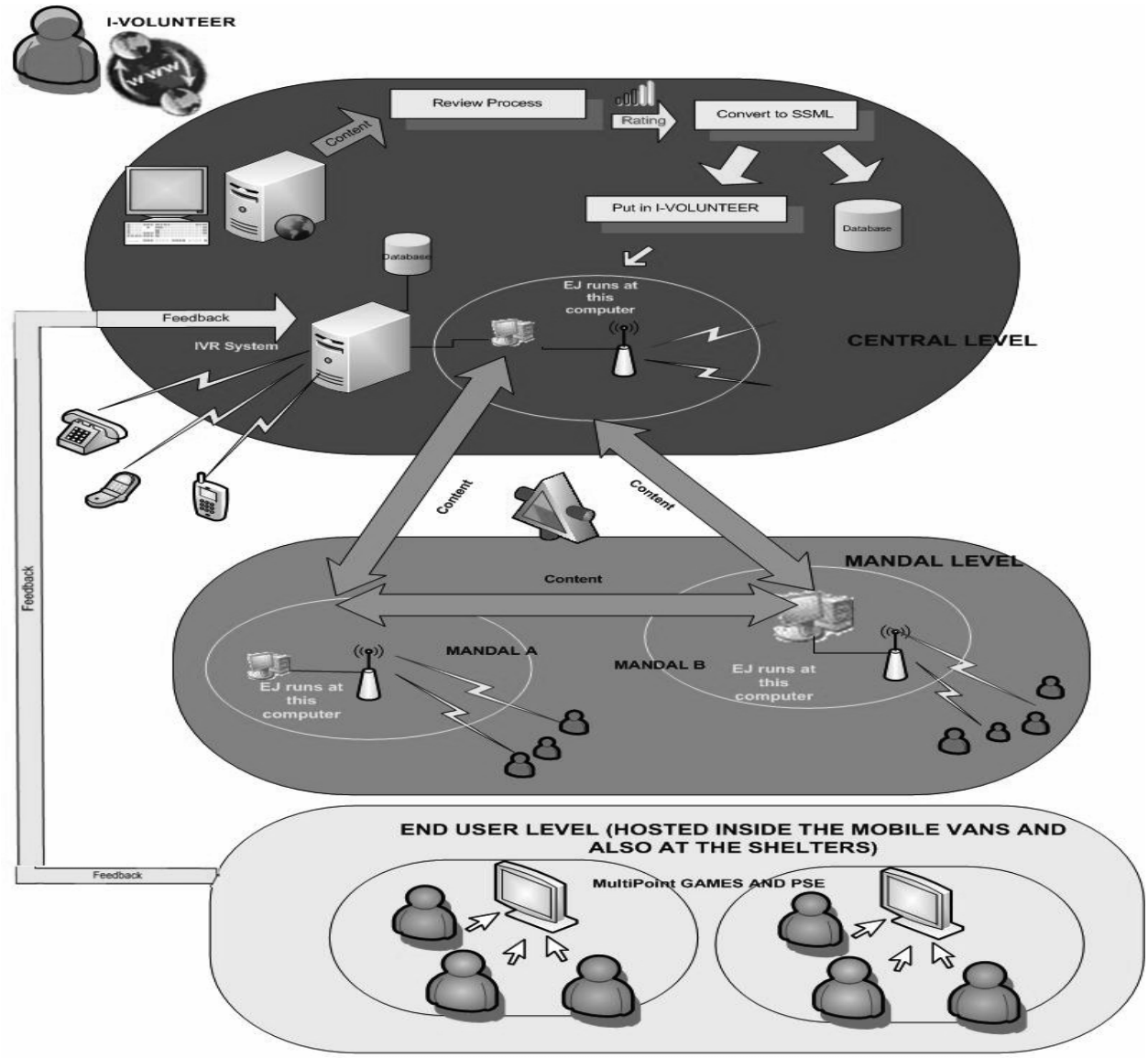


Figure 1. e- Education

2. Methodology

The different aspects of the paper are schematically represented in Figure 1.

Each of the components of the e-education is described below:

2.1. Central System

It is classified into two parts:

2.1.1. Talk-To-The-Teacher: Speech-Enabled IVR

This *Speech-recognition-capable layer* runs on the EJ module described below:

The teacher, from places where there is no availability of a computer or an Internet connection, is expected to call a Toll-free number, providing a facility to transmit real-time over telephone connected to transmitter/speaker. It ensures total community outreach irrespective of computer-availability [2].

2.1.2. Web Site: This has two parts:

I-Volunteer is involved in Community Generated Content. They are willing to share innovative content for uploading here. The content to be posted on the site and to be added to the repository has to be approved through a review and peer-review process. For this, a scheme inspired by Erdos Numbers is utilized. The highly rated members are graded '1' and people reviewed and trusted by them get the number '2' and so on. For an entry to be approved and posted, it must have a minimum number of points in a specific ratio of the members. This ensures the quality of the material.

In-Sync: The Shared Portal: This keeps all the mandals synchronized with each other, and provides a common platform to stay connected and updated [4].

2.2. Mandal System: This consists of the following eight modules which are described below:

2.2.1. Content storage and Transmission

All the instructive data is stored as *Instruction Definition File (idf),* which conforms to an *Instruction definition language (idl)*, an xml schema (extensible for future additions) defined by authors to describe each element of the audio content, including the instructive content, background scores, voices and other templates [5].

2.2.2. P-SCAN (Platform for Sharing Content Across Network: The P2P Module

The working of P-SCAN is given in Figure 2 below.

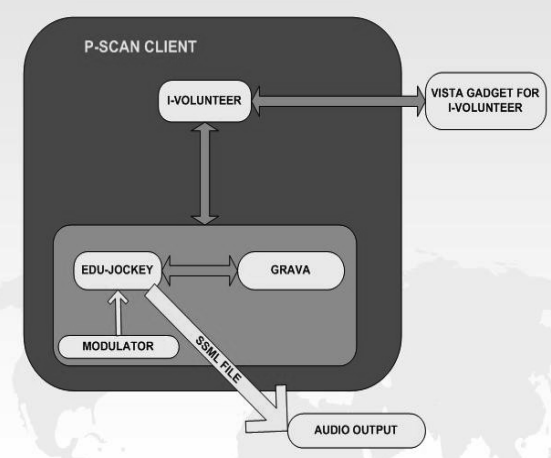


Figure 2: Working of P-Scan

This is a client built on peer-to-peer technology. This would be something like napster or limewire except that this is used to exchange educational content. This includes the presentations made in grava and also the EJ. Another dimension is added to grava (https://connect.microsoft.com/grava). Grava is the code name for a new set of tools from Microsoft's Education Products group that is designed to allow educator community to create and assemble material that will increase discovery and allow learners to at their own pace and learning style. It has an option to seamlessly upload it to I-Volunteer. Where GRAVA enables visual editing of files our EJ module enables creation and editing of audio files. These audio files are lessons which are text and are converted to audio (SSML). Vista gadget is updated using the feeds from I-Volunteer. In this context i-volunteer is something like a wiki hosting educational content of both EJ and grava, enabling to publish it to everybody. It could even open up a scenario where a same presentation could be modified by people at the same time.

2.2.3. Education Jockey (EJ): Content Mixing Module

This is a GUI to modify the *idf* file and allows mixing of the text content with voice, additional background scores and other related templates. The user can select the appropriate templates from a list of available options, already stored on the system. It also has a provision to add direct audio content that uses speech-to-text conversion.

2.2.4. Idol-Speaks: The Modulator

The idol teaches the stuff. Children can listen to their lessons through voices of their favorite celebrities. It has a range of voices from high-pitch female to bass-male and a facility to save voices from here to the EJ. It can generate new voices, modify existing ones and sort them according to the children's preferences.

2.2.5. Games Module

This module consists of a highly graphic multi-player DirectX games (with some help, we would love to use multi-sensor technology) that are *synchronized* with the radio broadcasts, and are updateable through XML feeds. For better accessibility, the hardware required for this module (specifically, a CPU connected to multiple consoles) is setup in a *mobile van* that tours all the villages.

2.2.6. PSE Module

This is a *fun-filled* activity for Personal and Social Education (PSE), providing a way to educate the children through interaction with animated characters with emotional-intelligence and role-playing capabilities that establish credible and empathic relations with the learners. The application adapts to the child's preferences for colors, games, subjects, voices, favorite celebrities, specific problem areas etc., so that his/her needs could be catered to efficiently, presenting a customized environment to the child [3].

2.2.7. The Broadcast Module

This parses the *idf* file (i.e., the text content mixed with appropriate voice tags, background scores etc.) to actual audio (given to the transmitter), making use of Text to Speech conversion facility [2].

2.2.8. Feedback Module

This module automatically gathers the data pertaining to the level of children's responses to the transmitted content. It generates reports on the number of children participating in various programs, their age groups, level of understanding and absorption capacity. These reports are sent to the central system in a timely manner, where they are interpreted. The content is then modified depending on the analysis done.

2.3. Last-Mile Connectivity

The refined and inexpensive technology of *Community Radios*, covering a range of 5 – 7 Kilometers radius is a revolution in India. Radio is the preferred choice, for its availability and affordable cost.

3. User Scenarios

The usage and usability of EJ:

Scenario 1: Transmission over the radio like how we propose to do in our system B.

Scenario 2: Peer to peer content sharing like system A. Using this people can work in groups, edit and share content together

Scenario 3: Creation of natural sounding, edited, customized audio books. (Could be published on http://www.freeonlinebooks.com)

3.1.System-A:

Scenario1: *Community contribution*

- Anyone who has access to the internet can log on to the I-Volunteer Wiki and write articles regarding any educative topic.
- For added convenience, he can download the I-Volunteer gadget and use it to make contribution.

Scenario 2: *Sharing content*

- Users can search for lessons on the P-SCAN network.
- The peer to peer client can be used to download lessons from the P-SCAN network.

Scenario3: *Users*

- Students can watch immersive multimedia lessons on their PCs. Multimedia lessons are made using GRAVA. They include videos, audio, animations and images.
- The EJ module extends the capabilities of GRAVA. It can be used to create audio books. EJ outputs lessons as SSML files that can be read out using text-to-speech tools.
- Optionally the EJ can output lessons as audio files. These can be loaded on to mp3 players, smart phones PMPs etc, so that the user can listen to them at his convenience.

3.2. System-B: *Talk to the teacher (IVRS)*

- Can be used by anyone with a phone.
- The IVRS system is built on Speech Server 2007.
- User calls the toll free number (provided by Satyam Foundation) and asks for the required lesson.
- Lessons are stored as SSML files on the server.
- The server uses a TTS application to read out the SSML files to the user.

Scenario 1:

Casual users can use this system to listen to topics of interest. He dials the toll free number, chooses the subject category and requests for a particular lesson, topic of interest or article. The server plays out the user's selection over the phone line.

Scenario 2:

Teachers in remote villages can use this as a tool for classroom teaching. They can call up, request for the lesson and turn on the speaker phone (or connect to an external speaker) so that the entire class can listen to the lesson.

3.3. Knowledge Casts (Broad cast module)

Scenario 1:

Teachers at the mandal level use the EJ to customize the lessons. They can schedule the broadcasts of different lessons. These lessons are transmitted via an FM transmitter connected to the mandal level PC.

Scenario 2:

Students in rural areas can listen to the knowledge casts using ordinary FM radios.

In between the knowledge casts, the broadcast schedule is played out so that the students may remember to tune in at the proper time.

3.4. Personal & Social Education

This is a very interesting way to impart social education to small children.

Scenario 1:

Children can play in a competitive mode. At each stage they are posed with two scenarios (for example how to cross the road), one right and the other wrong. If the right choice is made, then the player advances, else he goes back a few places.

4. Code

A sample code module is given below:

```
using System;
using System.IO;
using System.Net;
using System.Windows;
using System.Windows.Controls;
using System.Windows.Data;
using System.Windows.Media;
using System.Windows.Media.Animation;
using System.Windows.Navigation;
namespace PSE
{
        public partial class Window1
        {
    int[] pos=new int[3];
    //int score=new int;
    int player=0;
    int x=0,y=0;

    TranslateTransform trans = new TranslateTransform( 100, 300);

    public Window1()
                {
                this.InitializeComponent();
//trans.TryTransform(boy.TranslatePoint

                                // Insert code required on object creation below this point.
                }
private void clicked (object sender, System. Windows.Input.MouseButtonEventArgs e)
    {
      Random random = new Random();
      int num = random.Next(1, 7);
      textBox.Text = num.ToString();
      pos[player]+=num;
      move(pos[player]);
      playAnimation();
      readResponse();
      if(response==true)
      {
        score[player]++;
      }
      else
      {
        pos[player]-=num;
        score[player]--;
      }
       }
    private void move(int pos)
    {
      //calculate x,y coordinates from position
      x=(pos%3)*160;
      y=(pos/3)*158;
      Point p = new Point(200, 200);
      boy.TranslatePoint(p, (UIElement)start);
    }
       }
}
```

The above sample code is for Windows multimedia animation for educating the kids through games.

5. Our Vision

Authors being academic personnel with knowledge of the requirements, thought it fit to associate with foundations, which are already in the field and, accordingly, tied up with M.Venkatarangaiya (MV) Foundation. They had opportunity to interact with the children, which has helped us gain first-hand understanding about their predicaments. This provided them with the information, with regard to the approach required and the problems involved in the process of designing and evolving this work.

Authors could design a concrete model that would surely help realize the objective of e-education. They are delighted to note that the MV Foundation and the Satyam Foundation, after a detailed deliberation, have claimed it to be the solution to tackle illiteracy and have assured us not only the moral and material assistance, but also to extend the reachability of the module to the maximum extent. Thus, realizing vision of *"Achieving Total Literacy by 2020 (or earlier)"* for India.

6. Discussion

The present paper examines each problem of basic education systems in every dimension possible.

- The *hardware (radios)* at the end user level is easy-to-use, robust, portable & most importantly and inexpensive.
- *Quality of education*: At the lowest end user level, their knowledge is not confined to the knowledge imparted by their immediate teacher. The material generated undergoes a thorough review process. The synchronized games provided, substantiate their learning. Kids' awareness, social development and other major aspects of education are addressed.

7. Conclusion

E-Education is not limited by the availability of computers and internet. The system can

function irrespective of the availability of high-end gadgets, using basic infrastructure. The end user is not expected to be computer savvy or even expected to have minimum education. The user will find it very easy to interact with the system as it has a *wholly voice enabled GUI* that can be used for navigation through the application and this *GUI is customizable* to the user's language and preferences. Authors have only tapped the surface of the enormous potential such an application could have. It could be extended in every direction to provide News & Local Weather Reports, Night Schools for the elderly, Emergency Public Information and many more.

8. References

[1]) Patcharee Basu, Achmad Husni Thamrin, Shoko Mikawa, Keiko Okawa, Jun Murai, "Internet Technologies and Infrastructure for Asia-Wide Distance Education: 2007 International Symposium on Applications and the Internet (SAINT'07), January 2007.

[2] Eusebio Scornavacca, Stephen Marshall, "TXT-2-LRN: Improving students? Learning experience in the classroom through interactive SMS", 40th Annual Hawaii International Conference on System Sciences (HICSS'07), January 2007, pp. 5b

[3] Dawn Woodgate, Danae Stanton Fraser, "eScience, Science Education and Technology Integration in the Classroom: Some Practical Considerations", Second IEEE International Conference on e-Science and Grid Computing (e-Science'06), December 2006, pp. 138.

[4] Sami Huttunen, Janne Heikkila, "An Active Head Tracking System for Distance Education and Videoconferencing Applications", 2006 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'06) November 2006, pp. 30

[5] Marguerite L. Koole, "Practical Issues in Mobile Education", Fourth IEEE International Workshop on Wireless, Mobile and Ubiquitous Technology in Education - (WMTE'06), November 2006, pp. 142-146.

Region Based Image Clustering using Distributed K-Median Clustering

Aiyesha Ma Ishwar K. Sethi
Oakland University
Rochester, MI, USA
E-mail: @ ama@oakland.edu, isethi@oakland.edu

Abstract

Median or centroid based clustering is important in applications where retaining an object in the collection is important for the purposes of end user visualization, such as image browsing applications. Median based clustering is also important in applications where the mean of two object representations has no meaning. This paper adapts a distributed K-Median clustering algorithm, and demonstrates its applicability for a variety of object representations and distance measures.

The versatility of the distributed K-median clustering algorithm is demonstrated by clustering images using three different image representations and distance measures. The first uses global histograms, while the second uses a constraint based region comparison. The third uses a proposed combination approach comparing both global histograms and region sets, to overcome some of the individual limitations. Visual results of the three different approaches to image comparison are shown, and, for the test set used, the combination approach performed well.

1 Introduction

While content based retrieval of media assets has long been studied, especially images as evidenced by several review papers [16, 11], retrieval of media assets from distributed databases and peer-to-peer (P2P) environments is gathering interest with the increasing uses of P2P systems and collaborative ventures in sharing scientific and medical media.

In many image retrieval approaches, clustering is used to organize the images into similar groups either as part of the retrieval process or to reduce the amount of information presented when browsing by summarizing using relevant examples. Examples of clustering for both retrieval and browsing in non-distributed environments are found in papers by Sethi and Coman [13] and Stan and Sethi [17]. An example for browsing in a larger system is presented in [6]. Thus far however, little of this research in interactive browsing has occurred in P2P environments.

There are two primary benefits to clustering the images within a distributed environment. First, as with clustering in a non-distributed environment, the clustering result can be used to visually portray a global overview of the accessible images, and retrieval can be performed by selecting those images which lie within the cluster closest to a query image. Second, in a distributed environment it is beneficial to reduce the amount of communication between nodes, and by clustering in advance only the most relevant partial or summary information to a query need be sent.

The K-Means clustering algorithm is a well known and popular clustering technique, with many applications. This algorithm has the limitation that the mean vector is a new vector, and while sufficient for accomplishing the clustering, the vector may not be relevant in analyzing the results. Obtaining an object in the collection as the cluster representative is especially important in image or media applications where average or mean images have no meaning. This paper uses our earlier proposed distributed K-median clustering algorithm [7] which retains the cluster center as an object in the collection.

The purpose of this paper is to expand upon our earlier work, and demonstrate both the flexibility of the algorithm and its applicability to image clustering. Thus, this paper applies the proposed clustering algorithm to a collection of images and demonstrates that, in addition to simple global histogram representations with Euclidean distance, the algorithm can be applied to region based representations with a constraint based similarity measure.

Additionally, this paper proposes a combination distance measure using both global histogram and region level information. The proposed image distance measure appears to reduce the effects of the individual limitations.

Section 2 discusses some relevant background information. The clustering algorithm is presented in Section 3. The three image distance measures are discussed in 4, and the clustering results are shown in Section 5. A conclusion follows in Section 6.

2 Background

Content based image retrieval (CBIR) arose as the limitations of manual annotation became apparent. First, the time needed to manually annotate a large collection of images is unrealistic. Second, and perhaps more importantly, annotations are subjective, and vary according to any number of factors. In a networked environment, these annotations may also cause language barriers. Though P2P systems currently operate on keywords and filenames, these same problems arise and researchers have realized that CBIR techniques need to extend to P2P systems.

P2P systems fall into three categories: centralized (Napster), decentralized unstructured (Gnutella), and decentralized structured [14]. Research on content-based retrieval in P2P systems seems to focus on the decentralized arena. These systems use vector based indexing systems for the retrieval, but in addition they organize the network topology so that peers containing semantically similar documents are located near each other. Their aim is thus to improve query efficiency by first sending queries to those peers that are likely to contain that type of semantic content. General document retrieval is addressed in [19, 18, 12], while CBIR in P2P systems is addressed in [10, 9, 1, 4, 21].

Though much research has gone into information retrieval in P2P environments, little research has gone into developing methods for interactive browsing in P2P environments. In [3], the authors propose a browsing method for documents. Unlike the previously mentioned papers that focus on how to organize the network efficiently and how create and store the index, this paper focuses on developing a distributed K-median clustering tool, that can be used to cluster distributed documents in a method conducive to later visual browsing.

3 Distributed K-Median Clustering

In the P2P environment, there are two levels of computations: global calculations, and local calculations. While local calculations are computed individually at each peer, global calculations are either computed at a centralized server (as in the Napster model) or among the peers given a communication scheme. Since our focus is the algorithm, and not the communication scheme, the target environment here is the Napster model of a centralized server and index.

Briefly, in the distributed k-median algorithm, the server communicates the global cluster centers to the peers, the peers then calculate local cluster centers as medians and communicate these back to the server, and the server then updates the global medians as the cluster centers. Since global medians can't be directly calculated from local median information, an approximate global median is calculated. The following algorithm, Algorithm 1, was previously proposed in [7], and is similar to the K-Means algorithm proposed in [2].

The following notations are used: k, the number of clusters; nP, the number of peers; P, a peer; x_i, an image; C, a cluster; $Median(C_P)$, is the local median of the cluster at peer P; and $Median(C_G)$, is the global median of the cluster.

The distance from an image to a cluster center is denoted as $D(x_i, C)$. While [7] restricted the distance to the Euclidean distance between feature vectors, this paper uses no such restriction. This in essence changes the algorithm from a K-median algorithm to a K-clustroid algorithm.

For initialization in our experiments, each peer randomly selects $\lceil k/nP \rceil$ of its images and communicates these to the central server. The central server then randomly selects k of these images as the initial cluster centers, and communicates these to all the peers. If each peer is an expert on a set of topics, this simple initialization method should work fairly well. We reserve further study of distributed initialization methods as future work.

In this paper the 'Last Median' option is used for computing the local medians. In this approach each peer sends two representatives to the server: the local image nearest the last calculated global median, and the local median of the cluster. In addition, the number of local images nearest each of these representatives is tabulated and sent to the server. Thus information for peer P to communicate to the central server about cluster C, is $X_P(C)$, a set of two pairs,

$$\begin{aligned} X_P(C) = \{ \quad & (Nearest(Median(C_G)), \\ & \quad Closer(Nearest(Median(C_G)))), \\ & (Median(C_P), Closer(Median(C_P)))\} \end{aligned}$$

where $Nearest(Median(C_G)) = x_i$ such that $\forall x_i, x_j \in C_P, i \neq j$, $D(x_i, Median(C_G)) < D(x_j, Median(C_G))$. $Closer$ indicates all items in the local cluster that are closer to one representative than the other. Thus, this is defined for $Nearest(Median(C_G))$ as the set $\{x_i | \forall x_i \in C_P, D(x_i, Nearest(Median(C_G))) < D(x_i, Median(C_P))\}$, and for $Median(C_P)$ this is defined as the set $\{x_i | \forall x_i \in C_P, D(x_i, Median(C_P)) < D(x_i, Nearest(Median(C_G)))\}$.

The approximate global median for each cluster is then computed as the weighted median of the local representatives for that cluster. Thus, the approximate global median of all elements in cluster C,

$$Median(C_G) = Median(\{X_P(C) | \forall P\})$$

is computed by replicating each x_i, w_i times, and computing the median. For example, if $X_{P=1}(C = 1) = \{(P1x_1, 3), (P1x_2, 1)\}$ and $X_{P=2}(C = 1) = \{(P2x_1, 2)\}$, then the global median would be $Median(P1x_1, P1x_1, P1x_1, P1x_2, P2x_1, P2x_1)$.

Algorithm 1 Distributed k-Median Clustering Algorithm

Select initial cluster centers, $Centroid(C) \forall k$
repeat
  In Parallel do:
  for all $x_i \in P$ **do**
    for all C **do:** Compute $D(x_i, Centroid(C_G))$.
    end for
    Assign x_i to cluster C, where
    $D(x_i, Centroid(C_G))$ is minimized $\forall C$
  end for
  for all C **do:** Compute $X_P(C)$. **end for**
  Communicate $X_P(C) \forall C$ **then do:**
  End In Parallel
  for all C **do:** Compute $Median(C_G)$. **end for**
until centers are stabilized

The iterative process of recalculating global medians would continue until the cluster centers (or medians) stabilize under some criteria. In our experiments, stabilization was defined as no change occurring to any of the cluster centers from one iteration to the next.

In the context of a P2P environment where peers may join or leave, the full clustering procedure likely only needs to be performed on occasion. In general, as new peers are added to the system, the server communicates the current cluster centers, and the peers organize their local data accordingly and transmit their local representatives. Since the centralized server keeps a record of each peers' representatives, as peers are added or dropped the server can update the global medians accordingly and transmit those updates only after sufficient change. Furthermore, since the local representatives are each associated with a number of samples, this information can be used by the central server to decide which peer is most likely to have the closest response for a given query. This approach of querying the mostly likely peers first to reduce traffic is taken by several P2P systems, including [10]. We leave as future work the analysis of how often to update the cluster centers and of the reduction of traffic by first querying more likely peers.

4 Image Similarity

The benefit of the described K-Median clustering algorithm, is that any distance measure can be used, and the clustering method can be considered a generic clustroid based clustering method. While [7] restricted the demonstrated application to using a Euclidean distance, this section describes several distance measures between images. The first still uses the global histogram. The second is a region based approach, and the third is a combination approach using both the global histogram and region information.

4.1 Global Histogram

The feature vector for each image consists of a global histogram with 256 bins in the HSV (Hue, Saturation, Value) color space. The 256 bins have 16 bins in the hue component, 4 bins in the saturation component, and 4 bins in the value component. This feature vector was chosen to comply with the MPEG 7 specification of a color descriptor [8, 15]. The distance between images was calculated as the Euclidean distance between the histograms. This distance is somewhat naïve and better distance measures between global histograms have been proposed in the literature.

4.2 Region Based

To first extract the regions, segmentation was performed using online region growing [5]. Some parameters were set to perform the segmentation in an aggressive way, meaning many smaller regions were merged into larger regions. The reason for doing this was two-fold: 1) to reduce the number regions and thus reduce the computational complexity; and 2) reduce the chance of getting only a portion of an object, say for example only a part of a tree.

Each image was thus represented by the regions composing the image, and each region was represented by its location within the image (centroid), its color, and its eccentricity. Due to computational complexity, the constraint based region matching algorithm was applied only on the first order features, rather than the entire set described in [20]. Thus the weight estimation equation between regions in different images was changed to the following,

$$W_{ij} = \sum_{k=1}^{M} \sum_{l=1}^{N} S1(r_i, r'_j) S1(r_k, r'_l)$$

where r is a region in image 1, r' is a region in image 2, and $S1(r, r')$ is the similarity factor using first order features. M is the set of all regions in image 1 and N is the set of all regions in image 2.

Since the constraint based region matching returns a similarity value between 0 and 1, the distance between two images, I and J, is: $D(I, J) = 1 - Sim(I, J)$.

4.3 Combination

Since the test set of images consisted of digital photographs of primarily landscapes and large scenes, initial tests of the region based distance noted that the overly aggressive nature of the segmentation algorithm caused a loss in variance and color distribution information within each region. To compensate for this lost information, a combination approach is proposed here that uses both the global histogram and the region information.

To accomplish this, a similarity factor between global histograms was computed as follows,

$$Hs(I_1, I_2) = exp\left(-\frac{\parallel Hist(I_1) - Hist(I_2) \parallel}{2 \cdot 256^2}\right)$$

where I_1 and I_2 refer to the two images and the histograms were calculated in the same manner as described in the above subsection 4.1. Then, the image similarity was calculated as,

$$Sim(I_1, I_2) = Hs(I_1, I_2) \cdot S(R_1, R_2)$$

where $S(R_1, R_2)$ is the similarity between the region sets of the two images. Again, the distance between two images, I and J, was calculated as: $D(I, J) = 1 - Sim(I, J)$.

5 Image Clustering

As mentioned earlier, the target of this application is to facilitate indexing and browsing of the image collection over a distributed network. This section shows some of the clustering results, as the first stage of this larger goal. Depicted here are several set of images containing the cluster center, and the closest images within a cluster to the cluster center. Visualization techniques are not discussed here, since the intent is simply to show how the clustering might aid in reducing the problem size.

Originally, the images were clustered using global histograms. The results of this were previously mentioned in [7]. For the purposes of the discussion a couple of the results are shown here. As with any image comparison method, there are some instances where global histograms were sufficient for summarizing the images, and other instances where the summarization wasn't sufficient. Figure 1 shows both of these instances; two cluster centers and their nearest images within the cluster are shown. The top row depicts a 'SpringFlowers' cluster for which the six nearest images to the cluster center are fairly homogeneous. The second row depicts an 'Arborgreen' cluster, for which the nearest six images show less homogeneity, with several 'Football' images mixed in.

For the region based and combination methods, the test set consisted of 963 color images, from the University of Washington collection [http://www.cs.washington.edu/research/imagedatabase/groundtruth/]. As a result of the computational complexity induced by the region information, this set of images is smaller than that used when testing using only the histogram information. Images were distributed across 15 peers such that peers were experts, meaning a peer had many instances of a specific topic, and each peer contained varying numbers of images. Clustering was performed with K = 20.

While some groups of images were clustered in a reasonable way many clusters were quite poor when using just the region information. Figure 2 shows three cluster centers and the nearest 6 images within each cluster. The top row in the figure depicts a cluster with the center image falling in the category 'San Juan'. While the images nearest the center vary as to category, most fall in the description of 'water scene' or 'beach'. The other two clusters depicted illustrate one of the problems experienced with the region based approach. Many of the clusters had highly variable content, and this was a result of the aggressive nature of the region segmentation algorithm.

The combined approach is an improvement over both the individual global histogram approach and the constraint based region method. Four cluster results are depicted in Figure 3; on the left is the cluster center followed by the six closest images within the cluster.

The 'Football' cluster partially depicted in the top row had a total of 23 images all from 'Football' category within the collection. The collection had a total of 48 football images, of which 25 to 30 could be considered 'overview of field' scenes as opposed to cheerleaders, bands, or other close up or side views. The cluster only contained one image that was a side or close up view of the field, all the other images were overview scenes.

The 'Barcelona' cluster only had nine images total, and all were from the 'Barcelona' category of the collection. While the 'Barcelona' category contained many images only 13 of these were dark indoor scenes of a banquet and stage or dinner show. The cluster captured 8 of these 13 images, and only one image in the cluster was an indoor window scene instead of the banquet scene.

While the combined method performed well for most of the clusters, there were still a couple clusters that were more varied. The bottom row of Figure 3 depicts one such cluster. While the cluster center was an image of Cambridge, 40 of the 43 images in the cluster were either from the 'Cambridge' category or the 'SwissMountain' category. The two primary categories were nearly equally represented. Despite this lack of homogeneity within the cluster, most of the images containing snow in 'SwissMountain' category were placed in this cluster. The 'SwissMountain' category contained 27 images with snow (34 images total).

6 Conclusion

This paper presents a k-median clustering approach for use in a distributed environment, such as a peer-to-peer system. While the presented approach uses the Napster model of a centralized coordinator and index, the clustering method could be extended to less centralized models by deciding on a communication scheme.

While cluster initialization was not considered in this paper, choosing cluster centers in a distributed environment is perhaps one area that needs further consideration. Other

Figure 1. Global Histogram approach – Top row: SpringFlower; Bottom row: Arborgreen. Far left image is cluster center.

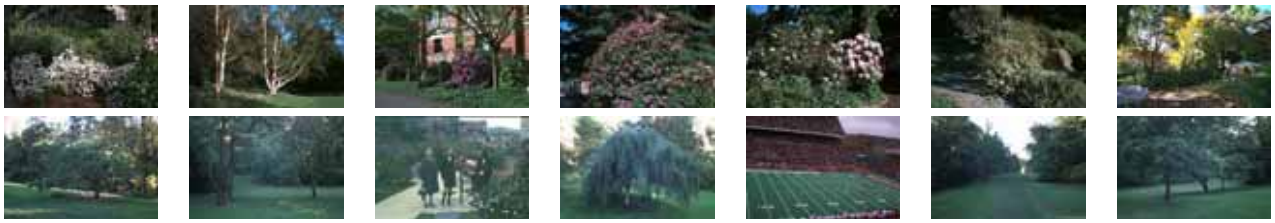


Figure 2. Constraint Based Region approach – Top row: SanJuan; Middle row: Arborgreen; Bottom row: Barcelona. Far left image of each row is cluster center.

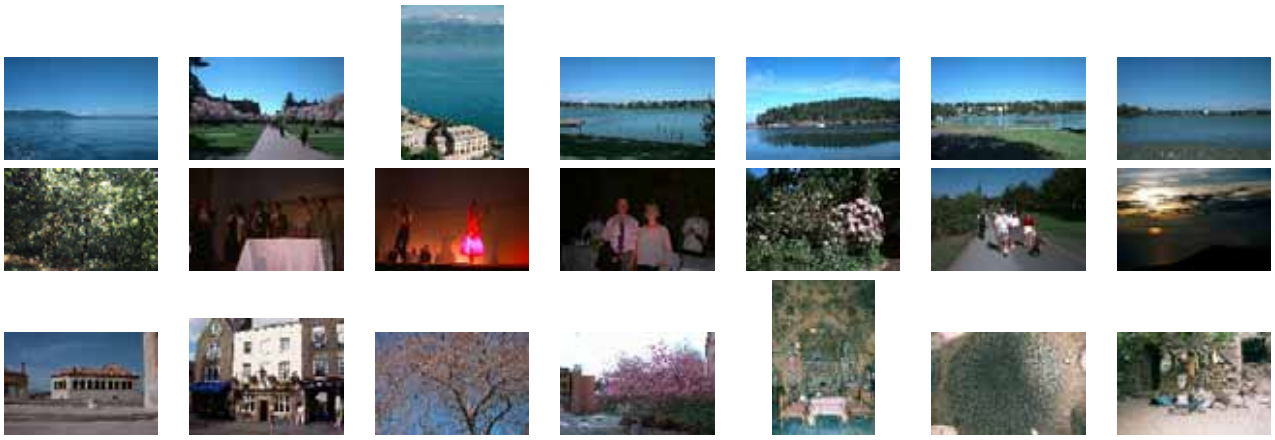
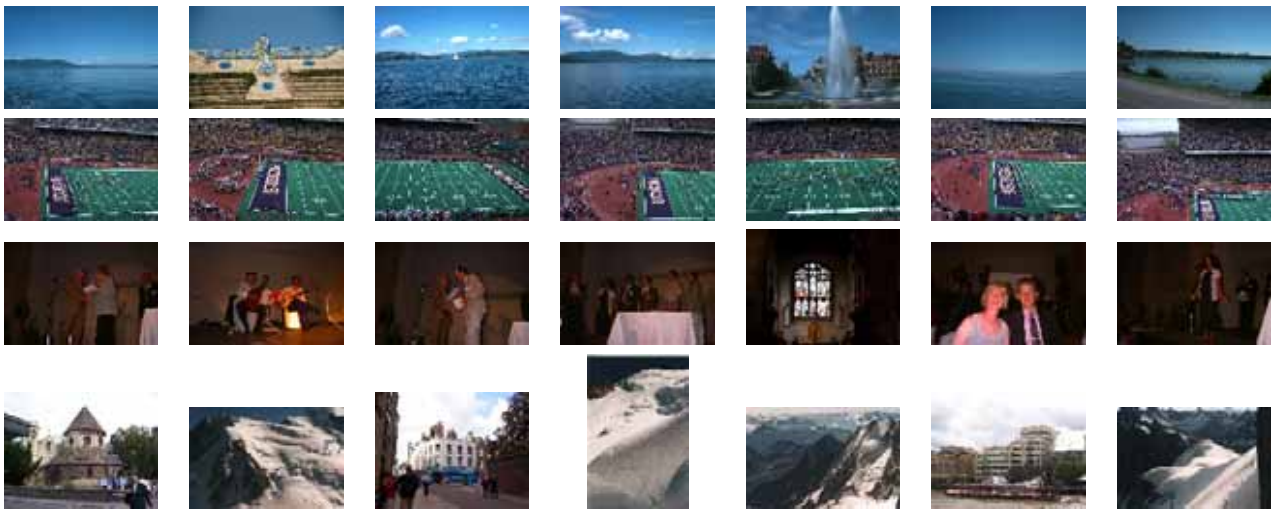


Figure 3. Combined Histogram and Constraint Based Region approach – Top row: SanJuan; Upper Middle row: Football; Lower Middle row: Barcelona; Bottom row: Cambridge. Far left image of each row is cluster center.



future work may include determining how often to communicate the local medians, and how often to update the global medians. Other issues particular to peer-to-peer systems are how to handle adding and dropping peers, and how many peer changes can occur before it becomes necessary to recompute the global medians.

While a global histogram approach to comparing images performs well in a number of cases (generally landscape), it has known limitations. The difficulty with a region based approach is the accuracy of the underlying segmentation algorithm combined with the computational complexity of comparing sets of regions. The proposed combined approach using both global histograms and region based comparison seemed to perform well. Future work may include changes to the representation of each region and possible binning of regions to reduce the computational complexity when using region information.

While this paper focused on using the K-Median clustering for a variety of distance measures, applying the cluster results is also important and future work will include developing a visualization system that takes advantage of the clustered images to display a global overview. As has been shown in non-distributed environments, displaying a global overview by showing the relationships between representatives can be an effective browsing tool.

References

[1] I. Blanquer, V. Hernndez, and F. Mas. A P2P platform for sharing radiological images and diagnoses. In *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Sept. 2004.

[2] I. S. Dhillon and D. S. Modha. A data clustering algorithm on distributed memory multiprocessors. *Large-Scale Parallel Data Mining, Lecture Notes in Artificial Intelligence*, 1759:245–260, Mar. 2000.

[3] G. Fischer and A. Nurzenski. Towards scatter/gather browsing in a hierarchical peer-to-peer network. In *P2PIR'05: Proceedings of the 2005 ACM workshop on Information retrieval in peer-to-peer networks*, pages 25–32, New York, NY, USA, 2005. ACM Press.

[4] I. King, C. H. Ng, and K. C. Sia. Distributed content-based visual information retrieval system on peer-to-peer networks. *ACM Trans. Inf. Syst.*, 22(3):477–501, 2004.

[5] M. Li, I. K. Sethi, D. Li, and N. Dimitrova. Region growing using online learning. In *Proceeding of the International Conference on Imaging Science, Systems, and Technology*, June 2003.

[6] E. Loisant, R. Saint-Paul, J. Martinez, G. Raschia, and N. Mouaddib. Browsing clusters of similar images. In *Actes des 19e Journes Bases de Donnees Avancees (BDA'2003)*, pages 109–128, Lyon, France, October 2003.

[7] A. Ma and I. K. Sethi. Distributed k-median clustering with application to image clustering. In *The 7th International Workshop on Pattern Recognition in Information Systems (PRIS-2007)*, June 2007.

[8] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, 11(6), June 2001.

[9] W. Müller, M. Eisenhardt, and A. Henrich. Scalable summary based retrieval in P2P networks. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 586–593, New York, NY, USA, 2005. ACM Press.

[10] W. Müller and A. Henrich. Fast retrieval of high-dimensional feature vectors in P2P networks using compact peer data summaries. In *MIR '03: Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 79–86, New York, NY, USA, 2003. ACM Press.

[11] Y. Rui, T. Huang, and S. Chang. Image retrieval: Past, present and future. In *International Symposium on Multimedia Information Processing*, 1997.

[12] O. D. Sahin, A. Gulbeden, F. Emekci, D. Agrawal, and A. E. Abbadi. PRISM: indexing multi-dimensional data in P2P networks using reference vectors. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pages 946–955, New York, NY, USA, 2005. ACM Press.

[13] I. K. Sethi and I. Coman. Image retrieval using heirarchical self-organizing feature maps. *Pattern Recognition Letters*, 20:1337–1345, 1999.

[14] K. C. Sia. P2P information retrieval: A self-organizing paradigm. Technical report, UCLA, 2002.

[15] T. Sikora. The mpeg-7 visual standard for content descriptionan overview. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, 11(6), June 2001.

[16] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), December 2002.

[17] D. Stan and I. K. Sethi. eID: A system for exploration of image databases. *Information Processing and Management Journal*, 39/3, May 2003.

[18] C. Tang and S. Dwarkadas. Hybrid global-local indexing for efficient peer-to-peer information retrieval. In *In Proceedings of the Symposium on Networked Systems Design and Implementation (NSDI)*, June 2004.

[19] C. Tang, Z. Xu, and S. Dwarkadas. Peer-to-peer information retrieval using self-organizing semantic overlay networks. In *SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 175–186, New York, NY, USA, 2003. ACM Press.

[20] T. WANG, Y. RUI, and J.-G. SUN. Constraint based region matching for image retrieval. *International Journal of Computer Vision*, 56(1/2):37–45, 2004.

[21] Z. Yang. Interactive content-based image retrieval in the peer-to-peer network using self-organizing maps. In *HUT T-110.551 Seminar on Internetworking*, Apr. 2005.

Appropriate Cutting Segments of XML Elements for Multiple Keywords Queries

Jeang-Kuo Chen*, Jen-Peng Hsu
Department of Information Management, Chaoyang University of Technology
jkchen@cyut.edu.tw, s9314630@cyut.edu.tw

Abstract

Keyword Search is an effective approach for most users to search information because they neither need to learn complex query languages nor need to understand the underlying structure of data. This paper proposes a method to abstract appropriate segments of XML elements for multiple keywords queries. The user only needs to input several related keywords without describing the hierarchical relationship among these keywords. From the associated DTD of an XML document, our method can abstract several logical XML element segments called ACS-D's. These ACS-D's then can be used to identify some practical XML element segments, called ACS-X's, from the XML document. As a granularity of XML instance data, each ACS-X contains all inputted keywords and the hierarchy architecture of the keywords. Users can browse interesting information from the returned ACS-X's in an appropriate range. Finally, we give two examples to illustrate how our method works.
Keywords: XML, DTD, Granularity, Keyword search.

1. Introduction

Keyword search is a popular information discovery method because user can find out data without learning any query language or knowing the underlying structure of data. A general search engine in the internet can provide user to query data through inputting keywords to find out web pages while these keywords may have no relationship. Recently, the eXtensible Mark-up Language (XML) has become a popular medium for data exchange or storage. Therefore, keyword search on XML documents has become one of the important researches. To speed up query on large amount of data, one of the most efficient methods is to build a good index. A well-constructed index allows a query bypass the need of scanning the entire data for results [9]. One of the difficulties for keyword search is how to determine the returned granularity of the results. The returned results of an XML document search may be the whole XML document or part of an element content [11]. If the returned granularity is too small, then the data may be insufficient to the user. An example is illustrated as follows. Suppose that the user inputs a keyword "Dazhong" to query an XML document, as shown in Figure 1. Two "publisher" elements can be found since their contents include the string "Dazhong". If only the two "publisher" elements and their contents are returned, as shown in the solid line frames of Figure 1, the user cannot identify the two "publisher" elements whether they are the publishers of magazine or book. If the returned granularity is too large, it may contain too much redundant data. Suppose that the user inputs two keywords "Wireless Network" and "John Wu" to query the data and requests the returned results having to include the two keywords. The returned results may be the root element "library" and its all sub-elements, as shown in the dotted line frame of Figure 1. The user is hard to find the interesting data in such a big range. Therefore, it is worth to study how to determine the appropriate granularity of search results.

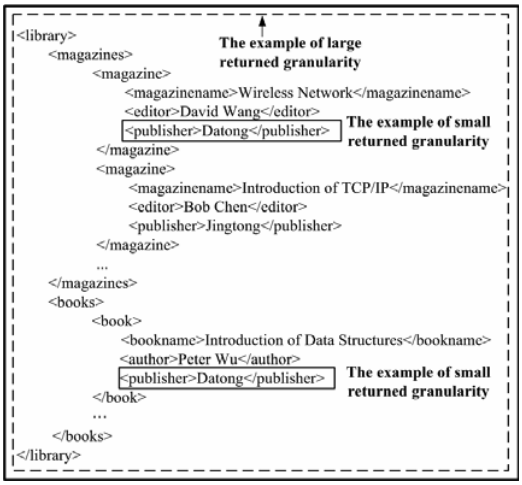


Figure 1. An XML document.

The motivation of this paper is to propose an approach to abstract an appropriate granularity for data query of multiple keywords in XML documents. The user only has to input several keywords to query data without describing the hierarchy among the keywords. Suppose that the keywords must appear in the contents of XML elements. From the associated DTD of an XML document, we can abstract several logical, tree-like XML element segments, called ACS-D's (Appropriate Cutting Segment of DTD). These ACS-D's can be used to identify some practical, tree-like XML element segments, called ACS-X's (Appropriate Cutting Segment of XML) from the XML document. As a granularity of XML instance data, each ACS-X contains all inputted keywords and the hierarchy of the keywords. Users can browse interesting information from the returned ACS-X's in an appropriate range. DTD or XML Schema can be used to produce ACS-D. Here we choose DTD because the structure of

DTD is easier to analyze than that of XML Schema. The name of an ACS-D root node may repeat. For example, the string "year" repeats at the root nodes of two different ACS-D's at the same level in the Figure 6. To solve the problem, we add two extra data items to the information of each ACS-D. One is the level value of each ACS-D root node. The other is the order value (from left to right) of the ACS-D root node among the siblings of the same level. To enrich the query result, we also consider that the XML element of complexity type may contain keywords. Thus, a special ACS-D is considered to have child nodes that are all complexity types rather than simple types. Our contributions are described as follows. First, we propose two algorithms and two ACS prototypes to abstract appropriate segments from XML documents. Second, our method excludes the problem of tag name repetition in DTD or XML Schema. Third, our method enriches the query result because not only the data in the contents of simple elements can be found, but also that of complexity elements can be found.

The remaining part of this paper is organized as follows. In section 2, we describe briefly the related work about XML keyword search. In section 3, we introduce the method of multiple keywords searching. In section 4, two examples are illustrated to show how our method works. Finally, we have a conclusion in section 5.

2. Previous Work

At present, there are two subjects for the research of XML keyword search. One is to add keyword search function to XML query languages, such as XML-QL, to facilitate data query [4,5,10]. The advantage of this way is increasing the accuracy of the search result while the defect is that the query syntax of XML query language is too complex to use for general users. The other subject is utilizing the characteristics of the XML document to perform keyword search [1,2,3,6,7,8,11,12]. Paper [3] designed a semantic search engine to search XML document data. This search engine has a simple query language suitable for general users and it returns semantically related document fragments that satisfy the user's query. XRANK system [6] can query XML and HTML documents and its DIL algorithm can find out all minimum elements containing all the keywords. Papers [1,2,7,8] provide the user to search data by inputting keywords on relational databases. In paper [2], a database is regarded as a graph of objects with edges representing relationships between the objects. Paper [8] designed the XKeyword System which is the extension of keyword search function in paper [7]. The XKeyword system can provide efficient keyword proximity queries on large XML graph databases. The ILE algorithm [12] can search for all the SLCAs (Smallest Lowest Common Ancestor) in an XML tree. The returned result is a set of minimum trees containing all keywords. There are two defects in [12]. First, the granularity of the returned result is large enough to contain much irrelevant data. Second, data may lose because some smallest trees are deleted accidently [11].

All the above mentioned researches of the second subject have certain contributions. However, the returned data is either too less or too much. In order to solve the problem of returned granularity, paper [11] proposed the concept of minimal information unit (MIU). When all keywords appear in any element content of an MIU, this MIU is regarded as a unit to return for user's reference. If keywords appear in different element contents of different MIU's, an ancestor-descendant relationship among these MIU's is checked whether the relationship exists or not. If existence, the involved MIU's are returned as a granularity to the user. The method [11] has some restrictions. First, it can apply to an XML Schema, but cannot apply to a DTD. Second, it uses the upper and lower limits of element occurrences in XML Schema to identify MIU's. If the numbers of element occurrences are too close to identify which one is dominant, the MIU is unable to be created.

3. The description of searching method

3.1. Structure of ACS

First we must define an appropriate granularity for returned data, called ACS (Appropriate Cutting Segment). The creation of ACS is described as follows. DTD or XML Schema can be used to produce ACS. DTD is selected because DTD structure is easier to analyze than XML Schema structure. Like the concept of MIU [11], an ACS is a certain element segment abstracted from a DTD. The structure of an ACS is an m-way tree with tree height 2. The abstraction of ACS depends on the type declaration of element in DTD. There are two prototypes of ACS. The first prototype is a general type with a non-leaf root and several leaf child nodes. For example, the subtrees circled by the dotted lines in Figure 2. The element corresponding to the root is a complexity type while that to the child node is a simple type. The second prototype is a special type with a non-leaf root and several non-leaf child nodes. For example, the subtree circled by a solid line in Figure 2. The elements corresponding to the root or the child nodes are all complexity types and the element corresponding to the root has no any simple-type sub-elements. The special type is considered to avoid losing data because keywords may all fall on complexity-type elements.

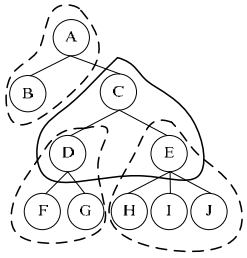


Figure 2. Two prototypes of ACS.

In addition to abstract ACS from DTD, we also need to abstract ACS from XML documents because real data exists in XML documents. This can be performed by using an ACS from DTD to match out several related ACS's from XML document. In order to distinguish an ACS derived from DTD or XML document, we call the former to be ACS-D and the latter to be ACS-X,

respectively.

3.2. Production of ACS

First we abstract several ACS-D's from a DTD that associates with the XML document. Delete some ACS-D's which do not logically contain all the inputted keywords. Then each ACS-D is used to match out several ACS-X's from the XML document. Likewise, some ACS-X's are deleted if they do not practically contain all the keywords. Three data structures are used to produce ACS-D. (1)T_D is an m-way tree derived from the DTD. The format of a node in T_D is (tag_name, ptr). The field tag_name is a tag name of an element. The field ptr is a pointer to a T_D node. (2)Q is a queue used to save each node in T_D for breadth-first search. (3)M_D is a 2D matrix for saving the original ACS-D's derived from DTD. Each row of M_D represents the information of a certain ACS-D. The schema of M_D is (LevelNo, SeqNo, rt, ct_1, ct_2, …, ct_m). The field LevelNo denotes the level number, monotonous increment from 1, of an ACS-D root, say, R. LevelNo is used to ignore some data to speed up search. The field SeqNo denotes R's order number, from left to right, of the nodes at the same level LevelNo. SeqNo is used to identify two nodes at the same level to have the same tag name. The field rt denotes the tag name of R. The other fields ct_i ($1 \le i \le m$) denote the tag names of an ACS-D child nodes. The following steps describe the process of ACS-D creation. Nodes in T_D are visited by breadth-first way.

Step1: Add the root node of T_D to Q.
Step2: If Q is empty, then stop.
Step3: Take out the first node P from Q; if P is a leaf node, then go to Step2.
Step4: Add all child nodes of P to Q from left to right.
Step5: If P has one or more leaf child nodes C, then P and C form an ACS-D tree structure; Add one row to M_D to represent the ACS-D and go to Step2.
Step6：If P's child nodes are all non-leaf nodes, then P and C form an ACS-D tree structure; Add one row to M_D to represent the ACS-D and go to Step2.

3.3. Multiple keywords searching

Given n keywords, k_1, k_2, ..., k_n, the following two kinds of ACS-X should be found and returned to the user. (1)All keywords exist in the same or several sibling nodes. As shown in Figure 3(a), each k_i ($1 \le i \le n$) may exist in a certain or several sibling nodes C_j ($1 \le j \le m$). (2)One keyword k_i exists in the parent node while the other keywords k_j ($1 \le i, j \le n, j \ne i$) exist in the same or several sibling nodes. As shown in Figure 3(b), k_i exists in P node and the others k_j exist in a certain or several sibling nodes C_l ($1 \le l \le m$).

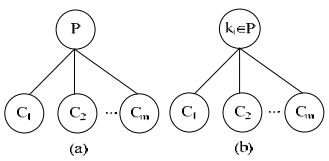


Figure 3. Two kinds of ACS-X.

An index tree T_X is built in advance to dynamically access XML element data. The schema of T_X node is (tag_ptr, ele_content_ptr, child_ptr_list). The field tag_ptr is a pointer to point to the tag name of an XML element. The field ele_content_ptr is a pointer to point to the content of an XML element. The field child_ptr_list is a pointer array to point to child nodes. A 2D matrix M_K (LevelNo, SeqNo, rt, ck_1, ck_2, …, ck_n) is used to record which ACS-D's include which keywords. The functions of the first three fields are the same as those of M_D. The other fields ck_i ($1 \le i \le n$) represent the keyword k_i appearing in the ACS-D. The process of finding out ACS-X in the XML document is described as follows.

Step1：For each keyword k_i, find the nodes m_i which includes k_i from T_X; Take out the tag names g_i of m_i to find out the corresponding ACS-D's D_i in M_D; Add the LevelNo, SeqNo, and rt values of D_i and k_i into M_K.
Step2：After all keywords are handled, delete some ACS-D's from M_K that do not contain all keywords.
Step3：Use the rows in M_K to find out the corresponding ACS-D's from M_D.
Step4：Use these ACS-D's to find out the corresponding ACS-X's from T_X and filter some ACS-X's which do not contain all the keywords; Finally, return all the remaining ACS-X's.

4. Example Demonstration

Two examples are illustrated to explain the process of multiple keywords search in an XML document. Figures 4(a)-(d) show an XML document, the related DTD, and their corresponding logically tree structures, respectively. The XML/DTD tree structure will substitute the original XML/DTD document hereafter for example explanation.

The index tree T_X for the XML document is shown in Figure 5. The tree structure T_D of DTD is shown in Figure 6 where the nodes circled by a dotted line denote an ACS-D. The abstraction of ACS-D is described as follows. Traverse T_D with the way of breadth-first search. First, visit the root node "bookstore" at level one. Because "bookstore" is a non-leaf node and its child node "category" is also a non-leaf node, both nodes form an ACS-D and one row ('1', '1', "bookstore", "category") for the ACS-D is added into M_D. Then visit the node "category" at level two. The node "category" is a non-leaf node and its child nodes "book" and "magazine" are also non-leaf nodes. Another ACS-D is abstracted and a row ('2', '1', "category", "book", "magazine") is added into M_D. The nodes in levels 3, 4, and 5 are checked with the same way. The final results of ACS-D are shown in Figure 7. Note that the two nodes at level 4 have the same name "year" in T_D. The SeqNo value can be used to distinguish the two different nodes.

The first example illustrates the case of all the keywords appearing in the contents of some simple elements. Assume that three keywords "XML", "David", and "Jingtong" are inputted to query data. First, find, from T_X, the nodes that contain "XML" and the answer is "bookname" as shown in Figure 5 (marked by a solid underline in subtree S). From M_D, find the ACS-D's that

contain the tag name "bookname". The ACS-D in the 7th row of M_D is found. The information of the ACS-D root node and the keyword "XML" are recorded as the first row ('5', '1', "book", "XML") into M_K. Next, find the nodes that contain the keyword "David" and the answer is "author" as shown in Figure 5 (marked by two solid underlines in subtrees S and U). From M_D, find the ACS-D's that contain the tag name "author". The ACS-D in the 7th row of M_D is found. The fact that ACS-D rooted at "book" contains keyword "David" must be appended into M_K. Therefore, the first row of M_K is changed to be ('5', '1', "book", "XML", "David"). Finally, find the nodes that contain the keyword "Jingtong" and the answer is "publisher" as shown in Figure 5 (marked by two solid underlines in subtrees S and T). From M_D, find the ACS-D's that contain the tag name "publisher". Two ACS-D's in the 7th and 8th rows of M_D are found. For the first found ACS-D, the first row of M_K is changed to ('5', '1', "book", "XML", "David", "Jingtong"). For the second found ACS-D, a new row ('5', '2', "magazine", "Jingtong") is added into M_K. M_K is checked when all the keywords are processed. The second row must be deleted since it does not contain all the keywords. The final M_K is shown in Figure 8. The data LevelNo (5) + SeqNo (1) + rt ("book") in M_K is used to find the complete ACS-D, the 7th row rather than the 3rd row, from M_D. With the data of the complete ACS-D, we can search T_X, beginning from level 5, to quickly find the corresponding ACS-X's. The results, two ACS-X's S, U, and their corresponding XML logical trees, are shown in Figures 9 and 10, respectively. The ACS-X U will be discarded because it does not contain keywords "XML" and "Jingtong" as shown in Figure 10. Finally, the XML element corresponding to the ACS-X S is returned to user as shown in Figure 11.

The second example shows the case of all keywords appearing in the contents of some complexity elements. Assume that two keywords "Computer" and "Language" are inputted. First, find, from T_X, the nodes that contain "Computer". The result is "category" as shown in Figure 5 (marked by a dotted underline in the first node at level 2). From M_D, find the ACS-D's that contain the tag name "category". The ACS-D's in the 1st and 2nd rows of M_D are found. For the first found ACS-D, a new row ('1', '1', "bookstore", "Computer") is added into M_K as the first row. For the second found ACS-D, a new row ('2', '1', "category", "Computer") is added into M_K as the second row. Next, find the nodes that contain the keyword "Language". The result is "category" as shown in Figure 5 (marked by a dotted underline in the second node at level 2). From M_D, find the ACS-D's that contain the tag name "category". Again, the ACS-D's in the 1st and 2nd rows of M_D are found. Therefore, the first and second rows in M_K are changed to ('1', '1', "bookstore", "Computer", "Language") and ('2', '1', "category", "Computer", "Language"), respectively. Since the two rows in M_K contain all keywords "Computer" and "Language", the final M_K is shown in Figure 12. The data LevelNo (1) + SeqNo (1) + rt ("bookstore") and LevelNo (2) + SeqNo (1) + rt ("category") in M_K are used to find the complete ACS-D's, the 1st and 2nd rows, from M_D. With the two complete ACS-D's, we can search T_X to find the corresponding ACS-X's. The ACS-D "bookstore" can be used to find an ACS-X while the ACS-D "category" can be used to find two ACS-X's. The three ACS-X's, A, B, C, and their corresponding XML logical trees are shown in Figures 13 and 14, respectively. The ACS-X's B and C will be discarded because they do not contain all the keywords "Computer" or "Language" as shown in Figure 14. Finally, the XML element corresponding to ACS-X A is returned to user as shown in Figure 15.

5. Conclusions

In this paper, we propose a method to abstract suitable element segments from an XML document as returned granularity for multiple keywords query. Our method has several advantages as follows. First, the user just only inputs the keywords without describing the hierarchical relationship among these keywords. Our method can automatically construct the hierarchy architecture of the keywords when the returned granularity is abstracted. Second, our method can be applied to both DTD or XML Schema. Therefore, our method still can be used to query the XML document that has only an associated XML Schema. Third, our method enriches the query result because not only the data in the contents of simple elements can be found, but also that of complexity elements can be found.

References

[1] S. Agrawal, S. Chaudhuri, and G. Das, "DBXplorer: A System for Keyword-Based Search over Relationship Databases," *Proc. of the 18th Int. Conf. on Data Engineering*, pp. 5-16, 2002.

[2] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S. Sudarshan, "Keyword Searching and Browsing in Databases using BANKS," *Proc. of the 18th Int. Conf. on Data Engineering*, pp. 431-440, 2002.

[3] S. Cohen, J. Mamou, Y. Kanza, and Y. Sagiv, "XSEarch: A Semantic Search Engine for XML," *Proc. of the 29th Int. Conf. on VLDB*, pp. 45-56, 2003.

[4] D. Florescu, D. Kossmann, and I. Manolescu, "Integrating Keyword Search into XML Query Processing," *Proc. of the 9th Int. World Wide Web Conf. on Computer networks*, Vol. 33, pp. 119-135, 2000.

[5] N. Fuhr and K. Grobjohann, "XIRQL: A Query Language for Information Retrieval in XML documents," *Proc. of the 24th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 172-180, 2001.

[6] L. Guo, F. Shao, C. Botev, and J. Shanmugasundaram, "XRANK: Ranked Keyword Search over XML Documents," *Proc. of the ACM SIGMOD Int. Conf. on Management of data*, pp. 16-27, 2003.

[7] V. Hristidis and Y. Papakonstantinou, "DISCOVER: Keyword Search in Relationship Databases," *Proc. of the 28th Int. Conf. on VLDB*, 2002.

[8] V. Hristidis, Y. Papakonstantinou, and A. Balmin, "Keyword Proximity Search on XML Graphs," *Proc. of the 19th Int. Conf. on Data Engineering*, pp. 367-378, 2003.

[9] M. Kaelin, Database Optimization: Increase query performance with indexes and statistics, TechRepublic, http://techrepublic.com.com/5100-10878_11-5146588.html?tag=search.

[10] A. Theobald and G. Weikum, "The Index-Based XXL Search Engine for Querying XML Data with Relevance Ranking," *Proc. of the 8th Int. Conf. on Extending Database Technology*, pp. 477-495, 2002.

[11] J. Xu, J. Lu, W. Wang, and B. Shi, "Efficient Keyword Search in XML Documents Based on MIU," *Proc. of the 11th Int. Conf. on Database Systems for Advanced Applications*, pp. 391-402, 2006.

[12] Y. Xu and Y. Papakonstantinou, "Efficient Keyword Search for Smallest LCAs in XML Databases," *Proc. of the ACM SIGMOD Int. Conf. on Management of data*, pp. 527-538, 2005.

```
<bookstore>
  <category>Computer
    <book>
      <year>2005
        <book>
          <bookname>Java and XML</bookname>
          <author>David</author>
          <publisher>Jingtong</publisher>
        </book>
        ...
      </year>
      ...
    </book>
    <magazine>
      <year>2006
        <magazine>
          <magazinename>Wireless Network</magazinename>
          <editor>Bill</editor>
          <publisher>Jingtong</publisher>
        </magazine>
        ...
      </year>
      ...
    </magazine>
  </category>
  <category>Language
    <book>
      <year>2006
        <book>
          <bookname>Studying Engilish</bookname>
          <author>David</author>
          <publisher>Daqian</publisher>
        </book>
        …
      </year>
      …
    </book>
    <magazine>
      <year>2007
        <magazine>
          <magazinename>Easy Speak English</magazinename>
          <editor>Mary</editor>
          <publisher>Tomorrow</publisher>
        </magazine>
        ...
      </year>
      ...
    </magazine>
  </category>
  ...
</bookstore>
```

Figure 4(a). An XML document.

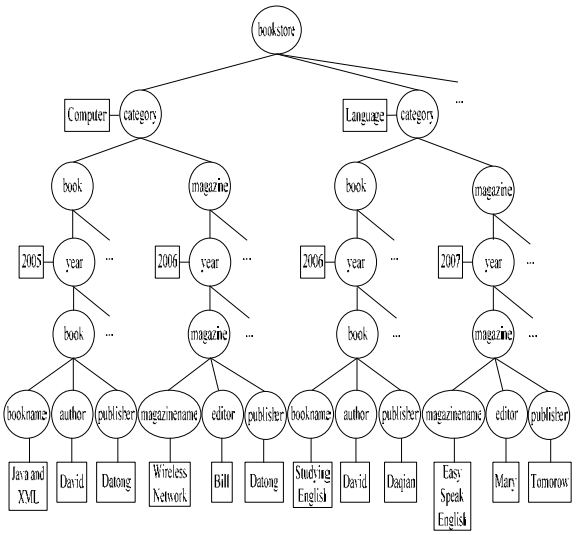


Figure 4(b). An XML tree structure.

```
<!DOCTYPE bookstore [
<!ELEMENT bookstore (category+)>
<!ELEMENT category (#PCDATA|book|magazine)*>
<!ELEMENT book ANY>
<!ELEMENT magazine ANY>
<!ELEMENT year ANY>
<!ELEMENT bookname (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT publisher (#PCDATA)>
<!ELEMENT magazinename (#PCDATA)>
<!ELEMENT editor (#PCDATA)>
]>
```

Figure 4(c). A DTD document.

Figure 4(d). A DTD tree structure.

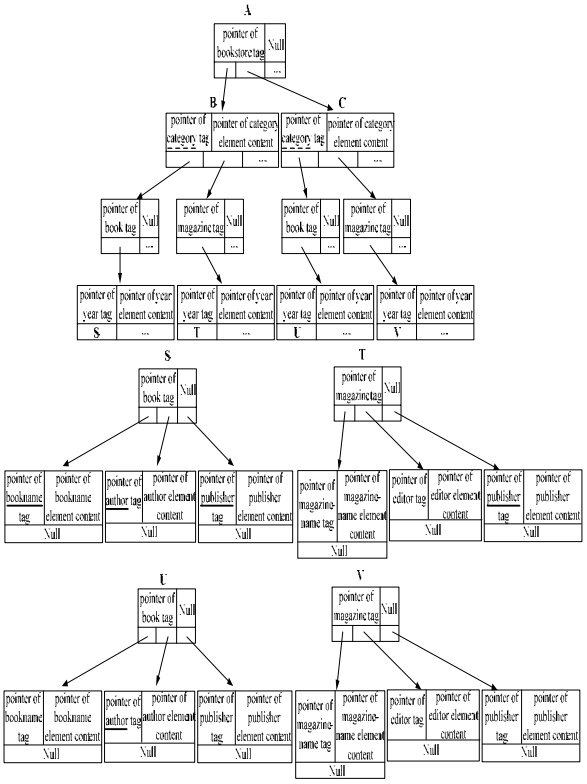


Figure 5. An XML index tree T_X.

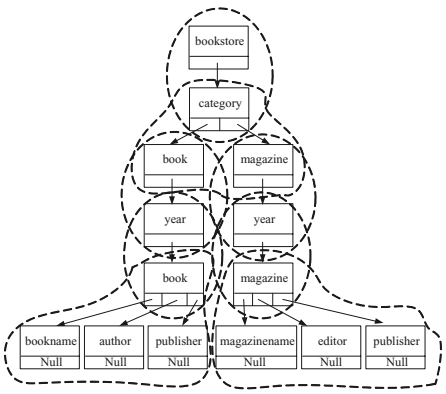


Figure 6. T_D of DTD.

LevelNo	SeqNo	rt	ct_1	ct_2	ct_3
1	1	bookstore	category		
2	1	category	book	magazine	
3	1	book	year		
3	2	magazine	year		
4	1	year	book		
4	2	year	magazine		
5	1	book	bookname	author	publisher
5	2	magazine	magazinename	editor	publisher

Figure 7. The ACS-D information in M_D.

LevelNo	SeqNo	rt	ck_1	ck_2	ck_3
5	1	book	XML	David	Jingtong

Figure 8. Final result of M_K.

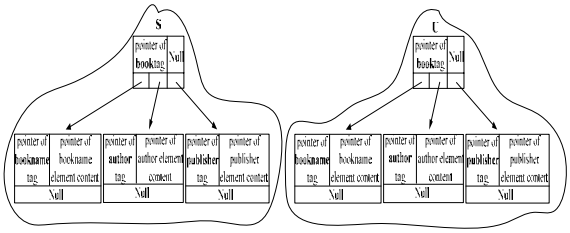


Figure 9. Two ACS-X's found in T_X.

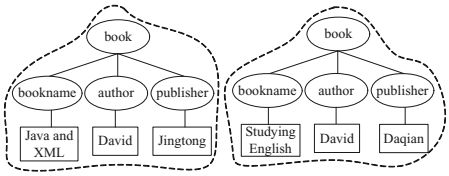


Figure 10. Two XML logical trees corresponding to S and U.

```
<book>
    <bookname>Java and XML</bookname>
    <author>David</author>
    <publisher>Jingtong</publisher>
</book>
```

Figure 11. Final returned result.

LevelNo	SeqNo	rt	ck_1	ck_2
1	1	bookstore	Computer	Language
2	1	category	Computer	Language

Figure 12. Final result of M_K.

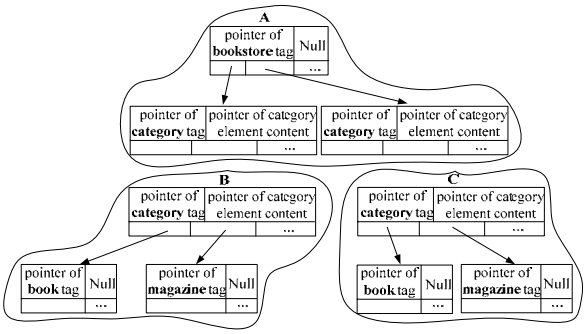


Figure 13. Three ACS-X's found in T_X.

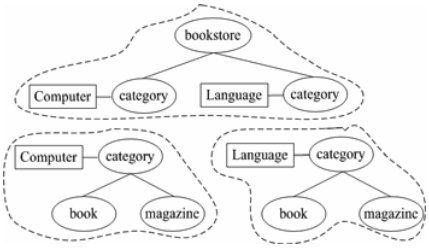


Figure 14. Three XML logical trees corresponding to three ACS-X's.

```
<bookstore>
    <category>Computer
        ...
    </category>
    <category>Language
        ...
    </category>
</bookstore>
```

Figure 15. Final returned result.

An Interactive Visualization System for Exploring Time-Series Data

Masahito Hirakawa, Kouji Nagamatsu[1], and Yuuki Ito[2]

Interdisciplinary Faculty of Science and Engineering, Shimane University
1060 Nishikawatsu, Matsue 690-8504, Japan
hirakawa@cis.shimane-u.ac.jp

Abstract

Management of time-series data is indispensable for future computing. In fact, lifelong is a keyword towards realization of high-quality health and social services in the coming years. We have therefore researched spiral-based information visualization for years. This paper presents several extensions of our previous spiral-based visualization. A new visualization scheme which visualizes both fluctuations of event occurrences and data values is proposed. In addition, the system provides a competence of mining a given data set for valuable features. Integration of spiral-based visualization and data mining helps the user to find interesting periodic and/or cyclic patterns in time-series data.

1. Introduction

Various sensors are available nowadays. They are connected and organize a sensor network which works to capture the information of an environment the user is situated, resulting in the realization of sophisticated and situated interaction between the user and the computing system. For example, it becomes possible the system turns the cellphone ring tone off by knowing that a meeting is going to start.

It is noted here the information from such sensors is a sequence of observations which are ordered in time. Of course, there exist a more variety of time-series data in, for example, finance, sales, industry, agriculture, health, physics, and meteorology [1].

Researchers have tackled for long years to have a deep understanding of time-series data [2]. One of the attractive features is a trend, where most attention has been paid to the underlying direction (an upward or downward tendency) and rate of change. In this paper, we focus on cyclic patterns which are another interesting feature and specific to time as we spend our time in a daily, weekly and monthly manner.

Statistical analysis and data mining techniques have been applied widely to get finding of cyclic features. They work effectively in such applications where the size of data is huge. However, capabilities of the existing techniques are still limited. We as human beings have better performance in finding meaningful patterns observed in time-series data under the situation that such data are presented in a sophisticated form. It would be reasonable to build a computing environment where the user taking a pattern cognition role and the system taking a data analysis role work together to see the good in others.

In this paper, we propose a visualization system for exploring time-series data with a capacity of mining them for valuable features. First, we extend the idea of spiral-based visualization we presented in [3] to enable the user to see both fluctuations of event occurrences and data values. A form of spiral staircases is adopted in this visualization scheme, where the interval of spiral circles changes depending on the period of event occurrences. Next, to cope with a large data set, a data mining technique is introduced. Possible trends and patterns in time-series data are extracted by the system. The user is freed from paying attention to all the data.

The rest of the paper is organized as follows. In Chapter 2, we refer to several related works as well as our previous work which gives a basis of this study. Chapter 3 discusses a basic idea of the proposed spiral-based visualization. Interactive functions provided in the system are also described. Chapter 4 concerns data mining facilities implemented in the system. Application of the system to dialysis treatments domain is presented in Chapter 5. Finally, in Chapter 6, we conclude the paper.

2. Visualizing Time-Series Data

2.1 Related work

Visualization of time-series data has been explored for a long time. The most traditional visualization techniques include sequential charts, point charts, bar charts, line graphs, and circle graphs.

[1] He is now with Murata Machinery, Japan. This work was done when he was a student at Shimane University.

[2] He is now with Techno Project, Japan. This work was done when he was a student at Shimane University.

Other modern techniques for visualizing dynamic visual patterns over time include information flocking boids [4], illustration-inspired techniques such as speedlines, flowribbons and strobe silhouettes [5], and animated 3D models [6]. However, all these approaches are focused on serial trends as handled in traditional visualization schemes.

Little study has been done for exploring periodic data, even though the periodicity is an important feature in our living. In [7], [8] a planar spiral, which is the path of a point in a plane moving around a central point while continually receding from or approaching it, is applied for visualization of a periodic structure in time-series data.

Here it is noted there are two types of periodic data [7]. One is pure serial periodic data having a single, continual dimension in which each period has equal duration. The other is event-anchored serial periodic data. Event-anchored data has periods with different durations, and the start of a new period is indicated by an event.

While trials in [7] and [8] are designed for pure serial periodic data, we are interested in event-anchored periodic data and have investigated so far, as explained in the following subsection.

2.2 Our previous work

We have developed a spiral-based visualization system [3]. It depicts periodic features in time-series data by two visualization expressions: a planar spiral and a helix spiral. The helix spiral is a coil formed by a wire around a tube. Although these two spirals are conceptually described as distinct objects, they can be presented as two different views of a single object. Specifically, the top view of a helix spiral presents its planar spiral.

Multiple spirals with colored square icons which represent events can be displayed in one window, as shown in Fig. 1. Linear views are given in the figure, as well as the planar and helix spiral views.

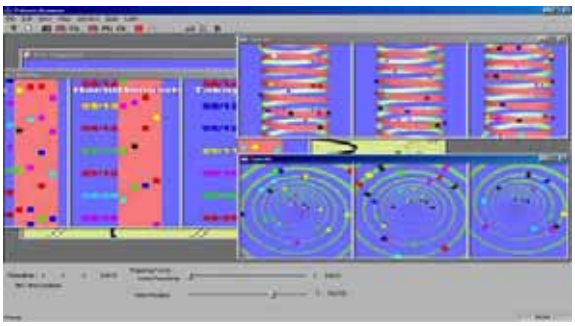


Figure 1 Multiple views of time-series data

A distinguished idea of our previous work is the capability of visualizing non-uniform periodic patterns. Specifically, the target events (i.e., icons) satisfying a certain condition can be aligned by properly adjusting the radius of each lap, as shown in Fig. 2. If the radius is short, it means the corresponding events occur frequently, and vice versa. As seen, this non-uniform spiral expression helps the user to find unique temporal patterns.

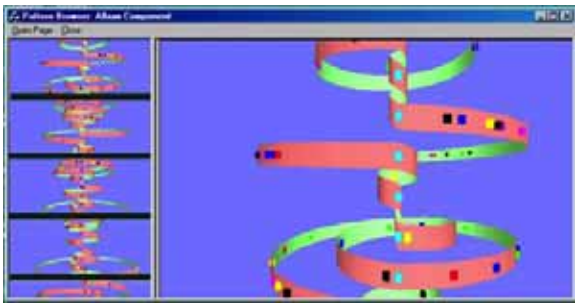


Figure 2 Non-uniform spiral

3. Extended Spiral-based Visualization

3.1 Basic idea

In the previous work, we focused on event occurrences since our main interests laid in periodic and cyclic patterns. Limited visual attributes (for example, color and shape of icons) are assumed for representation of other data properties. While this schema works properly, we were asked whether quantitative data can be presented as bars on a spiral or any other expressions. It would not be impossible to attach bars or protrusion, like in a 3D map, on the spiral. But, we were afraid that such extension disturbs the visibility of spirals, and decided to modify the visual schema.

Our proposal here is that the time dimension is presented by a time trunk. The time trunk grows vertically as time goes. Data elements are represented in a form of bars which encircle the time trunk. Their values correspond to the length of bars. Ideally, the shape is like a spiral staircase.

Figure 3 shows an example of the proposed visualization. The time trunk is the one which is placed in the center. Timeline is represented by a black line which winds around the time trunk. Events are expressed by icons. Here an event may last for a certain period. To make this visible, a line indicating the time duration is placed on the time trunk.

Like the previous work, the events satisfying a specified condition can be aligned. Vertical intervals are adjusted properly to meet constraints, resulting in the non-uniform winding of timeline, as shown in Fig. 4. The timeline is tightly winded when the associated events happen with a short interval, and vice versa. The time

trunk is colored for assistance depending on the time interval, that is, how often events occur. Red is for short interval, while blue for long interval. Meanwhile, in this figure, interpolation is applied for the given data elements to help the user recognize the change of values.

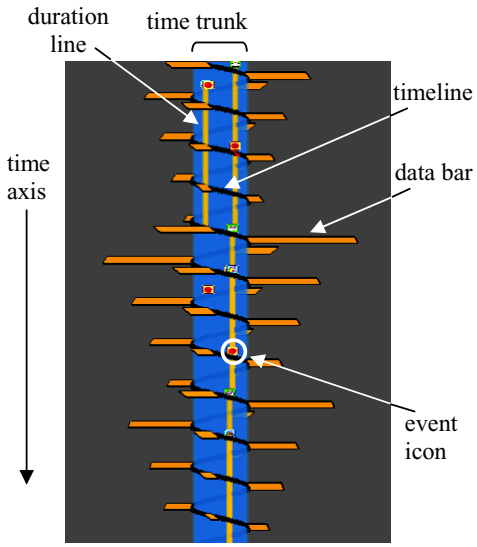


Figure 3 Basic representation

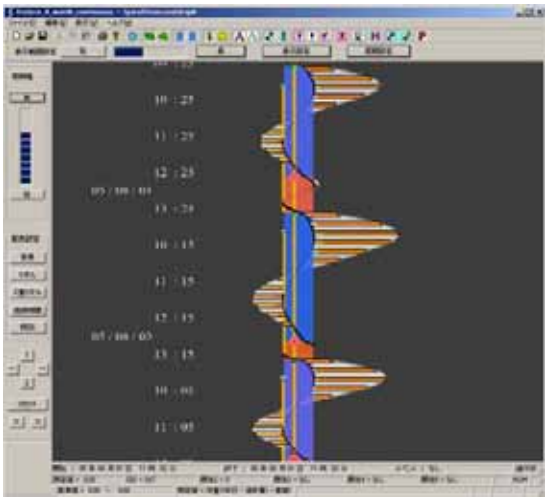


Figure 4 Alignment of certain event icons

The visualization example in Fig. 4 can be read as a short interval follows a long interval and their combination happens repeatedly. The user may wonder what the meaning of the combination is and check values at the boundaries to see whether there are any significant features. Actual application of the system to a medical domain will be explained in Section 5.

3.2 Visual options

The system provides several interactive facilities for helping the user to explore the given data, as listed below.

1) Change of views: A pseudo-planar view and a linear view of the spiral can be presented as well as a helix view, as shown in Fig. 5.

Though the pseudo-planar view corresponds to a cross section of the helix spiral and thus doesn't show all the data elements, the user can move the view forward and backward interactively.

Meanwhile, the linear view is interpreted as a stretch version of the spiral. A blue lane on the left corresponds to the time trunk (timeline). Data elements are presented on its right. Those elements may have an upper part value and a lower part value like the systolic and diastolic blood pressures. The linear view is suitable for the observation of a change of data values locally and comparison among multiple data sets

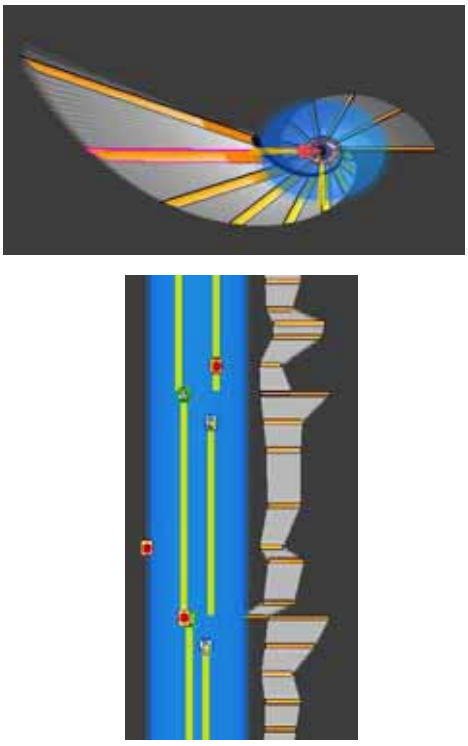


Figure 5 Pseudo-planar and linear views

2) Display of normal values/range: For a given data set, a normal value/range may be defined. Recommended blood pressure level is an example. The system can

present the normal value/range to help the user identify whether data values are good or not.

3) Multiple spiral display: The user would like to compare multiple spirals to see differences among them. The system allows the user to present multiple spirals at once. Further, the user is allowed to move their position up/down separately and overlap a spiral onto another.

4) 3D projection: In the visualization process, 3D coordinate values of a spiral are projected onto a 2D plane. The system adopts two transformations for this: parallel transformation and perspective transformation. Figure 6 shows an example of perspective transformation view, while the images in Figs. 3 and 4 are of parallel transformation.

There is a fact that perspective makes visual elements far away seem small. This may cause a trouble in comparing data elements. Therefore we adopt the parallel transformation as a default. Meanwhile, as you see in Fig. 6, perspective transformation overstates a change in data values and is good for evaluation sometimes.

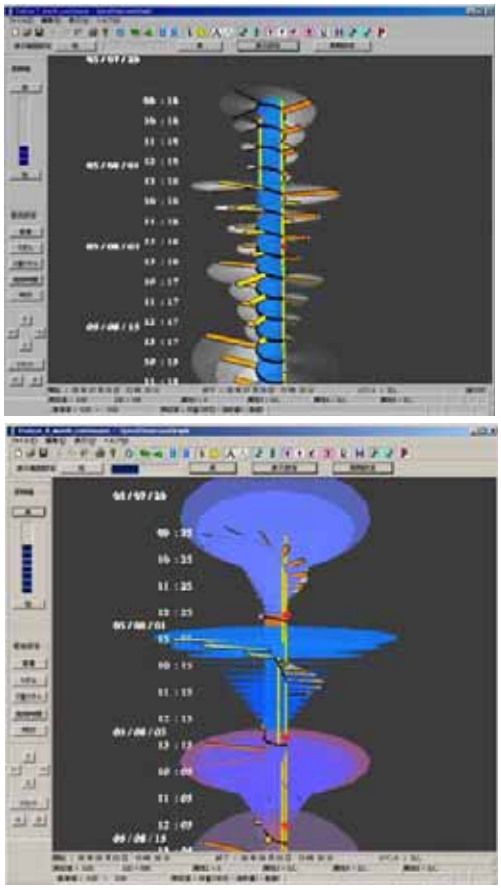


Figure 6 Examples of perspective transformation view

4. Assisting Pattern Discovery

Because of the fact that a time-series dataset contains a huge number of data objects in general, it is not practical to ask the user take a full responsibility for exploring all the data and identifying trends/patterns manually, even though a powerful and sophisticated visualization tool is available. We therefore implement a mechanism to extract underlying features and help the user to find meaningful trends and patterns with less effort.

Knowledge discovery with sophisticated interactive visualization is promising especially in such an environment that patients are responsible for managing their own health condition. Here temporal abstraction is a well-known technique in knowledge discovery or data mining, where time-series data are converted into a set of interval-based patterns [9].

One issue in realizing temporal abstraction is to decide time periods for pattern identification. Here it is not reasonable to decide time periods in advance. The system should rather provide a mechanism for allowing the user to control them on demand. Further, it is expected that each time period for pattern identification can differ since our behavior or physical condition, for example, is not stable and varies time by time.

According to this observation, we provide a mechanism of time period adjustment. The user first specifies a lower limit T_l and an upper limit T_u of time periods. The system repeats to have a time period lengthened from its minimum time period T_l by adding a certain time unit as long as its pattern (explained below) stays in the initial one.

For example, in Fig. 7, assume that t_s is the starting point and the initial pattern organized by the time period T_l is P_1. If the pattern changes from P_1 to another in the case when the time point exceeds t_p, the time period in this trial is determined to T_p.

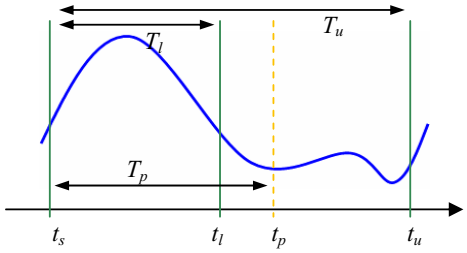


Figure 7 Adjustment of time periods

Next, for identification of patterns, we refer to the temporal abstraction primitives in [9] and define the following ones. Table 1 summarizes possible 32 abstraction patterns in reference to these primitives.

- State primitives: normal (N), low (L), and high (H)
- Trend primitives: stable (S), increasing (I), and decreasing (D)
- Peak primitives: peaks occurred (PF)

Table 1 Abstraction patterns

Pattern #	pattern	Pattern #	pattern
1	H & I	17	N -> L -> N
2	H & S	18	L -> N -> H
3	H & D	19	L -> N -> L
4	N & I	20	(H -> N) & I
5	N & S	21	(H -> N) & S
6	N & D	22	(H -> N) & D
7	L & I	23	(N -> L) & I
8	L & S	24	(N -> L) & S
9	L & D	25	(N -> L) & D
10	H -> N	26	(H -> N -> L) & I
11	N -> H	27	(H -> N -> L) & S
12	N -> L	28	(H -> N -> L) & D
13	L -> N	29	H & PF=N
14	H -> N -> H	30	N & PF=H
15	H -> N -> L	31	N & PF=L
16	N -> H -> N	32	L & PF=N

In addition, we prepare the following nine patterns based on min-max positions (see Fig. 8). The user can choose any pattern in Table 1 or Fig. 8, depending on what features the user is interested in.

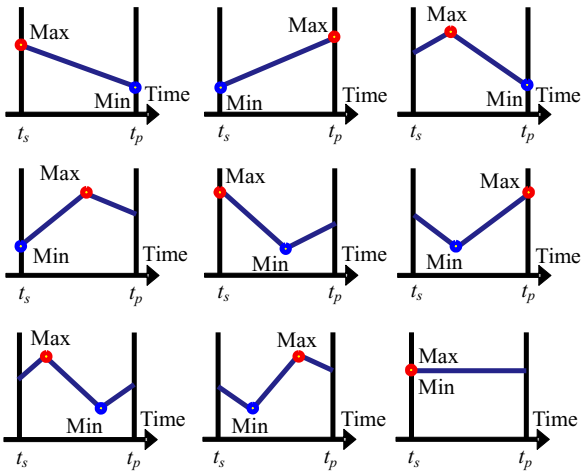


Figure 8 Patterns based on min-max positions

According to the pattern specified by the user, the system highlights the periods where the specified condition meets. Here it is allowed for the user to specify a series of patterns to be satisfied.

5. Application Scenario

We have applied the proposed visualization system for the analysis of data in dialysis patients. Here we focus on the relationship between blood pressure and feeling bad, since it is the most concerned issue in that domain. The following are some of the observations we have obtained through experiments.

Figure 9 shows a result explaining how blood pressure of a patient in each day changes. Green color belts tell the periods where the upper right pattern in Fig. 8 meets, that is, blood pressure goes down after a peak value. Mostly, the patient felt bad at such periods.

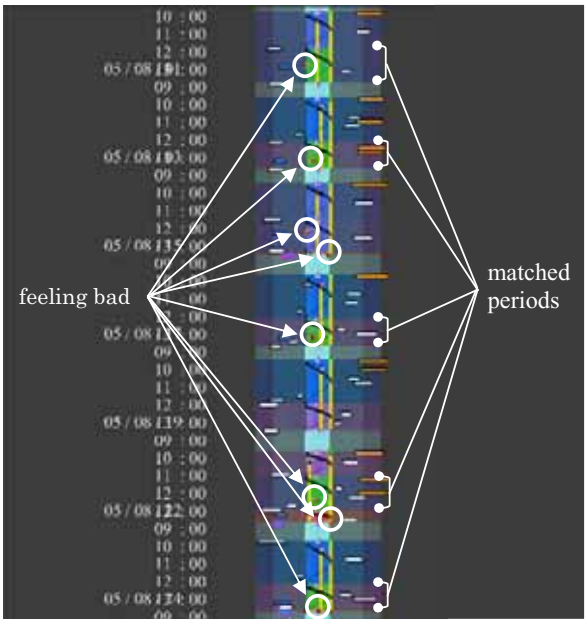


Figure 9 A spiral highlighting a relationship between blood pressure and physical condition

Figure 10 gives another observation which is obtained by applying temporal abstraction pattern 3 in Table 1. Highlights in green color, which correspond to this pattern, present that blood pressure (systolic) at the starting point is higher than a certain threshold and then gradually goes down along the treatment. The patient rarely feels bad in such case, while there are high possibilities of having trouble in other cases. Labels indicating the abstraction patterns are placed next to the time trunk for reference.

The experiments we have done so far are still in a preliminary level, and their feasibility must be checked carefully under the supervision of medical doctors.

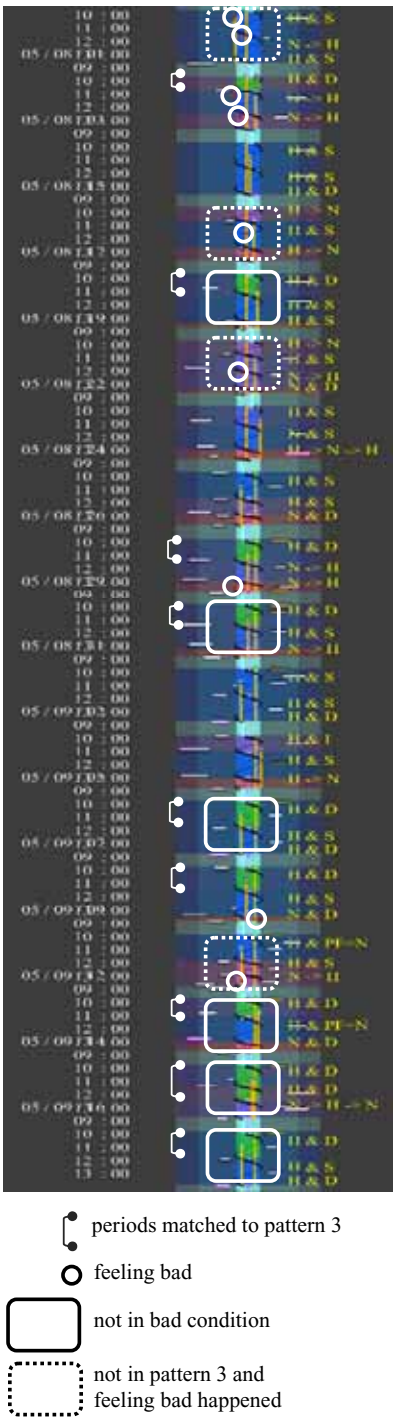


Figure 10 Result of applying the temporal abstraction pattern 3 in Table 1

6. Conclusion

In this paper we proposed techniques of visualizing time-series data in a 3D spiral form, aiming at spotlighting periodic aspects of such data. Specially, the system enables the user to observe fluctuations of event occurrences with data value change. In addition, the system provides a capability of mining a given data set to find specific trends/patterns. Integration of the spiral-based visualization and data mining facility helps the user to recognize interesting patterns hidden in the large data set. Application of the system to dialysis treatment data was also presented to demonstrate the usefulness of the proposed ideas.

References

[1] http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/
[2] http://www.stats.gla.ac.uk/steps/glossary/time_series.html
[3] Hewagamage, K. P., Hirakawa, M., "An Interactive Visual Language for Spatiotemporal Patterns," Int. J. Visual Languages and Computing, 12, pp.325-349, 2001.
[4] Moere, A. V., "Time-Varying Data Visualization using Information Flocking Boids," Proc. IEEE Symp. Information Visualization, pp.97-104, 2004.
[5] Joshi, A., Rheingans, P., "Illustration-inspired Techniques for Visualizing Time-varying Data," Proc. IEEE Conf. Visualization, pp.679-686, 2005.
[6] Farrimond, B., Hetherington, R., "Using 3D to Visualise Temporal Data," Proc. IEEE Symp. Information Visualization, pp.108-117, 2004.
[7] Carlis, J. V., Konstan, J. A., "Interactive Visualization of Serial Periodic Data," Proc. ACM UIST, pp.29-38, 1998.
[8] Weber, M., Alexa, M., Muller, W., "Visualizing Time-Series on Spirals," Proc. IEEE Symp. Information Visualization, pp.7-14, 2001.
[9] Nguyen, T. D., Kawasaki, K., Ho, T. B., "Discovery of Trends and States in Irregular Medical Temporal Data," Lecture Notes in Computer Science, Vol. 2843, Springer Verlag, Berlin Heidelberg New York, pp.410-417, 2003.

Retrieving time stamps for film scripts*

Carlos Teixeira
Lasige/DI/Universidade de Lisboa
Dept. Informática, FCUL, Campo Grande
1749-016 Lisboa, Portugal
cjct@di.fc.ul.pt

Ana Respício
CIO/DI/Universidade de Lisboa
Dept. Informática, FCUL, Campo Grande
1749-016 Lisboa, Portugal
respicio@di.fc.ul.pt

Abstract

Information retrieval has provided us powerful tools that can be useful for text alignment. A new parallel text alignment algorithm is presented in this paper. Perhaps the most relevant dimension lost in the film script is the real time connection with the film itself. The proposed algorithm was implemented in order to retrieve the subtitles time stamping for each dialogue line. This provides detailed alignments with the time signals that compose the film. Preliminary experiences revealed that this algorithm presents very promising results.

1. Introduction

Several types of textual information are produced during the process of films creation and commercialization. Besides video and audio signal, many programs and film distributions include subtitles. Not matching exactly speech transcription, as it would be done with an idealistic automatic speech recognizer (ASR), these transcriptions are often done by a human transcriber, which main concern is to keep the subtitle semantic value according to the available slot of time and visual space. Including subtitles serves mainly as a suitable form of presenting language translation or as a substitute of audio signal in any conditions where audio is not available. These conditions includes problems like availability of the audio signal, channel transmission, listeners environment (i.e. noisy or bad sound propagation) or the listener himself (i.e. deaf). On the other hand, if for some reason access to visual information is not allowed, translated subtitles could be used by a text to speech (TTS) system.

Additional real-time information can also be useful to overcome restrictions in visual or audio access. However, most of this additional information is not time aligned with the film itself. The film script is usually available as a separated text document which can be read apart from the film visioning. The script is structured into scenes and includes information from both dialogue and action. Dialogue lines are usually very close from what can be found in subtitles, being sometimes even closer to what was actually spoken in the audio signal. Each dialogue line begins with the identification of the corresponding character, which is usually unnecessary for the subtitles. Sometimes, a description of character behavior (visual expression and type of speech, etc.), for specific parts of discourse, is described between dialogue lines. Action information includes scene and shots identification, and several descriptions of the visual and audio scenario as well as the position of the characters. Often camera directions as well as shot boundaries are also included. Thus, character identification and its behavior while speaking dialogue lines, together with action information, can be a contribution from the script which is not available in the film and its subtitles. There is no explicit link between script and video, as the first one does not include quantitative time references. If time synchronization is made possible between these several streams, extra information can bring advantages in other usage contexts, as it has been done with subtitles.

Two main types of applications are envisaged, that will make use of the resulting aligned script. The first type will rely on the audio as a substitute of video signal in any conditions where video is not physically available. Similarly, as it was mentioned for audio, these conditions include problems with the availability of video signal, channel transmission, environment of the viewer (i.e. driving) or the viewer himself (i.e. blind). In such cases, information contained in the script could be helpful if transmitted synchronously with the use of TTS technology. The second type of applications aims to provide an integrated view of the process of film production, namely about what was originally in the mind of the scriptwriter. Current technology allows to foresee alternative views of the script as well as other

*This work was partially supported by FCT through the Large-Scale Informatic Systems Laboratory and Centro de Investigação Operacional – POCTI/ISFL/152.

texts, namely the possibility to access these documents synchronously with the visioning of the concerning film. The integration and navigation of video and other media is the central aspect in hypervideo, where video is not regarded as a mere illustration, but can also be structured through links defined by spatial and temporal dimensions [1].

The present work proposes an approach for automatic alignment of film script with its video signal, producing a synchronized script, and its integration in the visual domain together with extended browsing and querying facilities over the film. This paper is organized as follows. Next section summarizes previous related work. Section 3 describes the general approach for obtaining the time stamped script and the methodology used for parallel text alignment, focusing on the preprocessing of text pieces and on the alignment algorithm. Preliminary experimental results are given. Section 4 presents the tool developed for an enriched visualization of video. Finally some conclusions and perspectives of future work are presented.

2. Retrieving Related Contents for Video

Several approaches have been published for aligning text contents associated with video. A framework for aligning and indexing movies with their scripts is proposed by Ronfard and Thuong [5]. They propose a grammar for the script of a given movie. Structural entities, such as shots, scenes, actions and dialogs are extracted. Subtitles are extracted from the video stream using optical character recognition, producing a stream of time-stamped short texts. The alignment is performed by searching for the longest increasing subsequence of matched shots and subtitles. Besides subtitles and scripts, TV programs are sometimes transcribed for documentation on web sites. Gibbon [2] proposes an approach to use this type of transcriptions aligned with the video signal, to create improved hypermedia web pages for those programs. The time alignment is found by doing parallel text alignment of the transcripts with the subtitles of the same program. Gibbon refers difficulties arising at many-to-one sentence mappings. Text alignment is first done at the word level using a dynamic programming (DP) procedure. In a second stage, an alignment at the sentence level is searched for.

Aligning the script dialogues with closed captions is the basis to address character/speaker identification in [7]. A DP procedure finding the "best path" across a similarity matrix allows for defining an alignment at the word level. The approach combines the text alignment with audio segmentation to accomplish audio speaker identification.

Martone et. al. [4] propose the off-line generation of closed captions by synchronizing program transcripts with the texts (subtitles) produced by an ASR system. A DP procedure similar to the one used by Gibbon [2] is applied to perform text alignment at the word level. More recently, Martone and Delp presented two systems based on the alignment of program transcripts with ASR results [3]. One of them allows for locating speakers in news program, while the other one generates DVD chapter titles using the information of the closed captions.

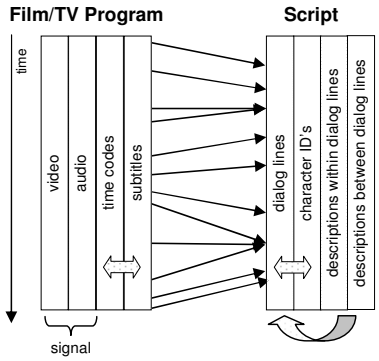


Figure 1. Aligning film with its script.

3. Parallel Text Alignment

In this section an approach is proposed for obtaining the time stamped script (TSS). The description of the global approach must consider two main blocks of information which are represented in figure 1: time synchronized information (video and subtitles); and the textual information structured according to the time line but without any detailed anchors to the real time of the movie (script). The arrows intent to represent the above mentioned anchors. In the present work, these time anchors are built exclusively between the subtitles and the dialogue lines. Also considering future alignments with texts with different time structures, a new text alignment algorithm is proposed in this paper.

The process begins by the preprocessing of two texts in order to obtain two streams of word sets: dialogs and subtitles. Then, the core alignment follows including two main phases. In first phase, the algorithm computes a heuristic assignment of subtitles to dialogs that maximizes the global similarity. The second phase searches iteratively for merging pairs of consecutive subtitles that produce local improvements. The last iteration occurs when no further improving merge can be found. In the end, due to several reasons explained in section 3.2, a small set of subtitles and dialogue lines remain unassigned or out of expected time order. A final post-processing phase (third phase) is used to assign these subtitles with unassigned dialogue lines. Each of these dialogue lines should be surrounded by already assigned dialogue lines that encompass the time order where

```
function one2one_assignment(st, dl, sim)
  pairs ← sort(sim) {sort pairs (i,j) in decreasing order of sim[i,j]}
  k ← 1
  repeat
    if is_free?(st_in_pair(k)) ∧ is_free?(dl_in_pair(k)) ∧
    sim[st_in_pair(k), dl_in_pair(k)] > α then
      assigned[st_in_pair(k)] ← dl_in_pair(k)
    k ← next k
  until all pairs were analyzed
  {k = m × n ∨ sim[st_in_pair(k), dl_in_pair(k)] ≤ α}
  return assigned

main ()
  Input st[1..m] {list of subtitles }
  Input dl[1..n] {list of dialogs }
  assigned[1..m] ← 0 {assignments of subtitles}
  {first phase}
  sim[1..m, 1..n] ← compute_sim(st, dl) {matrix of similarities }
  assigned ← one2one_assignment(st, dl, sim)
  {second phase}
  repeat
    improvement ← false
    i ← 1
    while i ≤ m ∧ ¬improvement do
      max_sim ← max{sim[k, assigned[k]], k = i, i + 1}
      best_s ← s : max_sim = sim[s, assigned[s]], s = i, i + 1
      merged_s ← merge(st[i], st[i + 1])
      if max_sim < compute_sim(merged_s, assigned[best_s])
      then
        st[i] ← merged_s {merge two consecutive subtitles}
        st ← remove_subtitle(i + 1){update subtitles list}
        sim ← update_similarities(st, dl, i)
        assigned ← one2one_assignment(st, dl, sim)
        improvement ← true
      i ← next i
  until ¬improvement
```

Figure 2. Core alignment procedure.

the concerned subtitle occurs. Finally, the resulting assignment is used to link time codes to the dialogs in the script providing the TSS.

3.1. Text Alignment Algorithm

Each dialog line and each subtitle is represented by a sequence of tokens. A token is defined here as a sequence of alphanumeric characters occurring between spaces or punctuation marks. A single token lexicon is build which will allow to refer to every token with an unique index. Descriptions of an emotion to be expressed by the character or a special setting of the scene, frequently appear as text between parentheses in the script. These may occur within and between dialog lines and are ignored in the alignment. Thus, each subtitle or dialogue line can now be represented by the respective set of indexes. As the vector space model will be used, the order of the tokens is not considered, and this set is represented in a vector with same size as the total number of tokens in the lexicon. Accordingly, each element of this vector contains the number of occurrences for a specific lexicon entry, weighed by its length.

Many to one assigning of subtitles to dialog lines must be produced. The algorithm proposed here uses the similarity measure of the vector space model [6] which has been adopted by many applications in the information retrieval area. Probably, the most relevant results were found in the text retrieval area where the model proves to be relatively language independent. One of the characteristics that can be seen as a problem is the fact that this model does not take into account the order of the tokens inside a document. However this did not seem to compromise the success of this model with such linear structures such as language.

A formal pseudo-code description of the core alignment procedure is given in figure 2. The algorithm aligns two lists: m sets of tokens from subtitles $st[i]$ and n sets of tokens from dialogue lines $dl[j]$ – typically $m > n$. The similarity of each subtitle with each dialog line is kept in a $m \times n$ similarity matrix and is given by:

$$sim[i,j] = \frac{st[i] \cdot dl[j]}{\|st[i]\| \ \|dl[j]\|}, \tag{1}$$

where $st[i]$ is the vector representing the i-th subtitle and $dl[j]$ is the vector representing the j-th dialog line.

The result of the core alignment procedure is the assignment of subtitles to dialogs given in a list denoted by $assigned$, where $assigned[i]$ contains the index of the dialog to which the i-th subtitle is assigned. Some subsequences of subtitles are eventually merged during the execution of the process and the current list st is updated accordingly.

In first stage, one-to-one assignment is performed by a greedy heuristic. The matrix of similarities sim is computed using equation (1). The values in this matrix are ranked building a list of ordered pairs (subtitle, dialog) sorted by decreasing order of the corresponding similarities.

Then, iteratively, assignments are made by looking first for the pairs with higher similarity values. The matching is done only if both elements of the pair were not previously assigned and if their similarity value is greater than a given threshold α. In that case, the $assigned$ list is updated accordingly. This cycle continues until a pair with a similarity value lower or equal to the threshold is found or all the pairs have been analyzed.

The number of subtitles is usually greater than the number of dialogs and, consequently, many subtitles remain unassigned by this algorithm. Considering a minimum value of similarity to accomplish a matching also increases the number of unassigned subtitles. Also, potentially, in the result, two consecutive subtitles may be linked to two dialogs located apart.

On the other hand, the process described above does not take into account the order of subtitles and dialogs in their respective streams. Considering these order restrictions in the time neighborhood of each subtitle, and to overcome the above mentioned problems, an improvement phase is executed. Iteratively, the algorithm searches for pairs of con-

secutive subtitles that can be merged together leading to an increased value of local similarity. This is done by going through each subtitle i and evaluating the impact of merging it with its successor. The merging of two subtitles is considered if it leads to a local improvement. In that case, the list st is updated, replacing the two sets of tokens by the concatenation of both into a single set. Finally, a similarity matrix is updated for another one-to-one assignment run. The whole process repeats until no eligible merges can be found.

3.2. Preliminary Experimental Results

The approach has been tested with episodes of a series produced by RTP[1], one of the Portuguese major television channels. The episodes have a duration of 45 minutes. The script of each episode includes an average of 3500 words, comprising an average number of 200 dialogue lines. The average number of subtitles for each episode is around 300. The algorithm presented above was coded in the C programming language and preliminary experiences revealed very promising results.

Figure 3 illustrates the evolution of the alignment between subtitles and dialogs during the execution of the algorithm for a given episode. The first plot is the result of the initial one-to-one assignment. Each point represents for each subtitle, in the X axis, the dialog to which an alignment was obtained, in the Y axis. The second plot displays the alignment at iteration 50 of the second phase. Most of the subtitles analyzed in this second phase are now assigned. The third plot corresponds to the last merging iteration (107). A few subtitles are still unassigned and a very few were wrongly assigned. Two situations leading to non assignment of subtitles were found. One relates with texts that do not exist in the script as, for instance, the first 21 subtitles referring to a summary of the previous episodes as well as a flashback to a specific scene in a previous episode (164-170.) Also, it happened that a single word subtitle reinforced its predecessor, although absent in the script, as for instance, in subtitles 274, 275 – "Anda, vamos embora." (Come on, let's go.); "Vamos!". The information about the authoring of the adaptation and edition of the subtitles also appeared at the end of the episode(297-299). Another situation concerned audio information that appeared transcribed in subtitles, as these were conceived for deaf people (263-270). The few false alignments, which occurred in less than 1% of the cases, were due to detached situations where the subtitles editor included in the same subtitle speech from different dialog lines. In these cases, the dialogs were very short. Final alignment, found after post-processing is presented in the last plot. The outlier cases described above, although specific for this episode, are similar to those found in other experiences.

[1]Rádio Televisão Portuguesa.

4. Enriched narrative film player and browsing

This section describes how to play and browse the narrative films integrating each of the pieces of information contained in the TSS. These pieces can be useful after being converted to the audio domain, specially in situations in which vision can not be used (see section 1). The use of the audio domain can be done with TTS systems. Audio synthesized dialogue lines sound necessarily very unnatural when compared to the original audio dialogues or to the traditional dubbing. However, the remaining information from the script (character ID, scene descriptions) is actually the main novel addition to consider and can be synthesized with lower quality requirements. Namely the emotional speech aspects which are essential for the dialogues, such as prosody, are not so crucial for this type of information. Actually, if a real speaker should pronounce this, his speech should be even less emotive than a narrator. To be efficiently discriminated, each piece of information in the TSS can be assigned to different audio spatial channels and, preferably, to different speaker voices. However, reducing a demonstration of this approach to the audio representation was found not very convincing, since it does not allow an easy evaluation for users who are not really restricted to the audio input. Considering this, at the present stage of this work and after implementing a robust algorithm for text alignment, the second concern is actually to provide a visual integration of the TSS with the film. A small video demonstration of the present work is available in: "http://www.di.fc.ul.pt/∼cjct/RTP/TSSdemo.zip". It shows 80 seconds of a scene from the 12th episode from the television series named "Quando os Lobos Uivam" (when the wolves howl), an adaptation of the novel of Aquilino Ribeiro with the same name made by Francisco Moita Flores for the RTP. In this scene, a couple of peasants (Rosa and Jaime) collect some bushes from an area they were recently forced to sell to a capitalist. A national guard named Modesto founds and menace them.

Figure 4 was obtained from the same portion of this film. The video is shown in the upper left corner and the name of the speaking character is shown right below. The subtitles are also presented bellow the video just according to common standards. Further bellow, a description about the specific shot can be found. All the right side of the display area was used for the script. This is shown as a scrolling text with new dialogue lines and related comments coming from below shortly after the corresponding subtitle was shown. It seems like the subtitles feed the scrolling script but, as mentioned before, these are often different from the dialogue lines and can sometimes include extra information

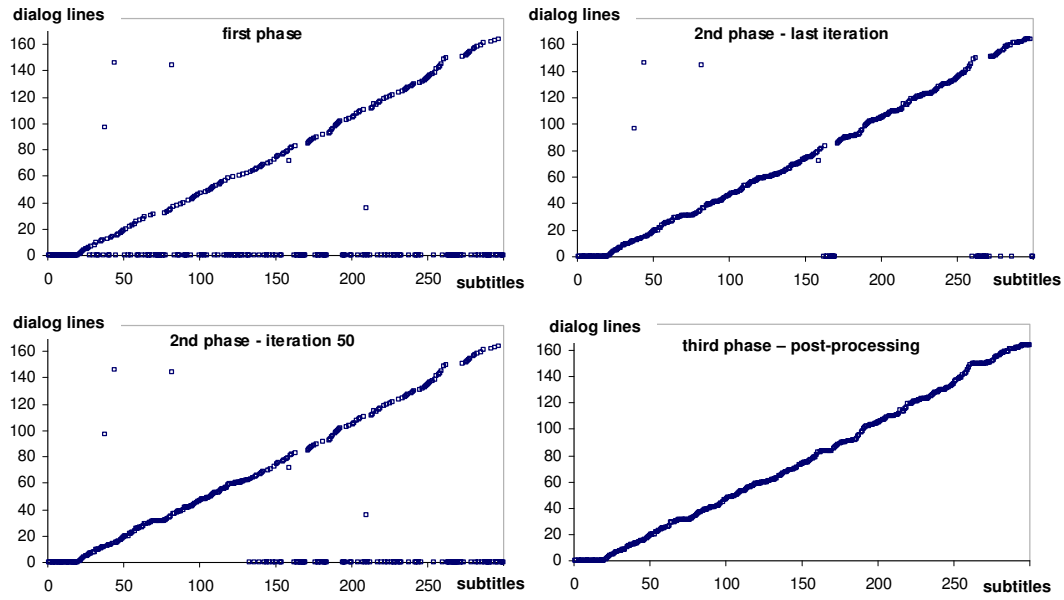


Figure 3. Evolution of the alignment.

Figure 4. Enriched narrative film.

Text unit	character ID	subtitle	shot description	dialogue line
begin	-1	0	-2	+1
end	+5	0	+6	+6

Table 1. Showing delays (in seconds).

about the actor expected behavior. Together with the dialogue lines, the corresponding shot description is included – the same way as it appears in the script and it was previously shown the bottom left corner. Each of these text streams have different showing times. However, all these timing sequences were built from the subtitles time codes. Table 1 shows the adopted time shifts for each piece of information in the script, taking the corresponding subtitle time code as reference.

The integration of all the texts in a single video allows browsing capabilities where the player can do fast forward, backward while all the other elements appear synchronously, namely the scrolling TSS. More than that, simple query facilities will be integrated allowing direct access to a particular shot, given some words spoken by a specific character. Besides character ID and subtitles or dialogue lines, other queries can be made using shot descriptions, scene names, expected character behavior and any combinations between these.

5. Conclusions

This paper describes the integration of scripts with subtitled films. A new specific algorithm for text alignment is presented and used in very preliminary experiments. This algorithm is practically language independent and robust enough to allow alignments between other related texts, not only with small sentence construction differences, but also with additional structural differences. This is useful for dealing with flashbacks and flashforwards of scenes in the film or the corresponding analepsis and prolepsis in the texts. Many of the approaches to text alignment in the literature are based in DP procedures [4, 2, 3] optimizing the edit distance between two text sequences or the longest common subsequence, working at word alignment level. These strategies give special emphasis to the order of words within phrases, as similarities evaluation takes into account the matching of consecutive words. However, in the present case, alignment at sentence level was preferred, to allow for word swapping within phrases and to provide flexibility to deal with more complex alignments.

Script alignment experiments with a wider collection of films in Portuguese and English are expected to be conducted in near future. These experiments will compare the performance of this algorithm with other known algorithms. Other texts will be used, namely resulting from behind the scenes documents.

In the present work, time anchors between film and its script are built exclusively between the subtitles and dialogue lines. However, future developments are foreseen that will enrich this alignment based on anchors between the video and the audio and the remaining information in the script - character identification and scene descriptions.

6. Acknowledgments

The authors want to thanks RTP, namely João Sequeira and the RTP Multimedia Director Francisco Teutónio Pereira for providing the videos and scripts used in the present publication.

References

[1] T. Chambel and N. Guimarães. Context perception in video-based hypermedia spaces. In *Proc. ACM Hypertext'02*, College Park, Maryland, June 2002. ACM.

[2] D. Gibbon. Generating hypermedia documents from transcriptions of television programs using parallel text alignment. In *Proc. Eighth Int. Workshop on Research Issues In Data Engineering, Continuous-Media Databases and Applications*, pages 26–33, 1998.

[3] A. Martone and E. Delp. An overview of the use of closed caption information for video indexing and searching. In *Proc. CBMI Fourth International Workshop on Content-Based Multimedia Indexing*, 2005.

[4] A. Martone, C. Taskiran, and E. Delp. Automated closed-captioning using text alignment. In *Proc. SPIE Int. Conf. on Storage and Retrieval Methods and Applications for Multimedia*, pages 108–116, 2004.

[5] R. Ronfard and T. Thuong. A framework for aligning and indexing movies with their script. In *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME 2003)*, volume 1, pages I21–24. IEEE, 2003.

[6] G. Salton, A. Wong, and C. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613620, 1975.

[7] R. Turetsky and N. Dimitrova. Screenplay alignment for closed-system speaker identification and analysis of feature films. In *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME 2004)*, volume 3, pages 1659– 1662. IEEE, 2004.

A Conceptual Approach for Active Surveillance of Indoor Environments

Vincenzo Deufemia, Massimiliano Giordano, Giuseppe Polese, and Mario Vacca

Dipartimento di Matematica e Informatica, Università di Salerno
Via Ponte don Melillo, 84084 Fisciano (SA), Italy
{deufemia, mgiordano, gpolese, mvacca}@unisa.it

Abstract

We propose an approach based on the Theory of Conceptual Dependency [16, 17] for intelligent video surveillance of indoor environments. The approach employs conceptualization to represent simple events and scripts to describe admissible situations of the specific application domain. Anomalies and suspicious events are detected by contrasting them against admissible situations. The conceptualization allows to hierarchically organize camera detected situations, making the reasoning process more intuitive. The approach supports video summarization of relevant scenes, and it provides an inference engine to handle complex queries.

1. Introduction

With the increasing need of security in today's society surveillance systems have become of fundamental importance. Video cameras and monitors pervade buildings, factories, streets, and offices. However, this has led to an increased demand for automatic systems supporting active video surveillance. These systems can be considered security support systems, as they are able to detect suspicious events and to raise alerts in order to warn security staff members.

The attempt to spread surveillance systems has led to new challenges for both hardware and software. A first issue is related to the high cost of cameras and the surrounding systems. The high cost of hardware used in traditional dedicated systems and the resources needed for the transmission and storage of images has limited the use of surveillance systems. Moreover, for an effective use of these systems organizations need to allocate security staff members to monitor the video coming from the different installed cameras, which has led to an increased demand for automation support. Thus, intelligent surveillance systems are required, capable of understanding situations and behave consequently.

Special kinds of video surveillance systems are those used in indoor environments, like homes, banks, or garages. These environments are characterized by a limited number and types of objects, actions and situations, which considerably reduces the challenges in the development of intelligent surveillance systems. Indeed, these considerations entail that a video surveillance can be purposive in its nature, as it knows beforehand the objects involved in the "surveillance". It follows that such a video surveillance system can use stereotyped knowledge for detecting both normal and anomalous events.

This paper presents an approach for an intelligent indoor surveillance system based on low cost hardware and techniques from Artificial Intelligence. In particular, we present a low-cost solution for video surveillance exploiting web technologies, open standards, and low-cost devices. Moreover, we exploit Artificial Intelligence techniques to enable the system "understand" events captured by the cameras. Our approach is based on the Schank's theory [16, 17, 18], a "non-logical" approach, which has been widely used in natural language processing area.

Two main reasons led us to use this theory in video surveillance systems. First, the presence of well-studied primitives to represent details about the actions; second, the possibility to use highly structured representations like scripts, which are a natural way to manage prototypical knowledge. Thus, we are able to associate different levels of meanings to a situation: conceptualization, scene, and script level, which allow us to deeply understand the current situations and to detect anomalies at different levels. Moreover, the structured information makes the design and reasoning process easier.

The paper is organized as follows. Section 2 briefly surveys Schank's theory. In Section 3 we discuss the system architecture, and the techniques employed for intelligent video surveillance. Section 4 discusses related works. Finally, in Section 5 final remarks and a brief discussion on future works conclude the paper.

2. Theoretical background

In this section we introduce some basic notions of conceptual dependency and scripts, which will be useful for describing the proposed video surveillance approach.

2.1. The theory of conceptual dependency

The theory of conceptual dependency (*CD theory*, for short) is a pictorial formalism developed by Roger Schank in the 70's [16] (see also [15] for a concise, but complete introduction), for representing complex events by elementary ones. A CD representation of an event (also called *conceptualization*) is composed of objects belonging to four classes linked together by rules. The classes of CD objects are ACT (actions), PP (Picture Producers), AA (Action Aiders), and PA (Picture Aiders).

The class ACT contains eleven elementary actions, like, for instance, PTRANS (Physical TRANSfer), indicating a position transfer of an object, or GRASP, representing the act of grasping an object by an actor. The class PP contains humans, animals, or objects. The role of classes AA and PA is to specify more precisely the semantics of the objects involved in the conceptualization. Examples of rules are $PP \Longleftrightarrow ACT$ or $PP \Longleftrightarrow PA$.The first rule states that PP is the agent of the ACT, whereas the second one states that PP is in the state PA. An application of these rules could be useful in describing the semantics of the movements of a museum warden called John by the following conceptualization

$$Warden \iff John \Longleftrightarrow PTRANS$$

A knowledge-based system using the CD theory (such as MARGIE [16]) contains a set of CDs representing general knowledge about a given domain. Basic reasoning tasks of such systems are:

- *to understand events*: The task of representing current events (*understanding*) using the stored knowledge is accomplished both reducing events to simple ones and instantiating the objects in the conceptualization with actual data (filling the slots);
- *to reason about events*: Once an event has been interpreted using the existing knowledge it is possible to make inferences and to supply the lack of information in occurred events.
- *to answer questions about occurred events*.

All these reasoning tasks on CD representations have usually been implemented as LISP programs.

2.2. The theory of scripts

Scripts were introduced by Schank and Abelson [17] to model the daily life stereotypical situations, such as "eating in a restaurant" or "taking a plane". They can also be used to describe a "theft in a museum", or a typical bank situation with customers waiting for a service.

Scripts are frame-like knowledge structures and represent prototypical knowledge. They contain sequences of *scenes* involving a set of objects (*Props*) and a set of people (*Roles*). Scenes contain general actions aiming to reach the same goal.

For example, a script RESTAURANT could involve tables, menù, food, and money as $Props$, and clients, waiters, and cooks as $Roles$. It can contain the scenes ENTERING, ORDERING, EATING, PAYING, and EXIT describing what generally happens when people *enter in* a restaurant, *order* dishes, *eat*, *pay*, and finally *leave* the restaurant. Each scene contains conceptualizations describing the events in the scene itself. For example, the scene EATING contains the action of the waiter giving the food to the client, and the one of the client ingesting the food.

Both scripts and their successor MOPs (Memory Organization Packets) can be considered as orderers of scenes [18].

Typical reasonings tasks with scripts are:

- *understanding*: The main function of a script is to interpret (*understand*) the occurred events using its own knowledge.
- *reasoning about occurred events*: Scripts help making inferences about occurred events, since they can be viewn as causal chains of events. Moreover, script knowledge can be used to supply the lack of information in occurred events, using default values. Scripts can also be used to foresee events.
- *answering questions about occurred events*.
- *summarization*: Scripts can be used in the summarization by selecting only the most relevant events [4].

Concluding, a script based understanding system has structures of knowledge representing expectations on the events usually happening in a given situation. Thus, from the computational point of view, script based understanding means to select a script and to fill its slots with the information coming from the input.

3. The proposed video surveillance system

The aim of a video surveillance system is to detect anomalies and to raise alert messages. In order to detect anomalous events, the proposed system tries to interpret a

scene based on its knowledge about "normal" situations, using the CD representation to describe single events and scripts for complex situations. Therefore, the proposed video-surveillance system is an intelligent system associating semantical representations to images. Similarly to many intelligent systems [17], this representation is built by stages, using a mixed bottom-up/top-down approach.

The architecture of the proposed system is shown in Figure 1.

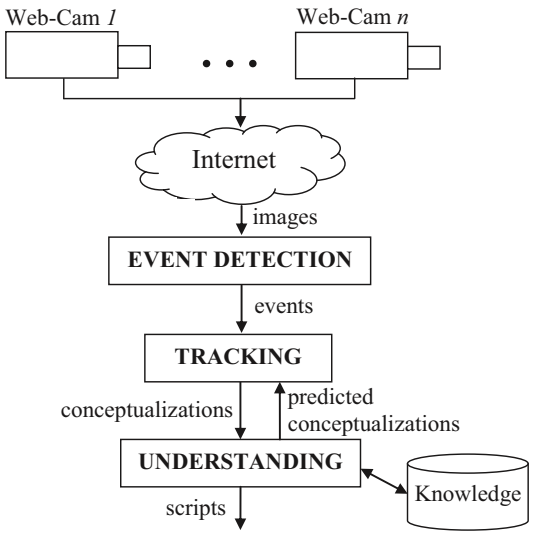


Figure 1. The system architecture

The aim of the *event detection* module is to acquire images from the camera and to detect "atomic" events (like entering a door or picking up a bag). The process of motion detection usually involves motion segmentation (i.e., segmenting regions corresponding to moving objects from the rest of an image) and object classification.

The *tracking* module aims to track moving objects from one frame to another in the detected events. This involves matching objects in consecutive frames using particular features. Successively, this module infers actions linking the recognized objects and the surrounding people. In particular, it builds conceptualizations and fills them with references to objects in the images. The output of the tracking module is a conceptualization describing the meaning of the tracked action.

The *understanding* module activates pertinent scripts and appropriate scenes from the script produced by the tracking module in order to identify possible anomalies. In particular, when a script is selected and instantiated, the understanding module takes the control and conceptualizations are built in a predictive way. In particular, when a script is activated, the conceptualizations belonging to the scenes that might occur are sent to the tracking module to work in a predictive mode. The output of the understanding module are scripts describing the occurred situations.

In what follows we provide a detailed description of each system module.

3.1. Architecture and Camera Level

We adopt internet as communication medium between webcams and the server on which the video surveillance system run. In particular, each webcam sends images to the central server, which selects and stores only images representing unseen scenes. In this way it reduces the needs of permanent storage resources. Since image analysis and acquisition can be very expensive in terms of system resources, especially with a high number of webcams, we have realized a multiple servers load balanced architecture.

When connected to the system webcams transmit images to the server at full speed (about 15 frames per second), by using the RTP protocol. The server acquires images coming from different webcams and processes each of them trough a codec chain. We have developed our system to be extensible, and to support the insertion of other codecs.

3.2. Event Detection

A fundamental step in video surveillance systems is motion detection, aiming to detect moving objects in the video stream. This entails the segmentation of each image through automatic image analysis techniques. The importance of this operation is due to the fact that the effectiveness of the segmentation process affects the performance of the other modules.

Many algorithms have been proposed for motion detection in video surveillance applications [7, 10]. We have employed the SGM (single Gaussian model) algorithm, proposed in the Pfinder system [21], which uses a Gaussian model to classify each pixel of the image. In particular, SGM uses an intensity-normalized color representation of each pixel in the image, and a multi-way Gaussian classifier to decide which foreground or background class each pixel belongs to, trying to cluster foreground pixels into blobs. Pixels in the current frame are compared to the background by measuring the log likelihood in color space. If a small likelihood is computed, the pixel is classified as active, otherwise it is classified as background.

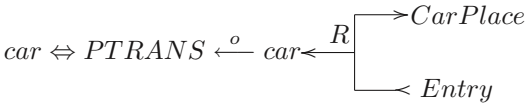
3.3. Tracking

This module tracks moving objects in the images produced by the Event Detection module. It accomplishes this task by employing a region-based tracking algorithm, tracking objects according to variations of the image regions corresponding to the moving objects [21]. In particular, the

algorithm uses small blob features to track humans in indoor environments. The body of humans is represented by a combination of some blobs representing various body parts such as head, torso, and the four limbs. Whereas, both human body and background scene are modeled with Gaussian distributions of pixel values. Finally, the pixels belonging to the human body are assigned to the different body parts' blobs using the log-likelihood measure. Then, by tracking each small blob, the moving human is successfully tracked.

This tracking algorithm is effective also for tracking non humans [10].

The information on the tracked objects are synthesized by constructing conceptualizations, which are given in output to the next module. As an example, the following conceptualization expresses the fact that a given car moves from the garage entry to a car place.



3.4. Understanding

The use of scripts for representing salient situations allows us to perform effective analysis operations and to produce logic consequences. As an example, it is possible to correlate different scenes and to determine if they contain anomalous events, such as an intrusion.

The understanding module applies the scene on the scripts representing the detected events. If an anomaly is identified an alert signal is produced.

The system has a knowledge base composed of a set of scripts, whose use is ruled by three operations: *selection*, *activation*, and *application*.

The *selection* operation selects one or more scripts from the knowledge base. These are candidates to the understanding of the current situation. This operation is performed by using the information (conceptualizations) coming from the tracking module.

The *activation* is the operation deciding the script (among the selected ones) to be used in the current situation processing.

Once activated, a script is *applied*, i.e., its slots are filled with the objects, roles, and actions from the situation being processed. At the end of the application the script becomes *instantiated*, and it returns a description of the meaning for the processed situation.

The activation is important since an activated script foresees the actions that will take place, and therefore it is useful to detect anomalous and suspicious facts and scenes.

3.5. Example

In order to understand what the understanding module has to accomplish, let us consider the video surveillance of a garage with many parked cars, with about hundred cars coming in and out during night and day. This kind of environment has a strong need of security, which so far could only be guaranteed by human operators.

Automated video surveillance systems seem adequate to accomplish the tasks of human operators in this kind of environment. In fact, such environment has a few roles (warden, client, thief, robber), a few props (car, desk, motorcycle). The actions are simple and repeated. The whole knowledge is prototypical.

The usual situations monitored in the garage are described by the following script, which provides more a concise and comprehensible description than isolated conceptualizations.

SCRIPT *$GARAGE*
Roles: warden, client, person
Props: car, desk
scene *CAR-ENTERING* car enters the garage
scene *PARKING*
car *parks*
car *stops*
scene *PEOPLE-EXIT-CAR*
the door *is opened*
the client *gets off*
the door *is closed*
(The following three actions can be iterated)
the door *is opened*
a person *gets off*
the door *is closed*
scene *TICKET*
A person *goes* near the desk
A person *takes* the ticket from the warden
scene *PEOPLE-ENTERING-GARAGE*
The client *enters* the garage
(The following three actions can be iterated)
A person *enters* the garage
scene *CAR-TAKING*
The client *opens* the doors
The client *enters* the car
(The following three actions can be iterated)
A person *opens* the doors
A person *enters* the car
scene *PAYING*
A person *goes* near the desk
A person *pays* the tickets
scene *CAR-LEAVING*
A car *leave* the garage

A car entering a garage let the selections of the scripts *$SERVICE*, *$PARKING*, and *$ROBBERY* from the knowledge base.

When a car enters the garage the scene CAR-ENTERING is activated. As a consequence, all the scenes in the script will be activated and instantiated according to the sequence in the script itself. In this case we have a normal flow of events. It is possible that deviations from the normal flow occurs. There are two kinds of deviations that are interesting from the point of view of a surveillance system: known deviations and alarming anomalies. The formers are simply exceptions (less frequent events), which can still be classified as normal ones. For example, events belonging to this category are both that of a person entering the garage to ask for information about fees, and that of a typical robbery attempt. The system deals with this kind of events using other scripts (or MOPs), as in the following.

SCRIPT *$ROBBERY*
Roles: warden, person
Props: car, desk
scene *PEOPLE-ENTERING-GARAGE*
(The following three actions can be iterated)
A person *enters* the garage
scene *CAR-TAKING*
The person *opens* the doors
The person *enters* the car
(The following three actions can be iterated)
A person *opens* the doors A person enters the car
scene *CAR-GOES-TOWARD-EXIT*
A car *goes-toward* the garage exit

The second kind of deviation is a serious one, as it cannot be framed neither in the normal flow nor in some track (i.e., alternative paths in the script) (see [17] for a more detailed discussion about the concept of track). In this case the system is authorized to raise an alert signal.

4. Related work

In recent years many different approaches have been proposed for analysis and recognition of motion patterns, and for behavior understanding in video surveillance systems. For a survey of video surveillance systems and related issues see the works of Dick *et al.* [5] and Hu *et al.* [7]. We will briefly review some works on activity recognition and understanding.

Hidden Markov Models (HMM) [11, 21] and Bayesian networks [13] are statistical methods largely used for activity recognition. A combination of both was used in the VIGILANT project for object and behavioural classification [14]. The Bayesian classifier was used for identification of object types, based on the object velocity and bounding box aspect ratio. A HMM was used to perform behavioral analysis to classify object entry and exit events.

Automaton have been used in activity recognition for representing the behaviors of the monitored objects. For instance, Bremond *et al.* use handcrafted deterministic automata to recognize airborne surveillance scenarios describing vehicle behaviors in aerial imagery [2], whereas Wada *et al.* employ non-deterministic finite automaton for multi-object behavior recognition [20].

Several grammatical approaches have been used for visual behavior recognition. Brand uses a simple nonprobabilistic grammar to recognize sequences of discrete behaviors [1], whereas Ivanov *et al.* describe a probabilistic syntactic approach to the detection and recognition of temporally extended behaviors and interactions between multiple agents [8].

Conceptual and knowledge based reasoning is considerably important in the recognition of objects and actions, or, in general, activities. The need for content representation in video surveillance systems naturally calls for the use of the well-known and tested representations from Artificial Intelligence.

Approaches for representing knowledge are usually divided into two main categories: logic and non-logical [9]. These two kinds of approaches are competitive, and the theoretical disputes attempting to prove the superiority of one with respect to the other has not finished yet.

A logic-based approach for video surveillance was recently proposed by Shet *et al.* [19]. In this work different kinds of logics have been used. The use of logic programming makes easier reasoning about events, but also allows to implement systems using well-known and tested programming languages.

On the side of non-logical approaches, Schank's theory has been used in content representation. In fact, his theory of scripts in the representation of the content of films was introduced by Parkes [12] (See also [3, 6]).

As we abide by the view of non-logical theories, we think that structured representations like scripts or MOPs can be useful also in the video surveillance field.

5. Conclusions and Further Research

In this paper we have proposed an approach based on the theory of conceptual dependence [16, 17] for intelligent video surveillance of indoor environments. The use of conceptualization has allowed us to represent in detail the actions detected by cameras and to use highly structured representations like scripts, which are the most natural way to manage prototypical knowledge.

In the future we intend to further investigate the refinement of the knowledge structures, with a structured and

dynamic knowledge base [18], which could lead to understanding situations in a more detailed way, and therefore, to a more accurate and active surveillance.

A further future work is to take benefits from the use of multiple cameras by applying data fusion approaches. This will allow us to solve important issues such as occlusion handling and continuous tracking.

References

[1] M. Brand, "Understanding Manipulation in Video", in *Proc. of 2nd International Conference on Face and Gesture Recognition (AFGR'96)*, 1996, pages 94–99.

[2] F. Bremond and G. Medioni, "Scenario Recognition in Airborne Video Imagery", in *Proc. DARPA Image Understanding Workshop*, 1998.

[3] M. Davis, "Knowledge Representation for Video", in *Proceedings of Twelfth National Conference on Artificial Intelligence* (*AAAI-94*), Seattle, Washington, AAAI Press, pp. 120–127, 1994.

[4] G. DeJong, "An Overview of the FRUMP System", in W.G. Lehnert, and M.H. Ringle, (eds.), *Strategies for Natural Language Processing*, Hillsdale, Erlbaum, 1982, pp. 149–176.

[5] A. R. Dick and M. J. Brooks, "Issues in Automated Visual Surveillance", in *Proc. of International Conference on Digital Image Computing: Techniques and Applications (DICTA'03)*, 2003, pp. 195–204.

[6] E. Hartley, A. P. Parkes, and D. Hutchison, "A Conceptual Framework to Support Content-Based Multimedia Applications", in *Proc. of European Conference Multimedia Applications, Services and Techniques (ECMAST'99)*, 1999, pp. 297–315.

[7] W. Hu, T. Tan, L. Wang, and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors", *IEEE Trans. on Systems, Man, and Cybernetics - Part C*, 34(3), August 2004, pp. 334–352.

[8] Y. A. Ivanov and A. F. Bobick, "Recognition of Visual Activities and Interactions by Stochastic Parsing", *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8), 2000, pp. 852–872.

[9] R. A. Kowalski, "The Limitation of Logic", in *Proc. of ACM Conference on Computer Science*, 1986, pp. 7–13.

[10] J. C. Nascimento and J. S. Marques, "Performance Evaluation of Object Detection Algorithms for Video Surveillance", *IEEE Trans. on Multimedia*, 8(4), 2006, pp. 761–774.

[11] H. Ning, L. Wang, W. Hu, and T. Tan, "Articulated Model Based People Tracking Using Motion Models", in *Proc. Int. Conf. Multi-Model Interfaces*, 2002, pp. 115-120.

[12] A. P. Parkes, "The Prototype Cloris System: Describing, Retrieving and Discussing Videodisc Stills and Sequence", *Inf. Process. Management*, 25(2), 1989, pp. 171–186.

[13] P. Remagnino, T. Tan, and K. Baker, "Multi-agent Visual Surveillance of Dynamic Scenes", *Image Vision Comput.*, 16(8), 1998, pp. 529–532.

[14] P. Remagnino and G. A. Jones, "Classifying Surveillance Events and Attributes and Behavior", *British Machine Vision Conference (BMVC2001)*, Manchester, 2001, pp. 685–694.

[15] E. Rich, *Artificial Intelligence*, 1983, McGraw-Hill.

[16] R. C. Schank, *Conceptual Information Processing*, 1975, Amsterdam, North-Holland Publishing Co.

[17] R. C. Schank and R. Abelson, *Script Plans Goals and Understanding*, 1977, Hillsdale, Lawrence Erlbaum Associates.

[18] R. C. Schank, *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*, 1982, New York, Cambridge University Press.

[19] V. D. Shet, D. Harwood, and L. S. Davis, "Multivalued Default Logic for Identity Maintenance in Visual Surveillance", in *Proc. of 9th European Conference on Computer Vision (ECCV'06)*, 2006, pp. 119–132.

[20] T. Wada and T. Matsuyama, "Multi-object Behavior Recognition by Event Driven Selective Attention Method", *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8), 2000, pp. 873–887.

[21] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body", *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7), 1997, pp. 780–785.

A Self-Organizing Approach to Mission Initialization and Control in Emergency Management

Shi-Kuo Chang[1] and Erland Jungert[2]

[1]University of Pittsburgh, Pittsburgh, PA, USA
[2]Linköping University, Linköping, Sweden
chang@pitt.cs.edu, jungert@foi.se

Abstract: ***One of the critical problems in emergency/crisis management is the initialization of the management operations, which must be started as quickly and as efficiently as possible. The quicker the resources and the efforts can be put together the more effective the management operations will become. Subsequently in the course of performing the management operations similar problems arise when new missions/operations must be started as a consequence of new demands or changing needs. To address these problems a self-organizing approach is proposed. In this approach, agents needed for a mission/operation that will be part of the emergency/crisis management operations are* self-organizing *to initialize and control their own operations. This approach is based on three basic concepts: a* Handbook of protocols*, a* service-oriented knowledge structure *and a* management cycle *for the continuous initialization and customizing of agents.***

1. INTRODUCTION

One of the critical problems in emergency/crisis management is the initialization of the management operations. Subsequently, in the course of performing the management operations similar problems arise when new missions/operations must be started as a consequence of new demands or changing needs. In both cases, the missions/operations must be started as quickly and as efficiently as possible because the quicker the resources and the efforts can be put together the more effective the management operations will become.

Of particular importance is the initialization process in the very beginning of the management operations, because in most cases it will have a strong impact on the operations, on the crisis itself and how it is completed. The initialization of not only the premier mission but also all subsequent missions during an emergency or a crisis must be handled effectively. The agents participating in these initialization processes must be adequately supported by the emergency/crisis management system because otherwise the operations cannot be properly carried out.

In this paper, our concern is how agents needed for a mission/operation that will be part of the emergency/crisis management operations can be self-organizing to initialize and control their own missions/operations. The proposed solution is based on three concepts: a Handbook of protocols, a service-oriented knowledge structure and a management cycle for the continuous initialization and customizing of agents. Together these concepts determine the self-organizing process of the emergency/crisis management operations. In the context of this paper, an agent may be a private person who serves as a first responder, a more specialized responder such as a fireman or a policeman, an expert in a knowledge domain such as chemistry, or a fully autonomous software agent. An approach that to some extent relates to this view is given by Visser et al. [3].

The paper is structured as follows. In Section 2 the general problem to be addressed is defined. The self-organizing process is discussed in Section 3. In Section 4 the service-oriented knowledge structure is described in detail. The Handbook of protocols is introduced in Section 5 and in Section 6 the control loop for the management cycle is discussed. Section 7 illustrates by examples the self-organizing initialization process. Finally in Section 8 the subsequent work will be discussed.

2. THE PROBLEM TO BE ADDRESSED

It is widely recognized that the first few minutes of an emergency/crisis are the most critical. Basically, in this situation the main problem is to first get an understanding of the actual situation and then allocate adequate resources and initiate a number of missions/operations; all with the purpose to deal with the emergency/crisis and bring it to a satisfactory resolution. Among the required resources there are human resources such as policemen and experts in a

variety of fields, which will be referred to as agents in the context of this paper. Agents may be of different types and they may play different of roles. Other resources may include organizations such as the fire department and material resources such as trucks and helicopters. However to get the management process and subsequently missions and operations of various types going requires human resources ranging from inexperienced first responders to a variety of experts. At the very beginning, the emergency/crisis organization can be seen as a loosely coupled system consisting of many different organizations, which must be turned into a smoothly working entity. To make this happen requires the proper initialization of a large number of tasks, which subsequently are called missions.

The problem is thus to determine how to conduct the self-organization of the initial processes of an emergency/crisis by means of generic agents (first responders) as well as expert agents, supported by existing service-oriented command and control systems. Processes in focus are the initial parts of the management operations, which if managed early and efficiently will allow successfully completion of the emergencies. Other processes of concern are the initial parts of missions identified during the management operations and which need to be managed by agents specifically appointed for these missions.

The problem is thus how to initialize the emergency/crisis management processes and other required missions/operations during a crisis under the above given circumstances in a self-organizing way.

3. THE SELF-ORGANIZING PROCESS

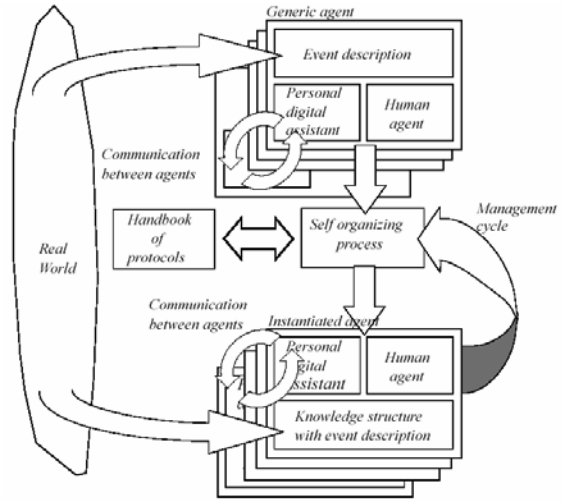


Figure 1. The self-organizing process including the Handbook of protocols.

In Figure 1 the self-organizing process including the Handbook of protocols is illustrated. The process shows how a generic agent can be turned into an instantiated agent. This process is carried out with the support of a Handbook of protocols. Originally a generic agent can be an ordinary citizen but it is also likely that he or she may have been given some kind of simple training by the community, and thus can be turned into to a first responder. In more complicated situations the generic agent may correspond to some professional type of agent such as a fireman or a policeman. The actions taken by these agents may vary considerably from one to another. However the expected activities of the agents are basically of the same nature, that is, to carry out some tasks that they have been ordered to carry out or forced into by circumstances out of their control.

Consequently the agents will collect as much information about the situation as possible and then consult the Handbook of protocols to determine what to do next. If they are somehow familiar with the situation they may be able to find the relevant solution on their own, i.e., they have "internalized" the Handbook of protocols. An example of such an internalized protocol is to call 911 (112 in Europe) and report an emergency/crisis event to the operator. To undertake the role of a 911-caller the generic agent is turned into an instantiated agent. In this simple context the agent does not do much more than this and once this task is finished the agent goes back to his/her original role as a generic agent.

Clearly, in more complex situations an agent needs to communicate with other agents who can either be generic or instantiated. At all times, the agent needs to consult the handbook. In the communication with other agents the personal digital assistant is used. The agent must go on collecting information from the real world to build up a relevant event description. Without knowledge about the on-going situation in the real world there is no way for an agent to start the self-organizing process. Consulting the Handbook of protocols alone does not help the agent in his mission.

The personal digital assistant in the proposed approach can be seen as a tool for communication and visualization of all available information. It also may include a number of decision support tools.

The self-organizing process in the proposed approach can be seen as part of a management cycle in which the instantiated agent(s) may repeatedly consult the handbook for further information. The instantiated agents can also start a series of mission control loops (see Section 6). As the instantiated agent performs the actual mission, besides the

personal digital assistant there are also other more powerful means available to the agent. Among them the service-oriented knowledge structure, which will support the collection of external information and the visualization of the actual event description. To do this, the agent will have a fairly large set of services available. These services make it possible to go on with the work and subsequently also to start the required control loop.

4. THE SERVICE-ORIENTED KNOWLEDGE STRUCTURE

The service-oriented knowledge structure shown in Figure 2 supports the agent in his/her work and make it possible for him/her to collect information about his/her mission and gradually build up and maintain the event description, which can be seen as the common operational picture. To perform the above tasks requires the continuous maintenance of the control loop. This also dictates a flexible and modular service-oriented system. A service-oriented approach also makes it feasible to reuse software that has been developed earlier, and to attach data and information sources in an independent way (see e.g. [1].

This depends on the fact that service-oriented components and modules only have weak bindings and therefore become independent of the data sources that are connected to the system. Besides, the modularity of a service-oriented architecture also makes evolutionary systems development possible.

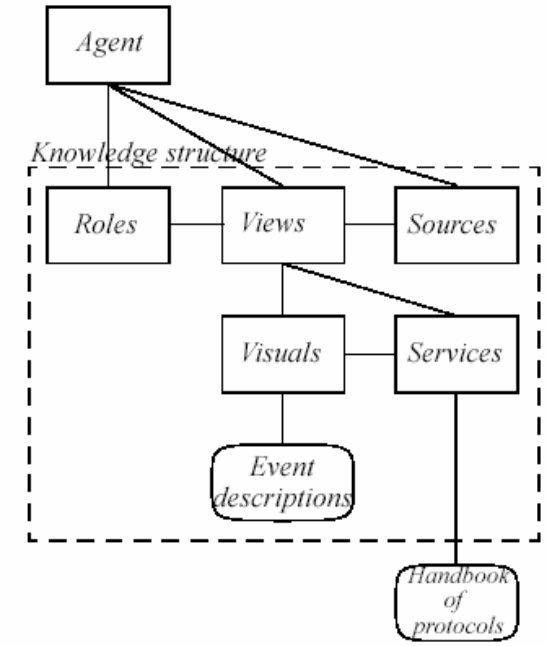


Figure 2. The service and role based knowledge structure for mission support.

A service-oriented architecture also gives the system a high degree of flexibility, which in turn makes it simpler to adapt to new applications, that is to changing needs and new requirements. Thus the usefulness of standardized interfaces between system components is to make the system interoperable with other external systems. Generally the service concept is not very well defined but within the field of information technology a service is defined as a function made available with a specified quantity and quality [4]. Within telecommunication a service is seen as the capacity to exchange information made available by a service provider [5].

The service-oriented architecture (SOA) comes out of the trial or needs to create weaker couplings between different system components. A service as defined by SOA is thus "a unit of work done by a service provider to achieve desired end-results for a service consumer" [6].

The service-oriented knowledge structure shown in Figure 2 the service and rule based knowledge structure for mission support illustrates how services can be tied to “views”. A view is defined as a combination of services and visuals. A visual in this context means the visualization of information collected by means of the services and may include images, maps, text and tables. In particular, certain visuals may include event descriptions of on-going missions, which may correspond to operational pictures mirroring the situation of the on-going crisis.

Generally, the relations between the different structures in Figure 2 are mostly of many-to-many type. Consequently, an agent may play many roles although not at the same time but over the time of the emergency/crisis. Within a role a particular agent may have a set of views available, which in turn may include a set of visuals and services. This will support the generation of the actual event description, which, as indicated, may include information of many types. Furthermore, from any given view and by means of any available service the agent is, at all times, allowed to consult the Handbook of protocols. By means of available services a user can also get further information from the sources, which makes it possible to update the event description and thus have a better understanding of the present situation and to take relevant decisions on a timely basis. Consequently, the service-oriented and role-based knowledge structure will give the agents a good situation awareness. The knowledge structure is, when applied, also a part of the control loop in the management cycle.

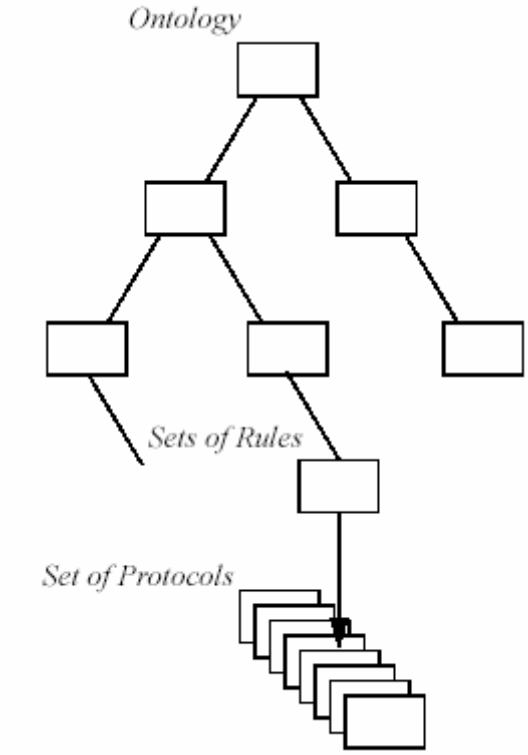


Figure 3. The knowledge structure of the Handbook of protocols.

5. THE HANDBOOK OF PROTOCOLS

The Handbook of protocols is the basic source for the agents in determining and understanding what to do (WTD) when initializing a mission/operation. The knowledge structure of the handbook is shown in Figure 3. When using the handbook the agent first must look for some available notion that is in focus for the operations and that is applicable to the initialization of the mission to be undertaken. The knowledge structure of the handbook supports this by means of an ontology, which further leads to a set of applicable rules and alternatives. Which rule should be fired by the system depends on the conditions supplied by the agent. The conditions identifying the rule will be specified during a dialogue between the knowledge system and the agent. Once the rule is identified the applicable protocol is determined. The protocol itself is presented as text. The protocol makes it possible for the agent to take proper actions in accordance with the given protocols.

The protocols can be specified using the active index [8, 9] with timing constraints.

6. THE MANAGEMENT CYCLE AND THE MISSION CONTROL LOOP

As mentioned above, the self-organizing process is part of a continuous management cycle for the instantiated agents to collect information, consult the Handbook and decide what to do next. Furthermore, the instantiated agents themselves may start and follow a *mission control loop* in their operations.

In the literature, many different types of control loops for command and control situations are proposed. For instance, there is the OODA-loop, used in connection to information fusion applications, basically for military applications, see [7]. However, here we apply the loop proposed by Hollnagel [2]. The motivation for this is because it is more general than the OODA loop. Hollnagel's loop is seen in Figure 4.

Basically, going through a control loop means that a controller, here an agent, develops a plan, which means that in the end an action of some kind must be carried out. For instance in the case of a forest fire the action can be to water bomb the area from an aircraft. This activity will produce some effects whose consequences must be determined. Thus information about the consequences of the action must be collected to guide the agent when the next step is planned. However the collected information will always include uncertainties of different kinds due to missing, incomplete or distorted data. These uncertainties are called *disturbances*. No matter what, disturbances will always exist and the controller must always take their existence into account before the next move is determined. However, to complete any mission efficiently will require control loops.

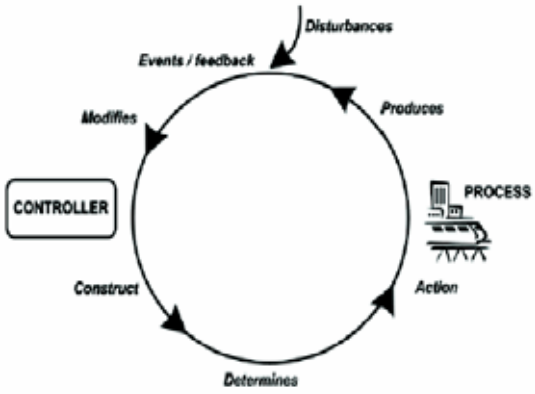


Figure 4. Control loop for emergency management.

7. MISSION INITIALIZATION AND CONTROL

The initialization processes discussed in this work include both the starting process of the management operations as well as most of the

subsequently determined (or planned) missions. Before these missions can be started the agent playing the role of commander of the management operations must appoint some agents for these missions. At the very beginning of the emergency a generic agent who may be the first person on the scene may voluntarily start the initialization process. Once an agent has been turned into, or has turned himself/herself into an agent with a specific role the initialization process is started. The process will go on just as long as it is required and will include at least one but sometimes several consultations of the Handbook of protocols. During the mission, participating agents may communicate with other agents either inside or outside their own organization. In any mission, with just a few exceptions, the responsible agent must start at least one control loop. Below, the two initialization processes will be discussed and illustrated.

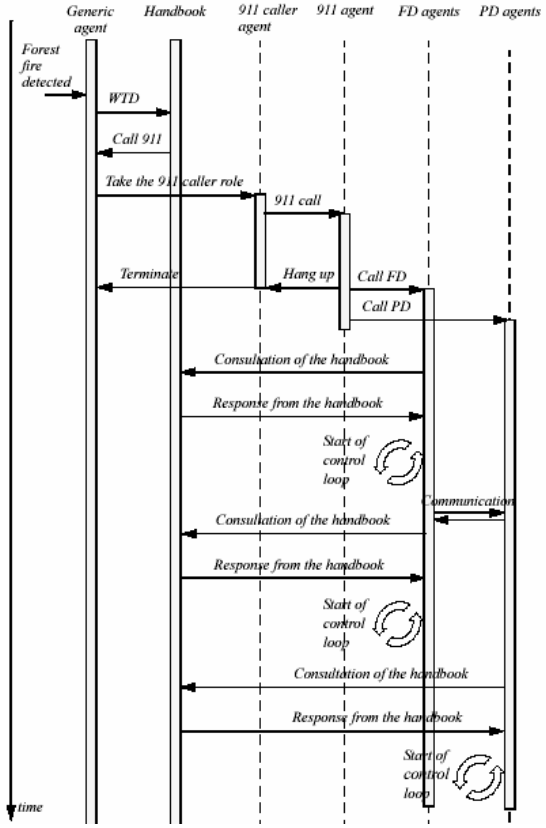


Figure 5. The initialization phase of a forest fire.

7.1 management initialization

The initial phase of an emergency (a forest fire) is illustrated in Figure 5 by means of a modified UML sequence diagram. The event may have started in some unidentified way but it is discovered by some individual who just happens to be present at the place. This individual corresponds to a generic agent and will act as a first responder. After having consulted the Handbook of protocols, which is available or accessible from his personal digital assistant, the generic agent turns into an agent with the role of a 911-caller. After the call is placed a 911-agent (operator) is activated. The two agents communicate and the caller transmits the event description. Once the event description is transmitted the call is finished and as the caller hangs up the agent goes back to his normal role. Thus this is an exception from the general approach as no control loop is created. The 911-caller agent does not exist anymore. The 911-agent, on the other hand, consults the Handbook, if required, and as a consequence engages in its role as an activator of other agents who in this case are the agents at the police and fire departments. In this particular case, there is also no control loop as the 911-agents primarily are just responsible for the activation of the agents made available by the community. In practice, however, the 911-agents quite often serve as point-of-contact between the engaged agents.

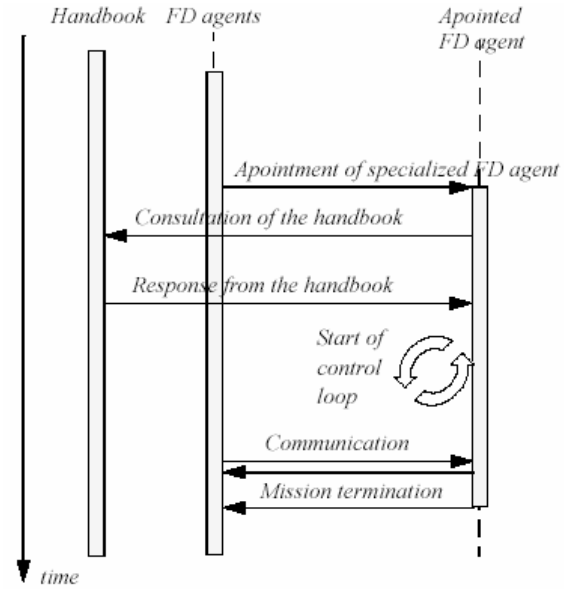


Figure 6. Initialization and control of a specialized mission.

The agents activated by the 911-agent now starts their operations by consulting the Handbook of protocols to make decisions involving plans for new missions to be undertaken, which again means that new agents must be appointed. So the operations go on, including frequent consultations of the Handbook followed by creation of new control loops and so on.

7.2 mission initialization

The missions following the premier initialization of the management operations are all subordinate missions, but they all follow the same general procedure, which again includes an activation step, consultation of the Handbook and the start of one or more control loops. This is illustrated in Figure 6 where a fireman is pointed for a certain mission, takes on a suitable role and consults the Handbook from which feedback is received and as a consequence a control loop can be started. During all these activities the agent may communicate with other agents as required by the events that occur during the ongoing mission. Once again when the mission is completed the agent returns to his original role.

8. DISCUSSION

The proposed approach outlines a framework of a self-organizing multi-agent system for emergency management. As pointed out in Section 1, the approach is based upon three basic concepts: a Handbook of protocols, a service-oriented knowledge structure and a management cycle for continuous initialization and customization of agents. Each of these concepts needs to be further extended and substantiated.

The Handbook of protocols can be as simple as an ordinary cooking book where each entry is a recipe (a protocol). Such simple Handbook may already be very useful in practice. For more sophisticated emergency/crisis management a more sophisticated Handbook, with a knowledge structure shown in Figure 3, needs to be developed. We plan to apply the active index approach [8, 9] and the IC cards user interface to the further development of the Handbook of protocols. Furthermore by extending the Handbook with learning capabilities it may be possible to support agents in future emergency/crisis management operations such that the Handbook also can serve as an instrument for *best-practice.*

The service-oriented knowledge structure is a knowledge structure that includes both static and dynamic information. For example the visuals may include videos, sensory inputs that are constantly changing. How to apply information fusion to efficiently and effectively handle both static and dynamic information in the service-oriented knowledge structure requires further investigation.

Last but not least the management cycle for continuous initialization and customization of agents depends very much on the particular requirements of different applications. A crisis such as a natural disaster, a medical emergency such as a person suffering a stroke, or a man-made crisis such as a terrorist attack, dictates different ways to structure the management cycle. The major types of these management cycles need to be categorized so that different types of resources can be listed, which can serve as the basis for developing different service-oriented architectures for emergency/crisis management.

REFERENCES

1. Jungert, E., Hallberg, N. Hunstad, A., *A service-based command and control systems architecture for crisis management*, International Journal of Emergency Management, Vol 3, No 2/3, 2006, pp 131-148.
2. Hollnagel, E. and Woods, D. D., *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*, CRC Press, New York, 2005.
3. Visser, A., Pavlin, G., van Gosliga, S. P., Maris, M., *Self-organization of multiple-agnet systems*, Proc. of the International workshop on Military Applications of Agent Technology in ICT and Robotics, The Hague, the Netherlands, 23-24 November 2004.
4. Rodosek G.D. (2003), *A generic model for IT services and service management,* IFIP/IEEE Eighth International Symposium on Integrated Network Management, pp. 171-184.
5. Gozdecki J., Jajszczyk A. och Stankiewicz R. (2003), *Quality of service terminology in IP networks*, IEEE Communications Magazine, vol. 41, issue 3, pp. 153-159.
6. He, H. (2003), *What is Service-Oriented Architecture?*, http://www.xml.com/pub/a/ws/2003/09/30/soa.html, latest visit 2004-01-27.
7. Hall D. L. och Llinas J. (Eds.) (2001), *Handbook of multisensor data fusion*, CRC Press, New York.
8. Chang, S. K. (1995), *Towards a Theory of Active Index*, Journal of Visual Languages and Computing, Vol. 6, No. 1, March 1995, 101-118.
9. Maresca P., Chang S. K., M. Pesce M. (2006) *Application of Active Index to the Management of E-Learning*, Rivista della societ` italiana di e-learning Rivista Si-el, vol. 3, n. 2, 2006, Novembre 2006, pp. 331-341.

Cooperative Support for Multimedia/Multimodal Applications for Max/MSP

P. Bellini, F. Frosini, N. Mitolo, P. Nesi, M. Spighi
Distributed Systems and Internet Technology Laboratory
DISIT-DSI – University of Florence, Italy
Via S. Marta 3, Florence, Italy, nesi@dsi.unifi.it

Abstract

In the areas of performance set up and multimedia/multimodal music training experiences, cooperative work is becoming a very useful support. It can be used to experiment and exploit new modalities of training and for shortening the activities for the organization of large performance integrating in a simple manner distributed systems for audio-visual processing and general control (motors, lights, actions, etc.). To this end, a flexible model to cope with groups, roles, tools, and large set of features is needed. For this purpose, this paper presents a flexible Cooperative Work Support to allow creating a large set of cooperative applications structurally supporting: group role and tool concepts, undo/redo, joining and rejoining, preserve causality and consistency of commands, etc. It can be used for creating a large variety of multimedia/multimodal applications. This work has been developed with for the I-MAESTRO STREP FP6 project of the European Commission for developing cooperative educational multimodal tools in the Max/MSP.

Keywords: Multimedia, Peer-to-Peer, cooperative work, CSCW, P2P, Max/MSP.

1. Introduction

Computer technology and high speed networks have made possible the development of cooperative tools for real-time interaction of multimedia and multimodal applications. These applications require group working and management, where each member may join or leave the selected group dynamically [6], [5].

For example, in order to allow the creation of distributed system of control and cooperative work among the people involved in a live performance (director, conductor, light director, audio processing director, musicians, etc.) they have to share a common area of work and to exchange information among their computers and control panel console. Similar issues are useful for groups involved in a cooperative experience of music training and/or for the preparation of an ensemble via rehearsals [3] or communicating expressiveness via interactive multimodal devices [4], [1]. In all the above mentioned cases, two phases are relevant: (i) the set up of the cooperative work community/environment, and (ii) the management of the cooperative activities in a simple and transparent manner allowing the creation of several workgroups in the same network among the people and tools involved (humans and computer users), and allowing the communications among them at logical level disregarding the effective implementation of the network (see Fig.1, in which a generalized scenarios is depicted).

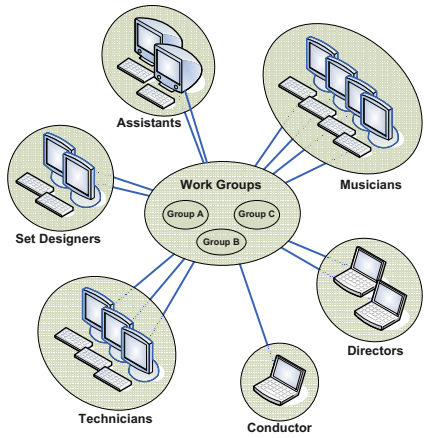


Fig. 1 – **General scenario**

This condition is typically synchronous, in which each performed change is made visible to all users. Tools for cooperative work can be of much benefit when discussing about common documents or graphs in real-time or quasi-real-time. The problems in those environments are due to: versioning management, undo, independency on the application layer as flexibility and reusability, causality of commands, consistency of commands and/or actions, undo/redo, rights control, history of changes, real-time response, flexibility in the group management, etc. Most of them depend on architectural solution, by commands and time management, and by communication protocols [3], [6], [7]. Despite to the vast literature exists on techniques and examples of cooperative tools and solutions [8], [9], [3], [10], no one of these solutions can be suitable to provide a satisfactory support for cooperative work on managing multiple communities/groups focused on performance management and/or for educational purposes of multimedia music, as explained in the following.

This paper is focused on presenting a Computer Supported Cooperative Work, CSCW, support for MultiMedia called COS4Media. COS4Media allows creating a large range of cooperative/collaborative applications, providing them a structural support for managing: groups, roles and tools concepts; undo/redo; joining and rejoining; preserve causality and consistency of commands; restoring of node status; absolute time synchronization, etc.

The COS4Media seems to be particularly suitable for enforcing collaboration into multimedia developed environments such as Max/MSP [1], EyesWEB [4]. On the other hand, it can be used to transform stand alone tools in fully distributed cooperative applications such as cooperative: doc editor, lessons class room manager, video player, metronome, white board, etc. The work presented has been performed in the I-MAESTRO STREP FP6 IST research and development project of the European Commission [2] for developing cooperative educational multimodal tools. In the I-MAESTRO project, Max/MSP has been chosen as the tool for audio processing in real time [1]. Max/MSP is a visual development environment for audio processing and multimedia, and user interface creation. It is more than fifteen years that composers, performers, software designers, researchers and artists are using it for creating interactive software for controlling performances and didactical tools. Max/MSP applications are called *patches* and they may run in real time on the Max players, thus allowing the production and the processing of music and events in real time. Max is highly modular, with most routines existing in shared libraries, and integrating third party module and components by means of *externals* and used in Max/MSP as native visual blocks.

The paper is organized as follows. In Section 2 the main scenarios and the conceptual model are presented. Section 3 reports an overview of the COS4Media architecture Computer Supported Cooperative Work Support for MultiMedia). In Section 4, some details about the usage of COS4Media into MAX/MSP applications are reported. Section 5 reports some applications. In Section 6, conclusions are drawn.

2. Conceptual Model and Scenarios

In a performance collaborative environment, several workgroups can be defined and set up. Each group put in communication a set of roles that share a set of tools establishing a certain number of collaborative relationships for communicating messages and commands. For example, three users and 5 computers are defining the cooperative environment in which a set of tools (humans and computerized) are creating a collaborative environment managing for example: TV camera, music scores, metronome, lights, motors, video synchronization, white board, a music score editor and/or viewer, slide show, and a chat.

In the COS4Media, the cooperative environment is modeled as a set of Groups. Each Group is modeled as a set of Roles. Each role can be covered by one (human or computerized) User and may be satisfied by one or more running Applications/processes. An Application can be a stand alone software application or can be a Logical Tool (e.g., metronome, motor control, music editor, light control, video grabber 1, audio grabber channel 2, video player, white board 45) into the software application (the software application can be a Max/MSP patch or a program). Thus for example, the status of the COS4Media collaborative environment can be: COS4Media = {Group G1, Group 2, ……..}

Where for group G1 we could have:

Group G1 = {Role Musician2, Role LightController,
Role GestureController, Role Director,
Role Musician2, ……. }
Role Musician1 (Violin I) =
{Tool Metronome, Tool Chat, Tool Score,
Tool HandGesture, Tool Lights, .. }
Role LightDirector =
{Tool Lights, Tool Score, Tool Chat, ..}
Role GestureController =
{Tool HandGesture, Tool Chat, …}
Role Conductor =
{Tool Motor, Tool Score, Tool Lights, ..}
Role Musician2 (Violin II) =
{Tool Motor, Tool Metronome, Tool Chat,
Tool Score, …}
Role Musician3 = ……
…….
Group G2 = …….
……

In this case, the GestureController is a computer based system (human supervised or not) monitoring and interpreting the Conductor gesture in order to change the Metronome and to send a video stream to remote musicians. A light Director is also collaborating in the environment, etc.

In such a collaborative performance environment, one of the Users may start by requesting the creation of the cooperative Group. The published Group presents the possible Roles that can be joined (for example, director, student, control 1, teacher, etc.). The joining to one of the role implies the automatic execution of the correct cooperative program (Max based or other) with the correct set-up (Max *patch* and/or with other parameters) on one of the computers involved in the cooperative environment.

3. COS4Media Architecture Overview

In the above mentioned scenarios a number of features are relevant: absolute time synchronization, the possibility of having undo/redo, the possibility of joining/rejoining the group obtaining updated status, fault tolerance about the information status, possibility of exchanging any kind of different artifacts, etc. Users

and tools that join or leave work groups at any time and the information about members have to be updated automatically whenever a configuration change happens. The COS4Media has to use information related to Groups, Roles, Tools for routing messages to the correct receivers and keeps trace of messages flows and occurred errors using log files.

The architecture proposed is fully distributed in the sense that all nodes of the collaborative environment have the same information and may play the role of master of group to allow saving the last final version of the artefacts produced. If the master leaves the group one of the others takes this role. As depicted in Fig.2, the proposed COS4Media is based on the following set of major components.

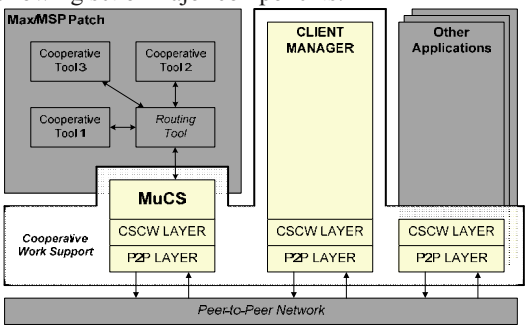


Fig. 2 – **General architecture of COS4Media**

COS4Media-CSCW: the layer realizing the CSCW facilities including support:

- to establish and manage workgroups of Users, with several applications and tools;
- to join and abandon a workgroup. Each workgroup can be joined by using its symbolic name obtaining updated information and the list of commands shared from the beginning;
- to resolve the messaging to peer nodes by using workgroup and roles.
- to manage different applications/tools of users joined with N:M relationships;
- to manage histories of changes/commands and perform undo/redo of commands. Including the resending of the status of command when a User re-joint or joint with delay the workgroup;
- to preserve causality and consistency of commands, reordering them according to absolute time stamp;
- to control rights of performing commands and actions, accepting commands only on the basis of the user profile;
- to collect commands, logs, events, errors and thus monitoring the whole distributed system.

COS4Media-P2P: the layer realizing the classical P2P facilities including support for:

- discovering the peers and thus of the users and applications of the CSCW;
- the absolute synchronization of the reference time on the different computers and thus allowing the execution of synchronous events and actions;
- exchanging files among peers;
- exchanging messages among peers;
- the activation/execution of programs on the peer node;
- hosting more peer nodes to the same physical computers.

To cope with the above features the COS4Media-CSCW has been designed by means of the classes depicted in Figure 3, [11]. CSCWLyaer Class wraps CSCW Service and provides methods for group management. It uses P2Player class wrapper to exchange information between peers. In order to make simple the communications among the P2P, CSCW and applications player the classes Invoker and Dispatcher have been created (their specialization are P2PInvoker, CSCWInvoker, etc.). Thus the Logger class is a singleton and allows to register all the actions performed. ActiveGroupList class wraps the list of groups currently working on the network. GroupManager class manages the specific group selected from the user.

CM, Client Manager: a tool based on COS4Media-CSCW and COS4Media-P2P that allows to

- public a Group structure;
- ask for the joining to other users and registering them;
- set up and put in execution the distributed cooperative Group in terms of Roles as applications on the several computers;
- monitor the activities of the COS4Media Layers, and of other applications of the cooperative environments (observing data, message, recovering errors, etc.).

The CM is supposed to be unique for each computer of the distributed system.

MuCS, Multimodal Cooperative Service: a library/module provided in the form of a standard Max external, that can be used by any patch in Max visual language.

Other Applications: The CSCW and P2P layers are also usable in other applications as libraries. These two layers allows to establish and management the cooperative work environment in a transparent manner. They can be exploited by a very large number of multimedia and control applications. They can be present in multiple instances on the same computer/peer as depicted in Fig.2. The applications and the tools are identifies in the cooperative environment with symbolic names.

COS4Media-CSCW maintains a list for messages history (e.g., commands and anti_commands, to perform the undo) since the creation of each Group.

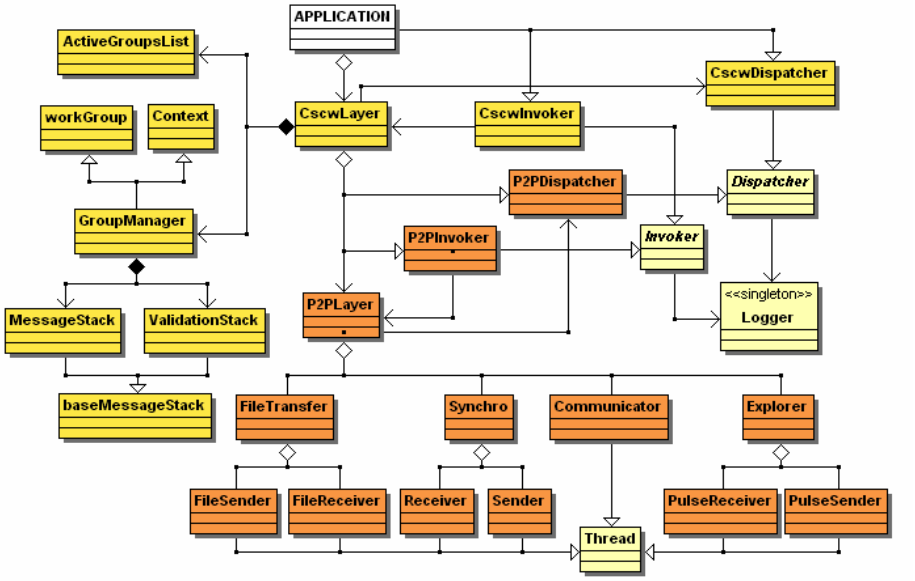


Fig. 3 – **Class diagram of the COS4Media-CSCW, COS4Media-P2P, main classes**

The automatic undo/redo feature is made possible for those applications that provide at the COS4Media-CSCW layer commands with their associated anti-command, which is a command that may invert the effect of the command itself. At application level the command may appear such as:

Message (Tool Score, …….,
Command (Insert Note(a,…) in C),
Anti_Command (Delete Note(a) in C))

Please note that anti_command related to actions on attributes of artifacts (color of a note) are obtained regenerating the previous status. Commands and anti_commands have to refer to symbolic references into the artifact such as the logic name of elements and attributes, or the logic position/location.

The applications may automatically generate the anti_command from a command, while the COS4Media is totally independent on the application and artefact structure and model. This permits to provide a collaborative support for a large range of applications independently of what they manipulate. COS4Media service has the capability to support time consistent execution of commands through the network using the COS4Media-P2P synchronization service, that periodically starts a synchronization procedure. For time critical applications, such as distributed sound playing, the user can also request an immediate synchronization procedure at any time.

In addition, the above mentioned list of commands/messages is kept time consistent using the messages *timestamp* signature that every messages carry by itself.

Network delivery delays may cause wrong messages sequences, in this case COS4Media aims to restore the time consistent sequence by calling anti-commands of wrong time messages and then redo in the right order the list of command (guaranteeing the consistency and causality of actions on the cooperative environment [6], [3]).

4. Cooperative Work Support in Max/MSP

The MuCS provides a set of functions for tools working inside a cooperative activity. The service is implemented as Max/MSP external. In order to use this service to create an object inside Max/MSP Designer Interface is mandatory.

In terms of the Max/MSP visual language, the MuCS is a box which provides a set of connection points for exchanging information in the Max/MSP environment as shown in Fig.4. The object presents three input and three output ports which are called inlets and outlets, respectively, in Max/MSP terminology (i.e., for sending and receiving information from the cooperative environment). The MuCS provides three different communications *channels* divided as services:

- *Time Critical Commands* (left side inlet/outlet): commands that nodes in the cooperative environment have to execute at the same time, such as starting a distributed tool;
- *Generic Messages* (middle inlet/outlet): messages that do not need to be strictly synchronized, like parameters setting;
- *CSCW Specific Controls* (right side inlet/outlet): work group join and leave, to get the number of online members, etc.;

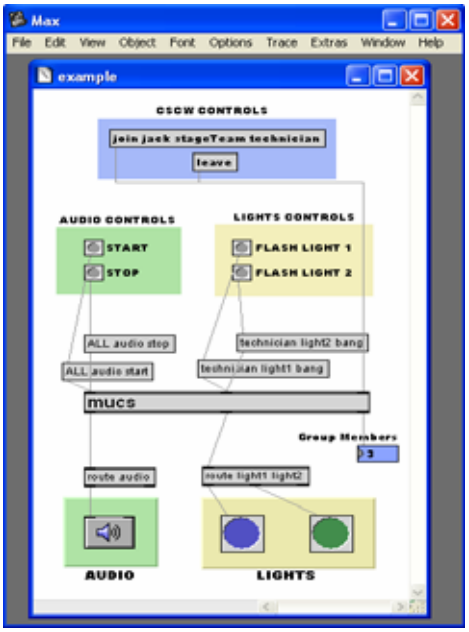


Fig. 4 - **Max/MSP cooperative patch example**

In order to show the effective capabilities of MuCS in virtualizing the cooperative activities (making simpler the communication of different tools among different patches/applications) the following diagram of Fig.5 is provided. The diagram depicts a use case in which one director and some technicians are joined together in the same work group. Two Max/MSP patches are active one for each role, including the MuCS external and the necessary logics. The director's patch contains several interactive buttons for manage, remotely, different light tools in the technician's patch. The situation is the following: Director wants to turn on the tool "light 1" in the technician's patch/application.

Director's side:

- User press button to turn of a light;
- the Patch generates the activation message and sends post into its MuCS;
- The MuCS extracts Role information from the message and invokes CSCW sending procedure;
- CSCW searches internal Role table finding users having the same role and for each uses P2P message service;
- P2P converts peer name in network address and sends the message.

Technician's side:

- P2P receives a message from a network address, converts it in peer name with the internal peer map and passes it to the CSCW;
- CSCW registers message and sender in the log file and passes the information to MuCS;
- MuCS reconstructs the Max/MSP message from data and uses the patch routing tool for correct delivering;
- routing tool uses the first word of the message for selecting the correct light tool and gives it the rest of the message;
- Tool "Light 1" is tuned on.

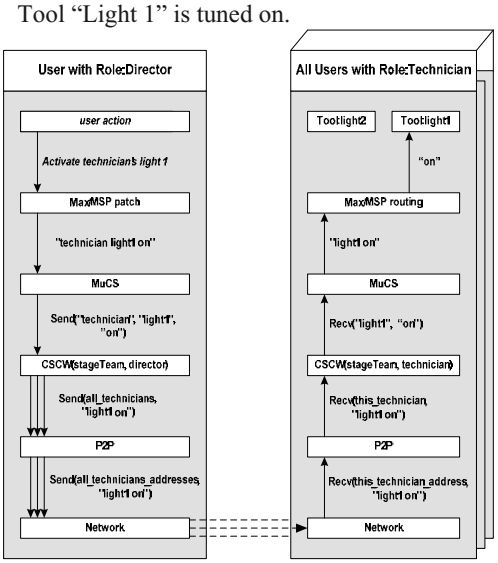


Fig.5 – **Virtualizing the message management**

5. Examples Application

Following figures show two different examples of COS4Media applications. The first consists of a cooperative music score editing tool (see Fig.6) see also [4], in the case of COS4Media data are fully distributed. User can choose and modified only one of the main score parts (in this case Violin or Cello) according to his own role, while he can follow any changes performed in all parts (Teacher role can modify both parts).

The Fig.7 shows an application in which a multimedia context allows the creation of cooperative sessions involving computers with Max/MSP environment using MuCS. It provides some features regarding:

- creating, joining and leaving a workgroup;
- starting/stopping video/audio in a synchronized mode;
- sending messages among different nodes of the cooperative environment and workgroups;
- painting in the whiteboard of the joined workgroup;
- chatting with the workgroup members.

All the previous cooperative actions/Tools can be shared with all Group users or with users that have an application using the same Tool.

6. Conclusions and Future Directions

This paper has presented COS4Media, a flexible Cooperative Work Support to allow creating a large set of cooperative applications structurally supporting cooperative: undo/redo (commands and anti-commands), joining and rejoining of groups, absolute synchronization, concepts of Tools to allow

establishing more flexible grouping, activating and monitoring the cooperative Group with CM application, preserve causality and consistency of commands, reordering them according to absolute time stamp, etc.

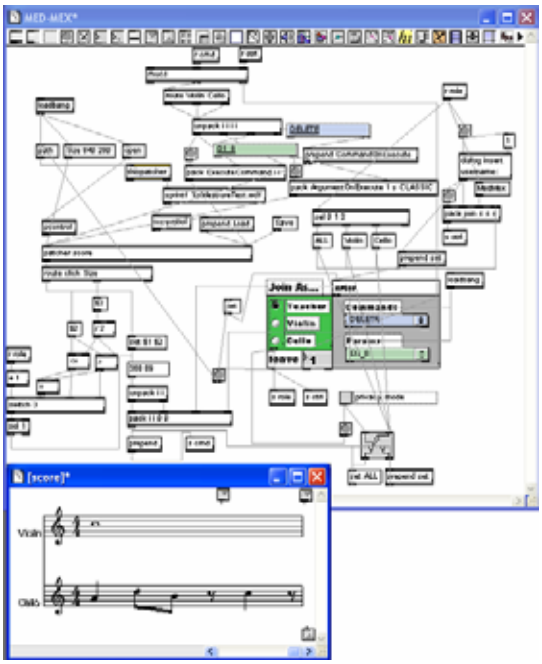


Fig. 6 - **Cooperative musical score editing**

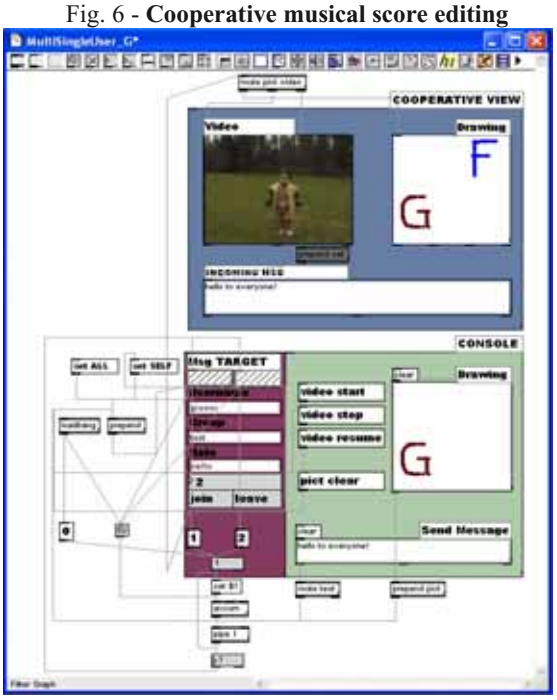


Fig.7 - **Cooperative multimodal session**

The COS4Media has been developed for the I-MAESTRO STREP FP6 project of the European Commission with the aim of creating cooperative applications for: (i) multimedia, (ii) gesture and posture recognition and tracking, (iii) music editing and lesson support based on MPEG ISO Symbolic Music Representation. COS4Media has been has used up to now to enforce collaborative working layer into Max/MSP, and also in stand alone applications for music editing.

Acknowledgements

The I-MAESTRO project is partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Programme for R&D (IST-026883, www.i-maestro.org). Thanks to all I-MAESTRO project partners and participants, for their interests, contributions and collaborations.

References

[1] Max/MSP: http://www.cycling74.com/

[2] I-MAESTRO EC IST project, URL: http://www.i-maestro.org,

[3] P. Bellini, Nesi, P., Spinu, M. B. (2002). Cooperative Visual Manipulation of Music Notation. ACM Trans. on Comp.-Human Interaction, Sept., 9(3):194-237,

[4] Camurri, G. Volpe, G. De Poli, M. Leman. "Communicating expressiveness and affect in multimodal interactive systems". IEEE Multimedia 12 (1), pp. 43-53, JAN-MAR 2005.

[5] Belini, P., Fioravanti, F., Nesi, P. 1999. "Managing Music in Orchestras". IEEE Computer, 26-34

[6] Borghoff U. M., AND Schlichter J. H., 1998 Computer Supported Cooperative Work, Springer, Berlin.

[7] Teege, G., 2000. "Users as Composers: Parts and Features as a Basis for Tailorability in CSCW Systems", Computer Supported Cooperative Work, Kluwer Academic Pub., Vol.9, 101-122.

[8] Dourish P., 1998. "Using Metalevel Techniques in a Flexible Toolkit for CSCW Applications". ACM Transaction on Computer Human Interaction, Vol.5, N.2, 109-155.

[9] Edwards W. K., AND Igarashi T., AND LaMarca A., AND Mynatt E. D., 2000. "A temporal model for multi-level undo and redo". In Proc. of 13th annual ACM symp. on User interface software and technology, San Diego, CA, USA, 31-40.

[10] Dennis C. Neale, John M Carroll, Mary Beth Rosson, Evaluation methods: Evaluating computer-supported cooperative work: models and frameworks, Proceedings of the 2004 ACM conference on Computer supported cooperative work CSCW '04

[11] Francesco Frosini, Giovanni Liguori, Nicola Mitolo, Paolo Nesi, DE5.1.1, revised, Model and Support for Cooperative Work and SMR for Music Education, DE5.1.1, IMAESTRO project.

A Black-box Testing Method for Multi-Agent Systems*

KaiYu Wan, Vasu Alagar
Department of Computer Science and Software Engineering
Concordia University, Montreal, Canada
ky_wan, alagar@cse.concordia.ca

ZongYuan Yang
Department of Computer Science
East China Normal University
yzyuan@cs.ecnu.edu.cn

Abstract

A Multi-agent system (MAS) is an open distributed system of agents, some that are autonomous and some that are reactive. They interact among themselves according to a set of norms and a variety of protocols. Due to the dynamic changes in the environment of the MAS there may not be a perfect correspondence between the actual global state of the agent's environment and the agent's representation of it. This disparity introduces a level of complexity in correctly programming the MAS. Thus it is essential to test the conformance of an implementation of the MAS to its high-level declarative architectural description, before the MAS is deployed for practical applications. In this paper we propose a black-box testing method that certifies the correctness of an implementation of MAS while identifying the faults, and omissions in it.

1 Introduction

A Multi-agent system (MAS) is an open distributed system of agents, some that are autonomous and some that interact among themselves according to a set of norms and a variety of protocols. An agent is obliged to respect its norm. The content of the norm establishes for the bearer of the norm the obligation to abide by it, and consequently to adhere to the prescribed behavior. In exceptional situations if an agent cannot abide by its norm, its norms will be re-instantiated so that it achieves its goals and provides expected services in a timely manner to its clients.

We use the term context to formalize the notion of a *setting in which an observation is made*. The setting can be either a system or environmental setting, and an observation in that may be attributed to one or more of the following: *event occurrence*, *message passing*, *agent state*, *environmental state*. An agent has *internal context awareness* (self-awareness) and *context awareness*. In particular a norm-aware agent must be aware of the context in which its norm is instantiated in order that it may react appropriately. In many applications this reactive behavior may also have strict time constraints. These factors must be accounted for in programming the MAS. Managing the complexity in the behavior model of the MAS is not a small task. Consequently, ensuring the correct behavior in an implementation of the MAS is quite hard. By correctness we mean that the program behavior is consistent with the architectural description of the MAS and the normative behavior of every agent is not violated in the program. The contribution of this paper is a method to test the correctness of an implementation of MAS.

It is hard to formally verify the MAS, because of the dynamic nature of the environment and the complexity involved in formalizing the environment in which the MAS has to function. Testing, based on the architecture specification of MAS, is a formal validation of the system and is an attractive alternative to formal verification. The test method that we propose in this paper takes into account the context in which the MAS functions and the dynamic agent configurations that are formed to solve specific problems.

1.1 Basic Assumptions on Agents

By agent we mean software agents governed by BDI (Belief, Desire, Intention) semantics. The set of beliefs is time-varying. An agent performs actions either because it wants to achieve its own goal or it is interacting with another agent in its environment. Desires are actions of the first kind. They are triggered by events initiated by the agent, not affecting any other agent in its environment. In our formalism we call such events as *internal events*. In an interaction, an agent performs a set of basic actions that are allowed by its set of goals and justified by its beliefs. We call the events that trigger interactions as *shared* (or *external*) events. An intention is a particular course of action to which the agent has committed. An action intended at an instance is to be committed at a future instance. Thus at any instance the intention of an agent is the set of all commit-

*This work is supported by grants from Natural Sciences and Engineering Research Council, Canada

ted but unfinished acts. Instead of describing each agent in isolation, we consider agent *types*. An agent type is characterized by a set of roles. All agents of an agent type play the same role, however they may have different implementations so that they may be used in different contexts. A generic classification of agent types [1] consists of *interface agent* (IA), *middle agent* (MA), *task agent* (TA), and *security agent* (SA). Agents are instances of agent types.

1.2 Basis of Testing

In this paper we raise and answer the following questions: (1) how to test the correctness of an agent implementation? (2) how to test the correctness of agent collaboration? (3) how to identify some behaviors in the specification that are not in the implementation?

A BDI agent is realized as a software component matching its behavior. A component corresponding to an agent is a black box. The component behavior, defined as input-output behavior is observed at its interfaces. If the observed behavior of the component is consistent with the abstract behavior, as defined by the BDI semantics, then we say the agent has a correct implementation. To test MAS we use a formal architectural model of it and define the component of the MAS as a composition of the components that correspond to the agents in the MAS. The behavior of the composite component is expressed as a composition of the behaviors of the components in the composition. If this behavior matches the abstract behavior of the MAS, as reasoned from its BDI semantics, we certify the correctness of the component-based implementation.

We model the BDI semantics as a transition system, extended with abstract data types, clocks, and predicates over data type and clock variables. The test cases generated from such a formal specification can detect these types of flaws as well as certify the correctness. An implementation under test (IUT) is either an implementation of a primitive component or an implementation of a composite component. A test case is a sequence of events generated from the BDI specification of the IUT, and it tests in the IUT that certain state is reached, or a certain functionality is correct, or a time constrained action occurs.

2 Architecture for Multi-Agent System

The software specification of an agent type is a component type $T = \langle Fr, Ar \rangle$, where Fr, called frame is a black-box view, and Ar, called architecture, is a grey-box view of the agent type. The black-box view defines the interface types, and the particular grey-box view Ar is a structured implemented version of Fr. An interface of a component is an instance of either *notifies-interface* type or *receives-interface* type. The architecture is primitive if its structuring is to be provided in an underlying implementation (outside the scope of component specification language). A non-primitive architecture includes several subcomponents *nested* to several levels. A specific implementation of a non-primitive structure is obtained by (i) instantiating adjacent level subcomponents, and (ii) specifying the interconnection between subcomponents by means of their interface ties.

In the design the goals of a BDI agent are identified as a list of requirements for the component. A component may be designed to function either autonomously or be reactive consistent with the intentions of reactive agents. A component is given sufficient *computational power*, *resources*, and *knowledge* to complete the *tasks* assigned to it from its environmental partners and deliver results within specified time bounds. Resources include its set of beliefs, and databases for transacting with other agents. A component communicates with other components in its environment through messages (events) at its interfaces.

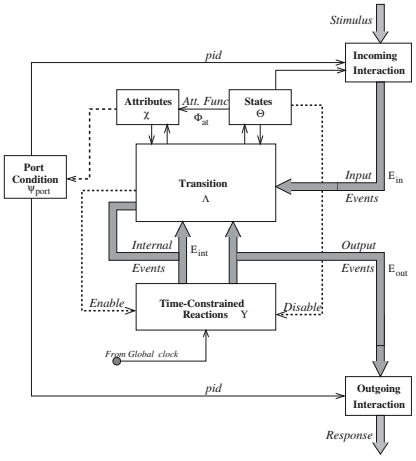


Figure 1. Agent Behavior

2.1 Component Behavior and BDI

An abstract interpretation of agent behavior is shown in Figure 1. The filled arrows in the figure indicate flow of events. A message is either an input event or an output event. An internal event corresponds to a task that the agent has to perform by itself. The fulfillment of the desires and communication tasks are governed by certain conditions, formalized as logical expressions involving the terms on the local attributes, and terms defined in the included abstract data types. In the process of executing its plans and fulfilling the desires, an agent intends certain actions. We call them constrained actions, or *time-constraints*. Time constrained actions are to be committed by the agent within

a window of time. Some of these constraints are made explicit in the statement of requirements derived from the plan. The dynamic behavior of the agent depends on its local context, it environmental context and/or the context, and is defined by the set of messages that can be received from or sent to other agents in a specific application.

The states (Θ), transitions (Λ), and time constraints (Υ) in the figure determine the behavior of an agent in fulfilling a task in a specific context. A message from the context is received by the agent only when the internal state of the agent, and the interface-condition where the message is received satisfy the specification for state change. The incoming message may cause its state to change, and may also require an internal computation, which updates the agent's beliefs, knowledge, and attributes, shown by the arrow labeled with 'Att.Func.' Some internal action or an output message may have to be performed in future according to certain constraints, including time constraints. The dotted arrow connecting the block of computation to that of time-constrained reaction signifies the enabling of one or more future actions due to a computation. This constitutes the intention part of the agent. Based on an internal clock, an outstanding event is performed by the agent, thereby generating either an internal event or an output message (or control message). By comparing this abstract interpretation of the component with the interpretation of Bordini [3] for AgentSpeak(L) programs we can be convinced of the faithfulness of the embedding of a BDI agent in the software component that encapsulates it.

2.2 Behavior of Agent Collaboration

The behavior of the collaboration is determined by the product of the transition systems corresponding to the agents. The behavior of a collaboration of several agents $A = \{A_1, \ldots, A_n\}$, where A_i is an agent is the set of (global) states G, $G \subset \Theta_1 \times \ldots \times \Theta_n$, where Θ_i is the set of states in the behavior of A_i. Informally, an event occurring at a global state $\theta = (\theta_{1i_1}, \ldots, \theta_{n,i_n})$, $\theta_{ji_j} \in \Theta_j$, is either an internal event occurring at a state θ_{k,i_k} of the agent A_k or it is shared by two agents, say A_j and A_k. In the former case, the next global state is the tuple $(\theta_{1i_1}, \ldots, \theta'_{k,i_k}, \ldots, \theta_{n,i_n})$, where θ'_{k,i_k} is the next state in agent A_k. In the second case, the agents A_j and A_k change their statuses simultaneously. That is,

$$\theta' = (\theta_{1,i_1}, \ldots, \theta'_{j,i_j}, \ldots, \theta'_{k,i_k}, \ldots, \theta_{n,i_n})$$

2.3 Agent Configurations

Unlike traditional systems wherein a collaboration is static, an agent collaboration is dynamic. We view a collaboration as a composition of several *configurations*, where a configuration is dynamically constituted in the system to solve a certain problem or a sub-problem in one domain. Agents in a configuration typically belong to the three types IA, MA, and TA of agents. The front end of agents in IA provides user interface, and helps user to represent their requirements and possible solutions into message sequence diagrams. At the back end, those sequence diagrams are translated into *Corde expressions*, which are defined later in this section. These Corde expressions are transferred to agents of type MA who analyze the *Corde Expression*, decide the configurations consisting of agents in TA and distribute the sub-tasks to the TAs. The task agents fulfill the given subtasks and inform the results to the MAs. The MAs combines the results from the TAs and give the feedback to the IA. The component corresponding to a configuration is a composition of components where each component is the implementation of an agent in the configuration. These configuration depend upon the structure of *Corde* expressions.

2.4 Corde Expression

A *Corde* expression is a sequence of messages combined by a set of constructs. These constructs are n-ary constructs whose operands are messages between agents. A *Corde* expression is equivalent to a configuration. Figure 2 shows the fundamental configurations, called atomic components, that correspond to *Corde* expressions discussed below.

Sequential Composition Construct $\gg$ The expression $a \gg b$ defines the sequential composition of messages a and b. That is, the message a is first executed, and using the result of its evaluation the message b is executed.

Parallel Composition Construct $\parallel$ The expression $a \parallel b$ defines a parallel composition of messages a and b. The result of execution is the merging of the results in time order.

Composition with no order $\circ$***and*** Given two messages a, b, the expression $a \circ b$ defines that messages a and b should be executed by the receiver agent, however the order of execution is not important. The result of execution is the set of results produced by the execution of messages a and b.

Nondeterministic Choice Construct $\wr$ Given two messages a, b, the expression $a \wr b$ defines that one of the messages be executed nondeterministically.

Priority Construct $\diamond$ Given two messages a and b, the expression $a \diamond b$ defines that message a should be executed first, and if it succeeds, the message b is to be discarded; otherwise, the message b should be executed by the receiver agent.

All the constructs have the same precedence, and hence a *Corde* expression is evaluated from left to right. To enforce a particular order of evaluations, parenthesis may be used. Any general configuration can be expressed as a composi-

tion of two or more atomic components.

3 Testing MAS Behavior

We consider *state coverage* and *transition coverage* as the two test adequacy criteria in our testing method. A *state cover* is a set of test cases required to identify each state in the component behavior to some states in the implementation. We are interested in a state cover *SC* of minimum size. A *transition cover* is a set of test cases required to identify each transition in the component behavior to some actions in the implementation. The transition $s_i \xrightarrow{e} s_j$ can be tested by the test case $t.e$ for every test case t in the state cover of the state s_i. We are interested in a transition cover *TC* of minimum size. The *test suite* for testing a component implementation is defined by $T = SC \cup TC$.

3.1 Step 1. Unit Testing

Each agent in MAS has a component design with its behavior defined by an extended transition system. We generate test cases from each transition system to test states and transitions (functions) in the implementation of the component. For each test case, a test is prepared and is input to the implementation. The result from administering a test is interpreted as follows:

Using State Cover

To test whether the implementation has a state corresponding to a state s in the specification, we use the state cover τ_s and its parameter list. The test $\tau_s \in SC$ is a sequence $(\lambda_0, \lambda_1, \ldots \lambda_i)$, where $\lambda_j \in \Lambda$, $s_0 \in \Sigma_0$ is the source state of transition λ_0 and s is the destination state of transition λ_i. The parameter list for τ_s is $(t_1, t_2, \vec{V}', \mathcal{W}')$, where

t_1 is the instance at which the implementation started at its initial state,

t_2 is the instance at which the implementation finished the test request,

$\vec{V}'$ is the set of variables in state s, for which values at instance t_2 are required, and

$\mathcal{W}' = \{\lambda \mid \lambda \in \tau_s\}$ is a subset of transitions in the test τ_s that are time constrained in the specification.

Assume that the implementation responds as follows:

$t_1 = a, t_2 = b$;

$\vec{V}^* = \{(V, v) \mid V \in \vec{V}'\}$, where v is the value of V in the implementation at instance t_2, and

$\mathcal{W}^* = \{x_\lambda \in \mathbb{R}^{\geq 0} \mid \lambda \in \mathcal{W}'\}$. For each $\lambda \in \mathcal{W}'$, x_λ is the time of invocation of its implementation.

The oracle will declare that a correct implementation for state s exists if all the following conditions are satisfied:

[finiteness] $a < x_\lambda < b < \infty, x_\lambda \in \mathcal{W}^*$;

[time constraint satisfaction] the time constrained predicate $tc(\lambda)$ for the transition λ evaluated at time x_λ is true: that is, $tc(\lambda)[x_\lambda]$ is true for $\lambda \in \mathcal{W}', x_\lambda \in \mathcal{W}^*$;

[functional correctness] the post condition $a_{post}(\lambda_i)$ for transition λ_i must be true when the values of state variables are substituted in the predicate $a_{post}(\lambda_i)$; that is, $a_{post}(\lambda_i)[V/v]$ is true when every $V \in a_{post}(\lambda)$ is replaced by its value v for every pair $(V, v) \in \vec{V}^*$.

The faults uncovered by state cover test are (1) time constraint violation along an execution path in the implementation, and/or not performing an intended action; (2) missing state in the implementation, which indicates that an update to the knowledge base has failed or an inference is not complete; and (3) inaccurate state implementation (inaccurate reasoning) or action fault.

Using Transition Cover

Let $\lambda = (s_i, e, s_{i'})$ be the transition in the specification for which we test the implementation. We assume that the implementation has already been tested for states s_i and $s_{i'}$. Hence, there exists in the implementation states s_i^*, and $s_{i'}^*$, which respectively correspond to the specification states s_i and $s_{i'}$. Let $\tau_\lambda = \lambda_0, \lambda_1, \ldots, \lambda_i$, $\tau_\lambda \in TC$, $\lambda_i = \lambda$, be the test case to test the implementation of τ_λ. The parameter list for this test is $(t_1, t_2, x_i, x_{i'}, \vec{V}_i', \vec{V}_{i'}', \mathcal{W}_i', \bar{e})$, where

t_1 is the instance at which the implementation started at its initial state,

t_2 is the instance at which the implementation finished the test request,

$\vec{V}_i'$ is the set of variables in state s_i, for which values at instance t_2 are required,

$\vec{V}_{i'}'$ is the set of variables in state $s_{i'}$, for which values at instance t_2 are required,

$\mathcal{W}_i' = \{\lambda' \mid \lambda' \in \tau_\lambda\}$ is a subset of transitions in the test τ_λ which are time constrained in the specification,

$\bar{e}$ is the set of specification events used in the implementation during the passage from state s_i^* to state $s_{i'}^*$.

Assume that the implementation responds as follows:

$t_1 = a, t_2 = b, x_i = a_i, x_{i'} = a_{i'}$;

$\vec{V}_i^* = \{(V, v) \mid V \in \vec{V}_i'\}$, where v is the value of V in the implementation at instance t_2,

$\vec{V}_{i'}^* = \{(V, v) \mid V \in \vec{V}_{i'}'\}$, where v is the value of V in the implementation at instance t_2, and

$\mathcal{W}_i^* = \{x_{\lambda'} \in \mathbb{R}^{\geq 0} \mid \lambda' \in \mathcal{W}_i'\}$. For each $\lambda' \in \mathcal{W}_i'$, $x_{\lambda'}$ is the time of invocation of its implementation.

The oracle will declare that a correct implementation for transition λ exists if all the following conditions are satisfied:

[transition correctness] every event in $\bar{e}$ which does not label a transition $\lambda_j \in \tau_\lambda$ should not be a specification event: for all $e' \in (\bar{e} \setminus \{e \mid \exists \lambda_j = (s_j, e, s_{j+1}), \lambda_j \in \tau_\lambda\}) \rightarrow e' \in \mathcal{E}_\mathcal{A}$, where $\mathcal{A}$ is the GRC under test;

[pre condition satisfaction] the guard for the transition must be true: $g(\lambda)_{pre}[V/v]$ is true when every $V \in g(\lambda)_{pre} \subset \vec{V}_i$ is replaced by its value v as given in $\vec{V}_i^*$;

[finiteness] $a < x_{\lambda'} < b, x_{\lambda'} \in \mathcal{W}^*$;

[time constraint satisfaction] $tc(\lambda')[x_{\lambda'}]$ is true for $\lambda' \in \mathcal{W}_i', x_{\lambda'} \in \mathcal{W}_i^*$;

[functional correctness] $a_{post}(\lambda_i)[V/v]$ is true when every $V \in a_{post}(\lambda_{i'}) \subset \vec{\lambda}_{i'}$ is replaced by its value v for $(V, v) \in \mathcal{V}_{i'}^*$.

When any one of the above condition is false, we declare the test case has identified a fault in implementing the transition. The faults that can be uncovered by transition cover test are (1) missing transition (a transition is not enabled); (2) extra transition (implementation has a functionality that is not part of the specification); (3) action fault (a goal is not reached); and (4) time constraint violation - an intention is not fulfilled.

3.2 Step 2. Pair Testing

We generate test cases from the product machine of the components corresponding to every pair of interacting agents in the MAS. For instance, an interface agent and a middle agent has to interact. The behavior of this interaction is the product of their individual behaviors. From the product machine, we generate the state cover and transition cover and use them to test the implementation of their combined behavior. The test results are interpreted as in Section 3.1.

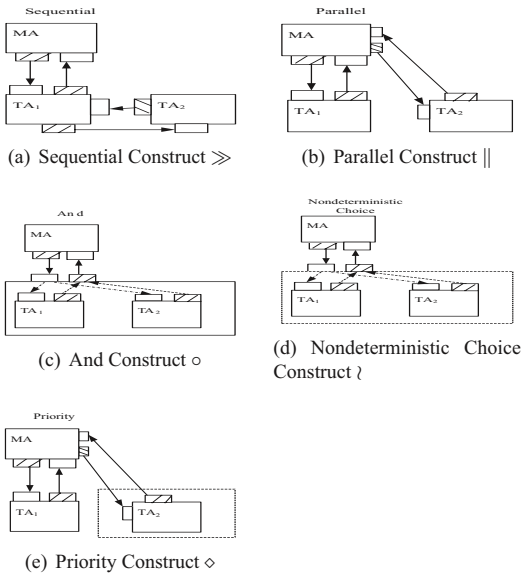
3.3 Step 3. Testing Fundamental Configurations

We can test the fundamental configurations by adapting the technique for "pair testing". A general configuration can be tested by splitting the configuration into fundamental configurations, testing the fundamental configurations independently, and putting these results together.

Testing the Component for Sequential Construct The atomic configuration in Figure 2(a) can be split into three sets of interacting pairs: $\langle MA, TA_1\rangle$, $\langle TA_1, TA_2\rangle$ $\langle MA, TA_2\rangle$. The implementation of each pairs is tested individually by the technique described in Section 3.2.

Testing the Component for Parallel Construct The atomic configuration in Figure 2(b) can be split into two independent sets of interacting pairs: $\langle MA, TA_1\rangle$ and $\langle MA, TA_2\rangle$. Although TA_1 and TA_2 are of the same type, it is necessary to test both the pairs $\langle MA, TA_1\rangle$, and $\langle MA, TA_2\rangle$. The technique described in Section 3.2 can be used to test their implementation.

Testing the Component for And Construct The atomic configuration in Figure 2(c) can be regarded as an interacting pair, one of them is the component *MA* and the other is the composition of components TA_1 and TA_2. Let *TA* denote the composition of TA_1 and TA_2. We test *MA* using unit testing method discussed in Section 3.1, and test the composition *TA* of components TA_1 and TA_2 by pair



(a) Sequential Construct $\gg$ (b) Parallel Construct $\|$

(c) And Construct $\circ$ (d) Nondeterministic Choice Construct $\wr$

(e) Priority Construct $\diamond$

Figure 2. Atomic Components of Corde Constructs

testing discussed in Section 3.2. We test the interacting behavior of *MA* and *TA* by applying again pair testing: that is, we construct the product of the behaviors of *MA* and *TA* and generate test cases from it to test the implementation of the interaction between *MA* and *TA*..

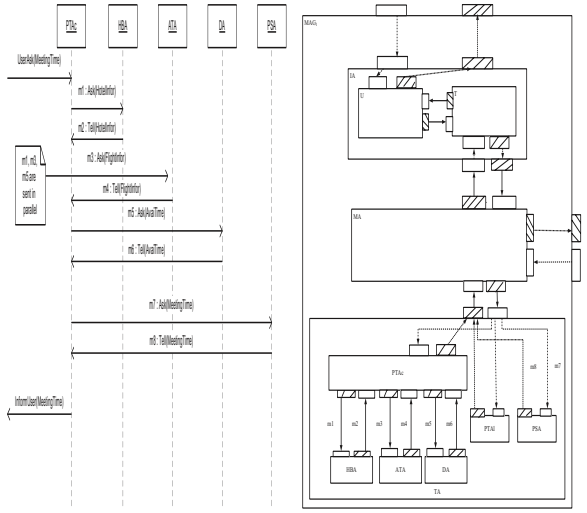
Testing the Component for Nondeterministic Choice Construct The atomic configuration in Figure 2(d) can be split into two independent sets of interacting pairs: $\langle MA, TA_1\rangle$, $\langle MA, TA_2\rangle$. The semantics of nondeterministic choice forces us to test both the pairs. The technique described in Section 3.2 can be used to test their implementation.

Testing the Component for Priority Construct The atomic configuration in Figure 2(e) can be split into two independent sets of interacting pairs: $\langle MA, TA_1\rangle$, $\langle MA, TA_2\rangle$. The semantics of priority forces us to test both the pairs. The technique described in Section 3.2 can be used to test their implementation.

3.4 Testing a General Configuration

In this section, we take the example discussed in [6] to illustrate our approach to testing a configuration of MAS. A simplified version of the *travel planning problem* (TPE) is as follows: *Caroline would like to meet Liz in London for preview receptions of an exhibition at the Tate Gallery. These will be held at the beginning of October. Both Liz and Caroline have other appointments around that time, and will need to travel to London from their homes in Paris and*

New York. Plan a solution. For simplicity of presentation, we only present the procedure for Caroline. There should be another part in the figure for Liz. The agent *IA* interacts with users, produces a sequence diagram as shown in Figure 3 (a), and translates the sequence diagram into the *Corde* expression $(\mathtt{m}_1 \gg \mathtt{m}_2) \parallel (\mathtt{m}_3 \gg \mathtt{m}_4) \parallel (\mathtt{m}_5 \gg \mathtt{m}_6) \gg \mathtt{m}_7 \gg \mathtt{m}_8$.



(a) Sequence Diagram for TPE (b) Configuration for TPE

Figure 3. TPE Example

After receiving the above description from the HAs, the MA creates the agent configurations as shown in Figure 3(b) and initializes the task agents in the configuration with the data required for solving the problem.

Using the test case generation methods described in the earlier sections we can generate test cases from the architecture shown in Figure 3(b) to test an implementation of this configuration. A step by step guide line follows:

1. The components *HBA*, *ATA*, *DA*, *PTA*1, *PSA*, *U*, *T*, *PTAC*, and *MA* are primitive. Test them using unit testing method.

2. Test the component *IA* using pair testing on the pair $\langle U, T \rangle$.

3. Test the interactions on the pairs $\langle HBA, PTAC \rangle$ $\langle ATA, PTAC \rangle$ $\langle DA, PTAC \rangle$.

4. Let *C*1 denote the configuration (as shown in Figure 3(b)) composed of the components $HBA, ATA, DA, PTAC$. Test the interaction between *C*1 and *MA*. Test the interactions between the pairs $\langle MA, PTA1 \rangle$ $\langle MA, PSA \rangle$. This completes the testing of the interaction between *TA* and *MA*.

5. Test the interaction between *MA* and *IA*.

6. Consider the composition of *MA* and *IA* as a component *C*2. Test the interaction between *C*2 and *TA*. This completes the testing of the component *MAG*1.

4 Conclusion

The main contribution of this paper is a black-box testing method for multi-agent systems. We encapsulated an agent as a component whose behavior satisfies the BDI semantics of the agent. By introducing a transition system extended with knowledge stores, guards that enforce normative constraints and time restrictions we model the BDI semantics as accurately as possible as software components. We stipulate that state cover and transition cover generated from such a transition system is sufficient to test the conformance of the implementation to the component-based architecture of agents. In the MAS architecture, an IA is trained to produce a *Corde* expression corresponding to a message sequence diagram. The MAs are given two kinds of expertise: (1) knowledge to understand and interpret message and (2) use the information in message and *Corde* expression to construct a configuration of task agents. Thus, the test cases are generated from dynamically constituted agent configurations in the MAS. We have explained how to interpret the test outcomes. We intend to automate the test case generation, testing, and interpretation of test results in future.

References

[1] V.S. Alagar, J. Holliday. *Behavior Extensions to FIPA Architecture and Specifications.* Proceedings of the 2nd International Conference on Active Media Technology (ICAMT), Chongqing, China, May 2004, World Scientific Publishers, 211-218.

[2] Vasu Alagar and Mao Zheng. *A Software Architecture for Multi-Agent Systems.* Computational Intelligence and Security (CIS2005), LNAI, Springer-Verlag, Vol. 3801, 303-312.

[3] R.H. Bordini, and A.F. Moreira.: *Proving BDI properties of agent-oriented programming languages*. Annals of Mathematics and Artificial Intelligence, 0: 1-30, 2004.

[4] Amira Regayeg, Ahmed Hadj Kacem, Mohamed Jmaiel. *Towards a Formal Methodology for Designing Multi-agent Applications.* MATES 2005: 153-164.

[5] C. Szyperski, *Component Software, Beyond Object-Oriented Programming.* Addison-Wesley, 1997.

[6] K. Wan, V.S. Alagar. *An Intensional Programming Approach to Multi-agent Coordination in a Distributed Network of Agents.* Declarative Agent Languages and Technologies III (DALT 2005), LNAI, Springer-Verlag, Vol. 3904, Page 205-222.

A NEW BROWSING MODEL BASED ON AJAX STRATEGIES

Walter Balzano
Dipartimento di Scienze Fisiche
Università di Napoli, Federico II – Italy
walter.balzano@gmail.com - http://people.na.infn.it/~wbalzano/

Maria Rosaria Del Sorbo
Dipartimento di Matematica ed Applicazioni
Università di Napoli, Federico II – Italy
marodel@gmail.com

Abstract
Web sites often include a huge amount of information, which might disorientate users and let them loose a full awareness about the subject of their surfing. Furthermore, required data might need several steps to be reached and it can be difficult to deep the knowledge without a considerable waste of time. This work is intended to reduce these difficulties, introducing a browsing model based on Ajax strategies. A new visual semantic is introduced through special double underlined link that means connections to "mini pages", special metadata which extend standard meta tags, usually present in web pages code. This model draws inspiration from an enhancement of metadata associated to web pages considered as a complex set of data. Such as in JPEG, PDF and MP3 files, also in web pages raw data, indispensable for the visualization, can be completed aggregating meta information. Added data allows the introduction of a double "miniature" of the page. This miniature is information and not graphical web page summary, edited in authoring phase. It can be shown directly as a preview from the hotwords linking to this web page with different detail levels: preview itself eventually includes links to other pages, which are shown with a lower detail level. Ajax technologies may accomplish these improvements without increase information load associated to this enrichment of data set associated to web pages: a model based on these technologies is presented, exploiting JSON (JavaScript Object Notation), a recent lightweight computer data interchange format. JSON is used for representing objects and other structured data over a network connection. JSON finds its main application in Ajax web application programming, instead of XML for asynchronously transmitting structured information between client and server. Finally a study case is shown as a testing benchmark for this model.

1. INTRODUCTION

Sharing information has built the Web success: everybody knows related link system that shows pages related to contents of a specific web page, the importance of search engines and other instruments to find and select required information in a large amount of data in World Wide database. Finding data might be complicated and the searching process might last long times.

A practical way to increase the user's awareness surfing the Web is an enhancement of capabilities to anticipate and complete the content of information related to a web page [1]. This foreseeing characteristic can be found in many web in graphical form as pop-up, little windows and canvas containing multimedia objects, presenting attractive and summary issue related to the source link.

For example we can remind of IntelliTXT technology that exploits special green double underlined hot-words, developing an on-mouse-over action to pop-up a window containing hyper medial advertising information. This and similar technologies take part in a innovative standard practice spreading in web pages, that is an enrichment of communication semantic with many new items.

Our work is connected to this trend in Web realizations and it presents a new browsing model, using Ajax technologies and proposing three main targets:

1. To offer to Internet surfers a more detailed vision and a wider and better contextualization;
2. To make required information reachable in lower number of steps;
3. To allow all the above listed benefits without a proportional or overwhelming load in terms of band occupation or information packets traffic.

2. EWF MODEL: EXTENDED WEB PAGE FORMAT DEFINITION

Theoretical and technological evolution trend expresses a strong need to describe information by additional information. This trend involves all data types such as simple raw data and complex data elements. For an

instance it is possible to consider raw objects evolution as pictures, sounds and various multimedia objects: attached to raw objects we can find additional elements representing a deeper description of the data. In other words, information of information, or metadata, is used to facilitate understanding, use and management of data. Few years ago the most part of data sets was just described by its file name: it was the case of audio, graphical, text and many other kinds of files. Conversely, nowadays the above mentioned types of file contain many metadata chunk referring to specific object characteristics: inside MP3 files we can find ID3v2 fields describing author, song name, genre and so on; inside JPEG files is plugged an EXIF record, storing recording date, copyright, comments and various other data; inside PDF files there is a supplementary record describing a summary, the keywords of the document, copyright notice and so forth.

Also more complex data can be provided by metadata [2]. It's possible to discern descriptive metadata, which only list characteristics of the file, from synthetic metadata, which are elaborated by determinated function to get proper targets starting from data included in the file [3].

We explicitly stress that metadata can be conceived not only as a short keyword list of the related file, describing the file itself, but also as a reduced and simplified view of the object [4]:

a) The first metadata type can add new elements to those already present in the file;
b) The second metadata type is a simple redundance respect to originary file, but in a reduced informative view.

This difference is pointed out in many metadata associated to graphic picture formats: it is possible to find description, date, time and author of the picture, providing really new information and therefore this information can be classified as metadata of descriptive type [6]. Conversely, synthetical type of metadata doesn't add any information, but it is a simple "miniature" of the file content. In many graphic and picture files it is possible to meet a reduced icon, showing an image preview: this preview doesn't add any new information to those already included in the main file, but it allows to get a synthetical and light version of it. In analogy to other file formats above listed and described, this work proposes a new Extended Web Page Format (EWPF), whose description will be provided in the following part of this section.

Metadata extension of web page includes two outstanding elements in addition to those usually existing in a standard web page, storable among XML file metadata: two information miniatures, which constitute two levels of summaries of the corresponding page. We want to stress that these miniatures have not graphical connotation, but they represent an information hypermedial abstract.

These miniatures have been conceived not only as a trivial shortned version of the corresponding web page, but also as a significant and possibly captivating preview of page contents. It's clear that miniatures consitute quantitative or qualitative information, descriptive or synthetical, as they can be managed by web authors depending on the peculiar design [5]. Miniatures aren't straight and automatically obtained by the browser, but they are provided and edited in authoring phase, together to other metadata present in meta tags. So author can write new words and include multimedial elements not present in the real web page: he has to take care of the only constraint of contextual coherence between web page and the relative miniatures, needed to save logical wholeness of the page and avoid information confusion. The picture illustrates a scheme of extended web page. In *Hidden data* section, shown in the first part of the page, there are the two miniatures described above, α e β, together to other meta tags: the first ones are hypermedial while the meta tags are tipically textual objects. Below Hidden data, *Visible data* are coded, which contain all objects really displayed in web page. In correspondence to web page extension through α e β miniatures metadata, web page visible data must include some special extended links, pointing just to miniatures. This special extended links inherit standard links functionalities, but enhance tooltip textual functionalities by a more general multimedial content. The appearance of an extended link is similar to a simple link, but the hotword can be, for example, double underlined. However hotword style of extended links can be customized and specified by author in CSS definitions. Many examples of these links can be found surfing the web, showing hypermedial contents.

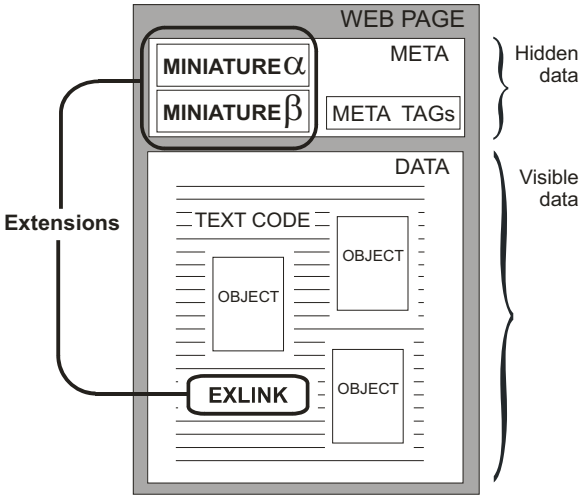


Figure 1: Extended web page format

3. EWF MODEL: HOW DOES IT WORK

Surfing the web, when users put the cursor over a web page extended link hotword, two actions may happen:

1. He may click on the hotword;
2. He may stay over the hotword without clicking it.

In the first case, the linked web page is totally displayed (WEBPAGE_1 in Figure 2). In the second case, a special extended multimedia tooltip will appear, including objects of different format, simple links and extended links too (MINIATURE α in Figure 2).

The user can now choose:

1. He enjoys his tooltip and then come back to the original web page;
2. He clicks simple or extended links inside MINIATURE α because he wants to display the related web page (WEBPAGE_2 in Figure 2);
3. He hovers the mouse over an extended link, displaying a second extended multimedia tooltip, less detailed then the first (MINIATURE β in Figure 2).

Figure 2 illustrates by a tree data structure all the cases listed above:

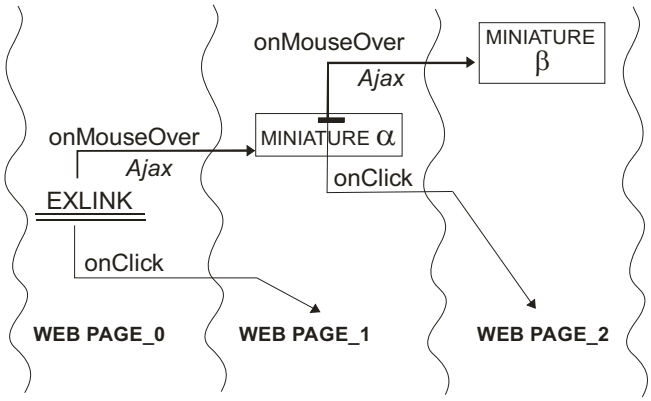


Figure 2: How does extended web page format works

After a Client request, displaying process starts when the web Server restitutes to the Client, through output of the parse.php function, a string object representing the miniature of desired detail level (Zoom = ALPHA or Zoom = BETA in figure 3).

Page request processing is managed by Ajax techniques that involve just data closely necessary to compose the miniature [7,8, 9, 10].

We stress that without Ajax innovative technologies a page request would involve the pre-fetch of a whole page: this operation would require an unnecessary flow of data, a information overhead and a resulting more confused operationIt's now understandable that β miniature is a sort of second level meta information, less accurate and smaller then α miniature first level meta information.

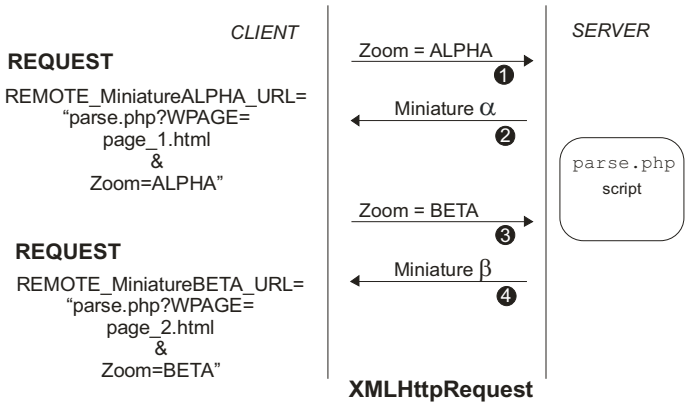


Figure 3: Client/Server `XMLHttpRequest` interaction

In other words a β miniature may be displayed only starting from a hotword included in an α page miniature and not directly from the starting page. α and β miniature objects are accessible by an external request

```
REMOTE_MiniatureALPHA_URL="parse.php?WPAGE=
page_1.html&Zoom=ALPHA";
```

characterized by Zoom parameter that specify the detail level of the displaying page:

```
switch (Zoom) {
   case ALPHA: α level miniature will be shown;
   case BETA: β level miniature will be displayed;
   default: the full page related to the link will be displayed;
}
```

Discrepancies between α and β are not only quantitative, depending on the higher or lower detail level of information in them contained, but also hierarchic, because only from a α miniature it is possible to reach a β miniature and not vice versa.

4. JSON IMPLEMENTATION

JSON (JavaScript Object Notation) is a recent lightweight computer data interchange format, used for the representation objects and other structured data over a network connection [11].

JSON finds its main application in Ajax web application programming, instead of XML for asynchronously transmitting structured information between client and server.In this section are presented the most important code fragments of a Json imlementation of the model described above.

The following code fragment is drawn from a CSS definition of extended link appearance:

```
<style>
.EXlink                <!-- Definition  -->
  { text-decoration: double-underline;
    border-bottom: 1px solid;
    color: green;
  }
</style>

                       <!-- Employment -->
<a href="newpage.html" class="EXlink"
onMouseover="javascript:miniature_A(this)">
MyHotWord </a>
```

Note that we show in bold the main action of this function. The second code fragment describes the operation of miniature_A

function but we outline just definition of the new data container MiniatureBETA_URL. Specifically:

```
       REMOTE_MiniatureBETA_URL =
"parse.php?WPAGE=page_2.html&Zoom=BETA";
```

So, the main difference between miniature_A(page) and miniature_B(page) functions is represented by Zoom level of the remote page requested: the first function achieves meta information about a remote page that is more detailed than the second function.
We explicitly repeat that an EXLINK can output so in full HTML page as in an ALPHA Miniature, with the difference that an EXLINK included in an ALPHA Miniature is managed by BETA and not ALPHA Miniature.
The simple data representation code of the current page

```
<script>
function miniature_A(page)
  {
     // ... Cross-Browser DHTML code to generate
     //     a graphical miniature widget ...

REMOTE_MiniatureALPHA_URL="parse.php?WPAGE=page_1.html&Zoom=ALPHA";

var REMOTE_MiniatureALPHA;
           // XMLHttpRequest Creation
       var http_request = new XMLHttpRequest();
       http_request.open("GET",

REMOTE_MiniatureALPHA_URL, true); // Ajax JSON Data Start
       http_request.onreadystatechange = function () {
               if (http_request.readyState == 4) {
                   if (http_request.status == 200)
                       { // Ajax JSON Data receiving
 REMOTE_MiniatureALPHA = eval("(" + http_request.responseText + ")");
                       }
                   else {
                          alert ("There was a problem with the URL.");
                        }
                                                  }
// Ajax JSON Data Stop
http_request = null;
          };
        http_request.send(null);
  }
</script>
```

So, the object REMOTE_MiniatureALPHA is a container of the received data, Ajax JSON Data Flow; In particular, in our model this object is a simple HTML data string which represent the miniature ALPHA.
If the REMOTE_MiniatureALPHA HTML string contains a hyperlink tag, then this link is managed by a similar function to miniature(page) called miniature_B. We skip the code related to this second

Miniature is made by Javascript Object Notation (JSON) formalism:

```
var MiniatureALPHA = { "description"    :
        "....Html Code....",  };
```

5. STUDY CASE

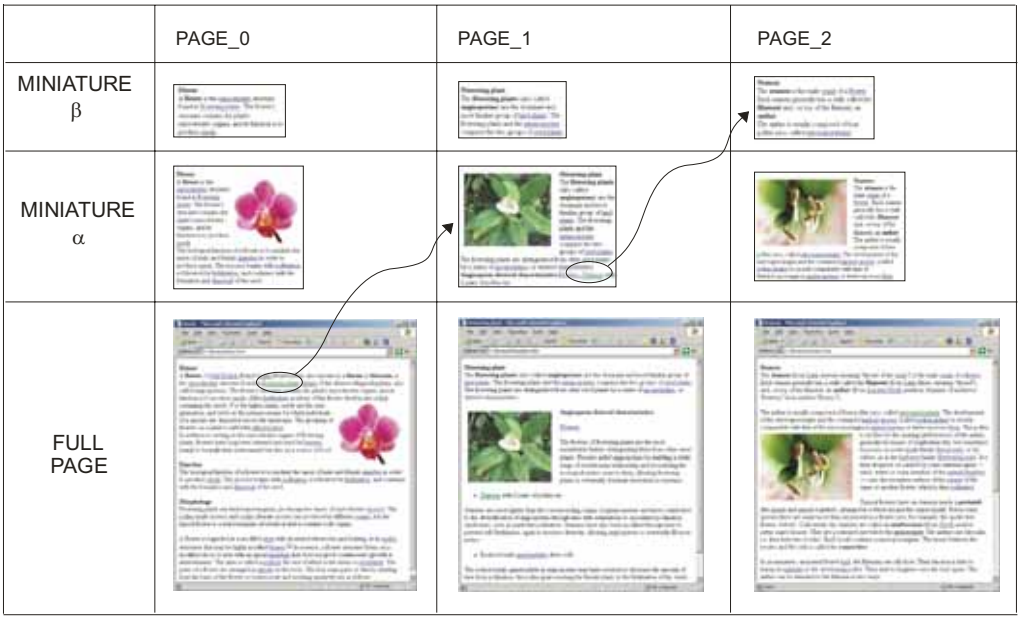


Figure 4: Example: pages with their metadata miniature objects.

Figure 4 shows an application of our browsing model. The pages are displayed together to their own minatures, in the fashion they are displayed when the user require them, acting on extended links on the original page or on a generalized tooltip. In Figure 5 we can observe real image of a double mouse-over action on extended links: the first tooltip shown on the original page is an α-miniature stemming from the exlink visible on the full page; the second shorter tooltip is a β-miniature deriving from a mouseover event on α-miniature tooltip.

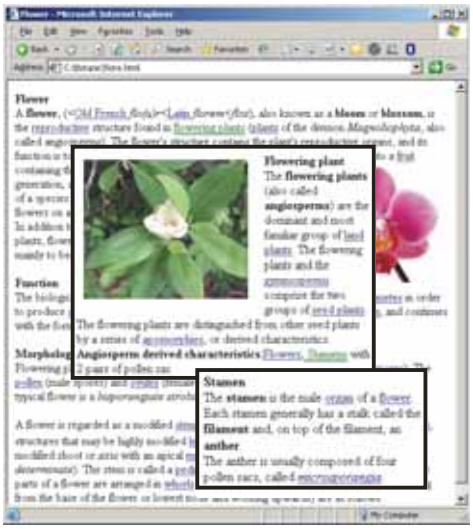


Figure 5: Real visualization of extended tooltip at the maximum level of displaying

We can observe that a β-miniature can't be originated directly from full page: this choice may be justified by a likely need of a more accurate information starting from a full page and a reduced information starting from an α-miniature tooltip, representing a lower level of information detail.

6. CONCLUSION

It is clear, specially considering the example just discussed, the potential benefit that might come from the introduction of this nested levels of information, built in web pages through an enhancement of meta information included in the page code as special metadata.

The drawbacks originating from this new and extended approach to tooltip objects and links can be found in a more toilsome authoring phase, because miniatures aren't automatically inserted in the web pages, but they require an author's accurate and coherent design and realization.

Nevertheless the user can take strong advantage by a global sight of the information included in the web site and by saving time to retrieve the desired information set.

Some issues can derive from the Ajax lack of a cross-browser operation , nevertheless this remains a captivating technique, for its typical light management of information flows. A possible development of this research will bring us to conceive a new way to represent information in a more smoothed fashion, as recommended by W3C, a deeper context awareness and a more comfortable surfing for the users.

REFERENCES

[1] Balzano W., Del Sorbo M.R. Scattered web vs. smoothed web model: a new concept for a hyperlink menu scheme Iadis 2005 *IADIS International Conference. Proceedings of the IADIS International Conference e-Society 2005*. Qawra, Malta, 27-30 Giugno 2005, Ed.Isaias,Karmakar.

[2] Joachims T. Transductive inference for text classification using support vector machines. In *Proccedings of 16th International Conference on Machine Learning (ICML'00)*, pages 200{209, 1999.

[3] De Vuyst, B. Bodard, K. Reaching for the site: meta tag litigation and access to internet-based information *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, Jan. 2003.

[4] Todd Stephens R. Knowledge: The Essence of Meta data *Meta-Tag: The Newest Member of the Meta data Family* http://www.rtodd.com/library/column/0803%20Matatag.pdf

[5] *Metatags: They're not just for search engines anymore*. June 17, 2001, from the World Wide Web: http://www.watchfire.com/resources/metatagswhite.pdf

[6] Knowledge: The Essence of Meta data *Meta-Tag: The Newest Member of the Meta data Family* http://www.rtodd.com/library/column/0803%20Matatag.pdf

[7] Smith K. Simplifying Ajax-Style Web Development Microsoft *Computer Magazine, Volume 39, Issue 5, May 2006, pp. 98-101.*

[8] Dailey Paulson L. Building RichWeb Applications with Ajax IEEE Computer society October 2005

[9] Mahemoff M. Ajax.Design.Patterns OReilly Jun.2006

[10] Eichorn J. Understanding AJAX Using JavaScript to Create Rich Internet Applications Prentice Hall PTR, Aug 2006

[11] Ranschaert B. Java JSON Tools, S.D.I.Consulting BVBA © 2006

http://www.w3.org/Metadata/

http://ltsc.ieee.org/wg12/index.html

A Method for Verifying Usability and Performance of a Multi-user Healthcare Embedded System

Chi-Lu Yang, Richard Weng
Institute for Information Industry,
Southern Information Service Division
Tainan, Taiwan, R.O.C.
E-mail: {stwin, bow4989}@iii.org.tw

Abstract

Developing high quality application systems is always an important task. As the average lifespan among human beings increases, the amount of aged population with chronic diseases also increases. Remote Home Healthcare Services (RHHS) is an efficient solution to improve the living quality of aged people and to resolve the growing cost in hospital-care and medical treatment facilities.

A goal of RHHS solution is to provide high quality systems which would directly be beneficial for those people with chronic illnesses at home. With this equipment, we could be able to take care of and monitor their health status. Therefore, usability and reliability of the equipment are key focuses during the development of quality RHHS systems. One common way to improve system quality is to perform system verification and validation. In this paper, we proposed a verification method that explains the strategies and executions for verifying healthcare systems. In performing the systematic method, we effectively reduced time and effort in re-testing and tracing bugs at the system integration and testing stage. The method could also be a guideline for persons who are in charge of testing similar systems.

1. Introduction

Developing high quality application systems must be an important objective in software engineering. In the past, improving test and maintenance techniques was considered an important stage for improving system quality. Many testing methods were proposed, for example: path testing, boundary value analysis, equivalence partitioning, transaction-flow analysis, etc. Nowadays, people further import the concepts of completeness and consistency checking. At an early stage in system development, it has been proposed to import requirement analysis into the scope of quality system development. Requirement traceability is also performed from the system analysis stage, through the whole system development life cycle, to integration testing, and even until the duration of system maintenance.

We have developed home healthcare equipment. Now through system verification techniques and methods, we would improve its usability and reliability.

In this paper, the first section inquires about some relative system verification techniques. Then, the next sections describe the architectural concepts of a healthcare system (HCS) and its internal characteristics. A system verification strategy with its meanings and focuses is shown in the fourth section while a method is described about test cases generation and execution in the fifth section. The testing results and discussions are recorded in the sixth section. Finally, the conclusions and future prospects are discussed.

2. System Verification Concepts

As information technology advances, the medicine methodology also progresses. Human lifespan is increasing. In 2000, the population of those over 60 years old around the world already exceeded 600 million [1]. By 2025, this number is expected to be more than 1.2 billion. The amount of aged population with chronic diseases is higher. The Healthcare Information and Management System Society [1] (HIMSS) showed that there are 800 million people with chronic diseases in the world. These people are going to raise budgets for medical treatment. In the USA, the budget will go from US$5 to 7 billion in year 2020. If people use information technology to develop high quality systems for home healthcare services, it will effectively reduce the cost of medical treatment.

Patients with chronic diseases need long-term care to reduce pathologies. If they are hospitalized over a long period of time, it is not only a financial and emotional burden on their families but also a waste of resources in hospitals. Therefore, one effective solution is to take care of chronic patients at home. In MIC's research [2], USA and Japan had confirmed a policy of digital home healthcare in medical treatment for the next ten years. It showed an imperative situation, wherein home healthcares are going to rely on innovation and application of information techniques.

When home healthcare equipment are provided with information technology and are combined with system testing techniques, they will be able to service end-users with higher quality. Some testing techniques are proposed in the following sections.

2.1. Test case generation

Crnkovic [3] introduced a component-based developing and testing method. Test cases were generated from unit functions in components. He tested one component by one

unit test case at a time and recorded its results in documents. However, while components were integrated into systems, the results would be treated as black-boxes in the system testing phase.

Wei-Tek [4] suggested three techniques to perform a unit test. They were specification-based test case generation, collaborative testing, and group testing, which could enforce the trustworthiness of the WS components before they were integrated into the new composite WS. In the specification-based test case generation, however, it did not mention possible test scenarios while components were integrated into systems.

Deng and Sheu [5] used UML methodology to analyze a software system. They proposed a Semantic Software Development Model (SSDM) to develop a system as the basis for software testing and maintenance, which were performed at several different levels of abstraction: algorithmic level, class level, use case level, and system level. They also systematically described a set of model-based strategies for test case generation. They generated test cases through the system executive path. However, in a multi-user system, those paths will interactively influence each other. System failures might frequently come from the firmware or hardware. Hence, the method might not guarantee reliability.

2.2. Test strategies

Embedded systems usually require a series of rigorous white-box, black-box, module, and integration testing before developers can release them to the market. Weyuker and Jeng [6] introduced analytic partition test strategies. Wei-Tek suggested a verification pattern (VP) approach to develop test scripts quickly [7]. They generated scripts based on classifying system scenarios into patterns (SP). The eight categories of SP could cover 95% when testing a commercial defibrillator. Those test strategies passed the structural testing. Nonetheless, it would be difficult to debug when developers need to find out which phase the bugs were in. In other words, it could rapidly test, but it may not debug quickly. Rohit Kapur [8] defined that a low-cost test requires satisfying three characteristics: predictable test, low-tech approach to test complex technology, and efficient test. This concept might also be applied successfully in testing a healthcare system.

2.3. Test methods

There are many test methods [9]. Some are suitable for software testing, like path testing, boundary value analysis, and equivalence partitioning. Some are suitable for software integration testing and system testing, like transaction-flow analysis, performance measurement, recovery testing, security evaluation, and user interface verification. Each method has its own suitable time and scope.

Aside from the abovementioned tests, people might also consider simulating testing environments, choosing testing tools, detecting and tracing bugs, etc. In this paper, we would like to focus on test case generation and test strategies, which we discussed in the start of this section.

3. Architecture Design Layout for a Multi-User Healthcare System

The Home Healthcare System (HCS) is designed to satisfy the need for home healthcare services. For RHHS, equipment should at least collect bio-signals, transmit messages, and perform human-machine interaction. It is a multi-user version. Its communication is through Ethernet. An architectural concept of HCS is described in Figure 1.

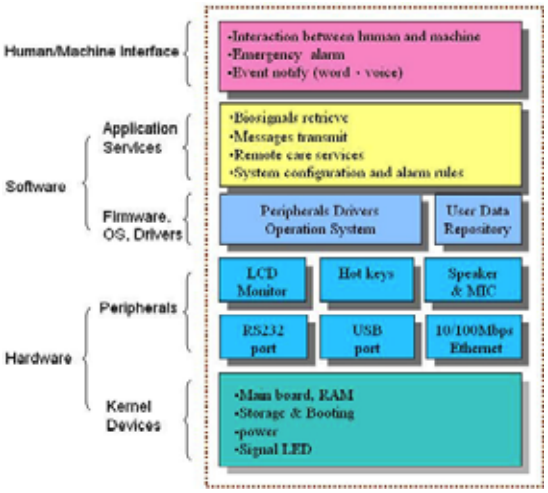


Figure 1. An architectural concept of HCS

While patients with chronic diseases use HCS to collect and measure bio-signals at home, HCS must respond with proper signals to the users. Sometimes, important messages should be sent to a bio-server for analysis. Furthermore, HCS is also a communicable equipment between remote nurses and patients at home. Remote nurses could interact with patients through the equipment. Another important function in HCS is the emergency button. When someone presses the button, a remote healthcare center will take urgent actions by SOP. Some characteristics designed in HCS are described as follows:

(1) *XML-Based multi-user architecture*. Through the XML technique [10], we designed a XML-Based multi-user architecture. Patients could use HCS to measure two bio-signals at the same time. This design could extend users and equipment. It could also reduce waiting time in a queue. On the other hand, nurses could oversee more patients in a limited time. XML-Based multi-user parameters are shown in Figure 2.

```
<?xml version="1.0" encoding="UTF-8"?>
<Configurations>
  <Config Name="BoxUser" Description="">
    <Param Name="User" Value="1#P001#Ken#COM1#OSTAR1#P001.TXT" Description="" />
    <Param Name="User" Value="2#P002#Jeff#null#OSTAR1#P002.TXT" Description="" />
  </Config>
  <Config Name="DeviceList" Description="">
    <Param Name="Device" Value="0#OSTAR#RS232CF.exe" Description="" />
    <Param Name="Device" Value="1#SPO2#SPO2.exe" Description="" />
  </Config>
  <Config Name="PortList" Description="">
    <Param Name="Port" Value="0#COM1" Description="" />
    <Param Name="Port" Value="1#COM2" Description="" />
  </Config>
</Configurations>
```

Figure 2. XML-Based multi-user parameters

(2) *Adaptive rules for unusual bio-signals*. Unusual and dangerous alarms are triggered by a patient's own adaptive rules. These settings could be set while the system is installed or updated remotely. The content of the rules is shown in Figure 3.

```
/* Filename: P001BLOOD.pe            */
/* P0001 Personal unusual rules for */
/* Heart and blood pressure          */
HLB=50    // Heart Lower Bound
HUB=90    // Heart Upper Bound
LPLB=60   // Low Pressure Lower Bound
LPUB=90   // Low Pressure Upper Bound
HPLB=110  // High Pressure Lower Bound
HPUB=140  // High Pressure Upper Bound
/* Every person has his own value   */
```

Figure 3. Adaptive rules for unusual bio-signals

(3) *Remotely update customized services*. If customers would like to add services, a system operator just needs to operate update functions in a server. After performing interactive Q&A, the server will mark a symbol on the client and prepare to update its services. It will not modify any hardware models while the system is executing remote update. Event handlers for remote updating are designed in Figure 4.

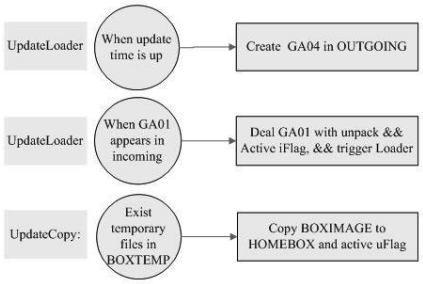


Figure 4. Event handlers for remote update

(4) *Secure messages transmission*. We needed to design highly reliable and secure transmission channels between HCSs and a bio-server. A point-to-point transmission mechanism was defined and named as the Healthcare Message Gateway (HMG).

In HMG, a sender was defined as MsgSender while a receiver was defined as MsgReceiver. Both MsgSender and MsgSender have special pre-package attributes (PPA) and common mapping tables (Figure 5). After the MsgReceiver obtains the messages from MsgSender, with a Secure lock code (Slock-code), it would decode messages using the PPA mapping table.

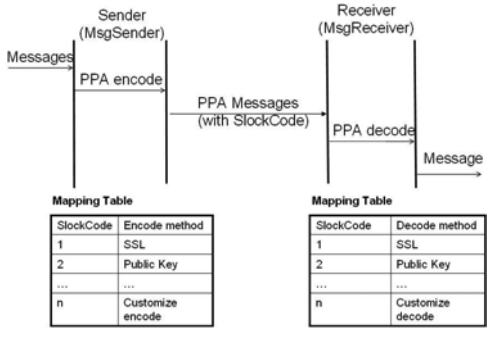


Figure 5. PPA in HMG

A complete message transmission in HCS contains sending and receiving. While messages are dealt with by PPA, a messageAck was first sent out by MsgSender. While msgAck was responded to by MsgReceiver, MsgSender will send PPA messages to MsgReceiver. If MsgReceiver successfully decodes the data, it will respond an Ack to finish this transmission. The flow is shown in Figure 6.

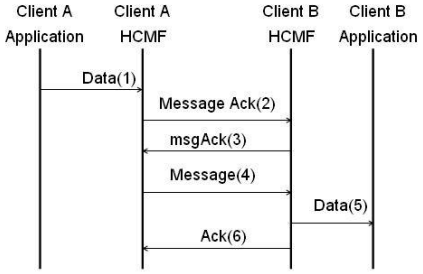


Figure 6. PPA messages transmitted in HMG

While Client A in Figure 6 has sent messages out in Step 2 or 4, but has not received Ack from Client B in Step 3 or 6, this transmission must be treated as a failure. In a failed situation, Client A must trigger a re-send mechanism while Client B should trigger a duplicate check mechanism.

(5) *Implementing system in components*. All functions in a system are component-based. The life cycle of components starts from design to implementation, whereas integration test and maintenance are contained in a system development life cycle and are traced in a requirement traceable mechanism [3].

HCS was developed through a waterfall life cycle model. We hoped to provide a complete, secure, reliable, and user-friendly HCS. Therefore, we tried to develop HCS with conscientious quality control processes. We began to consider a test plan in an early design stage and simulated all kinds of scenarios as much as possible.

4. Test Strategies

We divided system verification into four phases: (1) Unit testing phase: incoming verification and software workbench testing; (2) System testing phase: executing system integration and testing. It especially focuses on installation and un-installation testing, remote update testing, network security testing, and system applying scenarios testing; (3) Reliability testing phase: executing loading capability and stability testing; and (4) On-site validation: validating by end-users. Figure 7 shows their relations. Some details are described as follows.

4.1. Unit testing phase

In order to validate isolating hardware specification, we planned to perform incoming inspection verification. In this phase, focus will be placed on hardware and its surface and functional specs, LEDs standard, com ports, power voltage, memory capacity, etc.

Testers separated this task into visual specifications testing and functional guideline testing in orders. In the visual specifications testing sub-phase, testers would inspect hardware according to the defined front and back specs of HCS. In functional guideline testing, testers would inspect each function by test cases.

If it passes visual specifications testing and functional guideline testing, then it means that it validates isolating hardware specification.

The purpose of performing software workbench testing is to make sure of the correctness of independent software functions. Software testing would focus on whether or not

bio-signal values are properly shown and alarmed. It would also pay much attention to user interface functions and emergency signal handling.

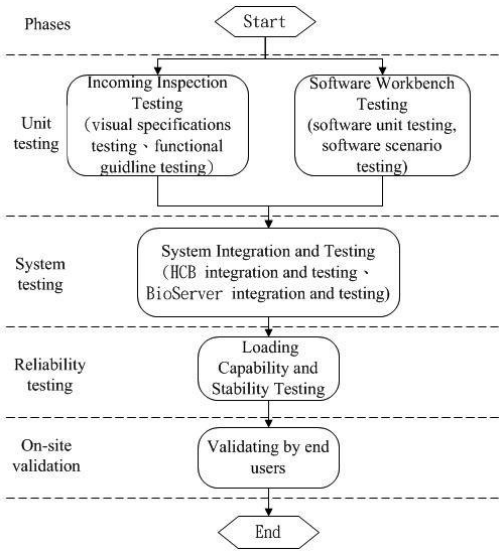


Figure 7. HCS four testing phases

There are two sub-phases in software workbench testing. First, we would focus on software unit test. In the implementing stage, developers program codes according to design documents. They also play the role of testers at the same time and cross test their codes. They are supposed to make sure of the correctness of others' independent unit functions. Second, we would focus on the software scenario test. Developers would test software scenario functions according to integrated scenario test cases. Testers would make sure that integrated functions and components could collaborate formally and as expected.

Finally, a test manager would check consistency between integrated scenario functions and what was written on the design documents. He is responsible for confirming implemented functions based on what was written on specs. Before performing this sub-phase, the test manager should control versions of source codes using a configuration change tool.

If it passes software unit test, software scenario test, and consistency checking, it means that the software in the workbench is verified.

If it passes hardware incoming verification and software workbench testing, it means it has finished the unit test phase and could enter the system testing phase. Before system testing phase, testers should deliver unit test reports which contain results of hardware incoming verification and software workbench testing.

4.2. System testing phase

The purpose of system testing is to ensure that HCB could be executed correctly. Before entering this phase, HCS must have passed the unit test. In this phase, testers treat unit functions as black-boxes. Therefore, they would try out the full HCS.

System testing was divided into four sub-phases. They were named as HCS integration testing, bio-server integration testing, system integration testing, and functional consistency checking. The scope is shown in Figure 8. Some details are described as follows.

(1) *HCS integration testing.* Testers would check HCS by system test cases. In this sub-phase, it would focus on the correctness of software installation and un-installation, remote update, and network security. It would also pay much attention to the soundness of HCS applying scenarios. Furthermore, it would also test the collaboration between HCSs and the bio-server which was simulated

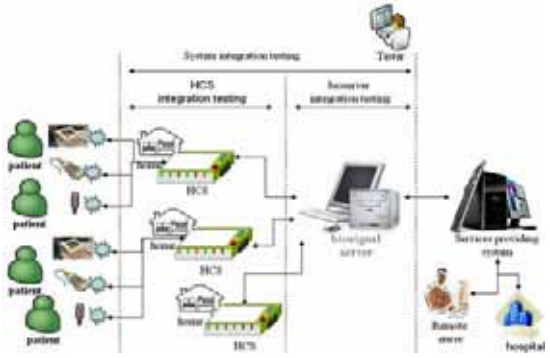


Figure 8. Scope of system testing

(2) *Bio-server integration testing*. Before this sub-phase, the bio-server should finish its own isolating system testing. In this sub-phase, it would test to cooperate with HCS, which was simulated. However, the details of design specs and test methods of bio-server are not included in the scope of this paper.

(3) *System integration testing*. In this sub-phase, testers would check collaborating functions between HCS and the bio-server by system test cases. It would focus on the architecture between HCS and bio-server. For example, interactive problems transmission, and remote system configuration update.

(4) *Functional consistency checking*. The purpose of functional consistency checking is to ensure whether or not the functions implemented are the same as the original customers' needs. The test manager should compare system scenario functions with what was written on specs.

If it passes system integration testing, network secure testing, and functions consistency checking, it means that the system functions are correct and could enter the reliability testing phase. This implies that HCS is ready for public demonstration.

4.3. Reliability testing phase

The purpose of the reliability testing phase is to verify whether or not HCS would work properly in the patient's house 24/7. We should put HCS into a loading and stable testing environment, which is an independent network. We would test HCS functions again using system testing cases in that environment. It is to ensure that HCS could work correctly and effectively in a real network.

In this phase, it would focus on three points. In a large services requested environment, HCS should first be able to work correctly. Second, HCS should still work with estimated performance, even in a jamming network. Third, HCS should recover into a normal status, after loading and stable testing is terminated.

If it passes the reliability testing phase, it means that HCS is a nice quality system. It could be an important healthcare equipment at home.

4.4. On-site validation

The purpose of this phase is to understand real situations while end-users operate HCS. Testers would collect feedbacks learned from end-users. These are used for improving the HCS's functions and expanding its market in the next version.

We would like to provide user manuals to related triers who carry HCS out in practice. We expect a total of 20 persons who are nurses or patients to be triers in a hospital. They will exercise HCS for two to three months and record their feedback on a weekly basis.

5. Test Case generation

There are three designed types of test cases: unit test case, integrated scenario test case, and system test case. The goal of unit test cases is to verify the correctness of unit functions, which were built from design documents. The goal of integrated scenario test cases is to verify requirements or special goals, which were combined from several relative unit functions. The goal of system test cases is to verify full executing system.

Test cases are generated at stages of the system development life cycle. Unit test cases are generated by design documents while integrated scenario test cases are created by requirement spec documents. System test cases are used to validate original customers' needs. One example about test cases generation is shown in Figure 9. Its descriptions are illustrated as follows.

Scenario	Requirement doc.(V2.0)		Design doc./Version(1.0)		Test case Spec.(ID)		
Original needs	Req.ID	Req. Name	Des.ID	Des. Name	Unit	Integrated Scenario	System
S001. Connect a tonometer (1st) to measure bio-signals. Then, show bio-signals on the LCD of box. If the measured values was unusual, then device sent out alarm by voice. Send out this measured record to bio-server.	001	Connect a tonometer (1st) to measure bio-signals (MB). Then, show bio-signals on the LCD of box.	001-01	Acquire MB from tonometer	TU001-1	TI001-1	TS001
			001-02	Acquire current status of user interface(UI). Try to transfer between different UI. Try to put new values and redraw UI.			
	002	Create measured results (MR) by XML, put it in outgoing folder.	002-01	Create MR by XML, put it in outgoing folder, let these ready to transmit.			
	003	If the measured values (MV) was unusual, then device sent out alarm by voice.	003-01	Acquire unusual rules of the patient's bio-signals	TU003-1		
			003-02	Determine if MV has a gab with rules, then sent out alarm by voice.			
	004	Send out these measured records to bio-server.	004-01	Monitor if there is any record waiting transmitting in outgoing folder.	TU004-1	TI004-1	
			004-02	Message gateway(MG) transmits records			
			004-03	MG executes re-transmitting	TU004-2		
				MG receives records and checks duplicates	TU004-3		

Figure 9. Test case generation

Original needs in the scenario are "S001-Connect a tonometer (1st) to measure bio-signals. Then, show bio-signals on the LCD of box. If the measured values were unusual, then device sent out alarm by voice. Send out this measured record to bio-server." S001 is analyzed at the requirement analysis stage and recorded in the requirement document, as follows:

- "001- Connect a tonometer (1st) to measure bio-signals (MB). Then, show bio-signals on the LCD of box."
- "002-Create measured results (MR) by XML, put it in outgoing folder."
- "003-If the measured values (MV) was unusual, then device sent out alarm by voice."
- "004- Send out these measured records to bio-server."

Then, 001 and 002 are designed into three unit functions, as follows: "001-01-Acquire MB from tonometer;" "001-02-Acquire current status of user interface (UI). Try to transfer between different UI. Try to put new values and redraw UI" and "002-01-Create MR by XML, put it in outgoing folder, let them ready to transmit."

Consequently, the scope and content of unit test cases are really generated. The unit test case ID is TU001-1, TU003-1, TU04-1, TU04-2, and TU04-3 in this example. Integration scenario test cases are generated by combining some unit test cases in requirements, as follows: TI001-1 and TI004-1. Finally, TS001, a system test case, is generated to map the original needs-S001.

Every test case includes the following columns: testing name, test case ID, test step (each step contains pre-defined input data/frame, pre-defined output (pass/fail criteria), practice output, etc.), environment requirement (software and hardware), testing results, special needs of testing process, pre-test case ID, tester signature, tester/developer manager signature, remark, etc.

6. Discussion

Testers have executed the test based on test strategies. There were 43 test cases in hardware incoming verification, wherein 22 were visual specifications test cases and 21 were functional guideline test cases. Eight problems were detected in the first testing cycle, where one was about visual problem, and the others were about functions among which there were five in firmware and two in peripherals. Details are described as follows:

1. OS Image did not support the Chinese environment.
2. There was no method about an automatic execution program when the boot-load was triggered.
3. Automatic canceling memory management could not work.
4. Could not set up the boot LOGO by myself when hardware is booted. It needed this function.
5. ADSL (PPPoE) protocol could not work formally.
6. COM port number 2 could not work formally.
7. Audio driver was not correct during memory mapping.
8. There was no proper hole on mechanism for speaker to spread voice.

Developers have revised 62.5% of the problems during second testing cycle. Numbers of the problems were 2, 3, 4, 5, and 8. They accomplished 100% problems during third testing cycle.

There were 31 test cases in software workbench testing, where 20 were software unit test cases and 11 were test cases about software scenario functions. Fifteen problems were detected during the first cycle, where seven were about unit independent problems, and others were about scenario functions, among which there was one heavy error caused in the design stage. The error is described as follows:

> "Remote update and setting mechanism were designed in an error way. HCS must restart twice to complete this operation. It didn't make sense. Designer had to change this pattern."

Developers have revised 87% of the problems (13

problems) during the second testing cycle. There were still two problems, which were "Combining data of address-book and message-gateway didn't accomplish." and "Remote update and setting mechanism were designed in an error way." In the third testing cycle, the developers have also accomplished these two problems. The problems were 100% revised.

In the unit testing phase, testers systematically checked and traced bugs in all sub-phases until bugs were revised. They also recorded those testing data. It was helpful for tracing the causes of bugs in system integration and testing stage. Comparing with a similar system we developed one year ago, we saved about 35% time and effort in testing and tracing bugs.

Testers have simulated a pure environment in the laboratory for system integration testing. Only one problem was detected among seven system test cases. Others had passed the test. The problem was "Remote update and setting mechanism with software installation couldn't work." They analyzed its causes and discovered that "Since they had changed patterns of remote update and setting mechanism, the software installation functions couldn't cooperate with them anymore." After reviewing few codes and parameters, this problem was successfully passed.

In system integration testing phase, another cross-system cooperating problem was detected, "when server was executing, it didn't handle files in incoming folder, which files came from message gateway in HCS." After ensuring that isolating HCS could execute properly, we could make sure this problem came from the cross-system. Its cause was that transmitting configuration settings in the server were different with those in HCS. Therefore, the server could not get the modifications of folders. When developers had revised it in their settings, this problem was cleaned away.

In the past, test strategies usually disregarded the unit testing in workbench. This time, however, we surely tried to cross test and recorded its intermediate details, to ensure the usability and integrability between unit functions of system. It implemented the useful task in software integration and system integration stages. We really saved a large amount of time and efforts in the system integration stage.

Performing an on-site validation phase will require users to carry out HCS in the actual setting and to collect their feedbacks. We need to plan service models and prioritize HCS first. This task belongs to the business services scope. Therefore, we will not discuss such details in this paper.

7. Conclusion

We proposed test strategies which were separated into several testing phases for different verification degrees. According to specification documents, it could generate test cases for all testing phases. Those test cases were used to perform different testing types. After performing testing activities, it reached a nice state of system verification. It could satisfy the system quality. The method does not only effectively reduce time and efforts with regard to re-testing and tracing bugs, but is also a useful guideline for persons who are in charge of testing similar systems.

In the future, we would like to build realistic systems partially and be validated by the end-users in communities and hospitals. It will be a cycle until the HCB product is introduced in the market, even until to end of the HCB life cycle. On the other hand, we will try to improve internal management and maintenance skills of the individual test cases, which are more systematic and traceable.

8. Acknowledgments

This research was supported by the Applied Information Services Development and Integration project of the Institute for Information Industry (III) and sponsored by MOEA, R.O.C.

9. References

[1]Sagitta Pan, "The Status of Digital Medical treatment and factories' tendency", AISP in MIC of Institute for Information Industry in Taiwan, August, 2006

[2]Bryan Lee, "Analyzing New applied System Model under Trend of Diversified Symphony and Transferred Model", AISP in MIC of Institute for Information Industry in Taiwan, Nov, 2006

[3]I. Crnkovic, M. Chaudron, and S. Larsson, "Component-based Development Process and Component Lift cycle", the International Conference on Software Engineering Advances(ICSEA'06), 2006 IEEE.

[4]W.T. Tsai, Y. Chen, and R. Paul, "Specification-based Verification and Validation of Web Services and Service-Oriented Operation Systems", Proceedings of the 10th IEEE International WORDS'05, 2005 IEEE.

[5]D. Deng, P. C.-Y. Sheu, T. Wang, and A.K. Onoma, "Model-based testing and Maintenance", Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE'04), 2004 IEEE.

[6]E.J. Weyuker, and B. Jeng, "Analyzing Partition Testing Strategies", IEEE trans. On software Eng., vol. 17, no.7, p 703-711, July 1991.

[7]W.T. Tsai, L. Yu, F. Zhu, and R. Paul, "Rapid Embedded System Testing Using Verification Patterns", IEEE Software, July-August 2005, p68-75

[8]R. Kapur, R. Chandramouli, and T.W. Williams, "Strategies for Low-Cost Test", IEEE Design & Test of Computers November-December 2001, p47-54

[9]Institute for Information Industry, CMMI Guideline (Level 3) VER-V1.0, MOEA, R.O.C. December, 2002.

[10]The World Wide Web Consortium (W3C), website: http://www.w3c.com

Towards A Formal Semantics for Distributed Multimedia Computing

Angela Guercio and Arvind K. Bansal
Department of Computer Science
Kent State University, Kent, OH 44242
{aguercio, arvind}@cs.kent.edu

Abstract

With the growth of the Internet there is a clear need to develop distributed multimedia languages that integrate asynchronous interaction, event based triggers, reactivity, communication, computation and visualization for transparent modeling, monitoring and human-computer interaction. In our previous research, we described an XML based distributed multimedia language TANDEM that integrates these capabilities. π-calculus models the reactivity, mobility, and computation model of concurrent programming languages that is lacking in current approaches of formal behavioral semantics of distributed multimedia languages that focus mainly on temporal aspect of languages. In this research, we describe a formal semantics of a class of such languages that support distributed multimedia computing by integrating a variation of π-calculus extended to handle multimedia streams and constructs, and semantics of trigger operation. The extended model has been applied to describe the formal semantics of TANDEM.

Keywords: Distributed, Internet, languages, multimedia, π-calculus, reactivity, semantics.

Introduction

As the Internet grows the need for distributed multimedia languages that integrates asynchronous and synchronous events, triggers, reactivity, computations, and multimedia visualizations has become very important for human computer interaction, automated monitoring and visualization. In our previous work, we described a distributed multimedia language TANDEM [8] and its conceptual model [7] that integrates transparent monitoring, visualization, event based triggering, communication and computation. In this paper, we give a formal semantics of these class of languages using an abstract model of TANDEM as an example.

In past semantics of a distributed multimedia system has been expressed such as temporal logic of reactive systems [5, 13], temporal logic of traces (TLT) by [14] — a linear-time variant of propositional temporal logic with quantitative constraints in the style of RTTL [1], and MDTL (Modular Distributed Temporal Logic) [10] that reasons about both synchronously and asynchronously communications modular object systems and extends the object-oriented specification language TROLL [6, 9] with a module concept. The concurrent model of π-calculus [12] models communicating behavior that integrates reactivity, process mobility and computation on concurrent interactive systems [11].

Current semantics representation of distributed multimedia systems need to integrate computation, communication and integration of asynchronous and synchronous behavior, and event based triggers. Similarly π-calculus needs to be augmented with interaction of asynchronous signals and groups of interacting multimedia streams, modeling of tandem events, the notion of triggers, and exception handling needed in event based multimedia computation.

In this paper we extend and integrate both the approaches to describe a formal semantics of distributed multimedia computing languages. We believe that the proposed model is general enough to express formal semantics of distributed multimedia computing languages. The paper is organized as follows: Section 2 describes the background and definitions. Section 3 briefly describes the conceptual model of TANDEM. Section 4 describes the formal semantics, and the last section concludes the paper.

2. Background and Definitions

For the notational convenience we will denote the lifted domain of a set S by $S^{\perp}$ where $\perp$ is the bottom element, Cartesian product of two sets S_1 and S_2 by $S_1 \times S_2$, finite mapping of a set S to a set R by $S \rightarrow R$, the i_{th} element of a sequence of data elements D by the symbol $\sigma(D, i)$, the i_{th} field of an n-tuple $T = <t_1, t_2, \ldots, t_n>$ by the symbol $\pi_i(T)$, and the power set of a set S by the symbol P(S). Two values v_1 and v_2 approximately matching each other with a similarity threshold δ are denoted as $v_1 \approx^{\delta} v_2$.

To model the temporal Boolean conditions, we use tri-state logic – *true*, *false*, and *undefined* (denoted by bottom symbol $\perp$). A current value is undefined if more time has passed since the user defined upper limit after the truth value (true/false) was established. We define $true \wedge \perp = \perp$, $false \wedge \perp = \perp$, $\perp \wedge \perp = \perp$, $true \vee \perp = true$, $false \vee \perp = false$, and $\perp \vee \perp = \perp$. We will denote this lifted domain of Boolean logic as $Boolean^{\perp}$.

A *stream s* is a pair of the form (s^A, s^D) where s^D is a sequence of elements containing the data and s^A is the set of attributes associated with the stream *s*. We use $\sigma(s^D, i)$ to denote the i^{th} frame (data_element) in the stream. A *synchronous signal* is characterized as $\tau - \varepsilon \leq [\sigma(s^D, i-1), \sigma(s^D, i)] \leq \tau + \varepsilon$ where τ is the period of the signal *S*, and $t \gg \varepsilon$, and ε is the *perceptual distance*, i.e. the maximum distance that can be tolerated before perceptual distortion

occurs. While τ is constant for strongly periodic streams, it fluctuates for weakly periodic streams.

The *frame* of a media stream is a set of objects O with spatial relationship between them. Let *P(O)* be the power set of objects, and *R* be the set of spatial relations, a frame is $(f \in P(O),\ P(f) \rightarrow R^{\perp})$ where R is the set of spatial relationships between the objects. Given a set of frames F and the set of natural N, the finite mapping $F \rightarrow N$ produces the time indexed data elements s^D of a stream *s*. Accessing a frame *f* in a stream *s*, is performed by the access operator that is defined as $\pi_1(\sigma(\pi_2(s), i))$ if $0 < i \leq ||s||$ otherwise $\perp$ where π_1 accesses the attribute elements of the stream, and π_2 accesses the data elements of the stream.

The synch points s^{sy} of a stream is a user defined *monotonic unique value array* to preserve the perceptual integrity of information conveyed by multiple streams. The synch point value $\sigma(s^{sy}, i)$ corresponds to the position of the frame corresponding to the i^{th} sync point.

An *asynchronous signal,* denoted by the symbol η, is produced by a reaction and can be received at any time and can have any length. We assume that an asynchronous signal is received without any delay after generation. Given a signal η and under the assumption that the signal occurs more than once $\forall_i [start_i(\eta), start_{i-1}(\eta)] > \theta$ where θ is the time required for the signal to be handled, and is the minimum distance that can be tolerated between successive asynchronous signals.

A distributed multimedia system is modeled as a communicating concurrent state machine in which multiple triggers are concurrently active on different remote sites. We distinguish two types of states: a *computational* state and a *multimedia* state. A *computational* state is a set of triples of the form (*identifier, attribute, value*), and an entity could be a stream, or a variable, or a constant, or a spatial constraint, or a temporal constraint, or mobile processes caused by migration of code over the Internet, or a channel between two computational units. A *multimedia state* M is a set of multimedia entities such as stream, asynchronous signal, partial conditions, attributes of media objects such as stream or asynchronous signals. A transition between multimedia states occurs if media entities in multimedia streams are transformed. A change of multimedia state generates changes in computational states as well. The transformation of a multimedia state involves the passage through many computational states where the multimedia state does not change.

3. An Abstract Model of TANDEM

The notion of perception is guided by the desire to exactly capture and render the real world phenomena. In real life, two concepts are very important: the *continuity* that contains the notion of temporality, and the *context* which expresses the spatiotemporal relationship between objects. Breakage of either of them causes lack of perception and comprehension. In multimedia reactive systems, continuity is guaranteed by the physical presence of the multimedia streams and their temporal relationship to each other that is guaranteed by the presence of multiple clocks and the presence of synch points at regular temporal intervals. The temporal logic of the system and the state behavioral semantics provides the behavioral rules for the language by describing the states and transitions between the states during computation. A distributed reactive multimedia system uses synchronous multimedia streams asynchronous signals (both data and triggers at irregular interval) represented as a pair (attributes, sequence of data elements) where the adjacent data elements are separated by a bounded temporal distance.

A distributed multimedia application is modeled as reaction graphs (see Figure 1) with five types of components: media sources, media sink, trigger units, transformers, and active repository.

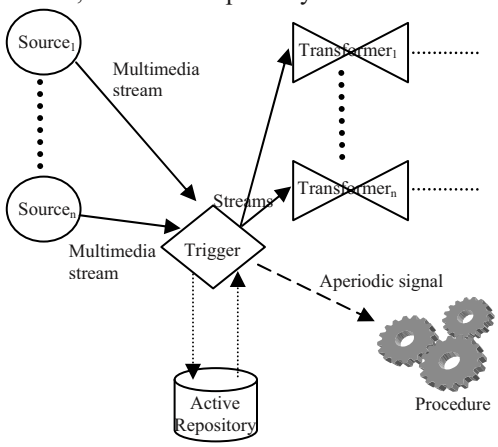


Figure 1. An abstract model of TANDEM

Multimedia streams are generated at media sources, and are visualized or archived at media sinks. Transformers transform the multimedia streams by providing synch points, alternating the periodicity, appending (or splitting) the streams, multiplexing (or demultiplexing) the streams, grouping (or ungrouping) a number of streams, or reducing the number of channels. Trigger nodes control the media streams and initiate reaction to the streams. *Triggers* provide a general mechanism for initiating actions or new events after a set of Boolean conditions are met. A trigger is associated with sets of multimedia stream groups and a set of conditions. A trigger reaction includes monitoring of external sensors, transforming the attributes of multimedia streams, redirecting a stream to a different destination, starting a new thread of computation, or a cascade of triggers and events. *Active repository* nodes are associated with trigger units. Active repository samples and analyzes the media content for the required conditions, and transmits the

outcome to the triggers. An active repository detects and archives *partial conditions* to match the conditions in loose order of occurrence. In order to satisfy the TANDEM events, there are two notions: history and the analysis of the coming streams. History is the time in past whose value is retained in the active repository, and is not forgotten if any of the event needs to retain. If all the events do not need a history then it becomes a forgotten past, and can never be recovered. The abstract operations in active repository are *match*, *analyze*, *compare*, *is_present*, *assert* and *reset.* The assertive actions of the triggers, transformer and the sink alter the multimedia state.

A *partial condition p* is either a simple condition having a tri-state logical value or a disjunction or conjunction of simple conditions. A *simple condition* is identified by matching the current value in media stream or asynchronous signal stored in the active repository as the trigger condition or by retrieving the already archived truth-value in the repository. Let V^S be an ordered set of the values of a media stream in time Δ and let V^R be the set of values archived in the active repository. Then the set of *simple condition* P is defined as a mapping from the Cartesian product of V^S and V^R to lifted domain Boolean domain :$(V^S \times V^R) \rightarrow^{\delta}$ $Boolean^{\perp}$. After receiving new values from the set of incoming media streams if a pair ($v^S \in V^S$, $v^R \in V^R$) is found such that $v^S \approx^{\delta} v^R$ then the simple condition is true. If no value $v^S \in V^S$ matches with any value $v^R \in V^R$ within the period Δ then the corresponding simple condition p ∈ P is set to false. Simple conditions exceeding the Δ limit are forgotten and their value is rest to undefined value denoted as ⊥.

A *transformer* transforms streams in the domain in streams in the range. A procedure is a computational unit that modifies variables-value binding of a computational entity. A *trigger unit* takes a set of synchronous and asynchronous signals and produces an asynchronous signal and sends the data along with the signal. Communication occurs on channels.

A TANDEM system is modeled as a concurrent state machine in which multiple triggers are active at the same time on different remote sites. We distinguish two different types of states: a *computational* state and a *multimedia* state. A *multimedia* state is a subset of a computational state, which involves exclusively media entities. A partial condition can be disjunction or conjunction of simple conditions or a time ordered sequence of partial conditions, or two partial conditions occurring in any order, two partial condition becoming truth in tandem. The satisfaction of the partial condition generates the event, which associates to the event generation a reaction. Given two Boolean conditions p, $q \in Boolean^{\perp}$, the time-limit in Boolean operations is limited by the relationship minimum(Δ^p, Δ^q) ≤ time-limit for reset ≤ maximum(Δ^p, Δ^q) since if a condition becomes true at time t = 0 then it would be forgotten after its time-limit unless refreshed.

4. Handling Reactive Tandem Events

A simple tandem event occurs when one or more time restricted partial conditions become true within a user defined time limit Δ. These simple Tandem events are chained together by a partial ordering sequencing to form a *composite event*.

A Tandem event is modeled as the disjunction of the conjunction of partial conditions. Given a set of partial conditions p_1, p_2, …, p_n, a Tandem event e, such that e(p_1, p_2, …, p_n) is true and Δ_j = [$begin_j$, end_j] be the interval that represents the time constraint associated with the partial condition p_j, the *time constraint* Δ for the Tandem event is defined as the interval max $\Delta_{j\ (1 \le j \le N)}$. The time interval is measured with current time treated as 0; the history, denoted by '$begin_i$' is treated as a negative number and the future time restriction, denoted by 'end_i' is treated as a positive number. Since time is based on clocks, and can be mapped on integers, we will represent time as integers for set based semantics. The time before the current time is *past*. Time before past is *history*. Past and history for the same partial condition are event specific and can only be removed if it doesn't occur in the past of any other event. The conversion from past to history is system specific, but is less than the synch point. The reaction time for any partial condition should be less than the minimum $\Delta_{j\ (1 \le j \le N)}$ otherwise it may effect the truthfulness of the constituting partial conditions. In other words, the reaction must occur in a finite time and this time is smaller than the overall constraint time Δ, and the starting time of the reaction coincides with the time in which the signal generated by the Tandem event starts.

Example 1. Let Δ_I = [*1, 4*] be the time constraint interval on p_i, then suppose that the user presses the button "Save" at the instant *t* = 3. This 'save' signal is received as an asynchronous signal, and is stored in the active repository. If a trigger module checks for the presence of the signal "Save" at instant *t* = 5 within the last 10 units, the active repository will find that the condition is true.

Let P be the set of all possible partial conditions, an *event* e ∈ ***E*** is takes N (N > 0) partial conditions P_1, …, P_N is modeled as the Cartesian product ($P^{\perp}_1 \times \ldots \times P^{\perp}_N$) → $Boolean^{\perp}$. The rationale for the use of the lifted domain $P^{\perp}$ is that partial condition is a set of many simple events, and using tri-state logic if any simple event is undefined due to lapse of the time interval Δ then the corresponding partial condition is undefined and modeled as the null symbol. The set of *Tandem events E* is modeled as a time constrained mapping (of the form ***E***: ($P^{\perp}_1 \times \ldots \times P^{\perp}_N$) $\rightarrow^{\Delta}$ $Boolean^{\perp}$) in which a time constraint Δ is enforced. The satisfaction of partial condition causes an event to be

satisfied, and causes a reaction in response. An asynchronous trigger signal is emitted by the trigger-unit when a set of partial conditions are satisfied that is value is *true* within the time constraint.

Both the trigger and the active repository use the physical clock as the system clock. Let τ be the period of the physical clock then, the period of the clock of a media stream S is $t = k \times \tau$, for $k \geq 1$. In order to discuss the time of the occurrence of a reaction, we distinguish three cases: (1) single synchronous stream, (2) multiple independent synchronous streams, and (3) multiple interdependent synchronous streams.

Given a media stream S with multiple asynchronous signals $\eta_1, \eta_2, .., \eta_m$, the synch points of S are given by $m \times \tau$, where τ is the period of s and $m \geq 1$. Given a set of partial conditions $p_1, p_2, \ldots, p_n$ and a Tandem event $e \in E$, such that $e(p_1, p_2, \ldots, p_n)$ is true, the reaction time of e, *React_time*(e), is given by $(j \times m \times \tau) + \varepsilon$ where j is the smallest value for which $(j \times m \times \tau)$ greater than or equal to the periodicity of the stream. In other words, the reaction time occurs at the first synch point of the stream encountered, plus an additional '$+\varepsilon$' allowed for perceptual distortion. In case there are multiple ungrouped independent streams $s_1, s_2, .., s_n$ and many asynchronous signals $\eta_1, \ldots, \eta_m$ a common synch point must be found for synchronization purpose. The synch point is computed again as the minimum of the synch points of the participating streams. No option is given this time to introduce a different value for the synch point. Finally, let us consider many dependent streams s_1, s_2, and s_n and many asynchronous signals $\eta_1, \eta_2, .., \eta_m$ are input to the system. When many streams are interacting, a group of streams is formed. By forming a group, the synch points for the whole group is computed as the minimum of the synch points of the streams participating in the group. (Eventually, a new synch point can be provided for all the streams of the group). The rationale for retaining the minimum of the synch points is to maintain the perceptual integrity of the constituting streams or group of streams. In the case there are multiple interdependent multimedia streams $s_1, s_2, .., s_n$, and asynchronous signals $\eta_1, \eta_2, .., \eta_m$ and given set of partial conditions $p_1, p_2, \ldots, p_n$ then *React_time*(e) $= j \times g + \varepsilon = j \times k \times \tau + \varepsilon$ where j is the smallest value for which $j(g) \geq d$ and $k \geq 0$.

The trigger operations that control the system are: *start_channel*, *close_channel*, *assert, reset, loop, hard_abort, weak_abort, suspend, resume*. The operation *start_channel (close_channel)* starts (closes) a channel between two valid multimedia agents (see Figure 2). The operation *suspend* temporarily suspends the current thread of activity, and the operation *resume* restores the suspended thread of activity. The loop construct can also be used to repeatedly render (or transmit) one frame of media stream at a time. The actions *hard_abort*, *weak_abort*, *suspend*, and *resume* are similar to the Esterel constructs [4] with a difference that the abort action is taken at media sync points instead of next iteration cycle. Under the assumption that each iterative cycle can render only one frame (data element) in a media stream at a time, sync points are multiples iterative cycle apart, and a media being rendered can not be stopped at the next iterative cycle without causing perceptual distortion.

5. Formal Semantics

Let MS be an element of the powerset of multimedia streams $\{S_1, \ldots, S_n\}$ used in a program. Let V be the set of the persistent entities Ψ that reside into the active repository. Then a transformer is a finite mapping $MS \rightarrow MS$, and a trigger module is a mapping of the form $(P(\Xi) \times P(S)) \rightarrow P(\Xi^{\perp}) \times P(S)$ where Ξ is the set of asynchronous signals and $P(\Xi)$ denotes the power set of asynchronous signals, and P(S) is the power set of the streams. Let *Agent* be the set of all different nodes – multimedia source, multimedia sink, multimedia trigger unit, multimedia repository, and multimedia transformer - in the TANDEM program, then a *channel* is described as a mapping $Agent \rightarrow Agent^{\perp}$ where the domain has an *out* port and the range has an *in* port. Let be x the name of a channel. The range is a lifted domain to take care of rendering. A multimedia agent *MMagent* transfers a multimedia entity ξ^M or receives a multimedia entity on a channel x, or performs a prefix action α (such as rendering) on the entity. Since the multimedia entities are multimedia streams, attributes related to multimedia streams, or asynchronous signals, define the effect of actions on these entities. There are 5 types of multimedia agents, namely, *MMSource, MMSink, MMRepository, MMTrigger*, and *MMTransformer* The extended π-calculus equivalent of multimedia agent based modeling is given in Figure 2.

MMagent::= $x(\xi^M) \mid \bar{x} < \xi^M > \mid \alpha.\ \xi^M$
MMSource::= x < MS>
MMSink::= x(MS)
MM*Repository* ::= x(MS) [*match*.MS | *analyze*.MS |
compare.MS | *present*.Var | *reset*.Var]
MMTrigger:= x (MS) [*iterate*.MS | *abort*.MS |
suspend.MS |*resume*.MS | *assert*.Var |
reset.Var] react$^{\Delta}$.P^M $\bar{x}$ < MS>
MM*Transformer*::= x(MS)*AttributeTranform*.MS $\bar{x}$< MS>
AttributeTransform.MS::= *append*.MS | *split.MS*|
multiplex.MS | *demultiplex.MS*| *downsample.MS*|
stretch.MS | *Trim.MS* | *group.MS*| *ungroup.MS*

Figure 2 : Modeling distributed multimedia systems

5.1 Semantics of repository operations

Match operation matches data elements in a stream with archived objects associated with a simple condition,

given a threshold and returns true if the object matches with a media entity in the stream. Let O = (*type*, *value*) be an object, then a match rule is (O, s, threshold) $\rightarrow^{\Delta}$ Boolean if $\pi_1(O) = \pi_1(\pi_1(\sigma(s^D, i)))$ and *matches*($\pi_2(O)$, $\pi_2(\pi_1(\sigma(s^D, i)))$, *threshold*) under similarity matching. *Analyze* analyzes takes *n* input streams and produces *m* multimedia bindings or variable bindings, or derives attributes of the media objects and archives to match simple conditions at a later time. *Compare* is a predicate that returns true if two values satisfy a user defined relationship, null otherwise. Compare tries to match first the type and then the data values as described: $(\{A\} \times \pi_1(s)) \rightarrow^{\blacktriangleright}$ Boolean$^{\perp}$ such that if $\exists k\ \pi_1(\sigma(\pi_1(s), k)) == \pi_1(A) \wedge \pi_2(A) \approx^{A} \pi_2(\pi_k(\pi_1(s)))\ \perp$. *Present* tests if the corresponding asynchronous signal η is true**.** *Assertion* sets the value of an entity. The entity can be an attribute or the data variable that could be bound to a video, sound, data element, signal values or a channel. *assert(Env, entity_index, newvalue)* $\rightarrow$ *(entity_index, new_value)* if entity(entity_index). If the entity is injected from Var then the value is of any type**;** if the entity is injected from Ξ, the value is a Boolean value**;** if the entity is injected form a set of channels, then the value is a new channel s' and assert binds s to s'**.** The reset of a partial condition occurs when the Δ constraint is not met or when a Tandem event succeeds. For those partial conditions which are not met at the end of the Δ time, the null value is mapped to false. The formal semantics of reset operation is $P \rightarrow^{\Delta} B^{\perp} \rightarrow$ {false} where the P is the set of simple conditions.

5.2 Semantics of the transformer actions

The append operation takes two streams s_1 and s_2, collects their attributes as pairs, and joins the data elements such that data elements of s_1 precede the data elements of s_2. The append operation marks the end of stream s_1 to facilitate the split operation in future.

$s_1 + s_2 = ((\pi_1(s_1), \pi_1(s_2)), s_1^D ; end; s_2^D)$
$\sigma(s^D, i) \in s_1^D$ if $0 < i \leq ||s_1^D||$ and
$\sigma(s^D, i) \in s_2^D$ if $||s_1^D|| < i \leq ||s^D||$

Splitting an appended stream s returns a pair of two streams s_1 and s_2 that are separated by end-marker.

split(*s*) = $((\pi_1(\pi_1(s)), s_1^D), (\pi_2(\pi_1(s), s_2^D))$
$\sigma(s^D, i) \in s_1^D$ if $0 < i \leq ||s_1^D||$ and
$\sigma(s^D, i) \in s_2^D$ if $||s_1^D|| < i \leq ||s^D||$

The *multiplex* operation colors and interleaves attributes and data elements of media streams to reduce number of channels. The interleaving is done between two adjacent sync points $[\sigma(s_j^{sy}, i - 1), \sigma(s_j^{sy}, i)]$. The data elements of the streams s_1 and s_2 when associated with sync points are modeled as $s_1^D \rightarrow s_1^{sy\perp}$ and $s_2^D \rightarrow s_2^{sy\perp}$. We also assume that $||s_1^D|| \leq ||s_2^D||$ and $||s_1^{sy}|| \leq ||s_2^{sy}||$.

multiplex(s_1, s_2) = $(((\pi_1(s_1), \pi_1(s_2)), s^D)$ where

$\sigma(s^D, \sigma(s^{sy}, i - 1), [\sigma(s^{sy}, i)]) = \sigma(s_1^D, [\sigma(s_1^{sy}, i - 1), \sigma(s_1^{sy}, i)]) \oplus \sigma(s_2^D, [\sigma(s_2^{sy}, i - 1),, \sigma(s_2^{sy}, i)])$

During *demultiplexing* the data elements with the same color are collected in the monotonic order of sync points to get the original streams.

demultiplex(s) = $(\pi_1(\pi_1(s)), s_1^D)$ *and* $(\pi_2(\pi_1(s), s_2^D)$
$s_1^D = \pi_1(\sigma(s^D, [\sigma(s^{sy}, i - 1), \sigma(s^{sy}, i]))$
$s_2^D = \pi_2(\sigma(s^D, [\sigma(s^{sy}, i - 1), \sigma(s^{sy}, i)]))$

Stretching synchronizes two or more media streams with different periodicity or fluctuating periodicity by duplicating frames, or by dilating the periodicity of one of the streams *s* by a scalar value. Duplication of frames gives audio distortion (stuttering). Perceptual distortion due to time dilation is less severe. Given a rendering time T and a stream *s*, such that $T > ||s|| \times \tau$ (stream's periodicity), stretching will transform a stream *s* to *s*' with a periodicity τ' such that $|T - ||s'|| \times \tau' | / ||s'|| \leq \varepsilon$ where ε is the perceptual distortion for individual frames.

Downsampling maps a media stream *s* with a periodicity τ to a stream *s*' by discarding some data elements at regular intervals to meet the rendering time constraint $||s'|| \times \tau \leq T$. Let the ratio be r be $||s|| \times \tau / T$ then frame deletion is achieved by shifting the $(i + j)^{th}$ data element to the i^{th} position such that $\sigma(s'^D, i) = \sigma(s^D, i + j)$ where $j = floor(i/r)$

Trimming is used when the length of a stream s given by $||s|| \times \tau$ is strictly grater than the allowed rendering time T. Trimming is achieved by transforming the stream s to a new stream *s*' such that all the data elements of *s* beyond the rendering time T map to a null value.

Grouping synchronizes two media streams s_1 (periodicity τ_1) and s_2 (periodicity τ_2) to get a composite perception. Under the assumption that equal number of data elements are rendered between any two sync points in the same stream, a pair of sets of sync points (s_1^{sy}, s_2^{sy}) are created such that $-\varepsilon \leq (\sigma(s_1^{sy}, i) - \sigma(s_1^{sy}, i - 1)) \times \tau_1 - (\sigma(s_2^{sy}, i) - \sigma(s_2^{sy}, i - 1)) \times \tau_2 \leq +\varepsilon$. Since a media stream may be involved in more than one groups with different sync points, each group has its own set of sync-points associated with streams within the group. If the two streams s_1 and s_2 already have two sync points s_1^{sy} and s_2^{sy} associated with them, then the set of sync points with minimum distance. *Ungrouping* is the reverse of grouping. The n-tuple of sync point sets associated with that specific group are deleted, and media streams in the group are not restricted to work in lock-step.

5.3 Semantics of trigger operations

We use the state logical behavior to describe the semantics of those constructs. Let α be the action taken that when applied transforms the multimedia state μ in μ', then a state transformation caused by an action α given the set of entities Ψ is written as

$$\mu \xrightarrow[\alpha]{\Psi} \mu'$$

We define a *streaming code* $k \geq 1$. The streaming code number encodes the reaction to an asynchronous signal, such as abort or suspend, performed on the streams samples between two synch points of a stream.

When k=1 the action is strong abortion, k>1 the action is weak abortion. We will denote the state after applying the actions in a single iterative cycle as μ^I. Under the assumption that the smallest data unit is a frame, the sync point for a media stream corresponds to m ($m \geq 1$) data units. Then the state transition for one traversing one sync point is $(\alpha^I)^m$. An asynchronous signal η that initiates a preemptive action has to wait 1 synch point to reach the new state μ'. However, if the abortion is strong, the streaming is interrupted at the first synch point (k = 1) of the stream and the control moves out of the loop. If the abortion is weak the streaming is completed after the current clip is over (k ≥ 1).

During suspension the current state is saved. However, the multimedia state is defined as disjoint union of the frozen state and the new state derived from the alternate thread of activity so that frozen state can be restored after *'resume*' action.

Table 1 describes the semantic rules for abortion, weak suspend, and resume. For week abortion k > 1 and for strong abortion k = 1.

abort	$\mu \xrightarrow[(\alpha^I)^m,\ k \geq 1]{\psi} (\mu^I)^{m*k}$ $\mu \wedge \neg\eta \xrightarrow[\text{Abort; } (\alpha^I)^m,\ k \geq 1]{\psi} (\mu^I)^{m*k} \wedge \eta$
suspend	$\mu \xrightarrow[\alpha]{\psi} \mu'$ $\mu \xrightarrow[\text{suspend; } \alpha]{\psi} \mu \oplus \mu'$
resume	$\mu^{susp} \oplus \mu' \xrightarrow[\text{resume}]{\psi} \mu^{susp}$

Table 1. Semantics of trigger operations

Conclusion

In this paper, we describe formal semantics of abstract constructs used in distributed multimedia languages through an example of TANDEM developed by us. The model extends π-calculus to handle multimedia entities, media streams, triggers, and abort and suspension constructs. We believe that future distributed multimedia language will be benefited by this model.

References

[1] R. Alur, T.A. Henzinger, "Logics and Models of Real-time: a Survey", in *Real time: Theory and Practice*, *LNCS*, vol. 600, Springer-Verlag, pp. 74-106, 1992.

[2] T. Arndt, K. Chang, A. Guercio, "Formal Specification and Prototyping of Multimedia Applications", *International Journal of Software Engineering and Knowledge Engineering*, vol. 10, no.4, pp. 377-409, 2000.

[3] A.F. Ates, M.Bilgic, B.Sarikaya, "Using Timed CSP for Specification, Verification and Simulation of Multimedia Synchronization," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 1, pp.126-137, 1996.

[4] G. Berry, G. Gonthier, "The ESTEREL Synchronous Programming Language: Design, Semantics, Implementation," *Science of Computer Programming*, vol. 19, no. 2, pp. 87-152, Nov. 1992.

[5] S.BjØrner et al, "Verifying Temporal Properties of Reactive Systems. A SteP Tutorial," in *Formal Methods in System Design*, Kluwer Academic Publishers, Boston, pp.1-45, 1999.

[6] A. Grau, J. K. Filipe, M. Kowsari, et. el.,, "The TROLL Approach to Conceptual Modelling: Syntax, Semantics and Tools", *Proc. of the 17th. Int. Conference on Conceptual Modeling*, vol. 1507 of *LNCS,* Springer Verlag pp. 277- 290, 1998.

[7] A. Guercio and A. K. Bansal, "A Model for Integrating Deterministic and Asynchronous Events in Reactive Multimedia Internet Based Languages," *Proc. of the 5th International Conference on Internet Computing*, Las Vegas, USA, pp. 46-52, 2004.

[8] A. Guercio and A. K. Bansal, "TANDEM – Transmitting Asynchronous Nondeterministic and Deterministic Events in Multimedia Systems over the Internet", *Proc. of the 10th Intl. Conf. on Distributed Multimedia Systems,* San Francisco, USA, 2004, pp. 57-62, 2004.

[9] P. Hartel, G. Denker, M. Kowasri et. el., "Information Systems Modeling with TROLL Formal Methods at Work, *Information Systems,* vol. 22, no. 2-3, pp. 79-99, 1997.

[10] J. Küster Filipe, "Fundamentals of a Module Logic for Distributed Systems," *Journal of Functional and Logic Programming*, *Special Issue on Logical Formalism for Program Composition*, vol. 3*,* pp 1-30, 2000.

[11] R. Milner, "A Calculus of Communicating Systems," *LNCS*, vol. 92, Springer-Verlag, 1980.

[12] R. Milner, "*Communicating and Mobile Systems; the π-calculus,"* Cambridge University Press, 1999.

[13] A. Pnueli, "Specification and Development of Reactive Systems", *Information Processing*, pp. 845-858, 1986.

[14] E. Vicario, "Engineering the Usability of a Visual Formalism for Real-Time Temporal Logic", *Journal of Visual Languages and Computing*, vol. 12, pp. 573-599, 2001.

Data Normalization and Fusion in Multibiometric Systems

Andrea F. Abate, Michele Nappi, Daniel Riccio
Dipartimento di Matematica e Informatica
Università di Salerno
{abate, mnappi, driccio}@unisa.it

Maria De Marsico
Dipartimento di Informatica
Università degli Studi di Roma "La Sapienza"
demarsico@di.uniroma1.it

Abstract— **Nowadays, the present trend in biometrics is towards multimodal systems. In particular, this work analyzes the combination of the two different biometries face and ear. The twofold contribution is given by the introduction of a new normalization function, namely the *mapping function*, able to overcome the limitations of commonly used functions, and by the definition of the *SRR* (System Response Reliability) index, whose purpose is to estimate the reliability degree of each system response.**

Experimental results in the final part of our work provide a positive feedback about theoretical statements within the body of the paper.

I. Introduction

Biometrics allows recognizing an individual based on physical or behavioural features and its potential is extremely high [2]. Most present systems rely on a single biometry; this makes them vulnerable to possible attacks, as well as little robust with respect to a number of issues. An example is the possible non universality of the chosen biometric feature, as in the case of subjects lacking in a limb for fingerprints recognition or of deaf-mute subjects for voice recognition. A multimodal system provides an effective solution, as flaws of an individual system can be compensated by the availability of a higher number of alternative biometries [6]. Nevertheless, using more biometries in a single recognition system requires to consider some critical aspects. First, each subsystem labels each subject with a numeric value (score) in the range [0, ∞]. However, such values generally come from measurements performed on different features and by using different scales and procedures: a direct combination of such values would give an incorrect result. According to this consideration it is clear the need to normalize the *scores* assigned by each biometry to individual subjects, before combining them using any fusion rule. The normalization module is a very important component in the context of multi-biometric systems [3]. On the other hand, due to the possibly different quality of data inputted from time to time to each subsystem, and to the possibly different accuracy of exploited recognition procedures, it could happen that not all responses are equally reliable. The definition of a measure for the response reliability of the single subsystems would be significant for fusing the single results in an overall final response; such aspect is too often neglected, though especially crucial. The main reason can be found in the considerable difficulty implicit in binding result reliability to input data quality. This would require an absolute metric to estimate the difficulty encountered by the system in identifying a subject, given an acquisition of his/her biometric features. Such metric could not be independent of the adopted classification. Moreover it should consider the relation between data used for enrolment and data used for testing. Finally, it is not possible to directly use an existing performance measure such as the system Recognition Rate, because it provides an estimate of the global ability of the system to recognize a subject, and not an evaluation for the reliability of a single testing operation. Even less it is possible to rely entirely on the single value of the distance between the probe biometric key and the biometric key of the first retrieved subject, as this information alone does not take into account the relation between the latter and all the remaining subjects in the gallery. In other words, similarity among gallery subjects should influence the results evaluation.

In this work we propose a solution to the above two problems by defining both a new *mapping function* normalizing scores from different subsystems, and a *system dependent metric*, namely *System Response Reliability (SRR),* based on the ability by a recognition procedure of separating genuine subjects from impostors in the most sharp and unequivocal way possible.

II. The Normalization Function

A number of different solutions have been proposed in literature to solve this problem. However, each of them presents some limitation. For example, the Min-max normalization technique performs a "shifting" of the minimum and maximum values in the interval between 0 and 1. Given a set of matching scores $\{S_k\}$, k=1,2,...,n , the corresponding normalized values are given by the (1) in Table 1. Such technique assumes that the minimum and maximum ever generated by a matching module are known.
The Z-score technique is the most widespread and uses arithmetic average and standard deviation of scores returned by the single subsystem. Normalized values are given by (2) in Table 1, where μ represents the arithmetic average of scores and σ is the standard deviation. The problem with Z-score is that it does not guarantee a common interval for

normalized values coming from different subsystems.

The Median/MAD technique uses the median and the MAD (median of absolute values); it is quite robust, and the normalized value from the matching module is given by (3) in Table 1, where MAD = median(|s_k – median|). As compared with Z-score normalization, Median/MAD is less effective, most of all when values have a non-Gaussian distribution; in such cases both median and MAD are non-significant estimations and the normalization technique neither preserves the original value distribution nor transforms the values in a common numeric interval.

In this section we propose a normalization function derived from the family of sigmoidal functions. A *sigmoidal* function is defined by the (4) in Table 1. Such function has the open interval (0,1) as codomain, so that it represents a possible mapping function from score values from different biometric subsystems into a single interval; as noticed before, such function would be applied before passing the score values to a fusion module. However we have two drawbacks: a) the distortion introduced by the function when *x* tends to the extremes of the interval is excessive; b) the shape of the function depends on the two parameters *c* and *k* that in turn strongly depend on the domain of *x* parameter. It is possible to reduce the distortion by deriving a new function *F(x)* from *f(x)*, with a pseudo-linear behaviour in the whole codomain though preserving the property such that $F(x) \in [0,1), \forall x$.

Table 1 Some of the normalization functions from the literature.

NORMALIZATION FUNCTIONS		
MIN/MAX	$s'_k = \frac{s_k - \min}{\max - \min}$	(1)
Z-SCORE	$s'_k = \frac{s_k - \mu}{\sigma}$	(2)
MEDIAN/MAD	$s'_k = \frac{s_k - median}{MAD}$	(3)
SIGMOIDAL	$s'_k = \frac{1}{1 + ce^{-ks_k}}$	(4)

Given *f(x)* we can find that it has 0 as horizontal asymptote when $x \to -\infty$ and 1 when $x \to \infty$. As noticed above, though *f(x)* shows a pseudo-linear behaviour in its central part, it introduces a non-linear distortion at the extremes, as can be seen in Fig. 1 (blue line). The first step is to compute the extremes of the exploitable region, i.e. the region where the distortion degree introduced by *f(x)* is still sufficiently small. Examining the third derivative of *f(x),* and in particular the points where it becomes null, it is possible to find out such extremes, which can be used for mapping the scores from the single biometries into the same interval [0,1). The third derivative of *f(x)* is given by:

$$f^3(x) = 6\frac{c^3k^3e^{(-kx)^3}}{(1+ce^{-kx})^4} - 6\frac{c^2k^3e^{(-kx)^2}}{(1+ce^{-kx})^3} + \frac{ck^3e^{(-kx)}}{(1+ce^{-kx})^2}. \quad (5)$$

It becomes null in two points:

$$xmin = -\frac{1}{k}\log\left(\frac{2+\sqrt{3}}{c}\right) \text{ and } xmax = -\frac{1}{k}\log\left(\frac{2-\sqrt{3}}{c}\right). \quad (6)$$

Let us assume that, in the general case, the maximum distance value returned by the biometric system is $xmax$, while the minimum value is $xmin$. Let us set up a system of equations using the two solutions and let us solve it with respect to the two variables *c* and *k*; the solutions obtained are:

$$c = \frac{2+\sqrt{3}}{e^{-\frac{xmin\cdot\log(7-4\sqrt{3})}{xmin-xmax}}} \quad (7)$$

and

$$k = \frac{\left[(xmax - xmin)\log\left(\frac{2-\sqrt{3}}{2+\sqrt{3}}\right) + xmin\cdot\log(7-4\sqrt{3})\right]}{xmax(xmin - xmax)} \quad (8)$$

In the specific case at hand the values in the domain of $f(x)$ represent the distances between feature vectors belonging to the subjects compared by the different biometries; they then fall in the interval [0, ∞[, so that we can fix $xmin = 0$, so simplifying the obtained solutions:

$$c = 2 + \sqrt{3} \quad (9)$$

and

$$k = -\frac{1}{xmax}\log\left(\frac{2-\sqrt{3}}{2+\sqrt{3}}\right) \quad (10)$$

We are left with a last problem. The function we are looking for must have the interval [0, 1) as codomain. Let us then define a new function $g(x) = f(x) - f(0)$ and compute its limit when $x \to \infty$:

$$L = \lim_{x\to\infty} g(x) = \frac{2+\sqrt{3}}{3+\sqrt{3}} \quad (11)$$

Our mapping function is then given by:

$$F(x) = \frac{1}{L}g(x) = \frac{1 - b^{\frac{x}{xmax}}}{ab^{\frac{x}{xmax}} + 1} \quad (12)$$

with $a = (2+\sqrt{3})$ and $b = (7-4\sqrt{3})$.

The function *F(x)* assures a pseudo-linear mapping for all values of *x* included in the interval [0, *xmax*]; if we admit

some distortion, it also yields to normalized values for x greater than $xmax$, still guaranteeing the constraint $F(x)<1$. This is essential in those biometric systems where the value of $xmax$ is not known in advance and we use an estimate of it. The graph of such function when $xmax$=400 is shown in Fig. 1 (black line).

When the value of $xmax$ is fixed in advance and $x \le xmax$ $\forall x$ it is possible to obtain a better result using the function

$$\overline{F}(x) = \frac{ab+1}{1-b} F(x), \qquad (13)$$

as confirmed by the corresponding graph in Fig. 1 (red line).

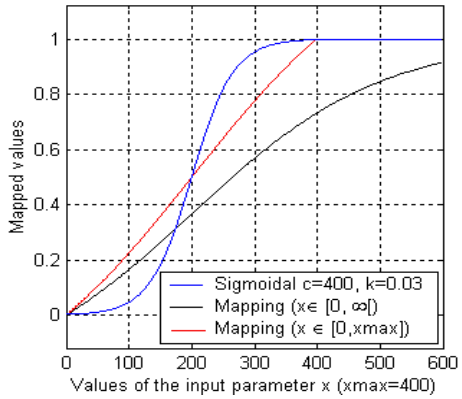


Fig. 1 Graphs of the sigmoidal function, of the mapping function when $0 \le x < \infty$, of the mapping function when $0 \le x \le xmax$.

III. System Response Reliability

Let A be an identification system and G its gallery of genuine subjects who were correctly enrolled. Assume there are at least n>0 acquisitions for each subject. Moreover, let p be a person to be identified. First of all, the system computes distances between p and all the subjects in the gallery: $d(p, g_i)$ where $i = 1, \ldots, |G|$. Such distances are then ordered increasingly, so that $d(p, g_{i_1}) \le d(p, g_{i_2}) \le \ldots \le d(p, g_{i_{|G|}})$.

We analysed two different measures: the relative distance between the scores of the two first retrieved distinct identities, namely SRR I, and the number of subjects near to the retrieved identity which are present in the gallery, namely SRR II.

We define the relative distance as:

$$\varphi(p) = \frac{\left(F\left(d\left(p, g_{i_2}\right)\right) - F\left(d\left(p, g_{i_1}\right)\right)\right)}{F\left(d\left(p, g_{i_{|G|}}\right)\right)}, \qquad (14)$$

where F is a suited data normalization function and g_{i_2} is the second distinct identity in the ordered list of distances. As experimentally observed, the relative difference tends to be small for impostor subjects and high for genuine ones, independently from the biometry and from the classification method. The second measure is computed in the following way. Given the normalized distance $F\left(d\left(p, g_{i_1}\right)\right)$ between the probe and the first identity retrieved from the gallery, the function $\varphi(p)$ is computed using the ratio between the number of subjects in the gallery having a distance from the probe lower than twice $F\left(d\left(p, g_{i_1}\right)\right)$ and the cardinality |G| of the gallery:

$$\varphi(p) = 1 - |N_b| / |G|, \text{ where} \qquad (15)$$

$$N_b = \left\{ g_{i_k} \in G \mid F\left(d\left(p, g_{i_k}\right)\right) < 2 \cdot F\left(d\left(p, g_{i_1}\right)\right) \right\}.$$

In both cases, the role of the data normalization function F is to normalize distances in the interval $[0,1[$ such that the function $\varphi(p)$ is independent from the specific function d used for the distance. We need to establish a value φ_k for the reliability index separating genuine subjects from impostor ones. Such value represents the point of maximum uncertainty. It varies depending on the biometric feature and on the classifier, so that it must be estimated from time to time for the single subsystem. The optimal φ_k is given by that value able to minimize the wrong estimates of function $\varphi(p)$, i.e. impostors with $\varphi(p)$ higher than φ_k or genuine subjects with $\varphi(p)$ lower than φ_k. SRR index can then be defined as:

$$SRR = \left| \varphi(p) - \varphi_k \right|. \qquad (16)$$

Notice that SRR gets high values both for $\varphi(p)$ much higher than φ_k (genuine subjects) and $\varphi(p)$ much lower than φ_k (impostors).

A. How to integrate SRR index into the fusion protocol

A multimodal system is composed of a number of subsystems each working with a specific biometry; as a consequence single subsystems will simultaneously produce a response and a reliability measure for it. The fusion module has then the additional task of carrying out an integration policy able to weight the single responses based on the corresponding reliability degree.

Different choices exist to perform integration, all equally sound in theory, yet possibly taking to very different results in practice. We find it worth then to analyse some of these hypotheses, considering their specific features and postponing to Section 4 their experimental evaluation. Let us assume to have a system S composed by N subsystems $T_1, \ldots, T_N$, each able to produce a list $T_i(1,\ldots,|G|)$ of $|G|$ subjects (where G is the gallery) ordered according to the distance $d_i(p,g_j)$ with $j=1, \ldots, |G|$ from the probe image p, and a numeric value srr_i representing an estimate of the reliability degree of its own response. In order to perform a

consistent fusion of data from the single subsystems it is necessary to guarantee the constraint $\sum_i srr_i = 1$ by normalizing the individual indexes using the formula:

$$w_i = \frac{srr_i}{\sum_{j=1}^{N} srr_j}, \ \forall\, i\,. \tag{17}$$

Moreover a consistent threshold th_i is estimated for each subsystem T_i above which we can consider its reliability satisfactory enough. The main integration policies that have been used are OR and AND. For the former, the system considers the combined response as valid only if at least one of its subsystems guarantees its own response with a reliability degree above the corresponding threshold, while for the latter the system considers the combined response as valid only if all of its subsystems guarantee their own response with a reliability degree above the corresponding threshold.

The problem that remains to be solved is to determine the optimal thresholds th_i for the single subsystems. In order to be able to automatically estimate a significant threshold for a given subsystem it is possible to rely on a number of observations. Let us assume that T_i subsystem has executed M times producing as many responses $\{T_i^1(1),\ldots,T_i^M(1)\}$ with their reliability measures $\{srr_i^1,\ldots,srr_i^M\}$ attached; the latter can be considered as the components of a vector $\overline{S}_i = \{srr_i^1,\ldots,srr_i^M\}$ containing the *history* of what happened till a certain moment. We observe that the value to assign to th_i threshold is strictly correlated to the features of vector $\overline{S}_i$, in particular to its average and to its variance. As a matter of fact, if $\mathrm{E}[\overline{S}_i]$ is an high value we can expect that system responses are reliable on the average and that the corresponding threshold is then proportionally high. However, an important role is also played from the variance $\sigma[\overline{S}_i]$ of $\overline{S}_i$ vector, that measures the stability of T_i subsystem; in particular, if the variance has a low value the system can be considered very stable in always giving reliable (unreliable) responses, while an high variance highlights an anomalous behaviour of T_i subsystem. The attractive characteristic searched for a given T_i subsystem is then that its vector $\overline{S}_i$ has a high value for the average and a low value for the variance; in such case it is possible to fix a high value for threshold th_i . We can formally summarize the above observations in the following formula:

$$th_i = \left| \frac{\mathrm{E}[\overline{S}_i]^2 - \sigma[\overline{S}_i]}{\mathrm{E}[\overline{S}_i]} \right|, \tag{18}$$

As a matter of fact such formula represents an indicative criterion to choose the thresholds for the single T_i subsystems, but also a rule for their dynamic update. In the next section we will experimentally show its validity.

IV. Experimental Results

The most used databases were FERET [5] and AR-Faces [4] for the face, and Notre-Dame [7] for the ear. Both for FERET and AR-Faces a subset of the whole database was selected; this is mainly due to the need of building a multimodal database where each subject had corresponding images for face and ear. For this reason FERET subset contains 100 subjects from each of the FAFB, FAFC and DUP I, probe/gallery, i.e. the first 100 subjects (with respect to the original label). In each test only onesubset was used. Such value was imposed because Notre-Dame contains more or less 100 subjects (114 to be precise). The choice of the first 100 labels, i.e. of a well identifiable subset, aims at facilitating the comparison of this study with possible future ones. The subsystems Face and Ear work in parallel and independently [1], while system performances are analysed based on Recognition Rate and Equal Error Rate (EER).

We start from describing experimental results related to the *mapping function*. The main goal of a normalization function is to map values from its domain to its co-domain [0,1], preserving the original input data distribution as much as possible. In order to evaluate the contribution provided by the new proposed normalization function, we studied the distortion degree introduced by the functions in Section 2, by comparing their performances when input data features are known in advance. In more detail, we chose the two following test functions: $f_1(x) = 2 \cdot (\cos(x) + 1)$ and $f_2(x) = 2 \cdot \log(x+1) \cdot (\cos(x) + 1)$ in [0, 2π] interval and applied the given normalization functions to both of them. In Fig. 2 a graphical representation is shown of how the different normalization functions discussed in Section 2 act upon the values of our two test functions, while in Table 2 numerical performances of the whole systems are reported. When minimum and maximum values are known, the normalization process is trivial; for this reason, we assumed to miss an exact estimate of the maximum value for the two functions $f_1(x)$ and $f_2(x)$ and we chose the average value in its place, in order to stress normalization functions even more. From Fig. 2 it comes out that *Min-Max*, *Z-scores* e *Median/MAD* do not comply with the constraint of having normalized valued fall within the [0,1] interval, moreover the last two also produce negative values besides introducing a sharp distortion in the original distribution. The *Sigmoidal* function guarantees a limited co-domain]0,1[. However it introduces an excessive distortion that is particularly evident for input values near to zero. Finally, our Mapping Function provides performances which are better when compared to all the preceding ones. It assures to have normalized minimum and maximum values within the desired range and also preserves input data distribution in a satisfying way, though relying on a maximum estimate which are far from the real value.

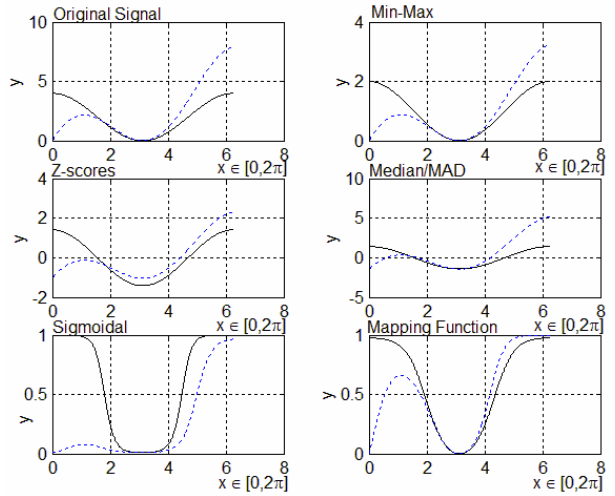


Fig. 2 The distortion introduced by the different normalization functions.

Table 2 Comparison between the performance of biometric systems for different normalization functions.

SYSTEM		PERFORMANCES				
		MIN MAX	Z SCORES	MEDIAN MAD	SIGM.	**MAPPING FUNCTION**
FACE	RR	93%	93%	93%	93%	**93%**
	EER	0.03	0.23	0.12	0.04	**0.03**
EAR	RR	72%	72%	72%	72%	**72%**
	EER	0.14	0.25	0.17	0.16	**0.14**
Face ⊕ *Ear*	RR	95%	93%	93%	94%	**98%**
	EER	0.018	0.23	0.11	0.02	**0.015**
	NRR	100	100	100	97	**90**

For the multibiometric architecture Face⊕Ear in Table 2, the subsystems have been combined using OR policy and SRR II reliability index. Results in Table 2 underline that the Face/Ear Recognition Rate remains unchanged across different methods, while the Equal Error Rate undergoes to significant variations, this is likely due to an increment of the FAR. The worst performance is given by the Z-scores approach that provides an EER above 0.22 for both single subsystems and the multi-biometric architecture. On the contrary, the proposed mapping function equals the optimal performances of the Min/Max approach, yet providing a high robustness with respect to the estimation of the maximum score. Indeed, while the performance of the Min/Max method degrades with inaccurate estimations of the maximum score, the mapping function still provides similar RR/EER values for large variations of the *xmax* parameter (see Section II). An example of how the effectiveness of SRR II index for face varies together with the value chosen for φ_k is given in the first graphic in Fig. 3; an asterisk underlines the point where the index precision is maximum. The second graphic in Fig. 3 shows which probes among the full set of 100 were correctly recognized (circle) and which were incorrectly identified (square). We can observe that squares are concentrated on the lower part of the graphic, underlining that $\varphi(p)$ function gets a low value, while circles are mostly located in the upper part, indicating a higher value of such function.

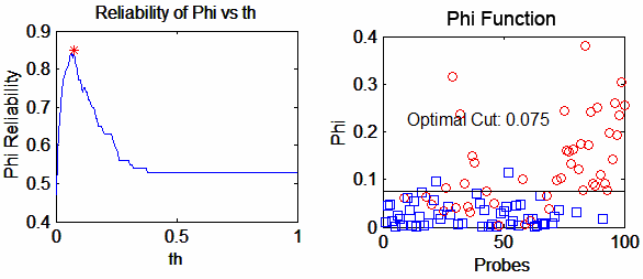


Fig. 3 Experimental evaluation of SRR index.

It is also important to verify the validity of the criterion for choosing the single th_i thresholds. To do this a comparison was performed between estimated thresholds (formula (18)) and optimal ones, the latter being computed by maximizing the number of system responses consistent with the value of SRR index. In this case the selected fusion rule is the AND, and the reliability index used is SRR II. For both face and ear, estimated thresholds are very close to the optimal ones, for all test sets (e.g.: EAR: 0.06/0.07, FAFB 0.6/0.5, DUP I 0.4/0.5, AR-LEFT LIGHT 0.5/0.7, SUNGLASSES 0.6/0.7). We tested both SRR indicators introduced in Section III, in combination with two different integration policies based on them, for data coming from single subsystems. Experiments were performed for the three different datasets extracted from FERET, combined with Notre-Dame ears database. The single facial recognition subsystem assures, on three different databases (FAFB, FAFC, DUP I), a Recognition rate of respectively 93%, 16% and 47%, and an Equal Error Rate of 0.03, 0.29, 0.19; the ear recognition module on Notre-Dame database provides a Recognition Rate of 72% and an Equal Error Rate of 0.15. Table 3 reports results obtained from the multimodal system Face/Ear using each of the integration policies (OR, AND) and varying the face database; the results are presented in terms of Recognition Rate (RR), Equal Error Rate (EER) and Number of Reliable Responses (NRR). Table 3 shows that on the whole all the integration policies guarantee an higher Recognition Rate than the one coming from face or ear (87% for DUP I with OR/SRR II instead of 47% e 72%), and at the same time considerably reduce the Equal Error Rate (0.015 for FAFB with OR/SRR II instead of 0.030 and 0.150), that in some cases even cancels out. The counterbalance of such recognition performance increment is the number of responses that are invalidated because considered as unreliable. However, it is worth underline that an unreliable response has not to be necessarily discarded or considered as wrong, but could represent a valid reason to perform a further check. This could even be more expensive,

like for example fingerprints; as such check would be performed only on a limited number of cases, the average computational cost of the multimodal system would be reduced anyway, with the advantage of an higher security degree. The first column shows that all responses are considered as reliable when fusion is performed without any SRR index. However, both RR and EER are significantly worse, showing its ability to discard little reliable responses.

Table 3 Comparison between the performance of the fusion rules.

Database		Statistics				
		None	SRR I		SRR II	
		OR	OR	AND	OR	AND
FAFB	RR	98%	99%	100%	96%	100%
	EER	0.028	0.016	0.003	0.015	0.000
	NRR	100	75	63	94	38
FAFC	RR	55%	76%	100%	84%	-
	EER	0.167	0.153	0.002	0.117	-
	NRR	100	85	2	74	0
DUP I	RR	75%	81%	100%	87%	100%
	EER	0.238	0.228	0.001	0.177	0.000
	NRR	100	91	18	84	22

From a quantitative point of view, we can observe from that OR always assures a higher number of responses considered as reliable, as expected. Finally, the two reliability indexes SRR I and SRR II have been tested on subsets extracted from AR-Faces database, combining face and ear using AND policy. The results of such experiment are shown in Table 4.

Table 4 Comparison between the performance of SRR I and SRR II.

Distortions on the Face		Statistics				
		Face	Ear	Face ⊕ Ear		
					SRR I	SRR II
Left Light	RR	83%	72%	RR	100%	100%
	EER	0.11	0.12	EER	0.001	0.000
	REL	93%	90%	NRR	37	30
Sad	RR	95%	72%	RR	100%	100%
	EER	0.07	0.12	EER	0.009	0.010
	REL	100%	90%	NRR	86	86
Scarf	RR	80%	72%	RR	100%	100%
	EER	0.17	0.12	EER	0.000	0.004
	REL	83%	90%	NRR	20	47
Scream	RR	47%	72%	RR	100%	100%
	EER	0.18	0.12	EER	0.002	0.000
	REL	83%	90%	NRR	19	10
Sun glasses	RR	87%	72%	RR	100%	100%
	EER	0.10	0.12	EER	0.020	0.020
	REL	87%	90%	NRR	87	77

The single subsystems performances in terms of Recognition Rate, Equal Error Rate and Reliability are reported in the first two columns of the table. As the database used for the ear is always the same, data are repeated on each row. The results obtained using the combined system with AND policy are reported in the following two columns; the reliability index which was used is on top of the column. Results in Table 4 further enforce our considerations. As a matter of fact, we can observe that the combined system guarantees in all cases better performances than single subsystems (in some cases like *scarf* we pass from 47% and 72% to 100% as for Recognition Rate). However, when the single subsystems are particularly stressed by input quality, the number of responses considered as reliable significantly decreases. This confirms the need for more biometries to further strengthen the system.

V. Conclusions

The paper faces crucial problems for a multimodal system, namely data normalization and system response reliability. In particular, we defined a new data normalization function, the *mapping function*, able to overcome the limits of the presently used ones. We also proposed two different reliability indices, providing the conditions for their actual exploitability within the single subsystems, such as the *a priori* identification of thresholds.

All the theoretical concepts introduced in the first part of this work are then validated in the section showing experimental data, through a number of tests performed on different databases chosen among those commonly used.

Future developments can certainly regard the extension to a higher number of either biometries or fusion models, so that a higher number of subsystems could be involved in our analysis.

References

[1] A. F. Abate, M. Nappi, D. Riccio, Face and Ear: A Bimodal Identification System, Proceedings of the Int. Conf. On Image Analysis and Recognition ICIAR 2006, Póvoa de Varzim, Portugal, September 2006.

[2] A. K. Jain, A. Ross, and S. Prabhakar, An introduction to biometric recognition, IEEE Trans. on Circuits and Systems for Video Technology, vol. 14, pp. 4–20, Jan 2004.

[3] A. K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, in Pattern recognition, Vol. 38, No. 2, pp. 2270-2285, 2005

[4] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class", in *IEEE Transans. on Pattern Analysis and Machine Intelligence*, Vol.24, No.6, pp. 748–763, June 2002.

[5] J. P. Phillips, H. Moon, A. S. Rizvi and P. J. Rauss, The FERET Evaluation Methodology for Face-Recognition Algorithms, in IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.22, No.10, pp.1090–1104, 2000.

[6] A.Ross, A. K. Jain, Multimodal Biometrics: An Overview, in Proceedings. of 12th European Signal Processing Conference (EUSIPCO), (Vienna, Austria), pp. 1221-1224, September 2004.

[7] Notre Dame Ear Database, http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html, 2007-01-07.

An Ambient Intelligence Application for Cultural Heritage

Gennaro Costagliola, Sergio Di Martino, Filomena Ferrucci, Paola Capone
Università degli Studi di Salerno
Via Ponte don Melillo, 84084, Fisciano (SA), Italy
{gcostagliola | sdimartino | fferrucci | capone }@unisa.it

Abstract

In this paper, we present an Ambient Intelligence application to provide a virtual guide for the visitors of an Italian botanical garden. The application exploits ubiquitous computing solutions (UltraMobile PCs) to present contextualized information to the visitors, ubiquitous communication technologies to get spatial awareness (through RFID active tags pervasively embedded in the environment) and communicate with a user-friendly interface to handle the interaction.

1. Introduction

Information technologies are widely recognized as very powerful tools to exploit cultural heritage. Indeed, if the main goal of early information systems for cultural heritage was limited to organize the enormous quantity of cultural artifacts, further development has brought new exciting opportunities in the field ([3], [6]). Nowadays, worldwide there are many projects and solutions aimed at improving the user experience, when visiting a heritage site, through some computing solutions. *Virtual guides* are a typical example of these applications, allowing the access to relevant cultural heritage resources through PDAs or Tablet PCs, resulting more effective than traditional audio/visual pre-recorded guides (see e.g. [2]).

However, every application in the cultural heritage field may pose challenging issues to computer science specialists, due to the variety of scenarios and problems that may arise, ranging from context awareness to multimedia databases, till human-computer interaction. Currently, many research efforts are devoted to the definition of less intrusive interaction paradigms, leading to more natural usages of computing devices within cultural heritage environments. The key ideas underlying the "*Ambient Intelligence*" (AmI) concepts seem to be an interesting way to tackle these issues.

AmI was envisioned by the IST Advisory Group (ISTAG) of the European Union in 1999, as "an exciting new paradigm of information technology, in which people are empowered through a digital environment that is aware of their presence and context sensitive, adaptive and responsive to their needs, habits, gestures and emotions" [8].

Within an AmI environment, users are mobile and communication is set up between individuals and objects in a transparent way. This communication is obtained via 'invisible intelligent devices', able to tag key spots in our surroundings. An AmI application concretizes this vision by merging together three focal technologies:

- *Ubiquitous computing*, i.e. integration of microprocessors into everyday objects like furniture, clothing, white goods, toys, even paint.
- *Ubiquitous communication*, i.e. enables these objects to communicate with each other and the user by means of ad-hoc and wireless networking
- *Intelligent user interfaces*, enabling the users of an AmI application to control and interact with the environment in a natural and personalized way.

In this research field it is posed the project "Ambient Intelligence@Giardini della Minerva" (AmI@GM), whose goal is to make available to the wide public the broad and precious historical and cultural patrimony of the *Historical Archive of the Salerno's Botanic*.

Indeed, from the ninth to the thirteenth century, the town of Salerno held the most important source of medical knowledge in Europe at the time, the *Schola Medica Salernitana* [13], considered the very first university of the world. In that period, many manuscripts were written about how to heal with the plants (e.g: [10]). Nowadays, these manuscripts are

spread in the most important museums all over the world, but the *Historical Archive of the Salerno's Botanic* aims to recreate a digital corpus of that knowledge in one of the most fascinating spots of Salerno, the *Giardini della Minerva* (Minerva's Botanical Gardens), which is the most ancient European botanical garden. To date, this digital corpus contains more than 10,000 images, arranged in a multimedia database.

The project described in this paper is meant to improve the spreading of this precious knowledge, using the most advanced informatics technologies, for a context-driven, ubiquitous presentation of its contents. To this aim, a specific application has been developed, intended to guide the visitors at the *Giardini della Minerva* (see Figure 1) in the discovery of the fascinating medieval books written in relation to the plants contained in this botanical/heritage site.

The difference with respect to the numerous products aiming to spread knowledge of the mediaeval books lies in the way the information is accessed. Indeed the developed application exploits the real environment of the *Giardini della Minerva* as a kind of information selector, allowing the user to go near one of the 500 visible plants, to see on an Ultra-Mobile PC how it was represented in the Middle Ages in the illuminated herbaria, as well as its curative effects.

Figure 1: A flowerbed in Giardini della Minerva

The result of the project is a product intended as a cultural and informatics reference for the multimedia productions, aimed at increase and diffuse the knowledge in the cultural heritage field.

The rest of the paper is structured as follows. In Section 2 there is a detailed description of the developed system. Section 3 reports on the adopted technologies and solutions, while Section 4 is focused on the designed User Interface. Some final remarks and future work conclude the paper.

2. The Giardini della Minerva and the digital Corpus

In this section we describe the Giardini della Minerva location and the multimedia database of digital images referring to the medieval herbaria.

2.1. The Giardini della Minerva and the related medieval manuscripts

The *Giardino della Minerva* is a living historical and architectonic site, placed in the heart of the old town of Salerno, in an enchanting place, rich of ancient tradition. Indeed, in these places the glorious Medical School of Salerno rose during the middle ages, well-known all over the World for its competences with the "*semplici*", i.e. the part of a plant with active therapeutic principle. Within the area of this garden, it is located the Historical Archive of the "Botanica Salernitana", which collects and studies the medieval texts concerning Salernitan medical and botanic subjects. Indeed, between the end of the XIth century and the full XVth the mediaeval manuscripts demonstrate the extraordinary meeting in Europe of the salernitan medical culture. The classification of 'simplicia' represents the basic fundament of the tractatistic medical literature of School of Salerno.

Currently, the archive keeps, among others, a cultural patrimony of inestimable quality but undoubtedly unaccounted, constituted by more than 10,000 images on slides of medieval herbaria (manuscripts of exquisite workmanship, splendidly illuminated), all based on the original "*Tractatus de herbis*" written by Plateario (an example is reported in Figure 2).

At present, the original herbaria are scattered into the main international museums, such as the ones in London and Paris. Due to their value and delicacy, it is very difficult for everyone interested in art or in medical issues to access these ancient books. Moreover, due to the distance between the museums and in the same museum, it is obviously impossible to consult at the same time more herbaria, for example in order to compare the representations of the same plant among authors.

Consequently there is a big limitation in the spread of the knowledge both of the philological-literary sciences and of the pharmacologic biology.

Figure 2: A page of a Manuscript

2.2. The multimedia digital repository

The modules described in the previous sections exploits a multimedia database, intended as the underlying versatile layer for the retrieval and the visualization of both the 10.000 medieval herbarium photos, both all the associated information.

To this extent, we choose the most appropriate set of metadata to structure the selected information. We choose the DublinCore [7] and its updates, described in next section. At the same time, due to the large amount of multimedia information, it was carefully designed a protocol for acquisition, paying much attention to the peculiarities of the old collection and to the project goal, together with the identification of the most suitable storage/server solutions.

Consequently, for each of the hundreds of plants in the garden, we stored the following information:

1. Its actual scientific and popular name and its medieval name.
2. A photo of the plant.
3. Its possible actual use in medicine and taxonomical classification.

Each plant was described in at least one medieval manuscript. For each page, we stored the following information:

1. Information on the Manuscript
2. A High-Resolution digitization (4800dpi) of the page in the Manuscripts.
3. A Low-Resolution, web-optimized digitization (4800dpi) of the page in the Manuscripts.
4. A transcription of the text (this because often is very difficult to interpret it)
5. A translation (in Italian) of the ancient text.
6. A summary description of the pharmacological effects associated by the author to the specific plant.

As a result of this step, we obtained a multimedia database, containing about 10,000 Hi-Res and 10,000 Low-Res photos.

2.2.1. Dublin Core

To allow the greatest flexibility in the data management, it was fundamental to define a database schema containing as much information as possible.

To this aim, we implemented the *Dublin Core* (DC) [7] metadata schema. The DC is a core set of 15 semantic definitions for describing resources likely to be useful across a broad range of industries and disciplines. It was defined by the *Dublin Core Metadata Initiative* (DCMI) [16], an organization instituted by the *National Center for Supercomputer Application* and the *Online Computer Library Center*. The DCMI is dedicated to "promoting the widespread adoption of interoperable metadata standards that enable more intelligent information discovery systems" [16]. The DC was successfully adopted by the *Computer Interchange of Museum Information* to describe the artworks of the principal world museums, such as the British or the Smithsonian.

By adopting the DC metadata, for each item in the database, it was possible to store, among other information, the title, the author, the date of creation, the size, the format, some describing keywords, and a link to one or more digital counterparts.

3. The AmI@GM system

The proposed system is as a pervasive computing-based virtual guide, accessible via an UltraMobile PC with an auto-localization module, meant to support visitors of the Giardini della Minerva in the discovering of its botanical information and medieval culture. In this section we present the main aspects of the developed system.

3.1. System Architecture

An overview of the overall system architecture is shown in Figure 1. It is based on a three-tiered model, encompassing a mobile front-end, a middle application server, and a database server. The client devices are some Ultra Mobile PCs, equipped with RFID readers, and communicating via Wi-Fi with a local middle-tier infrastructure. The multimedia database contained the digitization (both low and hi-resolution) of about 10.000 slides together with a textual description, while the spatial contextualization of the user devices was performed by exploiting the RFID technology. In order to make the product attractive for a wider range of intended users, the presentation of the information related to the images makes use of the adaptive techniques, basing on the profile of the current user.

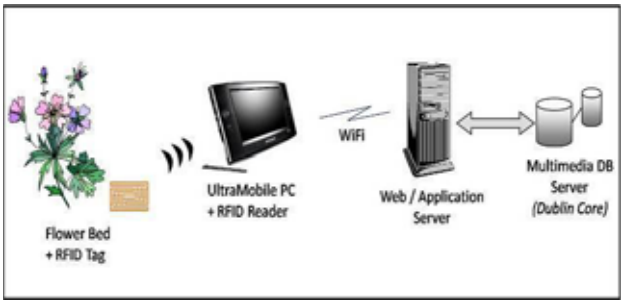


Figure 3: The System Architecture

Moreover, the intended execution environment (the *Giardini della Minerva* location) is disseminated of RFID tags, to provide spatial localization. Indeed, the idea is to associate to each flowerbed of the garden, an RFID tag containing a unique identifier.

Upon request, each visitor of the garden is equipped with an Ultra Mobile PC. The interaction between the application running on the MobilePC and the system is triggered by the detection of a RFID tag, providing the awareness of the user position within the environment. Once the RFID reader on the PC detects a tag, its unique ID is passed to the web server via the Wi-Fi connection. This allows the application server to automatically generate a dynamic page with the information regarding the flowerbed.

The three-tiered architecture led us to a more flexible system, where only a light user interface is stored onto the client devices. Indeed, the huge multimedia dataset of (eventually hi-res) images is stored on the back-end and sent on-demand to the client. Moreover, with this architecture, updates/upgrades in the profiling policies are easier to implement, since they can be modified only onto the Application Server.

3.2. Ubiquitous Computing Aspects

The main interaction device is an UltraMobile PC, i.e. an object whose size lies between a PDA and a standard notebook. Indeed it has a 7" touch-screen, no keyboard, Windows Xp Tablet-PC edition, and many connectivity options, such as the Wireless Network Adapter for IEEE 802.11b/g, Bluetooth, and a Compact Flash slot, used to interface a RFID reader.

The weight is less than 1 Kg, so these devices can be easily handled by visitors of the Giardini della Minerva.

These UMPC run a Win32 application, enlarge to full-screen, and embedding a web browser control. Clearly, all the standard Windows widgets have been hidden, to present a domain-specific user interface. Some screen-shots and a deeper presentation are provided in Section 3.4.

3.3. Ubiquitous Communication Aspects

In our system we employed two wireless technologies, to achieve the ubiquitous communication within the Giardini della Minerva.

Indeed, we needed a solution to locate in an automatic and transparent way the user position within the environment, and one to enable a wireless communication between the UltraMobilePC and the Web/Application Server. The first issue was addressed by using RFID technology, while the latter by employing a standard Wireless 802.11g network. In the following we describe these solutions.

RFID

Radio-frequency identification (RFID) is an automatic identification method, relying on storing and remotely retrieving data using devices called RFID tags or transponders. An RFID tag is an object that can be attached to or incorporated into a product, animal, or person for the purpose of identification using radio waves. Chip-based RFID tags contain silicon chips and antennas. Passive tags require no internal power source, whereas active tags require a power source.

RFID tags are categorized as either active or passive. Active RFID tags are powered by an internal battery; passive RFID tags operate without a separate external power source and obtain operating power generated from the reader.

We disseminated the flowerbeds of the Giardini della Minerva with RFID tags, as shown in Figure 4. In

particular, we used active tags, having a wider range, each holding a unique ID, useful to identify the flowerbed near the user. Indeed, this tag's ID is used as request parameter to the web server, to obtain as response a page with all the plants contained in the nearest flowerbed. Typically, each page contains about 20 species, identified by a photo and the name.

Figure 4: Displacement of Active RFID Tags within the flowerbeds

WIFI

The communication between the UltraMobile PCs and the infrastructure is managed by standard a Wi-Fi IEEE 802.11g network. Indeed, tree hotspots have been placed within the Giardini della Minerva, and are able to cover the entire environment with a wireless network.

3.4. User Friendly Interfaces

A key aspect of the ubiquitous computing is a spatial awareness of the system, to simplify the access to the right information, removing any unnecessary user navigational interaction with the system. In our case, the contextual awareness was used to filter the dataset of herbaria images. Indeed, when the user walks within the "Giardino della Minerva" and goes closer to a flowerbed, he/she will be guided to discover the set of embedded plants, through the presentation of a photo of that plant in the various seasons, of a photo taken by a contemporary herbarium and, the most enchanting aspect, of the illuminated herbaria images, showing how the plant was represented in the Middle Ages. Furthermore, if the user is interested, he/she can obtain enlargements of the selected page, being able to read the medieval text, describing about the curative effects speculated in the Middle Age on that plant.

Since the manuscripts were written in Latin, archaic French or Spanish, with a very difficult calligraphy, each page is coupled with its corresponding transcription and translation.

It is worth noting that the defined interfaces do not require any complex choice by the user, but all the interaction is driven by direct manipulation, since he/she simply clicks with his/her fingers on some images onto the screen to access the specific information. There is only one icon used within all the interfaces, used to go back one level, always placed in the lower-left corner of the page.

In particular, as stated previously, when the system detects an RFID tag, indicating that user is close to a flowerbed, a page is shown on the UltraMobile PC, showing the list of plants, with a photo of the species, its popular and scientific names.

Once the user has selected a specific plant by pointing on its photo, the systems shows a set of miniatures of all the medieval pages related to that plant, together with the name of the book (a screenshot is provided in Figure 5).

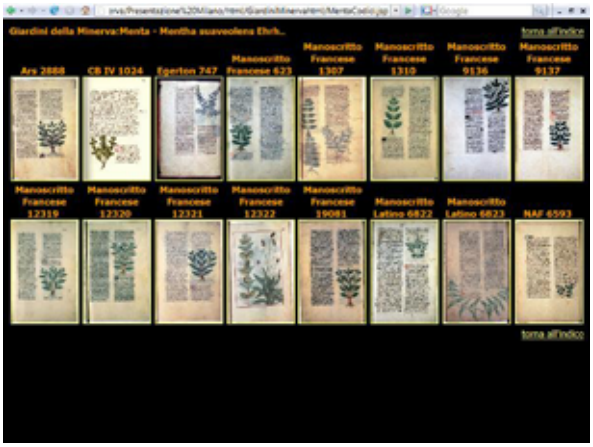


Figure 5: The set of Manuscripts, for the selected plant

Finally, the user receives a zoomed photo of the manuscript's page about the selected plant, a transcription of the page, and its translation (a screenshot is provided in

Figure *6*).

Figure 6: Detailed view of a Manuscript, together with its translation

4. Conclusions and Future Work

Diffusion of applications for cultural heritage is quickly rising. Nevertheless to ensure effectiveness of these systems it is necessary to provide users with appropriate user interfaces. It is recognized that traditional WIMP interfaces can turn out to be not fully adequate. Thus, there is the needing for novel interaction paradigms.

In this paper, we presented an ambient intelligence system we developed for one of the most beautiful and ancient Italian Botanical gardens. Indeed, the proposed system is able to reduce the needing for navigational interaction with the user, by exploiting contextual awareness to present the proper information. The spatial context is devised by some RFID active tags pervasively embedded in the flowerbeds of botanical garden), to provide a virtual guide for the visitors.

About future work, we are planning to perform some subjective and objective usability evaluations of the user-interfaces, to assess the effectiveness of the proposal.

5. References

[1] Advisory Group on Computer Graphics, Exploiting Virtual Reality Techniques in Education and Training: Technological Issues, Technical Report, http://www.agocg.ac.uk/reports/virtual/vrtech/toc.htm

[2] Augello A. et al., *MAGA: A Mobile Archaeological Guide at Agrigento*. In Procs of MobileGuide06, Turin, 18 October 2006, Italy.

[3] Bocchi, F. Bonfigli M. E., et al., *Virtual Reality and Cultural Heritage: Some Applications*, Proceedings of the EVA 2000 FLORENCE Conference, Florence, Italy

[4] Capone P., L'arte del vivere sano. Il Regimen Sanitatis Salernitanum e l'età moderna, Milano 2005;

[5] Costagliola G., S. Di Martino, F. Ferrucci, *Dynamic generation of educational virtual exhibitions*. In "Proceedings of the IEEE 1st International Conference on Information Technology: Research and Education (ITRE 2003)", New Ark, NJ, USA, 2003. IEEE Press, pp-198-202.

[6] Costagliola G., S. Di Martino, F. Ferrucci, F. Pittarello, *An approach for authoring 3D cultural heritage exhibitions on the web*. In "Proceedings of the Fourteenth International Conference on Software Engineering And Knowledge Engineering (SEKE'02)", Sant'Angelo d'Ischia, Italia, 2002. ACM Press, pp. 601-608.

[7] Dublin Core Metadata Initiative, web site: http://dublicore.org

[8] ISTAG Ambient Intelligence: from vision to reality. Last visited on May 28, 2007. ftp://ftp.cordis.lu/pub/ist/docs/istag-ist2003_draft_consolidated_report.pdf

[9] Mater herbarum. Fonti e tradizioni del giardino dei semplici della Scuola Medica Salernitana, a cura di M. Venturi Ferraiolo, Guerini e Associati, Milano 1995.

[10] Mattheus Sylvaticus, Opus Pandectarum Medicinae, O. Scoti, Venetiis 1523.

[11] Minerva Europe Project, 2001. Ministerial Network for Valorising Activities in digitalisation. IST Project n° 2001-35461

[12] Moore, P., Learning and teaching in virtual worlds: Implications of virtual reality for education. Australian Journal of Educational Technology, 1995.

[13] Schola Medica Salernitana. Description from Wikipedia, available at:

http://en.wikipedia.org/wiki/Schola_Medica_Salernitana

[14] Shneiderman B., 1998, *Designing the User Interfaces*, Addison Wesley.

An Integrated System for Easying the Access to Georeferenced Information on the Web

Fabio Pittarello
Università Ca' Foscari di Venezia
Dipartimento di Informatica
Via Torino 155, 30172 Mestre (VE), Italy
pitt@unive.it

***Abstract*—The association of data of different nature to geographical locations is becoming increasingly popular in a growing number of domains, including urban management, economics and social studies.**

Traditional geographical information systems (GIS) are powerful tools for processing georeferenced information, but in most cases they are still complex systems that require highly trained users for obtaining useful results.

This work proposes a novel web-based architecture for hiding to the users the complexity of GIS and for allowing them to interact with the underlying spatial engine through simplified visual interfaces that take care both of the query and of the presentation phases. Such integration represents an improvement in relation to most of the previous research work focused, respectively, on easying the query or the access to resulting data.

The system has been designed in collaboration with the researchers of Corila, a Consortium for the preservation of the Venice lagoon, that contributed to the definition of the system requirements.

I. Introduction

The knowledge of georeferenced data has become increasingly important in the last few years. Different domains ranging from urban studies to economics take advantage of such knowledge for monitoring complex processes and elaborating strategies. Unfortunately, in many cases the management of georeferenced data needs sophisticated tools and interfaces that require a long training for taking full advantage of the system. The result is that often the users can't take advantage of the system potential because of their inability to operate on complex interfaces. Users often include professionals that have a deep knowledge of their domain but are not skilled in computer science.

This work describes an integrated system for easying the access to georeferenced information on the web. The architecture of the system permits the user to access a remote repository of georeferenced information and take advantage of a sophisticated engine for performing spatial operations on such data, without being overwhelmed by sophisticated interfaces for the query and the presentation of information.

Two visual environments for the input and the output are integrated in a unified interface that completely hides to the user the complexity of the underlying GIS engine. Such integrated solution represents an improvement in relation to most of the research work for easying the access to GIS, that focuses only on the input or, alternatively, on the output interface.

Concerning the input, a visual environment compliant with the web standards is defined for composing visual counterparts of the variables and of the operators involved in the queries, hiding to the user the complexity of textual SQL-like languages. The input interface is based on the results of previous research work [1] [2] and has been functionally enhanced in order to let the users to use both discrete and continuous variables as query operands.

Concerning the output, a visual environment based on the Google Earth client is used to present the results of the query. Google Earth [3] is a web-based client-server system that enables the exploration of a 3D model of the globe. Google Earth can't be defined a full-fledged GIS, because it doesn't include interfaces and engines for composing and processing sophisticated spatial queries: basically its standard interface allows the users to retrieve addresses or paths connecting different locations. In spite of that, Google Earth is becoming increasingly popular because of the intuitive presentation of geocoded data on the Earth surface. Vector and raster information can be displayed by third-party developers on the top of Google Earth, allowing therefore to use it as a base for visualizing additional georeferenced data.

The management of georeferenced continuous data still represents an open issue, both for the processing and the access to such data [4]. This work uses one of the few engines available for the processing of continuous variables [5] as the back-end for an integrated, user-friendly, query and presentation environment. Such environment, which represents the primary focus of this work, enables researchers to make queries that involve also continuous variables and to visualize the results.

The work is organized as follows: Section II surveys related works. Section III introduces the general motivation for this work and its requirements. Section IV presents the interface and architecture for accessing and processing geocoded information, including continuous variables. Section V describes the application scenario, derived from environmental surveys on the Venice lagoon. Section VI draws the conclusions.

II. Related Work

This section considers related work, with a particular emphasis in the area of interfaces for querying GIS and in the area of the web access to geocoded information.

Research on visual languages and environments has been developed in recent years to help users to interact easily with computer systems avoiding to learn the tedious syntax of textual command languages.

Textual query languages derived from SQL allow skilled users to retrieve information stored in databases through text commands [6] [7]. The main drawback of this approach is that it requires the users to learn a complex syntax, exposing the unexperienced users to the risk of formulating uncorrect queries.

Research aiming at introducing visual counterparts for spatial queries has tried to solve such drawbacks, leading to a variety of proposals that include iconic [8], graph-based [9] and sketch-based [10] [11] [12] [13] approaches.

Concerning the operands for composing the queries, most of the approaches permit to use visual counterparts representing discrete objects (e.g., a point, a line or a region). Only in recent times a few researchers have started working on continuous variables. The paper [4] proposes a visual environment where the different syntactic parts of the SQL query are mapped to a set of areas that the users must fill in sequence with a set of visual counterparts of continuous and discrete operands, in order to generate a query. While the approach is interesting and contributes to fill in a significant gap in the domain of the visual representation of continuous fields, the proposal closely mimics the SQL language, resulting not intuitive for the users that are not familiar with its syntax. This work renounces to a part of the expressivity offered by the latter proposal, proposing a simpler query environment targeted to unexperienced users.

Concerning the operators for composing the queries, different classes of them have been defined, in order to perform sophisticated processing of data. In spite of that, only recently spatial operations have been formalized, that include also height and depth as a feature. As a consequence, only in recent times research about visual languages has focused on this issue: currently only a few visual environments allow to make queries about phenomena developing in a full 3D space.

One of such environments [1], to which the author of this paper contributed, allows to compose visual queries in an abstract 3D space where the operators and the operands are represented in a visual fashion. The underlying algebra is derived from the OpenGIS SQL proposal [7] and it is characterized by a set of topological, directional and metric operators, applied to operands positioned in a full 3D space. The implemented system allows the users to make queries with discrete variables, using in parallel one or more operators (e.g., *find all the churches with a crypt placed at north of the church and under the church level*). The current release of the system doesn't include the possibility to use continuous variables.

Concerning the access to geocoded information from the web, the situation has radically evolved in the last 10 years. At the beginning, GIS were proprietary systems with complex interfaces and modest opportunities to export data. With the advent of the web, a number of vendors (e.g., MapGuide by Autodesk [14]) offered systems with client components to embed in the web pages, implemented with Java or ActiveX technology. Such systems usually didn't offer any opportunity to export data to other formats for presentation.

A different approach, characterized by the ability to make a query from a standard web page and to present the result in the browser, using standard XHTML, is offered by other GIS, mainly open source products [15]. In most cases the web interface offers only a subset of the query and presentation potentialities of the full system. The paper [1] can be considered as an advanced example of such approach that uses web standards also for the query tool (i.e. a visual environment based on VRML [16] and XHTML).

The role of standards for the description of interactive 3D worlds for the net has become increasingly important in the last few years. VRML and its successor X3D [17] offer the opportunity to represent geometrical objects and interactive behaviors with a high degree of precision. The introduction of GeoVRML [18], an extension of the VRML language that then has been fully integrated in the X3D standard, has given the opportunity to represent geocoded information on the web using standard VRML/X3D browsers available for most operating systems.

X3D Earth [19], an interesting ongoing project launched by the Web3D Consortium [20], promises in the near future to bring to web users the richness of a 3D textured representation of the Earth together with the interaction potentialities offered by the X3D language.

A subset of the features of the X3D Earth project are already available in Google Earth [3], an innovative client-server architecture that enables the users to access a 3D model of the Earth and a set of georeferenced objects (e.g., representations of roads, locations, raster data, etc.) on its top. The client can be used also as a base layer for displaying additional classes of georeferenced information, using both visual tools and the proprietary *KML* language. Some researchers have started to take advantage of such feature, using the client as a front-end for displaying geocoded data generated by different services [21] [22].

Concluding, even though Google Earth lacks the interactivity of VRML and X3D worlds, currently it represents an unique opportunity for attaching georeferenced data to an Earth representation that can be easily accessed by any major desktop platform connected to the web. This is the reason why we decided to use this platform for the representation of the output of the queries.

III. Defining the Requirements

The general requirements addressed by this paper include:

- easy access to the geocoded data, both in the query composition and in the presentation of results;
- possibility to use a wide range of mathematical, metric,

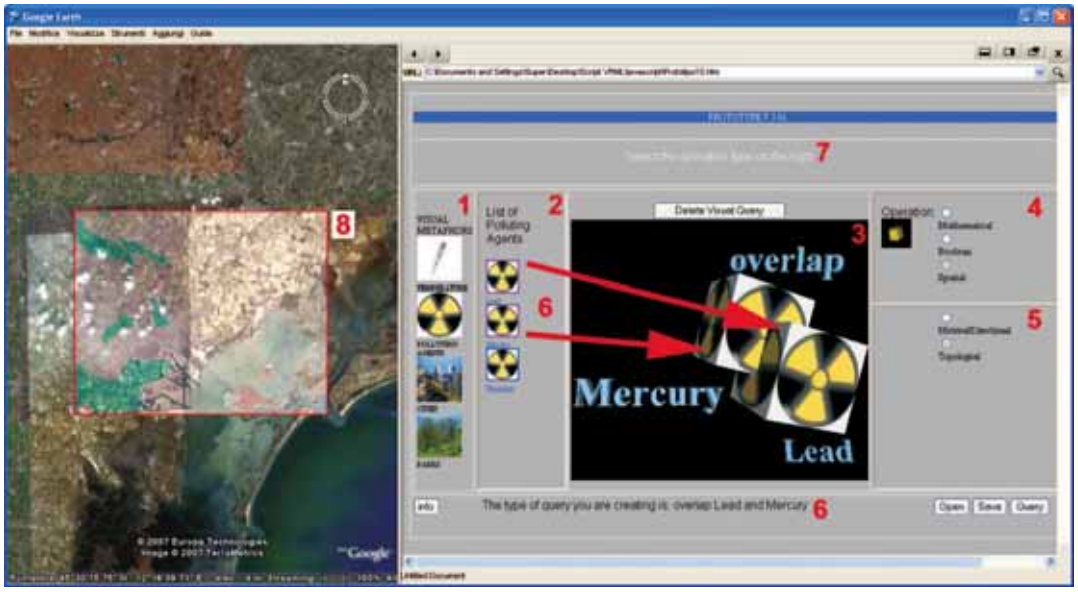


Fig. 1. The Integrated Visual Interface for Composing Queries (on the right) and Presenting Results (on the left).

directional, topological and boolean operators for processing the data;
- query and presentation of continuous data;
- web access to the remote GIS engine and minimization of the installation on the client side.

Such requirements intersect specific issues derived from a research activity related to the environment of the Venice lagoon. In spite of that, they are general enough to be consistent with the needs of other domains.

The environmental research is coordinated by Corila, a Consortium for the preservation of the Venice lagoon [23], with the contribution of the Departments of different Universities of the regional area. The Corila Consortium manages a large quantity of environmental data of different nature derived from experimental surveys of biologists, chemists and physicists that contribute to the research. Most data is associated to geographical locations (e.g., concentration of chemical elements, temperature, distributions of vegetables and animals, speed and temperature of water, etc.). A significant part of the surveys includes continuous data that must be stored, processed with different operators and visualized by the researchers participating to the research project.

There was an emerging request, stemming from different research groups participating to the project, of friendly interfaces for managing the insertion and the retrieval of data without needing a long training activity. Most researchers are expert in their domains, but don't have a deep knowledge of GIS nor they want to be involved in low level computer science issues.

Another requested feature was the possibility to retrieve data from remote locations, possibly minimizing the installation requirements.

IV. An Integrated Visual Approach for Easying the Access to Georeferenced Data

A. Hiding the Complexity of GIS

The proposal described in this work represents an answer to the requirements described above. The system is focused on easying both the query composition and the presentation of data.

In compliance with most previous research work related to visual query languages for the GIS domain, we implemented, for the input interface, a simplified and highly interactive environment for enabling the user to query the system manipulating visual counterparts of the variables and their relationships. A different choice was made for the output interface, conceived for showing the localization of the results on the territory. Such interface didn't require complex manipulation capabilities. Instead, it required the capability to navigate the georeferenced results superimposed to the territory map and to make simple selections of such data. Because of the current unavailability of a single technology for building both the input and output environments, we used different tools to achieve the result. In particular:

- the input phase takes advantage of a visual environment based on VRML and XHTML for composing the queries and for translating them into a textual string; such string is then passed to the GIS engine for processing;
- the output phase takes advantage of the Google Earth interface for superimposing the data to a representation of the Earth surface; the interface is used both for displaying the survey data and the query answers.

The interfaces have been integrated in a unified layout for the query composition and the data presentation (see Figure 1).

The choice of VRML and Google Earth for managing,

respectively, the input and the output phase is due to their different points of strength and weakness. While the VRML technology permits to manipulate 3D interactive objects for composing the queries in a visual fashion, currently doesn't offer a 3D Earth model for presenting the results. On the other side, Google Earth is a widely available 3D client for navigating the representation of the Earth at different levels of granularity and layering several classes of objects on its top, but currently doesn't permit to manipulate (e.g., move or rotate) the objects represented on the Earth's surface; therefore it can't be used for composing interactively the queries. Even the novelties of the last version of Google Earth, such as the possibility of adding objects complying with the new Collada specification [24] don't add interactive features, and therefore are not adequate to our goal.

The following subsections are focused on describing in detail the components of the visual interface; a description of the implementation architecture will follow.

B. The Input Phase: The Visual Counterparts for the Query Operators and Operands

The proposal described for the input phase stems from a more general research activity, to which the author contributed, aimed at improving the quality of interaction when accessing geographical data, with a particular reference to users that are not skilled in the computer science domain. In particular [1] [25] were focused on easying the composition of queries, translating the constructs of the OpenGIS SQL [7] into a set of visual metaphors that could be composed also by unexperienced users.

This work adds a further level of expressivity, allowing the users to use also continuous variables and introducing additional classes of operators.

The variables are mapped to textured and labeled 3D cubes, associated to survey data, named *geometaphors*; such visual metaphors are selected by the users in the 3D visual environment for composing the query.

The choice of mapping the variables to a 3D representation is due to the relevance of the third dimension for a significant number of operators (such as the directional operators, which include also height and depth). Such solution has already been evaluated in terms of usability and user satisfaction [26], showing that unexperienced users benefit from the application of such metaphor, allowing them to make complex queries without knowing any syntactical detail of a textual query language.

In this work the query operators have been selected matching the full set of operators available in the GIS engine chosen for the system with those ones deriving from a survey conducted among the potential users of the system. Such survey was focused on identifying the classes of operations that were considered more useful for the environmental domain. In particular:

- the metric, directional and topological operators, already considered by the previous works, have been confirmed for this system;
- the boolean and mathematical operators represent new additions, due to the processing of continuous variables and to the request of performing statistical operations on the geocoded data.

Because of the limits of the paper we don't have the space for an analytical description of all the operations; therefore we'll give a synthetic classification of all the operators involved. Each class of operators is visualized in peculiar way in the query interface:

- the *metric* operators are represented by a textual label placed next to the geometaphors (e.g., the relation *distance* is represented by a numerical label between two geometaphors);
- the *directional* and *topological* operators are represented by the spatial properties of the geometaphors disposed in the 3D scene (e.g. the relation *overlap* is represented by two overlapped geometaphors, the relation *disjoint* is represented by two separate geometaphors); complementary textual labels dynamically confirm the current spatial relation to the user that is composing the query;
- the *boolean* operators are visualized introducing a visual artifact between the geometaphors; the artifact is textured with a representation of the specific boolean operation and is marked by a textual label (e.g., the relation *and* is represented by a small cube textured with an icon representing the union of objects, placed between two geometaphors);
- the *mathematical* operators are represented by a textual label when the operation is applied to a single geometaphor; the operators are visualized introducing a visual artifact between the geometaphors when there are two or three variables (e.g. the relation *mean* is represented by a small cube textured with an icon representing the mean operator, placed between two or three geometaphors).

C. The Input Phase: The Visual Interface for Composing the Queries

The visual metaphors described in the previous subsection are composed in a 3D visual interface. The main part of the interface (see Figure 1 on the right) is represented by a 3D environment where the user can put from one to three geometaphors chosen from the left panel of the input interface and associated to the data available in the geographical database (Figure 1, labels 1 and 2).

The insertion requires a two-step procedure based on the selection of a class of objects (e.g., city, park, temperature, polluting agent, etc.) and the further selection of a specific instance (e.g., the city *Mestre* or the polluting agent *mercury*). The selection of a given geometaphor causes the visualization of the object inside the 3D area (Figure 1, label 3).

The right panel of the interface (Figure 1, labels 4 and 5) contains the classes of operators that the user can apply to the objects. Metric, directional, topological and boolean operators can be applied if there are two geometaphors in the scene; different types of mathematical operators can be

applied if there are one, two or three geometaphors in the scene. The system guides the user, visualizing only the classes of operators that can be applied in a specific situation.

After the choice of the operator type, the user refines the query composition, dragging the geometaphors along the three cardinal directions, in the case of spatial operations, or selecting a specific operator, in the case of mathematical and boolean operations.

Figure 1 describes a topological query with two continuous variables that represent the distribution of mercury and lead on the territory and the topologic operator *overlap*. The user searches for a simultaneous presence of the two agents, expressing the search through the visual composition shown in Figure 1. Two geometaphors associated to the two agents are moved in the 3D area and overlapped to mean the contemporary presence of the two pollutants. The relation is also confirmed by the label dynamically visualized over the operands.

A *query* button, available in the lower part of the screen (Figure 1, label 6), triggers the query execution.

The visual environment offers also some additional functionalities for enhancing the usability of the system:

- the point of view can be modified by the user, dragging the mouse over the 3D scene;
- an automatic zooming mechanism monitors the variations of the point of view and adjusts the visualization, in order to include all the objects available in the 3D scene;
- contextual help messages are visualized on the top of the window (Figure 1, label 7) for suggesting to the user the actions needed to complete the query;
- a textual version of the query in a pseudo natural language is visualized on the lower panel of the interface (Figure 1, label 6) and is updated in case of changes in the 3D scene;
- an *info* button allows experienced users to see the translation of the visual query into the actual string passed to the GIS server;
- additional buttons for canceling or saving the current query for future use are implemented.

D. The Output Phase: The Visual Interface for Presenting Information to the User

The presentation of data is obtained through the integration of the Google Earth client, that is used as a front-end for visualizing survey data and the query output.

Concerning the survey data, the Google Earth client is triggered when the user selects a geometaphor in the visual query interface. The user can then navigate the raster and vectorial data superimposed to the Earth representation using the standard commands of Google Earth. Additional details can be obtained clicking over the objects displayed in the scene. The Google Earth client is also activated after query processing by the GIS server.

Figure 1, on the left, shows the result of the topological query formulated on the right part of the interface. The result of the elaboration by the GIS engine appears as a semitransparent raster map (Figure 1, label 8) that shows, for a rectangular portion of the Venice lagoon, the locations where there is a simultaneous presence of mercury and lead.

E. The Implementation Architecture

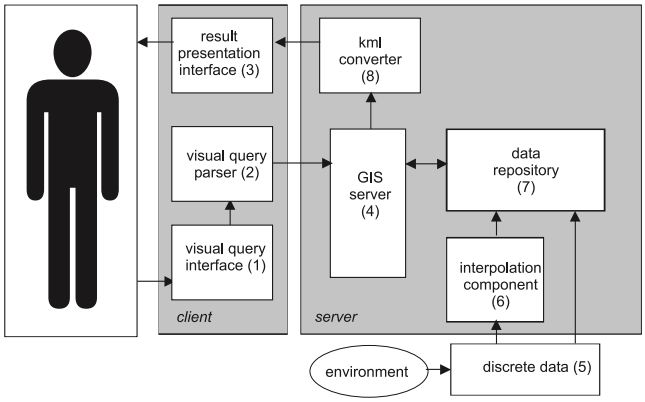


Fig. 2. The Implementation Architecture.

The implementation architecture (Figure 2) uses a number of components characterized by different technologies that have been integrated in order to comply with the initial requirements and to maximize the results in terms of the usability of the system. The result is available in the form of a web-based client-server application:

- the *visual query interface* component (1) has been implemented using a VRML based environment embedded in an XHTML layout, enhanced with CSS technology for presentation; the *visual query parser* (2) translates the visual query in a textual form and passes it to the GIS server;
- the *GIS server* engine (4) receives and elaborates the query, accessing the *data repository* (7); the geodata contained in such repository are both discrete and continuous ones; in the latter case the *interpolation component* (6) filters such data to obtain a continuous field. The result of the query is translated by the *KML converter* (8) and sent to the user interface for the presentation.
- the *result presentation interface* (3), based on the Google Earth client, receives the KML data and shows the result to the user.

The tools used for the development of the prototype include, on the client side, the Cortona SDK by Parallelgraphics [27] for the development of the 3D query environment and its integration with the XHTML/CSS interface and, on the server side, the ArcGis Server by ESRI [5] with spatial extensions for the management of continuous variables.

Concerning the choice of the GIS engine, a lot of time was dedicated to examine and compare the features of many commercial and open source products. Most of them were unable to process continuous data, which represented one of the initial requirements for our system. The exam of the literature didn't help us in this respect, because most of the research work about the management of continuous

georeferenced variables is focused on theoretical issues or on partially implemented systems. At the end, the survey led us to identify the ESRI GIS server as the most suitable for our purposes. The features of this commercial product include the possibility to derive continuous data from the interpolation of discrete elements and to manipulate them through a peculiar approach named *map algebra*.

V. Case Study: The Venice Lagoon

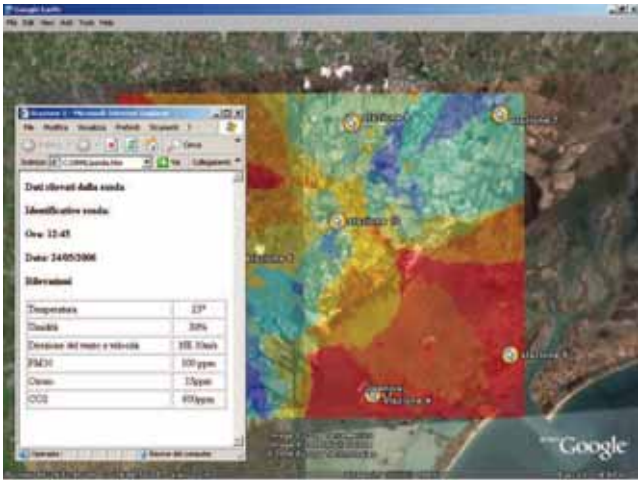


Fig. 3. Temperature Data and Survey Stations.

The collaboration with the Corila Consortium for the preservation of the Venice lagoon gave us the opportunity to focus on specific requirements and to direct the design and the implementation to the real user needs.

Currently, we cannot use the complete set of data collected by Corila, due to the ongoing redesign of *RIVELA*, the complex database hosting the survey data coming from the research units. The sample data that we used, however, are representative of the needs of the researchers, as stated by a survey of the Corila research program.

Because of the limits of the paper, we'll present only two examples that show how the integrated system may support the users in scenarios ranging from the analysis of the environmental data to the simulation of dangerous environmental situations. A third example, related to the same scenarios and focused on the composition of a spatial query with continuous variables and topological operators, has already been presented in Figure 1.

A. Example 1: Visualizing Discrete and Continuous Data from a Survey

The first example is focused on the access to survey data. In the current prototype this presentation is automatically triggered for visualization on the output interface when the user selects a given variable in the visual query interface. The semitransparent raster map superimposed to the sat image of Venice lagoon area represents the result of the interpolation on a temperature survey (Figure 3). The survey stations are labeled yellow landmarks; the user can click over the landmarks for displaying detailed information about the stations and data logged.

B. Example 2: Composing a Spatial 3D Query with Metric and Directional Operators

Fig. 4. A Query with Metric and Directional Operators.

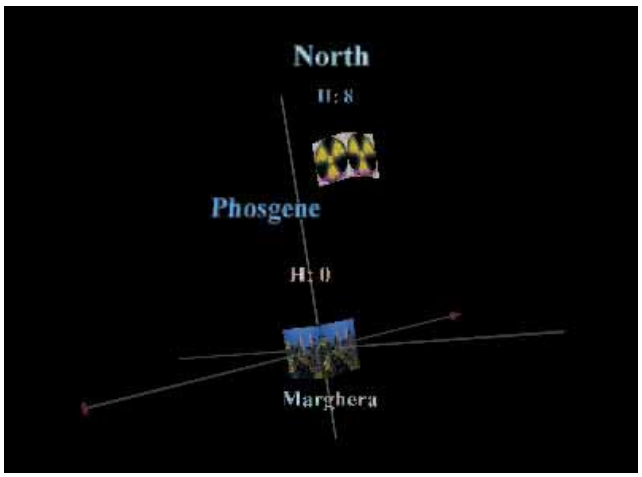


Fig. 5. A Query with Metric and Directional Operators - Zoomed View.

The example shows the composition of a query with both metric and directional operators. Both continuous and discrete variables are involved in such example. The underlying scenario is the simulation of a pollution numb originated by a leak of phosgene from the chemical plants that are situated near to the urban area of Marghera.

The responsible for the environmental safety of the Venice area needs to make some hypotheses in relation to the current direction of winds and to the height of the numb that represents a danger due to the polluting agent fall-out.

The user, on the basis of the current position of the polluting numb, asks to the system which parts of the numb can be considered as dangerous for the city of Marghera. As shown in

Fig. 6. A Query with Metric and Directional Operators: Presentation of the Results.

Figure 5, magnified for clarity, the geometaphors representing the city of Marghera and the phosgene numb are put in the 3D area.

The user, after the selection of the metric/topological operations type, moves the geometaphor representing the polluting numb along the North direction, then along the vertical axis to the height of 8 meters, that represents a security value for the fall-out.

Figure 6 visualizes a portion of the result, showing the part of the toxic numb (label 1) that represents a danger for the urban center of Marghera (label 2) in the context of the hypotheses formulated by the environmental expert.

VI. Conclusion

This paper has presented an integrated web-based architecture for enabling users that are not skilled in the GIS domain to access seamlessly georeferenced data and elaborate them in a visual environment. The prototype developed addresses the needs of the environmental research, because of the involvement in a specific project, but it is general enough to be used also in other domains.

The access to information is mediated by a set of visual interfaces that take care of the query and the presentation phases and permit the users to compose sophisticated queries without learning complex textual query languages.

The integration of the continuous variables in the visual metaphors for the query and the presentation of the results, is an additional feature of this work.

The prototype has been built according to the requirements of the users of the Corila Consortium and has been informally evaluated by a group of them, obtaining a first positive evaluation. The full implementation of the visual environment and its evaluation on the field is part of the ongoing development.

VII. Acknowledgments

I would like to acknowledge Santo Penna that contributed to the implementation of the prototype as part of his master thesis at the University Ca' Foscari of Venice. This work is supported by the Corila Consortium, in the context of a research project aimed at preserving the environment of the Venice lagoon.

References

[1] L. Paolino, F. Pittarello, M. Sebillo, G. Tortora, and G. Vitiello, "Webmgisql - a 3d visual environment for gis querying," in *Proceedings of International Conference on Visual Languages and Computing (VLC'03)*, 2003, pp. 294–299.

[2] V. DelFatto, L. Paolino, F. Pittarello, M. Sebillo, and G. Vitiello, "Webmgisql 3d iterating the design process passing through a usability study," in *Proceedings of HCI 2006 Engage, 20th BCS HCI Group Conference*, 2006.

[3] GoogleEarth, http://earth.google.com.

[4] L. Paolino, G. Tortora, M. Sebillo, G. Vitiello, and R. Laurini, "Phenomena: a visual query language for continuous fields," in *Proceedings of the 11th ACM international Symposium on Advances in Geographic information Systems (GIS '03)*. ACM Press, 2003, pp. 147–153.

[5] ESRI, http://www.esri.com.

[6] M. Egenhofer, "Spatial sql: A query and presentation language," *IEEE Transactions on Knowledge and Data Engineering*, vol. 6, no. 1, pp. 86–95, 1994.

[7] OpenGIS, "OpenGIS, Simple Features Specification for SQL - Revision 1.1," http://www.opengis.org/techno/specs/99-049.pdf, 1999.

[8] Y. C. Lee and F. L. Chin, "An iconic query language for topological relationships in gis," *International Journal of GIS*, vol. 9, no. 1, pp. 25–46, 1995.

[9] C. Traynor, "Putting power in the hands of end users: a study of programming by demonstration, with an application to geographical information systems," in *Proceedings of Conference on Human Factors in computing systems (CHI 98)*, 1998, pp. 68–69.

[10] B. Meyer, "Beyond icons: Towards new metaphors for visual query languages for spatial information systems," in *Proceedings of of International Workshop on Interfaces to Database Systems (IDS 92)*, 1992, pp. 113–135.

[11] D. Calcinelli and M. Mainguenaud, "Cigales, a visual query language for a geographical information system: the user interface," *Journal of Visual Languages and Computing*, vol. 5, no. 2, pp. 113–132, 1994.

[12] M. Egenhofer, "Query processing in spatial query by sketch," *Journal of Visual Languages and Computing*, vol. 8, no. 4, pp. 403–424, 1997.

[13] M. A. A. Portier and C. Bonhomme, "A high level visual language for spatial data management," in *Proceedings of Visual'99*, 1999, pp. 325–332.

[14] MapGuide, http://www.autodesk.com/mapguide/.

[15] MapServer, http://mapserver.gis.umn.edu/.

[16] VRML, "Virtual Reality Modeling Language Specification. ISO/IEC DIS 14772-1," http://www.web3d.org/x3d/specifications/vrml/ISO-IEC-14772-VRML97/.

[17] X3D, "X3D Specification. ISO/IEC 19775:2004," http://www.web3d.org/x3d/specifications/.

[18] GeoVRML, http://www.geovrml.org.

[19] X3DEarth, "X3D Earth Working Group, Web3D Consortium," http://www.web3d.org/x3d/workgroups/x3d-earth/.

[20] Web3D, "Web3D Consortium," http://www.web3d.org/.

[21] T. Smith and V. Lakshmanan, "Utilizing google earth as a gis platform for weather applications," in *Proceedings of the 22nd Int. Conf. on Interactive Inf. Processing Systems for Meteorology, Oceanography, and Hydrology*, 2006.

[22] W. E. Kilby and C. E. Kilby, "Aster multispectral satellite imagery and product coverage, british columbia - phase 2," in *Geological Fieldwork 2006*. British Columbia Ministry of Energy, Mines and Petroleum Resources and Geoscience BC, 2007, pp. 315–318.

[23] Corila, http://www.corila.it.

[24] Collada, http://www.khronos.org/collada/.

[25] M. Sebillo, G. Tortora, and G. Vitiello, "The metaphor gis query language," *Journal of Visual Languages and Computing*, vol. 11, no. 4, pp. 439–454, 2000.

[26] V. DelFatto, L. Paolino, and F. Pittarello, "Representing topological relationships by using 3d objects: an empirical survey," in *Proceedings of the International Workshop on Visual Languages and Computing (VLC 2006)*, 2006.

[27] Parallelgraphics, http://www.parallelgraphics.com.

Architectural solution for interoperable content and DRM on multichannel distribution

Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, Davide Rogai
DISIT-DSI, Distributed Systems and Internet Technology Lab
Dipartimento di Sistemi e Informatica, Università degli Studi di Firenze
Via S. Marta, 3, 50139 Firenze, Italy
www.AXMEDIS.org, nesi@dsi.unifi.it, http://www.dsi.unifi.it/~nesi
Office: +39-055-4796523, Fax: +39-055-4796363

Abstract

According to the new trends, final users are interested in acquiring content from different distribution channels and in using it on different devices/tools, while interchanging the content among devices and tools. This paper depicts the main scenario and reports a unified architecture supporting multichannel distribution and content interoperability. The paper presents the results produced in this area within AXMEDIS IST FP6, a research and development integrated project (Automating Production of Cross Media Content for Multi-channel Distribution) partially funded by the European Commission.

Keywords*: e-commerce, multichannel distribution, interoperable DRM, iDRM, MPEG-21.*

1. Introduction

Final users are requesting more functionalities from content and content distributors. Content distribution services are grounded on a large set of technologies for content formats, connections and digital transmission, content processing and adaptation, content protection, and for Digital Rights Management, DRM. See for a general overview [Koushanfar et al., 2005]. In terms of content formats and DRM, many solutions are available on the market such as Apple i-Tunes, Microsoft Windows Media DRM, Adobe DRM, Intertrust, and OMA (Open Mobile Alliance, [Iannella, 2002]), and others see [Lin et al., 2005]. Most of these solutions have relevant limitations on the content formats and interoperability, since they support only a limited number of media formats, devices and distribution channels. Others have DRM mechanisms which allow exploiting/controlling only a limited number of rights on the digital content and therefore they allow establishing only a limited number of business models. Despite the large number of offered solutions, none of them seem to be generally accepted.

Business and final users are becoming more and more interested in using more complex digital content (e.g., interactive content with several kinds of media inside: audio, video, games, document, etc., SMIL, HTML, SCORM, etc.) [Bellini et al. 2006]. They expect to receive this kind of content from different distribution channels and to use it on different devices/tools, according to different business models (e.g., renting, pay per play, pay per use, all you can eat, passing the content to friends while receiving some bonus). The present state of the art is dominated by the lack of interoperability among different: content formats and DRM solutions (different licenses, protocols, protection models, etc.), distribution channels, devices and tools, accounting information, etc.

Attempts to solve this situation have been the creation of open DRM standards; such as those proposed by MPEG-21 [Wang et al., 2005], and/or by OMA [Iannella, 2002]. On the other hand, market standards such as i-Tunes, Microsoft DRM, can hardly be substituted by these new open standard solutions which suffer from a lack of interoperability among them. This is due to the assumption each of them make as to become the universally "Adopted Standard". Among the emerging standards, the MPEG-21 is likely to be the most promising, it can deal with: license (formalized in some REL, Right Expression Language), event reporting for logging the actions/rights performed/exploited on the players, and the formats for content packaging called Digital Item Declaration (DID) [MPEG-21 DID] and the protection information called Intellectual Property Management and Protection (IPMP).

In this paper, the focus is on the problems lying behind any enabling of interoperability among multiple distribution channels. The studies and the solution reported in this paper have been worked out for AXMEDIS (Automating Production of Cross Media Content for Multi-channel Distribution) Integrated Project FP6 of the European Commission (http://www.axmedis.org), [Bellini and Nesi, 2005]. The

AXMEDIS consortium consists of: European digital content producers, integrators, aggregators, and distributors, collecting societies, together with information technology companies and research groups.

2. Digital Rights Management Scenarios

At present, there is a large number of content formats ranging from basic digital resources (documents, video, images, audio, multimedia, etc.) to integrated content packages such as: MPEG-21 [MPEG-21 DID], SCORM [Mourad et al., 2005], OMA. Packages can wrap digital resources with other related information (e.g., metadata, identification codes) so as to make them ready for delivery. Such solutions are more flexible, if compared with proprietary solutions where the DRM can be applied only to the single resources. The typical content production and distribution scenario, which synthesizes the most relevant phases from content packaging to content distribution, is shown in Figure 1.

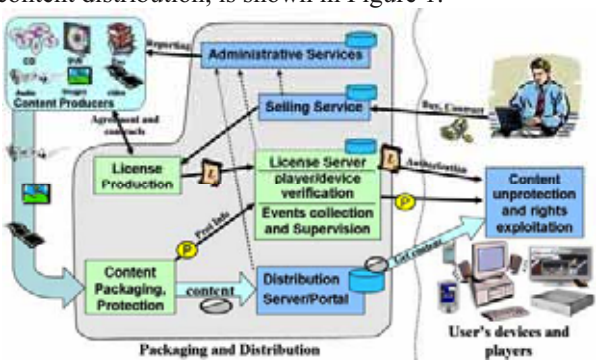


Fig.1–A simplification of a typical DRM based Content Production and Distribution solution.

The distributor establishes a contract/agreement with the Content Producer (this is a simplification). The distributor may protect the produced content package to keep a certain level of control about the exploitation of rights. The content distributor may make business by allowing consumers to access the content through specific licenses. The latter ones describe the set of rights granted to the consumers (the rights are the actions that can be performed on the content, e.g., play, print, copy, etc.). The digital resources are packaged in order to get a package for the distribution which is protected by using some algorithms. The protection information (P) has to reach the final user in order to allow him to unprotect/open digital resource and/or the package, when the user is authorized to do it. This kind of information is typically called Protection Information or IPMP as in MPEG-21. Both license and protection information is needed to exploit the rights on the content. They can reach the content via different paths or together. If they arrive together, the protocol from the License Server to the Player/device results to be simplified; and license production cannot be delegated to third parties.

Once the content package has reached the consumer's player, he/she may exploit the acquired rights, e.g., "a play". In order to permit the performance of authorized actions only, the player/device has to verify if it can be authorized according to the license, for example contacting the License Server. The License Server processes the license database to verify if the authorization can be granted to the player. In alternative, a copy of the license can be sent to the player. Every time a right is exploited, the involved distributors/ producers may need to have back an evidence for their Administrative Services (e.g., for sending the bill). This information can be easily recovered in solutions that constrain the player/device to contact the License Server for each grant authorization. In those cases, a sort of Action Log record (Event Report in MPEG-21) can be produced by the player/device. The information on the user Action can be used for recovering statistics about content usage, to adjust the service, to make market analysis, etc.

According to the above short presentation of the typical scenario, it can be stated that: content formats, license, protection information, and action log, strongly influence the architecture and protocols of the distribution channel, thus creating differences which may limit the interoperability among the different channels and DRM models. The difficulties are mainly due to the differences among the different DRM solutions and devices, to the information transferred among the involved entities and to the related protocols used to perform authentication, granting the authorization, accessing to the license, accessing to the protection information, requesting license production, reporting actions log, etc. All these details are strongly relevant for the interoperability among different DRM distribution channels even if they are based on different tools, formats and protocols.

3 AXMEDIS Distribution Services

A description of the AXMEDIS architecture can be recovered from [Bellini et al, 2005] and the full specification from the AXMEDIS portal. The following AXMEDIS architecture for multichannel distribution of interoperable content and DRM is supported by a number of service tools for: Authorizing and Managing Licenses, Supervising and controlling activities, as well as for User and Tool Authentication. The models and tools mentioned in the scenarios can be adopted by any actor of the value chain without any limitation; therefore "producer" and "distributor" have to be intended as roles

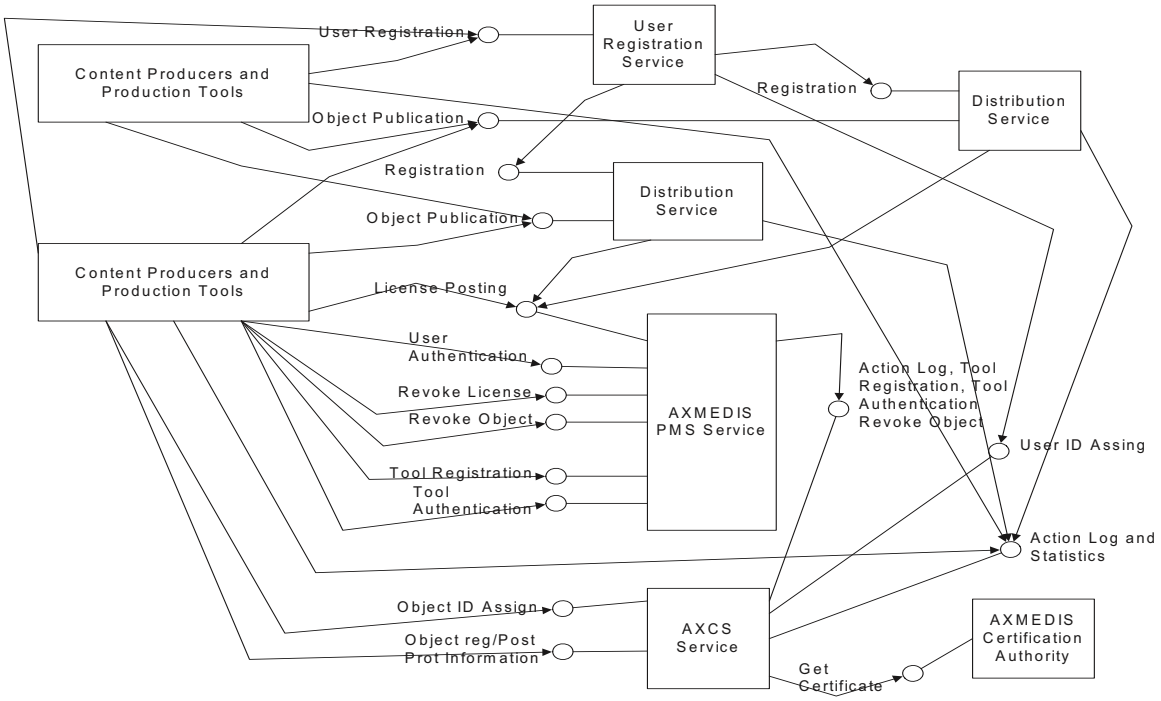


Figure 2 – DRM Services, Producer's point of view and its tools, a Business to Business case from Producers to Distributors. (Please note that for one of the Producer we have not depicted all possible connections)

rather than effective entities. The tools mentioned above refer to generic functionalities, as described below.

The main rationales which led to the definition of the AXMEDIS tools and web services are reported along with the presentation of the most relevant scenarios for B2B and B2C, respectively.

Authorizing and Managing License Service

The AXMEDIS PMS (Protection Manager Support) Service is mainly a License Server which allows granting the authorization to the players on the basis of the licenses. It is the primary access for a large set of capabilities exploited by the content producers such as: license posting, verification, revoking; revoke of objects, tools and users, etc. Some of these services are delegated to the AXCS (AXMEDIS Certifier and Supervisor). Most of the AXCS services are conveniently made accessible by the PMS since it establishes a secure communication channel with the Production Tools and Final User Tools. All the above services have information (such as certificates, protection information, grants, action logs, etc.) that should pass via a secure channel in order to avoid simple sniffing attacks.

Supervising and Control Service

The AXCS (AXMEDIS Certifier and Supervisor) Service provides a number of services for producers and distributors such as: object ID assignment, user ID assignment, access to Action Log reports and statistics about events and user usage.

In addition, the AXCS provides a number of services to the PMS such as: storing of the actions logs (grants, posting of licenses, and other events), tool registration and authentication, posting of protection information, revoke of object, tools and users. The AXCS collects and keeps the information regarding the registered objects, users, devices, etc., and therefore it allows the management of black lists. It also stores the Protection Information of each protected object, resource and the list of actions performed on them, the so called Action Log database. Each Action Log describes an action performed on a given content/resource, by a given user, on the basis of a given license, etc. The whole set of Action Logs allows to produce the reporting and the production of statistics.

Registration Portal and Service

It is used to register users (final and business) in collaboration with the AXCS and the AXMEDIS Certification Authority. The registration portal allows collecting information about the users, regardless of the assignment of their unique ID and provides certificates. The PMS and AXCS does not have any personal and private information about the user and see them only via their unique ID.

3.1 Business to Business Scenarios

In Figure 2, the point of view of the Producers with their production tools and two related Distributors is presented. The Producer typically performs the

following operations as to the above mentioned services and tools.

User Registration for business users as content producers and distributors. The registration provides them a unique ID and a certificate released by a Certificate Authority.

Tool Registration to register the tools used by the producer for producing content object packages, licenses, and to revoke licenses. In general, different users may use the same tool, so that the certificate should be managed at the User's level, whether this distinction is needed by the control and/or revocation process and tools.

User and Tool Authentication operations performed to verify if the connection is established with the right user and tool. A simplification can be performed in the case of final users if User and Tool are considered a unique element, e.g., the Device. The authentication based on the certificate allows also storing the keys to establish a protected channel for exchanging sensible information.

Object Publication, Object ID Assignment, Object Registration/Posting of Protection Information. In most cases, the producer asks for the production of protected objects to the distributor, moving the not protected objects to the latter. In this way, the producer delegates also part of the control about the exploited rights. In other business cases, the activity of content protection is performed by some intermediate producer. The scenario reported in Figure 3 is more complete and offers several advantages. The protection is directly performed by the producer who can distribute the protected objects to more distributors (even with different DRM and protection models).

License Posting. The license provided by the Producer defines the rights that he has deployed to a distributor (also called distribution or mother license). These licenses typically include the possibility of (i) object reselling according to different business models, (ii) object adaptation to make the object suitable for different distribution channels and devices, (iii) object encapsulation, (iv) object interrupt for advertising, etc. Therefore, the licenses, that the Distributor produces (to enable the rights' exploitation for the final users), depend on the licenses the Distributor bought from the Producer. Once the license is posted on the PMS (License Server), the corresponding action should be communicated to the AXCS in the form of Action Log (reporting the event).

Revoke License. This operation can be performed to terminate the contract that the Producer has with a Distributor/User. The License Server, such as the PMS Service, has to manage a black list of revoked licenses.

Revoke Object. This operation allows invalidating a specific object by including its unique ID in a black list. This action disables the distribution of the Protection Information related to that object.

Action Log and Statistics Accesses allows getting detailed reporting and statistic information about the exploitation of rights and other events. The Producer is interested and it is authorized to have the detailed information about the exploited rights for the produced objects. The detailed reporting refers to the Action Log records which were produced according to some produced objects. The Action Log record for each exploited right includes also: content owner ID, Producer ID, Distributor ID, Object ID, User ID, right, device ID, date and time, etc. Some kinds of information (such as the User ID) cannot be delivered to the Producer, in order to respect the privacy.

3.2 Business to Consumer Scenarios

Figure 3 depicts the point of view of a DRM Client and Player with respect to the distribution of content performed by two Distributors and other DRM services. In this case, the activities are performed by means of the so called DRM Client and Player that includes the device/tool which the user performs all actions from. Any User may have/use one or more devices/tools supporting interoperability. The DRM Client/player typically performs the following operations.

User Registration and Authentication to register the user on the circuit of the Distributor(s). The Distributor needs to identify the User so as to provide him with the bill or to get the payment in advance (both models are possible). The second case allows the User to maintain anonymity with respect to the distributor, yet limiting the flexibility of the business model that can be agreed among these business actors. The registration in this case leads to produce a unique ID for the User and to deliver a Certificate in order to establish protected communication and authentications.

Tool Registration and Authentication to identify the tool used by the final user and to certify that they are trusting tools. The distinction between user and tool allows managing multiple users exploiting the same device for accessing content and exploiting rights such as in archives, schools, content factories, families, etc.

Query (browse, selection), Get Object and Set Contract. The final user may go to any Distributor to make some query for selecting the content objects. In order to enjoy the content object, some rights have to be acquired by the user setting a contract with the Distributor according to some business model. In any case, according to the established contract, the Distributor has to issue a corresponding License, and to post it on the License Server (PMS). Once the license is posted, the corresponding action may be logged into the AXCS in the form of Action Log record (reporting the event).

Do Payment, Dispose Payment, Delegate Get Payment. According to the business model, the Distributor starts the process to obtain the payment from the user. This is typically performed by delegating some external Payment and Compensation Service to get the payment from the User (Delegate Get Payment). The Distributor periodically verifies the status of the payment to enable the service to its client. A similar service can be used by the Colleting Societies to dispose the revenue sharing among the value chain actors involved in the rights sold by the Distributor (Dispose Payment).

Grant Authorization, Content usage, rights' exploitation. Finally, the user who obtained the content can try to exploit its rights, for instance making play on its DRM-enforced player. This action implies to contact the PMS to establish a secure connection in order to get the Grant Authorization. Once the grant is provided, the corresponding action is communicated to the AXCS in the form of Action Log record.

4 Unique Backoffice DRM Interoperability

By using the AXMEDIS framework and tools, it is possible to set up a large set of different architectures and configurations, which allows managing at the same time and in harmonized manner several different distribution channels, while supporting also production and distribution on demand. The AXMEDIS solution allows realizing multichannel interoperable channels supporting both MPEG-21 and OMA for distributing content via satellite data broadcast towards PC and/or STB, via Internet towards IPTV and/or STBs, via mobile channel to moves, etc., as demonstrated by AXMEDIS demonstrators (www.axmedis.org).

The AXMEDIS framework allows automating the activities of production and distribution by scripting these activities and putting them in execution on the Content Processing GRID platform: the so-called AXMEDIS Content Processing. This solution allows setting up different interoperable distribution channels with a unified back-office. For example, (1) AXMEDIS/MPEG-21 content format distribution towards PC and STBs exploiting functionalities of both the AXCS and the AXMEDIS PMS; (2) OMA distribution according creating OMA packages and accessing to OMA license server according to the so called Separate Delivery; (3) a channel can be based on Microsoft Windows Media DRM, creating MS content

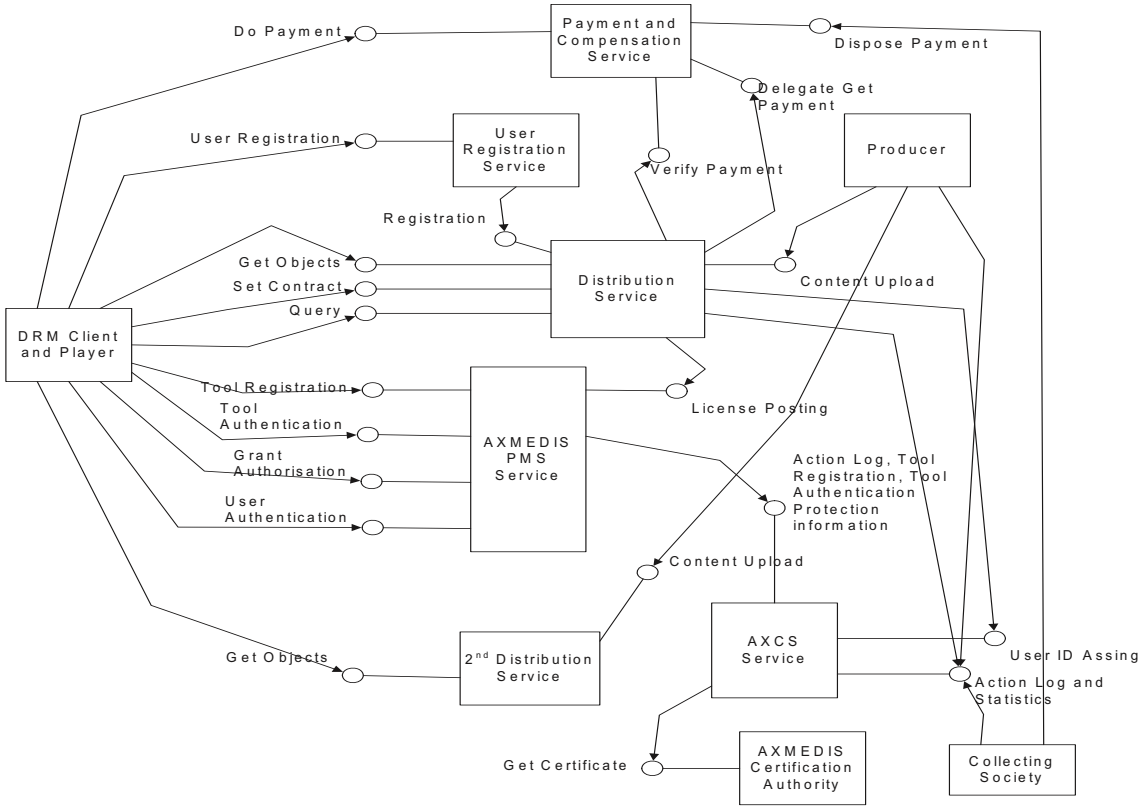


Figure 3 – Exploiting DRM Services, point of view of final user with its DRM client and player, a Business to Consumer Scenario. (please note that, for one of the distributor not all the connections have been reported).

and posting licenses on related License Server. These channels may have a unified or separate portal with services for content promotion and distribution in streaming and/or download. On the basis of the AXMEDIS back office tools and Rules [Bellini, Bruno and Nesi, 2006] specific services/adapters can be set up by intermediate business partners for transcoding content packages and/or licenses among different non interoperable channels, so as to make them interoperable (for example converting MEPG-21 REL license into OMA DRM license). The above mentioned rules may include adaptation of content, of metadata and of license. In some cases the content cannot maintain the same functionalities in all platforms and therefore some parts of the license have no sense for some devices.

5. Conclusions

This paper has presented the architecture of AXMEDIS for multichannel interoperable content production, protection and distribution. The architecture has been defined as a result of an analysis of a large set of different content models, distribution solutions and DRM architectures, and supports the harmonization of the MEPG-21 with OMA and Windows DRM. The analysis has been performed to define common bases for their interoperability and to define a unified architecture supporting different models and solutions of interoperability, thus supporting and facilitating the transition from the present status and solutions to a possible unified and standardized model.

The paper reported an overview of the AXMEDIS architecture and tools which are a solution to automate, accelerate and restructure production and protection processes. AXMEDIS architecture supports multichannel and interoperability by allowing to set up services for the migration of content from different formats (by setting up some adaptation services), among different channels and from different DRM formats [Bellini, Bruno and Nesi, 2006]. Such different channels may have their specific devices and business models and their specific DRM solutions. This paper described only a part of the whole AXMEDIS framework and architecture which is addressing many other problems and critical points. AXMEDIS solution is mainly based on MPEG-21 and OMA models and it provides and stimulates the usage and the exploitation of the developed features for creating many AXMEDIS compliant tools and solutions, while making the core aspects and solution accessible in the form of AXMEDIS Framework. The full specification of AXMEDIS tools and WSs can be accessed on the AXMEDIS portal.

6. Acknowledgements

The authors would like to thank all the AXMEDIS project partners (ANSC, AFI, EUTELSAT, Giunti ILABS, HP Italy, FHGIGD, TISCALI, XIM, ACIT, EXITECH, University of Leeds, University of Cataluña, University of Reading, BBC, TEO, ELION, TIM, SDAE, etc.), the Expert-User-Group and all affiliated members for their contributions, supports and collaborations. The authors would also like to express their thanks to the European Commission IST DG-INFSO FP6 for partially funding AXMEDIS.

7. References

1. Bellini P., Bruno I., Nesi P., ``A language and architecture for automating multimedia content production on grid", Proc. of the IEEE Int. Conf. on Multimedia & Expo (ICME 2006), IEEE Press, Toronto, Canada, 9-12 July, 2006.
2. Bellini P., Nesi P., “An architecture of Automating Production of Cross Media Content for Multi-channel Distribution”, Proc. of the first Int. Conf. on Automated Production of Cross Media Content for Multi-channel Distribution, 30 Nov - 2 Dec 2005, Florence, Italy, IEEE Computer Society press.
3. Bellini P., Nesi P., Ortimini L., Rogai D., Vallotti A., ``Model and usage of a core module for AXMEDIS/MPEG21 content manipulation tools", Proc. of the IEEE Int. Conf. on Multimedia & Expo (ICME 2006), IEEE Press, Canada, 9-12 July, 2006.
4. Iannella, R., “Open Digital Rights Language (ODRL)”, Version 1.1 W3C Note, 19 September 2002, http://www.w3.org/TR/odrl .
5. Koushanfar, F., Inki Hong, Miodrag Potkonjak, “Behavioral synthesis techniques for intellectual property protection”, ACM Trans. on Design Automation of Electronic Systems (TODAES), Vol.10, Issue 3, ACM Press, July 2005.
6. Lin, E.T., Eskicioglu, A.M., Lagendijk, R.L., Delp, E.J., “Advances in Digital Video Content Protection”, Proceedings of the IEEE, Vol.93, N.1, pp.171-183, January 2005,
7. Mourad, M., Hnaley, G.L., Sperling, B.B., Gunther, J., “Toward an Electronic Marketplace for Higher Education”, Computer of IEEE, pp.58-67, June 2005.
8. MPEG Group MPEG-21 DID, “Introducing MPEG-21 DID, Digital Item Declaration”, www.chiariglione.org/mpeg/technologies/mp21-did/
9. Wang, X., De Martini, T., Wragg, B., Paramasivam M., Barlas C., "The MPEG-21 rights expression language and rights data dictionary", IEEE Transactions on Multimedia, Vol.7, N.3, pp.408-417, 2005.

Exploiting MPEG-21 File Format for cross media content

Pierfrancesco Bellini, Paolo Nesi, Davide Rogai

DSI-DISIT, Department of Systems and Informatics, University of Florence
Via S. Marta, 3 - 50139 Florence, Italy, nesi@dsi.unifi.it, http://www.dsi.unifi.it/~nesi

Abstract

Cross media and hypermedia content formats and related tools have to take into account aspects related to the protection and management of intellectual property right along the value chain from content production to content distribution up to the final user. In this view, the relevance of MPEG-21 standard is growing also for the B2B manipulation of digital objects, using the nested levels of protection in the B2B exchanges. When studying how to map the produced nested Digital Items on the File Format standard several difficulties have been faced since the MPEG-21 have limitations in managing protected nested levels that are very relevant for the B2B activities. The paper discusses them and proposes solutions that have been realized in the AXMEDIS models and tools. AXMEDIS is a research and development integrated project of the European Commission.

Keywords*: MPEG-21, content distribution, File format, nested protecting levels, B2B packaging. .*

1 Introduction

The solutions for digital content distribution and e-commerce are mainly based on the state of the art of multimedia content modeling, packaging, protection and distribution. At present, there exists a large number of content formats ranging from basic digital resources (documents, videos, images, audio, multimedia, etc.) to integrated content packages such as: MPEG-21 ISO [4], WEDELMUSIC [2], SCORM [11], OMA (Open Mobile Alliance), TV-AnyTime Forum [7], etc. These integrated content formats try to wrap different kinds of digital resources/files in a container/package with their related information (e.g., content metadata and descriptors, relationships among resources, etc.) and they make such resources ready for delivery (streaming and/or downloading), in plain (clear-text) and/or protected forms. In fact, some of the above mentioned solutions are enabling a large range of business and transaction models and provide some integrated DRM (Digital Rights Management) solutions to cope with Intellectually Property Right (IPR), such as those based on MPEG-21 REL (MPEG Rights Expression Language) [15], OMA ODRL (Open Digital Rights Language) [8], and others [10], [9].

In AXMEDIS (Automating Production of Cross Media Content for Multi-channel Distribution) [1], the MPEG-21 DI (Digital Item) has been adopted as a basis to realize a platform to produce, protect and distribute cross-media content in XML and binary formats. The AXMEDIS Data Model has conceived the so called AXMEDIS Object (the unit of distribution) as an MPEG-21 DI, with specific structuring, metadata, descriptors and capabilities for nesting levels of object composition and protection. AXMEDIS supports both DI representation XML files according to DIDL (Digital Item Description Language) and ISO (International Standard Organization) media File Format [6], [5], [14]. The work presented in this paper has been developed in AXMEDIS which is a large research and development Integrated Project FP6 of the European Commission DG INFSO (http://www.axmedis.org), [3].

This paper discusses an experience in using MPEG-21 Part 9 [14] to represents complex digital items, much more complex objects with respect to those presented as example in the standard. DIs. The issues are related to the above mentioned needs where MPEG-21 IPMP (Intellectual Property Management and Protection) is used for protecting sub-parts (DIDL Item elements). The target of the protection can be an Item element which includes protected Item elements. In this paper, a solution for representing such DIs in terms of File Format (FF) has been outlined. This paper aims at explaining about how the MPEG-21 Standard (Part 9) can be used (extended in terms of semantics) to model complex objects by using the solutions proposed in AXMEDIS. The solutions proposed impacts on the realization of DIs representing multimedia presentations, based on HTML and/or SMIL (as presentation models) and on digital resources and links and nested DIs [3].

2 MPEG-21 File format and protection issues

The standardization of the file format of MPEG-21 has basically reused the ISO media standard produced for

MPEG-4 [14]. This choice produced a very simple specification for which the MPEG-21 description (DIDL [12], IPMP [13], etc.) has to be included in the XMLBox inside a MetaBox as in the example of Figure 1.

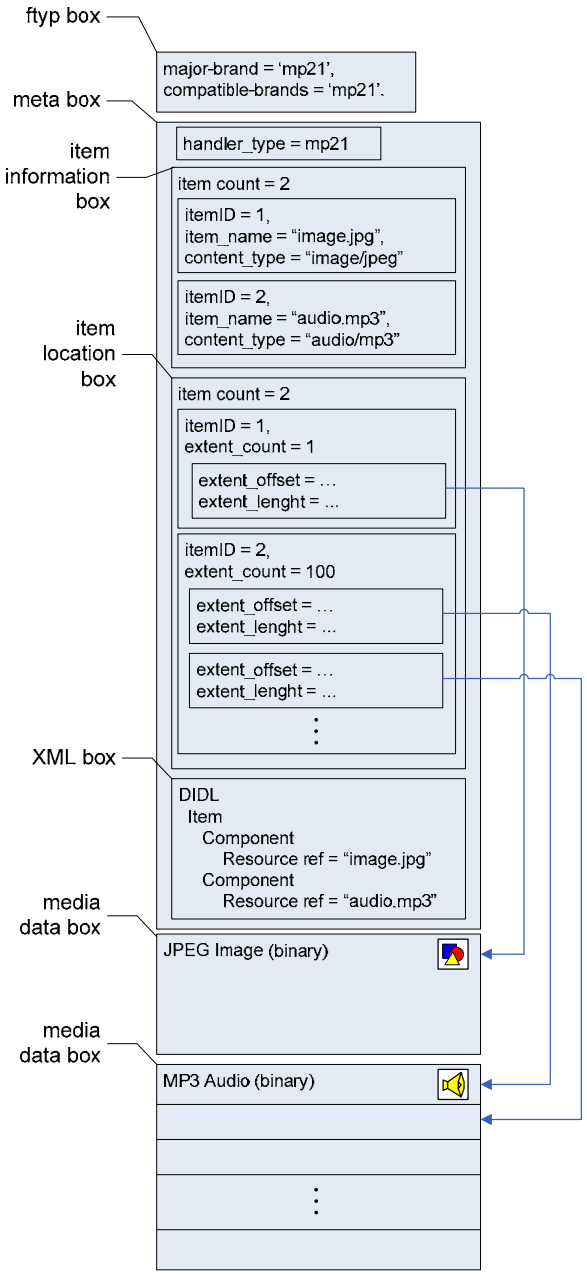


Figure 1 – File format example

In that manner, a placeholder for MPEG-21 based description has been reserved; media components can be located in proper MediaDataBox and referred by the XML structure using references. A DI that includes an image and an audio file can be represented in MPEG-21 FF as depicted in Figure 1. Please note that DIDL structure refers to "*image.jpg*" and "*audio.mp3*": the proposed semantics requires a player to resolve those references by accessing to the proper MediaDataBox. For that reason the ItemInformationBox is populated with the media in order to be accessed by the player (i.e., loaded at the opening). The ItemInformationBox is linked with the ItemLocationBox on the basis of the "itemID", the second one represents a look-up table for selecting the file portion that includes a given resource.

The XML Box sketched in Figure 1 is reported in the following box with more details.

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns:xi="http://www.w3.org/2001/XInclude"
xmlns:dii="urn:mpeg:mpeg21:2002:01-DII-NS"
xmlns="urn:mpeg:mpeg21:2002:02-DIDL-NS"
xmlns:ipmpdidl="urn:mpeg:mpeg21:2004:01-IPMPDIDL-NS"
xmlns:ipmpinfo="urn:mpeg:mpeg21:2004:01-IPMPINFO-NS"
xmlns:dc="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<Item>
    <Descriptor>
        <Statement mimeType="text/xml">

    <dii:Identifier>urn:axmedis:00000:obj:…axoid…</dii:Identifier>
        </Statement>
    </Descriptor>
    <Descriptor>
        <Statement mimeType="text/xml">
        <AXInfo xmlns="urn:axmedis:01">
            … missing B2B related metadata…
        <AXInfo>
        </Statement>
    </Descriptor>
    <Descriptor>
        <Statement mimeType="text/xml">
        <dc:Description >
            … missing DublinCore metadata…
        </Description>
        </Statement>
    </Descriptor>
    <Component id="poster">
        <Resource mimeType="image/jpeg" ref="image.jpg"/>
    </Component>
    <Component id="music">
        <Resource mimeType="audio/mp3" ref="audio.mp3"/>
    </Component>
</Item>
</DIDL>
```

The advantage of producing a MPEG-21 file in the File Format (FF) version, and not as a XML file, allows supporting interoperability and the possibility of including media resources in their original binary format. The solution only based on XML (see above) may refer at outside resources, losing in this manner, the possibility of delivering the collection of files as a unique "package". An alternative can be to embed resources into XML by encoding them in base64, enlarging in that manner the size of 30%. The advantage of the FF based solution consists in the fact that resources are directly accessed, even without the need of processing XML, since the information which is needed to access and to render a resource is the mime-type (to select the proper parser/renderer) and

the location to begin reading them and their end in terms of offsets inside the file. This data is in the FF MetaBox.

When everything is in clear-text/XML simple actions are required to produce a FF DI: to ideally detach the resource from the structure and substitute with proper references, then to write the XML and the media in separate portions of information.

2.1 Representing protected DIs

When DI part is protected, problems arises on how to represent content in the FF. MPEG-21 standard defined IPMP Components as the protection elements of a DIDL-based structure in order to protect a DI element the proper IPMP element has to replace the corresponding DIDL [13].

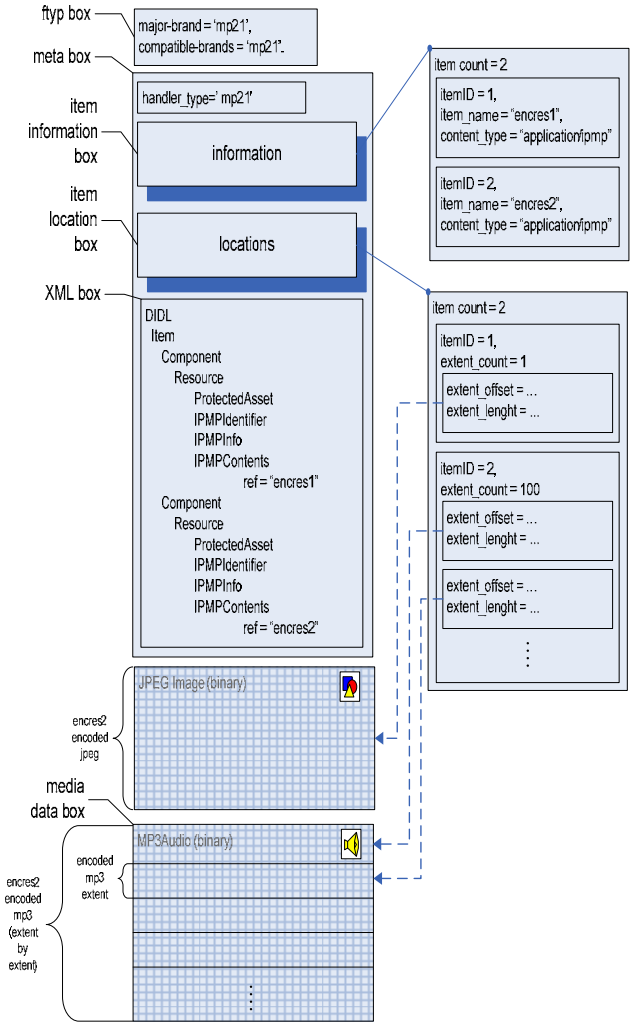


Figure 2 – Simple protection (only digital resources)

In Figure 2, the result of protecting the resources that have been embedded in the MPEG-21 file is shown. In this case, all the information is contained in the XMLBox and the player accesses at the resource just after it acquires how to process (e.g., unprotect) it with the proper IPMPTool (described in the IPMP Info). A variation is also shown about how to encode the digital resource: using the extents of the clear-text resource, the encoded one can be divided into extents in order to allow on-line unprotection and rendering mechanisms. In Figure 2, resources have been represented in the XML Box with the proper tag ProtectedAsset which can be embedded inside a Resource in order to notice that the content inside have to be processed by a suitable IPMP Tool before obtaining the clear-text digital resource according to the mime-type.

Problems arise when the protection is applied on other elements together with the digital resources. In AXMEDIS, the DIDL Item is the basic distribution element and in this case the complete Item content has to be protected by replacing it with an IPMP Item. A solution can be to protect the whole XML element with the embedded digital resources that have to be included "by value" and encoded with base64.

A solution can be realized by performing a two-step protection process (see Figure 3): first encoding the digital resources as in the first case (see Figure 2), then protecting the Item element, that is now linked to the digital resource with references, in the corresponding IPMP Item. In this last solution, the accessibility of digital resource is preserved, and a box-like protection is performed, thus protecting metadata and other elements with related advantages for protecting the enriched metadata. Please note that the IPMPInfo element has to be defined for the Item elements and for the ProtectedAsset elements.

The solution proposed in Figure 3 appears to be a good compromise between box protection and accessibility of the resources. In any case, this solution is still incomplete since does not consider the case of nested Digital Items, where an Item element contains other Item elements, and different protections are applied to each level. That case is the most promising use case for B2B transactions in which integrators and distributors are interested in repackaging content in nested level imposing their protection model and protecting at the same time their added value (e.g., presentation information, metadata, gluing information, etc.). On this regard, Figure 3 shows that protected content is directly included in the XML structure, this is possible by implicitly encoding the

result of the IPMPTool processing in base64 (that ensures output is a set of text characters that can be included in a XML text node) as depicted by the following XML Box extract.

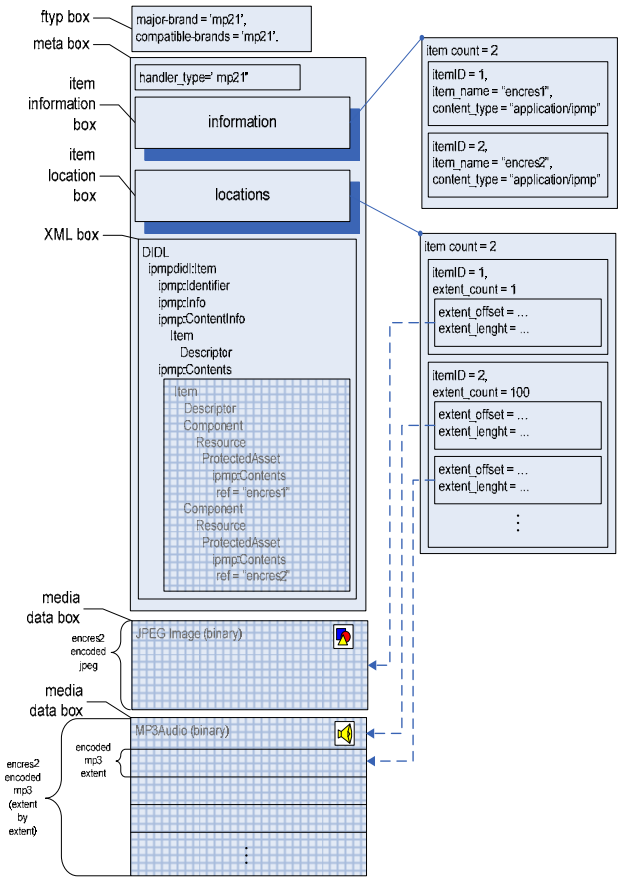


Figure 3 – Item element protection (two-steps protection, resources plus IPMPcontents)

Please note that the IPMP section into the XML box contain the protection information regarding the protected segment into the XML box itself.

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns:xi="http://www.w3.org/2001/XInclude"
xmlns:dii="urn:mpeg:mpeg21:2002:01-DII-NS"
xmlns="urn:mpeg:mpeg21:2002:02-DIDL-NS"
xmlns:ipmpdidl="urn:mpeg:mpeg21:2004:01-IPMPDIDL-NS"
xmlns:ipmpinfo="urn:mpeg:mpeg21:2004:01-IPMPINFO-NS"
xmlns:dc="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<ipmpdidl:Item>
      <ipmpdidl:Identifier>
      <dii:Identifier>urn:axmedis:00000:obj:…axoid…</dii:Identifier>
      <ipmpdidl:Info>
            … omitted IPMPTool information …
      </ipmpdidl:Info>
      <ipmpdidl:ContentInfo>
            … omitted Content metadata …
      </ipmpdidl:ContentInfo>
      <ipmpdidl:Contents>
/9j/4AAQSkZJRgABAQEAyADIAAD/2wBDAAUDBAQEAwUEBAQFBQ
UGBwwIBwcHBw8LCwkMEQ8SEhEPERETFhwXExQaFRERGCEYGh
0dHx8fExciJCIeJBweHx7/2wBDAQUFBQcGBw4ICA4eFBEUHh4eHh4
eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4e
Hh4eHh7/wAARCAEKAMgDASIAAhEBAxEB/8QAHAAAAQUBAQEAA
AAAAAAAAAAAAwECBAUHAAYI/..... ....
… result of the protection of clear-text <Item><Descriptor> etc…
      </ipmpdidl:Contents>
</ipmpdidl:Item>
</DIDL>
```

As proposed in the solution depicted in Figure 3, when the protection is applied, the link between the container and the digital resource is lost, because the reference is hidden by the protection. In this case, it is impossible to determine which resource belongs to the Item element represented in the XMLBox. This situation based on nested DIs provokes the lost of the association between encoded digital resources and the corresponding Item elements. If the extraction of a sub-part of the DI (i.e., Item element) is required during the DI processing by any tool (e.g., Authoring tool, automatic production, etc.), a complete unprotection of all nested levels in order to select which digital resource has to be moved together with the XML portion.

The same problem arises when a tool requires protecting an Item element which includes IPMP Item elements. As stated above, in our B2B model and use case, to protect the result of an integration of different DIs means to encode the whole DI (i.e., the Root Item Element) with an IPMP Tool. This requirement imposes to process also the attached digital resources with the same encoding. Therefore, the two-steps protection is performed at any level: in this way, if a tool needs to render a given digital resource has to unprotect all the higher nested levels. In the following section, a solution for managing compounded DIs, structured with nested levels that could be protected is presented, highlighting the lack of flexibility of the MPEG-21 FF without the proposed solution.

This case shown is of great importance for B2B activities (as in the AXMEDIS platform and tools) since a business requirements state that: "*distributable media has to be protected and can be used for integration. When integrated a new protection has to be applied to the obtained content in order to protect the value of the integration itself*".

3 Proposed mapping of nested protected DIs

The AXMEDIS solution allows associating structure elements with encoded digital resource on the basis of AXMEDIS Object Identifiers (AXOIDs). Any encoded digital resource is named using the container's AXOID as its prefix (i.e., axoid0000_res0). AXMEDIS requires also that the produced DIs, if protected, has to expose in a proper placeholder a

description of everything is contained inside (and for protection is not directly visible). The ContentInfo element of IPMP components has been designed to include information about protected content: in AXMEDIS the identifiers of the inner levels of the nested structure and related metadata are reported in this element.

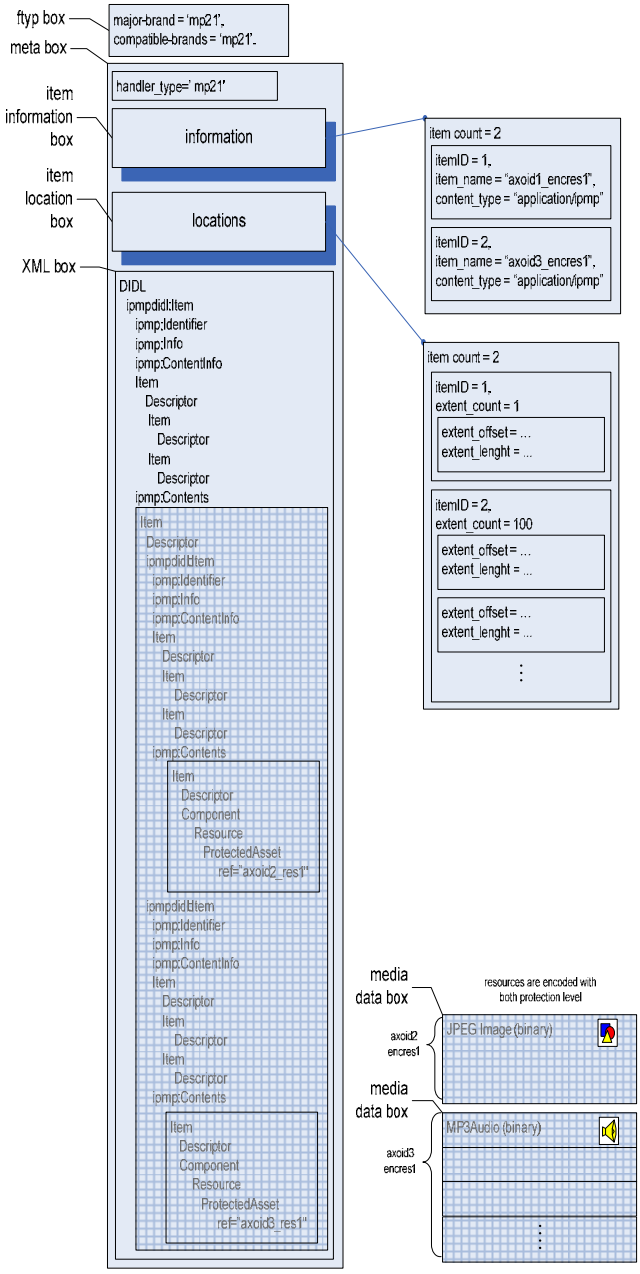


Figure 4 – Two nested levels DI with a protection at each level

The two solutions (i.e., AXOID-based reference, and AXOIDs of protected content) allow to always associate DI structure with encoded digital resources as depicted in Figure 4. The Figure 4 shows how to represent a DI, which contains an outer Item element that, in turn, contains two other Item elements, and the latter contains a digital resource each. The presented DI has been created by starting from the composition/reuse of two MPEG-21 DIs modeled as depicted in Figure 3. In their separate models, each of them was endowed of protected digital resources and protected IPMPDIDL contents into their corresponding XML Box.

When these two MPEG-21 objects are compounded together to create a new composite object (e.g., a collection, an album, etc.), their MediaDataBox are protected with an additional nesting level (enforcing the protection of the integrator, increasing security and the controllability). In addition, the IPMPDIDL contents containing the XML box of the components is protected as well as a unique blob hiding the details of the decomposition and thus avoiding the direct access to the components without activating the unprotection imposed by the integrator.

The proposed solution balances between box-like protection and digital resource accessibility. The most evident drawback is that is implicitly assumed that a given digital resource has been protected with all the IPMP Tools that have been used for the container boxes. For example, *axoid1_encres1* in Figure 4 has been protected with the IPMP Tools that are specified by the protection of the first (the root IPMP Item element) and the second nesting level (IPMP Item that contains *axoid1_encres1*).

The above solution has been adopted for the limitations of the MPEG-21 format. In fact, there is no way of specifying directly in the FF which protection has been used in terms of MPEG-21, but only in terms of MPEG-4 IPMPX, supposing probably of hosting into the MPEG-21 packages only MPEG-4 resources. The drawback of this approach in formalizing the FF is to lose flexibility with respect to what can be done by using XML. The MPEG-21 standard FF excludes some of the technologies that have been included by MPEG-21 by confining it in the XML box, limiting the usage of the MPEG-21 as a general packager. The IPMP components would require a more intensive integration with the FF specification in order to really map on that binary format what can be obtained by using XML. Thus these problems have been solved by using the above mentioned solution in AXMEDIS MPEG-21 solutions and tools. The proposed solution allows obtaining nested levels of protection preserving fast access to digital resources.

4 Conclusions

AXMEDIS is proposing a solution and platform, based on MPEG-21 standard that address also B2B content distribution and exploitation requirements. When studying how to map the produced DIs on the File Format standard several problems have been faced. The proposed solution is capable, with the present structure of the MPEG-21 standard FF (File Format), of solving the above mentioned problems, permitting direct access to the information. The AXMEDIS platform includes editor, authoring tools and players for PC, PDA, and mobiles. All tools are accessible for free from www.AXMEDIS.org .

The hardest challenge is when DI is structured as a nested container and at each level protection can be applied by using IPMP components. In this case, two solutions have been designed: i) encoding inner level by value in the XML ii) separating digital resources in the proper MediaDataBox while applying on them all the protections required on the basis of the level in the DI they belong to. The first is not efficient in terms of digital resource access and makes useless the File Format with respect to a simple XML document. The second is based on the implicit assumption that digital resources are encoded with all the IPMP tools of the upper nested levels. The proposed solution allows obtaining nested levels of protection preserving fast access to digital resources.

5 Acknowledgements

The authors would like to thanks to all AXMEDIS partners (among them: TISCALI, Universitat Politecnica di Catalonya, University of Leeds, University of Reading, HP, EXITECH, BBC, TEO, ELION, EUTELSAT, ACIT, AFI, SDAE, SIAE, ETRI, FHGIGD, SEJER, EPFL, VRS, etc.) including the Expert User Group, for their contributions and collaborations. A specific acknowledgment to EC IST FP6 DG INFSO for the partial funding of AXMEDIS.

6 References

[1] AXMEDIS Framework, “Framework and Tools Specifications”, www.axmedis.org .

[2] Bellini, P., Barthelemy, J., Bruno, I., Nesi, P., Spinu, M., “Multimedia Music Sharing among Mediatheques, Archives and Distribution to their attendees”, Journal on Applied Artificial Intelligence, Vol.17, N.8-9, pp.773-796, 2003.

[3] Bellini, P.; Nesi, P.; Rogai, D.; Vallotti, A., “AXMEDIS tool core for MPEG-21 authoring/playing”, Proc. of the first International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution, IEEE Computer Soc. Press, Nov. 2005, Florence, Italy, AXMEDIS 2005.

[4] Burnett, I.S., Davis, S.J., Drury, G. M., "MPEG-21 digital item declaration and Identification-principles and compression", IEEE Transactions on Multimedia, Vol.7, N.3, pp.400-407, 2005.

[5] Coding of audio-visual object Part 12: AMENDEMENT 1: File format extension and guidelines, ISO/IEC FDAM 1 14496-12:2004.

[6] Coding of audio-visual object Part 12: ISO base media file format, ISO/IEC FDIS 14496-12.

[7] Hulsen, P.; Kim, J.-G.; Lee, H.-K.; Kang, K.-O., “Delivering T-learning with TV-anytime through packaging”, Proc. of the IEEE International Symposium on Consumer Electronics, pp.614-619, Sept. 1-3, 2004.

[8] Iannella, R., “Open Digital Rights Language (ODRL)”, Version 1.1 W3C Note, 19 September 2002, http://www.w3.org/TR/odrl .

[9] Lee, J., Hwang, S.O., Jeong, S.-W., Yoon, K.S., Park, C.S., Ryou, J.-C., “A DRM Framework for Distributing Digital Contents Through the Internet”, ETRI Journal, Vol.25, N.6, pp.423-435, December 2003.

[10] Lin, E.T., Eskicioglu, A.M., Lagendijk, R.L., Delp, E.J., “Advances in Digital Video Content Protection”, Proceedings of the IEEE, Vol.93, N.1, pp.171-183, January 2005,

[11] Mourad, M., Hnaley, G.L., Sperling, B.B., Gunther, J., “Toward an Electronic Marketplace for Higher Education”, Computer of IEEE, pp.58-67, June 2005.

[12] MPEG-21 Part 2: Digital Item Declaration Language, ISO/IEC 21000-2:2005.

[13] MPEG-21 Part 4: Intellectual Property Management and Protection Components, ISO/IEC 21000-4:2006.

[14] MPEG-21 Part 9: File Format, ISO/IEC FDIS 21000-9.

[15] Wang, X., De Martini, T., Wragg, B., Paramasivam M., Barlas C., "The MPEG-21 rights expression language and rights data dictionary", IEEE Transactions on Multimedia, Vol.7, N.3, pp.408-417, 2005.

Effects of Agent Communications on System Performance in Medical Organizations

Yasir Khalifa§, Shi-Kuo Chang§, Douglas B. Fridsma ¥

Computer Science Department (khalifa@cs.pitt.edu, chang@cs.pitt.edu)§
Biomedical Informatics Department (fridsma@cbmi.pitt.edu)¥
University of Pittsburgh, Pittsburgh, PA 15260

Abstract. *In this paper, we use network analysis and collaborating agents to test, analyze, and evaluate the effects of communication policies on the performance or response in medical organizations using simulation tools. The goal was to learn how these simulations compared to real-world organizations, both, in emergency and routine environments. In the system evaluation, we incorporated the expertise of health care professionals and real-life medical situations. The use of computer communication protocols facilitated the simulation of agent interaction and teamwork in a hospital setting relevant to high and low levels of emergencies. The research adds a mechanism for health care field to determine emergency medical policies and procedures.*

1. Introduction

In a large organization, there is a high level of uncertainty as to whether or not communication among agents has been received. Depending on the events in an organization, the use of communication affects the uncertainty and performance of the organization. For example, in an organization with a dynamically changing environment and a high rate of unexpected events, communication reduces uncertainty and increases performance. On the other hand, in an organization with a stable environment and a low level of uncertainty, communication could potentially add more time delays and would affect the efficiency of the organization. Because studying real organizations is expensive and difficult, there is a need to develop new methodologies to test and analyze the effects of communication on organizational performance. In this research, we created a simulation tool that uses organization theory [1], discrete event simulation [2] & contract nets [3] to investigate how acknowledgement of communication affects organizational performance and test this tool using data from real-world organizations. This research expands our understanding of communication in organizations, & provides an application of contract nets to computational organization theory. This research uses an information-processing approach to examine organizations in the context of organizational simulation, taking advantages of computer network theory and agent communications with acknowledgment policies. The presence of software to analyze agent distributed computer networks [4] enables us to apply network analysis approaches to information processing theory of organization. Agents and activities in organizations resemble computer nodes and processes in a distributed network. A network analysis technique, for example contract net, considers intelligent communications among its agents. The level of contract net's acknowledgment procedure could be more complex; furthermore, contract net is useful because it is not only uses acknowledgments, but it also represents interactions between agents to examine more complicated semantics and to resolve unexpected events or exceptions [5]. We can also learn about human agents from computer or artificial agents.

2. Condition C & STAT Page Protocols in Health Care

In health care organizations, some unexpected events could be critical. For example, as shown in Figure 2, a patient is suddenly experiencing difficulty breathing which could be a cardiac problem and may lead to a cardiac arrest. The patient indeed is experiencing a cardiac arrest and requires an immediate surgery. Condition C (Crisis) is a state in which the patient could be prevented from dying. To prevent the patient from getting into a fatal state (condition A--cardiac arrest, patient is dying), there is a need for efficient communications. A surgeon is available, but the anesthesiologist is not found! In this environment, acknowledgment during the communication becomes critical in order to find the next available anesthesiologist, and is likely to result in preventing the patient from dying, and consequently in reducing mortality rates.

Condition C and STAT page procedures are used frequently to communicate in emergency situations within various health care settings such as nursing homes, ICU, or regular hospitals. When a condition C call is made, the condition C Team that consists of a number of doctors, nurses, and technicians will be automatically directed to the patient or emergency site [6]. There is a guarantee that at least one member of the team will be immediately present at the patient site, making the behavior of *condition C* similar to *communications with acknowledgments, ack*. Using *ack* among the crisis team members effectively prevents the transition from condition C to condition A. STAT page on the other hand represents communications without acknowledgments since we don't know whether or not a doctor received the page call. In addition, condition C compares to concurrent processing and STAT page compares to sequential processing as shown in Figure 3. In order to know which protocol is more efficient (ie, reduces mortality rate) and under what states or conditions one is better to use than the other, we will study the efficiency of each protocol under various conditions as they compare to real-life situations. From the condition C and STAT page example in particular, and from other emergencies in general, acknowledgment of communication has effects on coordination and system performance. The interdependence between ack policy and a high number of

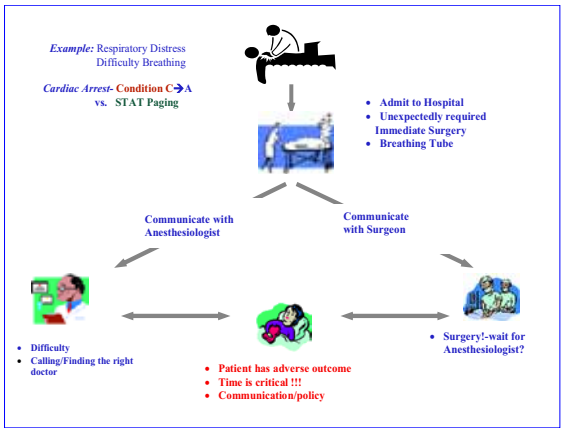


Figure 2. Communications in Health Care Organizations

exceptions makes achieving optimal performance more complex [7]. In a stable

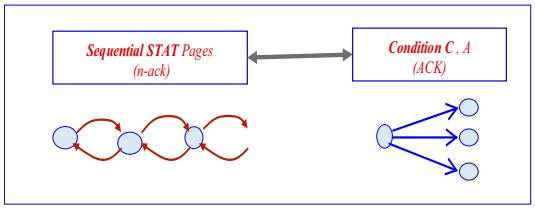


Figure 3. Condition C & STAT Page Protocols

environment, acknowledge communication has fewer effects on coordination and performance due to lower level of exceptions. Furthermore, regardless of the domains of a disaster,

Communication effects on Cost and Quality		ack (Condition C)	n-ack (STAT Page)
Stable environment *Routine organization*		↓ Cost ↑↑ Quality	↓↑ Cost ↑↑ Quality
Unstable environment *Non-routine organization*		↓↑ Cost ↑↑ Quality	↓ Cost ↓↓ Quality
Performance	*Quality*	*Number of resolved exceptions*	
	Cost	*Total wait and processing times*	

Figure 4. Communication effects with and without ack

immediate and efficient responses are necessary to improve the organizational coordination [8].

3. Development of a Test Methodology

To study, test, and analyze the effects of communication on the performance of an organization, we created a simulation tool called MedSim that uses organization theory, discrete event simulation, and contract nets to investigate how acknowledgment of the communication affects the organization's performance, and test this tool using data from real-world organizations. MedSim enabled us to study the communication effects on performance in routine and emergency environments. We combined the information-processing approach with multi agents in a contract net to study the effects of communications on the performance of organizations. We integrated information processing and contract nets approaches into the simulation model, MedSim, which is described later in this paper. Furthermore, we simulated models with and without acknowledgment of communication, and evaluated how these models work in the simulation system in comparison to the real world. The simulation results were close in their behavior to the performance behavior shown in Figure 4, where routine & non-routine organizations were used with acknowledgment & no-acknowledgment communication policies (*ack, n-ack*). In the next two sections, we illustrate the combined systems between information processing view of organizations and network analysis using contract nets to communicate between collaborating agents.

4. Information-Processing Resource Allocation Model

The procedure for assigning an activity to a responsible agent will be used for resolving exceptions [9]. It is considered a re-assignment or re-allocation process of activities. The assignment may be made according to different policies such as:

a) Work Load, b) Skill Level, and c) Lowest-in-Hierarchy. For example, assignment according to workload will be used practically within a communication channel to see if an agent is available according to the current workload. The other policies are useful when the focus of the study is based on the characteristics of actors or activities regardless of the relationship with communication policies and their effects on performance. Figure 5 shows an assignment

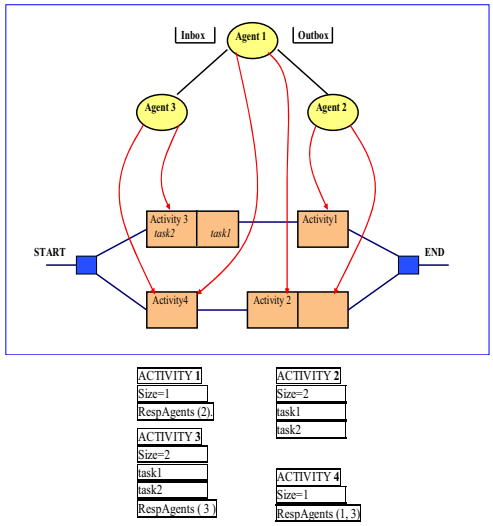


Figure 5. Agent-activity resource allocation model

model with variable size of activities distributed according to the agent-activity-exception assignments. In general, in the modeling of the simulation system MedSim, various policies will be used for resolving exceptions. As shown in Figure 5, various situations of agent-activity assignment may be made, which are illustrated according to the following scenarios:

(a) Single-agent to single-activity is satisfied by agent2-activity1
(b) Multiple-agent to single-activity--agents1, 3 – activity4.
(c) Single-agent to multiple-task-activity--agent3 – activity 3.
(d) Multiple-agents to multiple-task-activity--agent1, 2 – activity2.

Processing activities depends on the responsible agent's workload and the size of an activity [15] [3]. For example, as shown in Figure 5, an agent responsible for multiple activities and an activity may include multiple tasks. Each agent has an inbox and an outbox. The number of tasks in the inbox indicates the number of tasks waiting for processing. The number of already completed activities is stored in the agent's outbox. Exceptions are unexpected events that require resolution by establishing communication among agents with matching role or domain, according to the policy used [10]. The next section describes communications between agents using contract nets with background information and characteristics.

5. Contract Net Model in Health Care

This paper presents an expansion of contract nets from various disciplines into health care organizations. The dynamically changing nature of health care organizations, the flexibility of agents being hybrids as they are free to join or leave a group or location within the organization, and the fact that there are condition C and STAT page protocols that compare to sequential and parallel processing as one-to-one and one-to-many agent interactions, all of these factors together, make contract nets an ideal tool to investigate interactions between agents in order to resolve exceptions in health care. Figure 6 represents a contract net view in a hospital setting where a patient was experiencing a sudden respiratory distress which may lead to a cardiac arrest.

The communication using a contract net could result in preventing a transition from condition C to condition A. When agent1 is looking for agent2, an anesthesiologist, the time is critical. There is a time threshold imposed until agent2 is found and designated the work in order to perform the surgery. In the model, agent3 is found, which matched the domain needed (anesthesia), and the agent is available. The next section describes the use of contact nets for condition C and STAT page protocols.

5.1 Contract Net Modeling for Condition C and STAT Page Protocols

In this section, we illustrate the interactions in the contract net at various levels for the condition C and STAT page protocols as they relate to communications with and without acknowledgments. In the case of condition C (Figure 7.a), each agent is expected to reply with "msg=ACK, Done" or "msg=ACK, Failure", after which the communication terminates. Failure to receive a reply from each agent in the crisis team (condition C Team), means the communication will be considered incomplete (similar to n-ack policy—STAT page, as in Figure 7.b). The idea is to have every agent in the temp buffer/crisis team reply to the call. However, the reply does not have to commit. It could be "ack, Done" or "ack, Failure". In the case of "msg=Failure", a substitute agent from the list of condition C Team must be found in order to continue the contract net procedure.

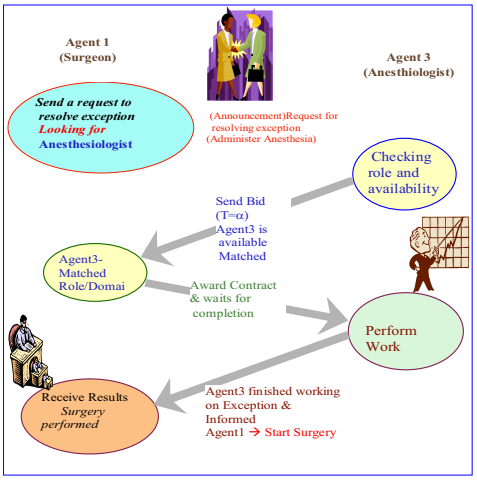


Figure 6. Contract net in health care

In either case (STAT page or condition C), there is no guarantee that every agent replies to an announcement call. A way around it would be to restart the contract net communication every time it is incomplete within a pre-specified time threshold. The side effect of this approach is time overhead, which could become critical during the transition to condition A (patient is dead) from condition C (patient about to die) in real-life. A Sender agent broadcasts an announcement call requesting exception resolve to a sequential set of agents. Each agent is specified by role, required-skills, responsible-for activity/task attributes, and domain. Once a match of these attributes is reached, the set of selected agents will be saved in a temp buffer to start contract net communications with each agent from the set of agents until an available agent is found (COMMIT message is received).

First, we start with the announcement level for both protocols, and then show the entire interaction levels (task bid, task award, and results) for condition C and STAT page. The various protocols as they apply to Condition C and STAT page as communications with/out acknowledgments, and then map them into the actual design of the system, which will be illustrated in the next section.

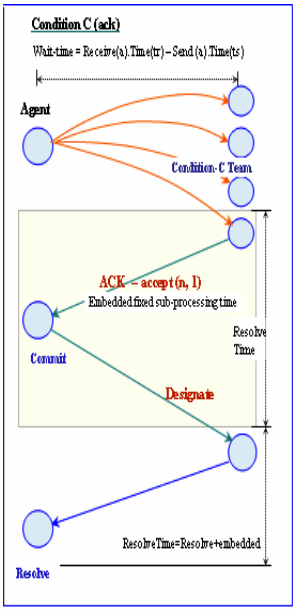


Figure 7.a. Contract net- Condition C Interactions

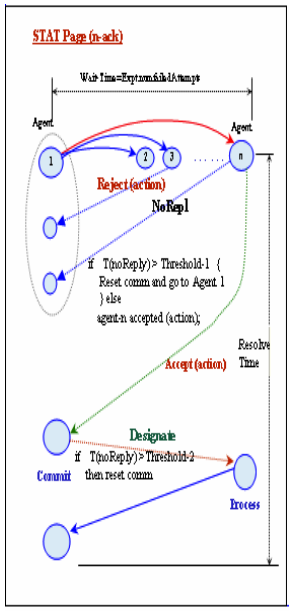


Figure 7.b. Contract net – STAT Page Interactions

At the announcement and bid levels, when a sender-agent (Initiator) sends a message to the receiver-agent (Responder) as "action=resolve-exception?" then depending on type of communication protocol used, STAT page or condition C, the receiver-agent could propose, refuse, or action not understood, as the interaction is made with one agent (STAT page) or with multiple-agents (condition C Team). At the task-bid level, the receiver-agent (Responder) sends one of the following actions (COMMIT, BUSY, NU-- Not Understood) within a specified time deadline threshold, as shown in Figure 8. The sender-agent (Initiator) could receive one of the actions shown in the same Figure. If the Initiator agent received a commit action from the Responder agent, then the Initiator could do one of the actions in Figure 8, which means that the Initiator may accept by designating the exception resolve to the Responder, otherwise the action would be rejected. Finally, when the receiver-agent is designated an action to resolve in a pre-specified time, the receiver-agent may send one of the actions shown in the last listed actions of Figure 8, where the message is an acknowledgment action that could be "done, failure, or cancel", in which case the contract will be re-started for a new communication channel. The sender or receiver agent's inbox is increased by one if the exception-size=1-task, or increased by the number of tasks. To resolve an exception, we start the contract net communications. When agents communicate, the wait time of an exception in STAT page is computed from the number of failed attempts to resolve the exception, the time to start the contract net, and the time a responder agent commits to resolve the request, including the time with no-reply, and whether the agent is idle or not, the system time stamps the agent each time it checks its status. In condition C, the wait time of an exception is computed from the time of starting contract net until an agent starts resolving an exception (a positive delta time is assigned randomly and added to the exception's wait time that is drawn from a range of values to behave close to a real-life condition C-Team).

```
Sender-agent (Initiator): receive (action(receiver-agent, msg=COMMIT))
                          receive (action(receiver-agent, msg=BUSY))
                          receive (action(receiver-agent, msg=NU))

Initiator Agent:  send (action(receiver-agent, msg=DESIGNATE))
                  receive (action(receiver-agent, msg=REJECT))

Receiver-agent(Responder): send (action(sender-agent, msg=ACK=DONE))
                           send (action(sender-agent, msg=FAILURE))
                           send (action(sender-agent, msg=CANCEL))
```

Figure 8. Agents (Sender, Initiator, Receiver Interactions)

9. Discussion and Future Work

The simulation model used information processing theory of organizations, discrete event simulation, and contract nets to investigate how acknowledgement of communication affects organizational performance. In particular, we modeled condition C and STAT page protocols to study their effects on performance in stable and emergency environments. While testing the system and generating results, we applied several verification and validation steps starting with retrospective analysis to see how the simulation model performed. We then conducted face validity by interacting with health care professional about the results and expectations to see close to real-life the results were, and finally we split data from real-organizations into two disjoint data sets, test and training data, each of which was about 50% of the original size. We considered the training data set as a basic reference to validate the behavior of the simulated output and to see how close it followed the behavior represented by

the real-data. This was done by training the algorithm model on the training data set and then testing the model independently on test data set. This process ensured that the simulation model represented real systems behavior closely enough to provide useful results and to see how the simulation engine would behave with real data in comparison to the simulated data and vice versa. We conducted hundreds of simulations runs and analyzed results starting with the low and high number of exceptions in both emergency and routine environments which we categorized by low and high exception rates. We studied and analyzed the communication effects in stable and emergency environments. The study showed that in a stable environment without ack (STAT page), the system behaved in such a way that when the number of exceptions is low, the wait-time and processing time are smaller than using ack (condition C). However, using the training data (simulated) and comparing results to test (real) data (which was originally the entire real-data split into two equal sets, 50% training and 50% test data), the analysis showed that the behavior of simulated data was close enough to real data. The simulation outputs also showed that the use of condition C with high level of exception rates did impose cost constraints by increasing the exception wait-time and processing-time at the cost of providing service quality.

Finally, improving communication effects on performance is a valuable exercise for any type of organization. In particular, health care organizations need improvements due to the dynamically changing nature of unexpected events. Resolving exceptions of different levels of uncertainty requires efficient communication protocols that provide better outcomes

a) Socially, such as saving mortality rates, providing efficient medical care, and

b) Economically, such as designating the proper physicians to patients in routine and emergency environments in order to resolve unexpected events efficiently. These improvements could be achieved by applying computer science concepts to social science. Several new questions were raised while conducting this research and performing experiments. These could be the basis for future work on this topic such as adding more abstraction to the system to cover multiple domains per agent, automating the contract net component in order to add audio driven by language between agents and larger KB close enough to real-life interactions, and adding a learning component to generate learning rules that could provide useful patterns among the system parameters which could be useful to social and computer sciences.

Bibliography

[1] Kunz J.C., Rivero C.E., Levitt R.E., "Simulating Work Processes and Organizations to Engineer Strategic Projects", Proceedings of ETCE/OMAE 2000 Joint Conference: Energy for the New Millenium, February 2000, New Orleans.

[2] Banks J., Carson J.S., Nelson B., Nicol D.M., "Discrete Event Simulation", Prentice Hall International Series in Industrial and Systems Engineering, 3rd edition, 2001.

[3] Smith R.G., "The Contract net Protocol: High-Level Communication and Control in a Distributed Problem Solver", Bond, A. H. and Gasser, L. (Eds), Readings in Distributed Artificial Intelligence, 357-366, Morgan Kaufman, Los Altos, CA, 1988.

[4] Riley, G., Ammar, M., Fujimoto, R., Perumalla, K., Xu, D., "Distributed Network Simulations Using the Dynamic Simulation Backplane", Proceedings of International Conference on Distributed Computing Systems 2001 (ICDCS'01).

[5] Khalifa Y., "Organizational Medical Simulation System: Effects of Intelligent Multi Agent Communications on System Performance", University of Pittsburgh, August 2004.

[6] Fridsma, D.B., "Error Analysis Using Organizational Simulation", 2000 AMIA Annual Symposium, 2000.

[7] Perrow C., "Complex Organizations: A Critical Essay", (3-edition), New York, Random House, 1986.

[8] Malone T.W, Crowston K., "Interdisciplinary Coordination Study". ACM Computing Surveys (CSUR). March 1994. Volume 26, issue 1.

[9] Decker K., Sycara K., Williamson M,"Modeling Information Agents: Advertisements, Organizational Roles, and Dynamic Behavior", Proceedings of the AAAI-96 Workshop on Agent Modeling. AAAI. 1996.

[10] Tu S., Musen M., "Modeling Guidelines and Protocols: A Task-Based Approach", Stanford Medical Informatics Stanford University, School of Medicine, 2000.

[11] Buchanan B.G., "Rule Learner Induction System: Prediction Analysis", University of Pittsburgh, Computer Science Department, Pittsburgh, PA, 2002.

[12] Khalifa Y., Chang S.K., et al, "A Prototype Spatial-Temporal Reasoning System for Emergency Management", Proc. Of Int'l Conference on Visual Information Systems VISUAL96, February 5-7, 1996, Australia, 469-478.

Proceedings
International Workshop on
Distant Education Technology
(DET 2007)

Co-Editors

Paolo Maresca, University Federico II, Napoli, Italy
Yasir Khalifa, University of Pittsburgh, USA
Xin Li, Google Inc., USA

Learning by doing: vLab, a virtual laboratory for Computer Engineering education

Roberto Laschi
Dip. Elettronica Informatica e Sistemistica
Università di Bologna, Italy
rlaschi@deis.unibo.it

Anna Riccioni
Dip. Elettronica Informatica e Sistemistica
Università di Bologna, Italy
ariccioni@deis.unibo.it

Patrik Suzzi
Dip. Elettronica Informatica e Sistemistica
Università di Bologna, Italy
patrik.suzzi@studio.unibo.it

Abstract

In this paper we present the design and implementation of an advanced virtual lab aimed at supporting teaching and learning in Computer Engineering courses. We first outline the requirements an educational tool must meet to represent an effective and comprehensive support for students through the whole process that leads them to achieve conceptual understanding and technical skills. The described virtual lab, based on a constructivist approach and suitable for use in blended-learning contexts, responds to a general model that allows its application in different educational domains. In the second part of the paper we describe the implementation of vLab 2.0, a prototype built following the proposed directions and the results of a prior experimentation within the Information Security Technologies course hold at the Engineering Faculty of the University of Bologna. vLab 2.0 relies on Eclipse technology in order to guarantee openness and extensibility, and to access a set of advanced features which concur in providing support for building a comprehensive and sophisticated tool for teaching and learning.

1 Introduction

Information and Communication Technologies are constantly advancing and becoming more pervasive. Nowadays, one of the most challenging issues is represented by the individuation of new, complex and effective ICTs application. Education is a sector which can greatly benefit from ICTs improvements and proper adoption.

Our research project aims at defining and implementing an integrated tool to support teaching and learning in Computer Engineering courses. The ideal framework in which this instrument can be effectively used is blended learning, here intended as a combination of two different approaches: synchronous traditional face-to-face lessons and asynchronous, individual experimentations in a virtual laboratory.

One of the main skills a computer engineer must achieve during his studies is the ability to design a well dimensioned software system, able to satisfy the defined requirements in a specific domain. Once that the specifications and the domain have been settled, the engineer starts an incremental process which cycles through four main steps:

- analyze the problem
- sketch a solution
- verify how the model fits the requirements
- design, implement and test a prototype.

What we want to do is to define the requirements an educational tool must meet in order to represent an effective and comprehensive support for students through the whole process that leads them to achieve conceptual understanding and technical skills.

Modeling frameworks and simulators have a strategic importance in helping students to gain greater insight into subjects [1, 2, 3]. Many authors have pointed out that the use of simulation tools often reinforces learning, offering a proper support for verifying theory, and contributes to improve students' performances in various disciplines [2, 3]. Especially in Computer Engineering, modeling skills are traditionally taught relying on CAD tools [3, 4].

To achieve technical proficiency and to improve practical programming skills students often benefit from the employ of advanced IDEs. They can acquire better understanding of subjects by quickly designing, simulating and then testing and extending systems [2].

Our purpose has been to integrate these different functionalities in one comprehensive educational tool. In our perspective, this tool should be projected towards educational and not industrial objectives. This makes it suitable to focus on relevant didactic aspects and to suggest an educational, partial code to students as a

starting point for software development, so that they can complete, modify and reuse it in different and more complex contexts. Following this direction, our virtual lab helps users to:

- understand and solve problems related to key theoretical concepts
- solve problems through the design and construction of new artifacts or processes
- gain and improve the technical skills their future professional roles will require.

One of the main facets of the project is its suitability to be used in different formative domains. This feature is achieved through the adoption of one of the best-known abstraction in computer engineering field: the block diagram [5, 6].

2 Requirements

In 2005 we developed a first prototype, vLab 1.0, taking the course on Information Security Technologies hold at the Engineering Faculty of the University of Bologna as our reference case study.

vLab 1.0 GUI was composed by three synchronized areas: the main one shows a graphical representation of the system to be analyzed and allows students to configure and run experiments. The second area is dedicated to prompt feedback reporting: for each component part of the system whose execution successfully ended, notes about data dimension and execution time are presented. On the other hand, for those components that could not complete their execution, usually because of a wrong parameters setting, errors are shown in red. Finally, a third view displays an educational sample code related to the ongoing exercise. A two-year experimentation conducted by making vLab 1.0 available to around 200 students confirmed the effectiveness of the chosen approach and served as a successful acceptance test, but also pointed out vLab 1.0 main limit: its lacking support to students' autonomous experimentation, based on self-constructed experiments. In line with these results, we focused on constructivist theory, which assumes that students should be provided with an open and extensible environment: they have to be actors, other than active, in their own learning process [7]. For this purpose it is fundamental to provide students with various instruments aimed at covering the different functionalities which can have a central role in the learning process [7, 8, 9]. Figure 1 shows how to reap these broad goals, outlining the major actors and features involved.

The sketch represents the workbench dedicated to the graphical modeling of the current experiment as the central aspect within the virtual lab. The workbench includes a library of primitive components, that students can use to construct their own experiments. It is also possible for users to load a predefined exercitation, chosen among the available sets, and configure and run it. According to this approach, tutors and teachers are charged with the responsibility of creating a starting collection of experiments that learners can set and run; but then learners too can cover the instructor's role and modify or build their own trials from scratch. Thus saving and restoring features in a standard format, so that new built experiments can be shared between users, assume a primary importance within the virtual lab.

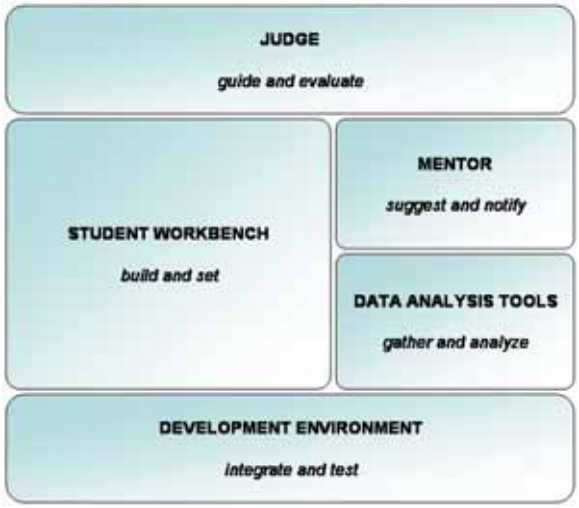


Figure 1. Actors and tools role in the virtual lab environment.

The workbench should also let users switch between different detail levels, in order to analyze different sides of the same topic. This can provide a more effective, customizable and comprehensive didactic experience and is central to gain a major compliance with different learning styles. Especially for computer science students it is important to have the ability to access both the systemic and the algorithmic view within a same solution, while referring to our case study such a facility enables to switch from a security service perspective to a mechanism-oriented one.

Another issue within the virtual lab is represented by help instruments. In fact, activities can be introduced and proposed by an intelligent tutoring system, the "judge" in figure 1, which evaluates users' performances and suggests new tasks according to students achievements and objectives. Furthermore, each step conducted within an experiment can be followed by a "mentor" aimed at pointing out possible errors and suggesting how to correct them. Users' understanding and reflection can also be improved by allowing the access to external additional resources, such as technical documentation or reference standards.

The virtual lab also includes gathering and analysis tools whose function is to help users in collecting experimental data and interpreting them. Finally, users can switch at any time to a development environment integrated within the lab to complete,

modify, debug, test and run a program, starting from a partial educational code automatically produced during an exercitation.

With these features, the virtual lab can really support the constructivist approach and learners' active and primary role in their own educational experience. Building and sharing new knowledge can, in fact, pass through:

- the construction of new experiments
- the modification of predefined experiments
- non trivial experiments settings
- the interpretation of gathered experimental data relying on the featured analysis tools
- the discovery or application of new analysis methods
- accessing multiple external information sources.

Furthermore, users can extend the virtual lab in order to comply with their changing needs and share their new knowledge with other users, for instance adding their new built schemas to shared libraries, or implementing special analysis tools for specific needs.

3 Eclipse technology

The implementation of a prototype adherent to the exposed constructivist approach has requested the design of an appropriate architecture and the adoption of a proper technology suitable for its realization.

The architecture of an effective educational tool must be open and extensible. Eclipse plug-in architecture and extension point mechanism perfectly address this major issue [10]. Relying on Eclipse technology is crucial also in order to access a set of basic and advanced features essential for providing a complete and sophisticated tool for support teaching and learning in higher education. Eclipse technology simplifies the creation and coordination of synchronized multiple editors, views and perspectives and makes straightforward to integrate a fully-featured IDE customizable for different programming languages. Furthermore, within Eclipse Community several projects are being developed as general purpose plug-in which can effectively be integrated within our educational tool.

Table 1 outlines the mapping between the requirements previously discussed and the proper technical solutions selected within Eclipse technologies.

Requirement	*Solution*
Judge	BPEL
Student workbench	GEF-based editor
Mentor	Ganymede
Data analysis tools	BIRT and special purpose plug-ins
Development environment	JDT

Table 1. Mapping between requirements and technical solutions.

Figure 2 outlines the resulting vLab 2.0 architecture, which is represented by a stack where each tier depends upon, uses and integrates the level below.

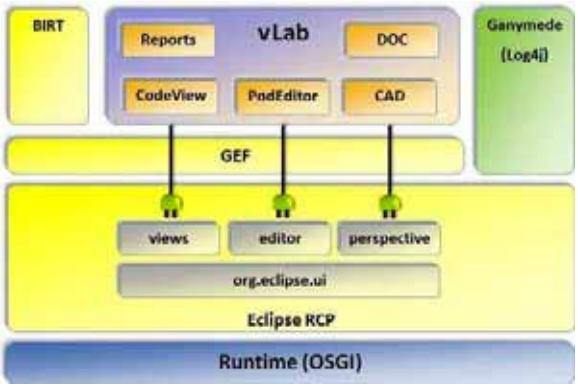


Figure 2. The virtual lab architecture based on Eclipse technology.

The OSGi runtime, responsible for the coordination of multiple and different software components, is the foundation of the framework [11]. The Eclipse Rich Client Platform relies on OSGi services to provide a specific implementation of the standard, replacing the generic OSGi bundles with Eclipse plug-ins [12]. It also offers a new set of facilities, collected in the org.eclipse.ui package, essential for the Eclipse Development Framework and tailored to support external contributes through the extension point mechanism. The vLab feature leverages different resources to provide its added functionalities: among them, the Eclipse RCP, the Graphical Editing Framework [13] and several projects, like BIRT [14], or third-parties plug-ins such as Ganymede [15], built to integrate Log4J within Eclipse applications. Accordingly to Eclipse technology main characteristics, vLab is supplied with a set of extension points in order to be open to future expansions.

3.1 The black-box model

The exercitations within the virtual lab are based on a logical model that reflects the block diagram abstraction.

The block diagram key elements are blocks and connections. The blocks have an associated function or system of functions, and can have inputs, outputs and parameters which are graphically represented by nodes. The connections are oriented and must respect some basic rules related to data flow to be considered syntactically valid: each input or parameter node can only be a target for at most one connection, while output nodes can propagate their value towards multiple destinations being the source of several connections.

To cope with increasing schema complexity, the box abstraction has been added. Box represents a

boundary which can isolate a sub-graph providing it with an external interface through pins. Box pins are called "adapters" because of their double function: other than performing data flow, they can allow users to accomplish the possibly required data type conversions. Boxes can also be obscured to hide their inner structure: thus, they represent the needed support for implementing the hierarchical design approach and for switching between different detail levels within the same schema. Finally, boxes can be used to introduce recursion and iteration.

Since blocks can execute, and so change their state, it's necessary to define a color code to discriminate between different states. Orange blocks are missing input data, while yellow blocks are fully configured and ready to be executed. A similar semantic occurs within nodes, too: a node whose associated value is null is colored in red, and becomes green when the handled data assumes a defined value.

The following figure illustrates the depicted model.

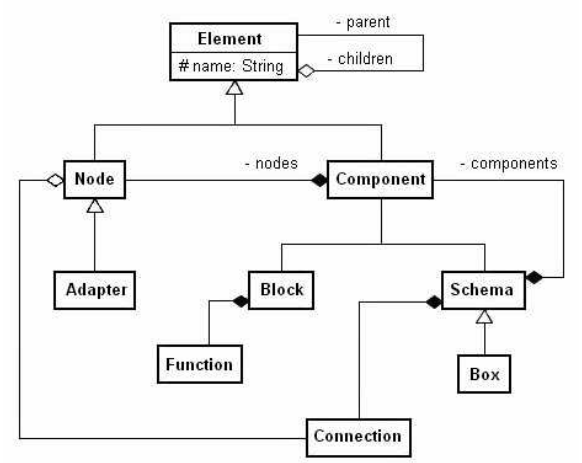


Figure 3. The underlying abstract model.

3.2 Plug-ins

The primary way users have to interact with the virtual lab is through a graphical representation of the exercitation based on the depicted logical model. For this reason, the GUI is composed by the workbench centered on a GEF-based editor, and by several views. Each view is aimed at a specific task, and synchronized with the underlying model.

The editor includes a palette that displays the available components and allows vLab users to build their schemas by mean of drag and drop operations. Users can also load a predefined exercitation and then modify it. The Eclipse outline view provides an alternative tree representation of the schema elements.

Once that the schema has been completed, users can switch from the modeling phase to the simulation of the built system, setting and then executing all of the components. The Eclipse property view supports users in configuring the required parameters, thanks to a business logic which allows to assign a value only to disconnected input or parameter nodes: in fact, a connection between two nodes represents a channel that propagates the source data value to the target.

When simulating a system behavior, it is possible to trace and record a series of characteristic values which can be used to perform tests and evaluations, other than to refine system dimensioning. Engineers usually base their efficiency evaluations on time and data dimension: for this reason the virtual lab can trace, for each block, the associated function execution time and input and output data dimensions. Once that these information have been saved, they can be displayed within a BIRT report: BIRT plug-in, which can be easily integrated with the virtual lab, allows users to represent gathered data according to various layout and to complete them with aggregated and derived information. As a consequence of the mechanism of extension points, the virtual lab can also be extended with other special purpose plug-ins, when tracing time and data dimension is not enough.

After modeling and simulation support, it is important to provide the third requested feature: automatic educational code generation. To achieve this purpose we implemented an extension of Eclipse basic view, aimed at integrating the outline of the graphical schema with the representation in form of partial code of each schema element. Saving the generated content as a file with the appropriate extension enables users to switch to the proper development perspective.

Another key feature is represented by the online interaction support, which is offered through various services: the first one is the *"judge" plug-in*, currently in a planning phase, which may rely on the BPEL project [16, 17]. A central role is also covered by the *prompt feedback reporting view*, based on Ganymede plug-in and aimed at immediately notify users with errors or significant info related to the current activity. In order to simplify the construction and test of new schemas, Ganymede displays "info"-tagged messages to give positive or neutral feedback, "warn" messages to point out incomplete configurations or not allowed operations, and "error" message to report exceptions obtained during run-time simulation. The third component involved in offering a complete interaction support is a comprehensive *help online* documentation, related both to the virtual lab usage and to the components available in the palette. Finally, it is important to provide users with a documentation *perspective*, that allows students to consult additional and external resources using a browser view and a suggested and extensible link list.

3.3 Screenshots

According to the requirements previously outlined and to the correspondent technical solutions individuated in the Eclipse technology framework, we integrated our research with the development of vLab 2.0, a prototype of advanced virtual lab.

Figure 4 shows a very simple example of how the virtual lab works. To make the screenshot more readable, it has been focused on only three views: the GEF-based editor with its palette, the code view and the XML view. These views usually appear in the exercitation perspective together with other facilities.

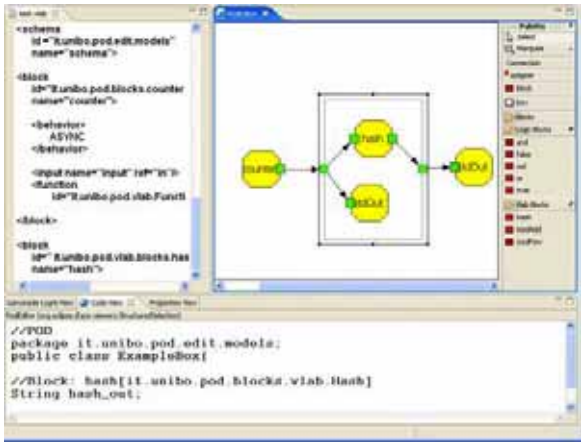


Figure 4. vLab 2.0 screenshot (exercitation perspective): editor, code view, XML view.

The schema represented in the editor is composed of four primitive components and a box with two adapters, which could be obscured to hide its content. The editor palette is organized in different categories related to libraries which can pertain to various domains; a special category can be reserved for the student's self-constructed new components.

Accordingly to the Model-View-Controller pattern, the various views within the virtual lab all rely and operate on the same underlying model in order to keep themselves synchronized. During a work session within vLab, the shared model structure or the data it holds change following users' actions: the system must then propagate the suitable notifications to the interested recipients, in order to make them able to properly respond to the modifications occurred. The common model can be represented through an abstract tree that, when building the XML description of a graphical schema, is translated into a corresponding Document Object Model.

Figure 5 focuses on a different combination of views, still referring to the exercitation perspective: the right part of the screenshot shows the "mentor", implemented by a log view based on Ganymede plug-in. The log view contains messages tagged as "info", "warn" and "error", each one represented in a different text color, respectively green, orange and red. These messages have been interactively produced during the modeling and the simulation phases, in consequence of specific actions performed by the user.

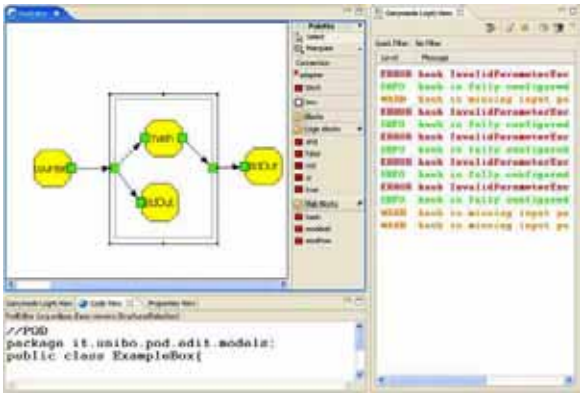


Figure 5. vLab 2.0 screenshot (exercitation perspective): editor, XML view, log view.

vLab 2.0 offers also other perspectives, each one providing different features through various plug-ins. For example, a BIRT report can be used to display the data collected during simulations. Furthermore, the virtual lab will soon integrate a special purpose plug-in aimed at performing FIPS-140-2 tests to evaluate bits casuality and another one focused on frequency distributions analysis.

3.4 Delivery

vLab 2.0 will be extensively tested within the 2008 edition of the course on Information Security.

Students will be able to access vLab 2.0 through AlmaChannel, the University of Bologna e-learning platform [18, 19]. Different distributions will be available for download: a feature-based one, for those users who already have Eclipse platform installed on their computers, and a Rich Client Platform version, suitable for who prefers to deal with a ready-to-run product [12]. The RCP will probably integrate SourceForge Dr.Java plug-in in order to simulate a simplified Java Development Environment [20]. The availability of different distributions, one of which simpler to use, is desirable in the experimental context: the course on Information Security Technologies has a mixed target composed of students with various starting informatic competences.

4 Conclusions and future work

In Computer Engineering education a central issue is to make students achieve strong design and programming skills. Providing students with advanced tools aimed at support them during modeling, testing, evaluation and implementation of software systems can

help reaching this objective. Such a tool, to be effective, must be complete, highly customizable and offer an integrated support for the various step the learning process goes through. Our research led us to design vLab, an advanced tool aimed at satisfying the described requirements.

The system we are currently developing, vLab 2.0, relies on Eclipse technology. The plug-in architecture is essential to supply extensibility and to reuse third-parties tools compliant with our needs. Finally, Eclipse provides support for designing different solutions for delivery and customizing packages with different features dedicated to specific targets.

Our future efforts will be addressed both at improving vLab adding new advanced features and at dealing with new open and challenging issues. A key point is represented by users' activities evaluation, and is strictly connected to the presentation of customized feedback and the suggestion of new tasks based on users' profiles, formative objectives and activity history. This research direction will also imply an integration with the standard SCORM [21]. Furthermore, we are planning to analyze BPEL suitability to define and manage users' portfolios and curricula.

An interesting and useful achievement would be to build new component libraries to adapt vLab 2.0 for being adopted also in contexts other than Information Security Technologies. A suitable new frame is represented by Logic Circuits, as we experimented following Eclipse GEF tutorial based on logic diagrams. Finally, it is possible to make vLab 2.0 provide educational code in various programming languages, other than Java.

Bibliography.

1 Chung G. K., Harmon T. C., Baker E. L., (2001). The Impact of a Simulation-Based Learning Design Project on Student Learning, IEEE Transactions on Education 44 (4): 390-398.

2 Büchner P., Nehrir M H., (1991). A Block-Oriented PC-Based Simulation Tool For Teaching and Research in Electric Drives and Power Systems, IEEE Transactions on Power Systems 6 (3): 1299-1304.

3 Allwood J. M., Cox B. M., Latif S. S., (2001). The Structured Development of Simulation-Based Learning Tools With an Example for the Taguchi Method, IEEE Transactions on Education 44 (4): 347-353.

4 Ma J., Nickerson J. V., (2006). Hands-on, Simulated, and Remote Laboratories: A Comparative Literature Review, ACM Computing Surveys 38 (3): art. n. 7. DOI= http://doi.acm.org/10.1145/1132960.1132961.

5 Mano M. M., Kime C. R., (2004). Logic and Computer Design Fundamentals, Pearson Prentice Hall.

6 Stallings W., (2006). Cryptography and Network Security, Pearson Education.

7 Black J. B., McClintock R. O., (1995). An Interpretation Construction Approach to Constructivist Design, in B. Wilson (Ed.) Constructivist Learning Environments.

8 Guidorzi R., (2003). e-Learning Projects at Bologna University: an Overview, Proceedings of the International Conference on Networked e-Learning for European Universities.

9 Guidorzi R., Diversi R., Colin M., Lodoli G., (2006). A Constructivist Approach in Designign an e-Learning System Identification Course, Preprints of the 7th IFAC Symposium on Advances in Control Education.

10 Gamma E., Beck K., (2003). Contributing to Eclipse: Principles, Patterns, and Plug-Ins, Addison-Wesley.

11 OSGi Alliance, (2005). OSGi Service Platform specifications, Release 4. Available from http://osgi.org.

12 McAffer J., Lemieux J. M., (2005). Eclipse Rich Client Platform : Designing, Coding, and Packaging Java Application, Addison-Wesley Professional.

13 Moore B., Dean D., Gerber A., Wagenknecht G., Vanderrheyden P., (2004). Eclipse Development Using the Graphical Editing Framework and the Eclipse Modeling Framework, IBM RedBooks.

14 Peh D., Hannemann A., Hague N., (2006). BIRT: A Field Guide to Reporting, Addison-Wesley Professional.

15 Ganymede plug-in: http://sourceforge.net/projects/ganymede/.

16 BPEL project: http://www.eclipse.org/bpel/index.php.

17 Eclipse Italian Community: http://www.dis.unina.it/ECLIPSE/index.php?option=com_content&task=view&id=12&Itemid=28&lang=en.

18 Vicari D., (2004). AlmaChannel Project, Citam internal report, University of Bologna.

19 AlmaChannel portal: https://www.almachannel.unibo.it/portale/index.htm.

20 Dr.Java plug-in: http://drjava.sourceforge.net/.

21 Advanced Distributed Learning (2006). SCORM 2004 3rd Edition. Available from http://www.adl.gov.

Multimedia Knowledge Eclipse Environment: an Eclipse Tool for Multimedia Application Design

Paolo Maresca
Dip. Informatica e Sistemistica
Università di Napoli Federico II, Italy
paolo.maresca@unina.it

Giuseppe Marco Scarfogliero
Università di Napoli Federico II, Italy
g.scarfogliero@studenti.unina.it

Lorenzo Sorrentino
Università di Napoli Federico II, Italy
l.sorrentino@studenti.unina.it

Abstract

In our approach for designing distributed intelligence systems, each object is enhanced by an index cell. Such an object is called a Tele-Action Object (TAO). Index cells behave like agents, however there can be numerous index cells. Objects enhanced by Index Cells (IC) can perform actions by themselves. Therefore, intelligence is distributed to these tele-action objects. Objects may also contain multimedia data. An active index consists of a network of index cells and can best be used in a distributed intelligence system. A lot of distributed knowledge can increase the overall intelligence of an intelligent system. This paper demonstrates the application of TAO and active index to the Eclipse Environment in order to distribute intelligence. A Multimedia Knowledge Eclipse Environment for Index Cell has been developed for these reasons. In this paper we will discuss also the advantages of using eclipse environment in order to manage distributed intelligence systems application showing an e-learning system.

1 Introduction

The problem of designing of an e-learning system is a particular case of the more general problem of designing a distributed multimedia system, in which communication and intelligence are linked together, using advanced techniques in artificial intelligence to bridge the gap between them. The solution is to distribute intelligence to the nodes of a complex system providing only a low level of intelligence for each object at each node.

A distributed computing system is therefore also a distributed intelligence system. The key application areas of such distributed intelligence systems include e-learning, tele-medicine, digital library and community network.

In our approach for designing distributed intelligence systems, each object is enhanced by an index cell. Such an object is called a tele-action object. Index cells behave like agents, however there can be numerous index cells. Objects enhanced by index cells can perform actions by themselves. Therefore, intelligence is distributed to these tele-action objects. Objects may also contain multimedia data.

Active index consisting of index cells can best be used in a distributed intelligence system. A lot of distributed knowledge can increase the overall intelligence of an intelligent system.

In this paper, we will demonstrate the application of active index and TAO to the Eclipse Environment in order to distribute intelligence. A Multimedia Knowledge Eclipse Environment for Index Cell ($MKE^2 4IC$) has been developed for these reasons. As an example of application we can consider the management of e-learning activities in e-learning courses. An e-learning system is also a distributed intelligence system, where the instructors and students are intelligent human beings. What we want to demonstrate is how to augment human intelligence by an active index using an Eclipse Environment.

2 TAO and Index Cell Architectures

A multimedia environment is composed of several media type that could be connected by temporal and/or spatial links. These links describes the media ongoing into the multimedia environment. It is necessary that the media are organized by an unique model. The model used is a result of a collaboration between University of Pittsburg and University of Napoli Federico II and University of Salerno [Arndt 2003, Maresca 2003, Guercio 2002, Arndt 2002, Guercio 1998]: the model is named Tele-Action Object (TAO). The models had a lot of enhancement [Piscitelli 2005, Pesce 2005, Sorrentino 2006, Scarfogliero 2006].
The state of the art is following illustrated.

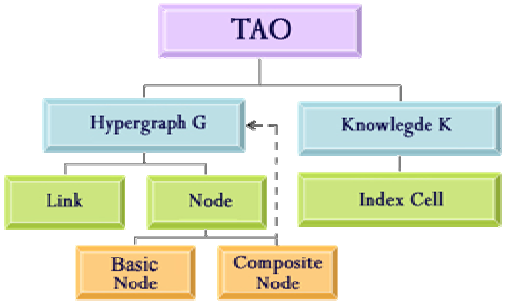


Figure 1. TAO Structure.

A Tele Action Object, can be defined as a multimedia object, having a static structure G and a dynamic structure representing knowledge associated to the object.

As we have seen in the last section, TAO is completely defined by the pair (G, K). Where the hyper-graph G is defined by the pair (N, L) with N representing the node set and L the link set. Each node represents another TAO or an icon and each link represents a relation between TAO's or hyper-graph nodes. Then the generalized icons represents the hyper-graph nodes and the spatial, and/or temporal operator are translated by links.

The knowledge structure is used in order to describe the environment in which the hyper-graph objects are defined and used. The knowledge structure allows an object to leave the passive state in order to assume the active one. In the active state the object can answer automatically to the input coming from the system outside.

This Knowledge structure has been implemented using Index Cells. Index Cells are Mealy FSM whose structure is showed below in fig. 2.

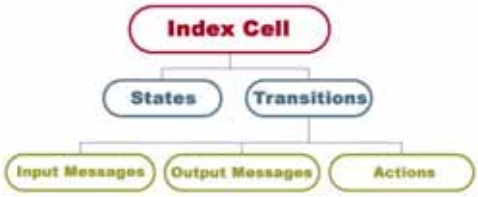


Figure 2. Index Cell Structure.

The diagram of a index cell is analogous to that one Mealy FSM, constituted, therefore, by states and transitions between them.

The model, moreover, previews that to every transition can be associated the Input Messages, which defines the condition of superability of the same transition, the Output Messages, that are sent from the Cell to the other System of Index Cells, when the transition becomes superskillful, realizing therefore a communication by exchange of messages, and elaborative actions, that are true computation executed when the state change.

This model has been completed during a research jointly conducted by Pittsburg University and University of Napoli Federico II. The model use XML in order to define a multimedia application based onto TeleAction Objects. For these reasons the model has been named TAOXML.

3 The Multimedia Knowledge Eclipse Environment architecture

As we have said, a multimedia system can be seen in terms of intelligent objects. Every object reacts differently depending on the input that it receives from the outside. The mechanism of answer to the stimuli can be realized associating a private knowledge to the TAO through the Index Cells. Then, the software that we have built must enable us to manage both the static description of the multimedia application (in terms of TAO objects) and the definition of Index cells net representing the dynamics structure.

Figure 3 shows the architecture of MKE^2 software in terms of use case diagram.

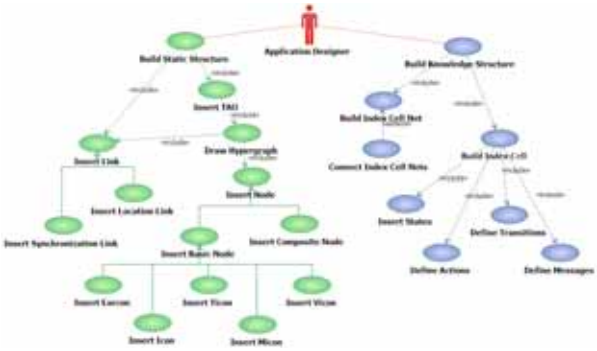


Figure 3. MKE^2 Use Case Diagram.

Particularly you can identify two principal use cases: Build Static Structure and Build Dynamic Structure. The first one describes the operations to be carried out in order to define the multimedia structure. The second one enable us to define the dynamic structure in terms of index cells. All the operations reported in fig.3 need to have a great user interaction by using a software editor. The authors have used the well known Model-View-Controller architecture pattern in order to build the editor. This model is the most useful one if you want to leave the business logic separated from presentation logic.

The Model-View-Controller architecture is shown in figure 4. It shows also the interconnection among components due to the complexity and relations among them.

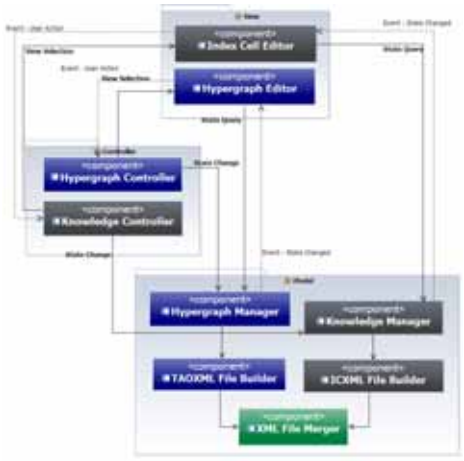


Figure 4. Model-View-Controller architecture.

In details we have the following components description:

- The Static Sub-System: The aim of this subsystem is to furnish the tools in order to build the multimedia static application in terms of TAO objects. The TAO architecture is translated into XML.
- The Dynamic Sub-System: The aim of this subsystem is to furnish the tools in order to buid the multimedia dynamic application in terms of Index Cells. The Index Cells Dynamic application is translated into XML.
- The XML File merger subsystem: This components has the aim to merge static and dynamic xml files generated by using the editor.

The architecture shown in figure 4 suggest the implementation of three plug-in using Eclipse architecture. Each one has been developed separately but the XML file merger subsystem that, up to now, is under advanced developing phase. All three components composes the MKE^2 integrated environment. The MKE^2 eclipse environment has high advantages if compared with a same environment developed with other architectures. In fact, the eclipse runtime enable us to design the knowledge net assigned to a TAO without loading the design tools for the TAO hyper-graph saving a lot of memory space.

4 Multimedia Knowledge Eclipse Environment

Following we will show the Multimedia Knowledge Eclipse Environment. The Eclipse software system integration suggests us the Graphical Modelling Framework in order to realize MKE^2 plug-ins. The reason of this choice is due to the separation between graphical data (icon, link etc) and semantic data (what the icon and links mean). Graphical Modelling Framework manage these separation using two different models. The Graphical Editing Framework is devoted to the management of Graphical data and another model is used to manage the semantic data. The same GMF models manage also MVC patterns according to the project design.

The XML File Merger subsystem is, up to now, in advanced developing phase.

All the components have been constructed using RUP construction Phase and Rational Software Architect tool using a Model Driven Software Development (MDSD) [Booch 1999, Maciaszek 2001].

5 Managing intelligent systems using MKE^2

There are a lot of advantages using Eclipse environment to cope complexity of intelligent systems. An active index can be used in order to manage e-learning activities in e-learning courses. An e-learning system is also a distributed intelligence system, where the instructors and students are intelligent human beings.

Eclipse platform could be useful since the modularity of the architecture and the huge availability of existing plug-in in the repository of the community. We are thinking therefore to develop applications of this type using the MKE^2 described architecture.

6 Multimedia System example description: An E-Learning System

We want to demonstrate the Multimedia Knowledge Eclipse Environment (MKEE) as an e-learning systems engineering tool able to realize the following multimedia system used jointly from the university of Naples Federico II and the university of Pittsburg (USA).

Let us consider an adaptive e-learning system that uses a web technology in order to manage a distance virtual course. Lessons has been organized by using an hypermedia structure composed of 5 multimedia (Lesson1 to Lesson5). The system must both distribute various difficulty lessons being based on the virtual class proficiency/deficiency and to make possible the access of course contents. The contents hypermedia structure, is passive, the access to it transforms in a active way.

Such structure can be transformed in an active one associating to it some index cells. To such aim it is opportune to describe the wished behaviour of the system, so that they can be modelled to opportune index cells that answer to such requirement: the system must distribute a lesson of greater difficulty, with respect to the actual one, when the threshold of proficiency (three students exceeds the verification tests of the current lesson) is caught up. In the same way the system must distribute a lesson of smaller difficulty when the threshold of deficiency (two students doesn't pass the verification tests of the current lesson) is caught up. Figure 5 show the multimedia system.

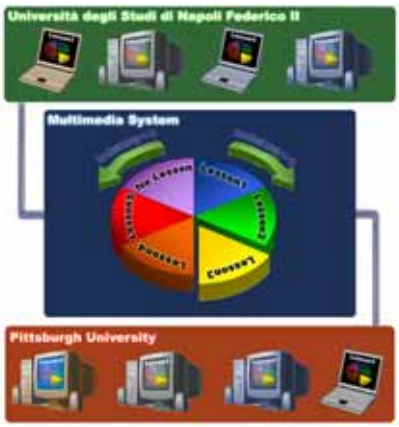


Figure 5. MKEE dynamic subsystem

System must trace all the events relating to proficiency/deficiency virtual class. Particularly it must be able to advise professor when it is necessary to rebuild the course and to block the distribution of its contents because the virtual class passed the test of the greater difficulty lesson or virtual class didn't pass the test related to the smaller difficulty lesson.

7 System design

The described system can be realized using TeleAction Objects (TAO):

- **Proficiency TAO.**
  The multimedia object is composed of:
  - A document accessed by the students that have overcome the attached verifications to the multimedia lessons with profit. It's aim is to communicate students the test result.
  - A Proficiency Index Cell (PIC), that if triggered, cause the proficiency level increase of an unity. When such level reaches the threshold of default proficiency (3), the Index Cell sends a message to the teacher, with the purpose to inform him that an enough number of students has reached the necessary proficiency level to increase the difficulty of the disbursed contents. The PIC sends besides a message to the Self-Adjustment Cell, with the purpose to increase the level of difficulty of the disbursed lesson.

- **Deficiency TAO.**
  The multimedia object is composed of:
  - A document accessed by the students that don't have overcome the attached verifications to the multimedia lessons. It has the assignment to communicate students the test result.
  - A Deficinecy Index Cell (DIC), that if triggered, cause the deficiency level increase of an unity. When such level reaches the threshold of default deficiency (2), the Index Cell sends a message to the teacher, with the purpose to inform him that an enough number of students has reached the necessary deficiency level that makes necessary to decrease the difficulty of the disbursed contents. The DIC sends besides a message to the Self-Adjustment Cell, with the purpose to decrease the level of difficulty of the disbursed lesson.

- **TAO Multimedia Lesson.**
  The main part of the Multimedia system is composed of:
  - A complex multimedia learning document, composed of five lessons on the same matter with increasing difficulty. Only one of them must be accessible in dependence of the general course of the virtual class.
  - Self-Adjustment Index Cell (SIC), that receiving the message harder, increases the difficulty of the learning material of learning. In the same way, when it receives the message easier, it decreases the difficulty of the learning materials with the purpose to make them simpler. Such Index Cell has besides the assignment to interrupt the disbursement of the lessons and to send a notification to the teacher in the two cases limit previously described (Lesson 5 and proficiency threshold. Lesson 1 and deficiency threshold).

The multimedia documents constitute the hypermedia structure (statics) of the system, the istances of index cells described previously are the knowledge structure (dynamics) of the system. Their integration has shown in the figure 6.
The system dynamic behaviour is described by the interaction diagram in among cells index (see figure 9), which show the message exchange among the cells that constitute the system knowledge structure.
Figure 7 shows also that more simple index cells has been added to the interaction diagram. They are necessary in order to comprehend the whole multimedia system behaviour: the index cells named Level1 to Level5 represent the knowledge structure associated to lesson multimedia documents. They have to send a message of output increment with the parameter increase:boolean=TRUE to the Proficiency Cell Index, if the student has gotten a positive result from the test; to the Deficiency Cell Index, if the student hasn't gotten an enough score. The Professor Index Cell, receives the notification messages sent by both the PIC and DIC, whose assignments are to notify to the Teacher that the level of difficulty of the multimedia lesson disbursed is respectively increased or decreased. The Professor Index Cells receives also messages of EasierLesson or HarderLesson from the Self-Adjustment Cell, when the disbursed contents are revealed respectively too much difficulties or too much simple for the virtual class.

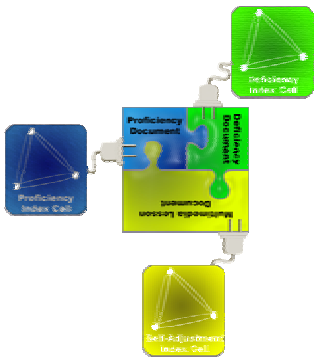


Figure 6. Tao-plug

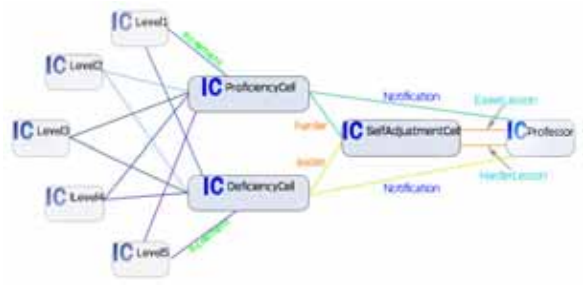


Figure 7. System dynamic behaviour

8 MKEE as e-Learning systems engineering tool.

The Multimedia Knowledge Eclipse Environment is the ideal tool to realize the illustrated system, because it allows to model its double structure (static and dynamic) through practical visual editors and to get a description of the whole system in TAOXML, a simple language obtained by the fusion of the language of the TeleActions Objects and the XML. The language of the TeleActions Objects, furnishes the semantic constructions to describe a multimedia system and the interactions among its elements; the language XML, allows instead to express such concepts through precise syntactic constructions, in the full respect of the multimedia ontology which constitute the base of the same language of the TeleActions Object, furnishing so consistence and rigor to the used model. It appears besides evident as a description in XML of the modeled system is able suffered to be used from a CASE tool for the automatic generation of the necessary code to the implementation of the system.

9 System Implementation using MKEE.

The use of an eclipse based development environment, allows us to treat the multimedia system as a new Tao Project, that is represented by a folder in the project explorer view of eclipse environment.

MKEE allows to separately manage the static and dynamics part of the system, dominating so the intrinsic complexity related to the double structure. For such a reason two different editors based on Eclipse GMF technology has been used.

The Multimedia Knowledge Eclipse Environment for Tao (MKEE4TAO), illustrated following, allows the multimedia designer to define the whole system static structure, in terms of TeleAction Objects, in simple and intuitive way. All it takes is selecting in the tool palette the elements that you want to insert and to position them on the Canvas, through drag and drop Operation. It is necessary, besides, with the purpose to give consistence to the system, to plan the ownerships of every inserted element, using the Property view, which allows a practical insertion of the values using a tabular structure. It is possible to sail the whole Canvas using the Outline view in the angle in low to the left (see fig. 8).

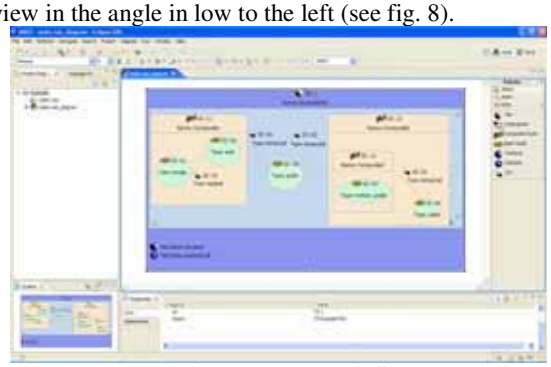


Figure 8. MKEE4IC

In analogous way it is possible to use the Multimedia Knowledge Eclipse Environment for Index Cells (MKEE4IC), (see figure 9) to draw the index cells that constitute the system knowledge structure.

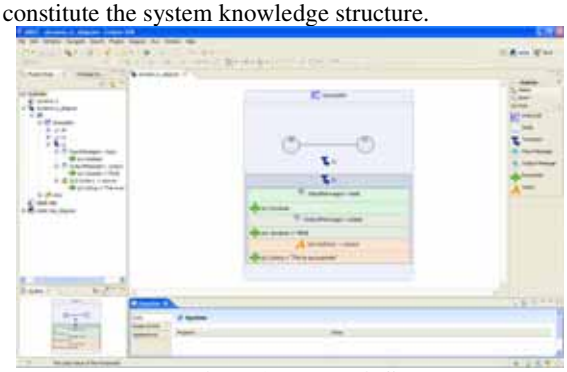
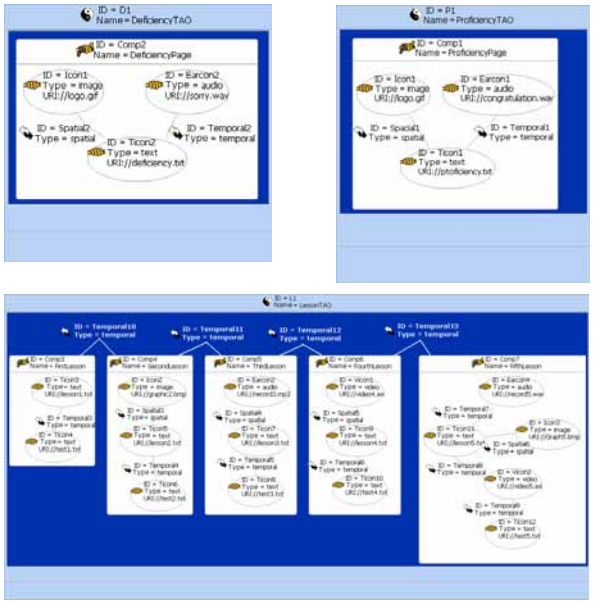


Figure9. MKEE4IC

Saving the project, the eclipse environment automatically produces two files, each for every structure. The first one with extension *.TAO contains the static TAOXML, the other one with extension *.IC contains the dynamic description of system. The following diagrams are drawn for the described objects:



10 Conclusions

In our approach an application of TAO and active index to the Eclipse Environment has been done to make possible to distribute intelligence in an intelligent distributed system.
Designing distributed intelligence systems, means to enhance each object with one or more index cells. Such an object is called a tele-action object (TAO). Index cells behave like agents and objects enhanced by index cells can perform actions by themselves. Therefore, intelligence is distributed to these tele-action objects. Objects may also contain multimedia data. An active index consists of a network of index cells (IC) and can best be used in a distributed intelligence system. A lot of distributed knowledge can increase the overall intelligence of an intelligent system. In conclusion we have demonstrated in this paper how to augment human intelligence by an active index and how to manage intelligent systems using an eclipse environment platform, realizing a distance learning system.

Bibliografy.

1 Arndt, T., Chang S.K., Guercio, A., Maresca P. (2003). An XML-Based Approach to Multimedia Software Engineering for Disctance Learning, Journal of Distance Education Technologies 1 (1): 40-62.

2 Maresca P., Arndt, T., Guercio, A., (2003). Unifying Distance Learning resources: The metadata Approach, Journal of Computers 13 (2): 60-76.

3 Chang H.J., Hou, T.Y., Hsu A., Chang S.K., (1995b). The Management and Application of Tele-action Objects, Journal of ACM Multimedia Systems 3 (5-6): 204-216.

4 Chang S.K., (1996). Extending Visual Languages for Multimedia, IEEE Multimedia 3 (3): 18-26.

5 Davis, M., Denning, K. (2000). Online learning: frontiers in the creation of learning communities. Proceedings of the Conference: Networked Learning 2000: Innovative Approaches to Lifelong Learning and Higher Education Through the Internet, University of Sheffield.

6 Guercio, A., Arndt, T., Chang S.K., (2002). A Visual Editor for Multimedia Application Development. Proceedings of the 22nd IEEE Conference: Distributed Computed System Workshop 2002.

7 Arndt, T., Chang S.K., Guercio, A., Maresca P. (2002). An XML-based approach to multimedia software engineering for distance learning. Proceedings of the 14nd International Conference on software engineering and knowledge engineering,2002.

8 Guercio, A., Arndt, T., Chang S.K., (1998). Visual Tool for a Multimedia IC Development Environment (MICE). Proceedings of the IEEE Symposium on Visual Language, 1998.

9 Chang S.K., CS2310 Exercise 2, http://www.cs.pitt.edu/~chang/231/231ex.html (2006)

10 Chang S.K., Towards a Theory of Active Index, http://www.cs.pitt.edu/~chang/365/ic.html (1995)

11 Scarfogliero G.M. (2006). A Multimedia Knowledge Eclipse Environment per le Celle Indice: Progetto ed Implementazione. Ph.D. Thesis, University of Napoli Federico II.

12 Sorrentino L. (2006). A Multimedia Knowledge Eclipse Environment per la gestione del TAOXML: Progettazione ed Implementazione. Ph.D. Thesis, University of Napoli Federico II.

13 Pesce M. (2005). La conoscenza applicata al TAOXML environment. Ph.D. Thesis, University of Napoli Federico II.

14 Piscitelli M. (2005). Un ambiente di ontologie multimediali basato su TAOXML. Ph.D. Thesis, University of Napoli Federico II.

Distributed e-learning with TAO through Eclipse and Grid computing

P. Maresca[1] , P. Donadio [2]

[1] Dipartimento di Informatica e Sistemistica (DIS), University "Federico II", Napoli
paomares@unina.it +39 081 7683168
[2] Alcatel-Lucent Italia - Optical Network Division (OND) – ALTA member
pasquale.donadio@alcatel.it +39 0828 398 694

Abstract

With the increasing popularity of the Grid Computing, both in Grid resources and Grid network environment, it is getting more imperative to provide a distributed mechanism that can enable multimedia content diffusion in a flexible and autonomic way. In this research we evolve the Tele Action Object (TAO) concepts to allow parts of the computational process to be executed remotely, in a distributed environment based on Grid. This paper proposes an active Grid architecture and middleware for a new generation of e-learning applications, based on self-managed and distributed Index Cells (IC). On the basis of the defined architecture we have implemented a prototype of Grid based TAO that validates our thesis. The prototype is realized integrating two open-source technologies: "GLOBUS" of the Globus Alliance and "Eclipse Callisto".

Keywords*: Tele Action Objects, Grid Computing, Autonomic systems, Eclipse, E-learning.*

1. Introduction

The widespread Internet is the living environment of multimedia services based on Grid computing technologies. A basic premise of Open Grid Services Architecture (OGSA) is that everything is represented as a service, which is a network-enabled entity that provides capability to users. In addition there is a network paradigm shift dictated by the need of rapid and autonomic service creation, deployment, activation, and management combined with context customization and customer personalization. Such a motivation can be traced in different organizations and research activities as well as market forces. Since Grid provides an infrastructure for the interoperability of legacy applications in distributed environments, research people have started using Grid concepts to integrate and enhance knowledge creation and visualization systems.

In this paper we propose a new paradigm for Distributed Knowledge Management [1] based on Service Oriented Architecture [2], using Eclipse and Grid technologies. The tool used to add programmability to Grid-enabled creation and visualization systems is extension to the widely used supporting tool, Eclipse, which is also a powerful supporting tool for Eclipse Italian Community. Moreover an application of the architecture has been done into e-learning domain. The organization of the paper is as follows. In section 2 we introduce background and related works of the Tele Action Objects paradigm. In section 3 we position a layered architecture based on Grid that enhance the actual TAO in a distributed fashion. In section 4 we propose the Ubiquitous and Autonomic - Index Cell model (UA-IC) that enhances the actual Active Index model solving in a brilliant way the well-known TAO synchronization problem. Section 5 presents the Globus agnostic middleware integration in Eclipse that enables the TAO on the Grid and Web infrastructure. On the basis of this infrastructure, in Section 6 we show a short example of use of Distributed - TAO in e-learning.

2. Background and related work

The landmark Journal of 'Visual Language and Computing' by S.-K Chang [3] sparked the development of a number of important visualization and multimedia systems. Many of these were based on the notion of dataflow whereby the overall visualization process is broken down into smaller parts, or modules, which are connected in a network. Data flows through this network, being transformed finally into multiple media resources represented by images, videos and audio. Key contributions

were the MICE system described by Guercio et al [4], and the elegant reference model for Active Index by S.-K Chang [5]. An important application of visualization in distance e learning is to share and view data in a distributed environment acting a quasi-broadcasting system for knowledge management and diffusion. This paper by Maresca et al [6] is a key contribution in understanding the distinction between using visualization for on-line computational steering, and for off-line post-processing of data.

3. Distributed TAO: a layered architecture based on Grid

We see the multimedia visualization process as an ordered sequence of work tasks where results from one work task are input to a subsequent work task. This is commonly realised as a TAO dataflow process as presented in the traditional reference model [7]. The novelty introduced in this paper is a new paradigm, the Distributed TAO (D-TAO) that reuses the basic principles of TAO (Active Index) and adapts it in a Grid-based environment. In this scenario the D-TAO dataflow pipeline could be viewed at different levels of abstraction: a top level independent of resources and low levels where resources are progressively bound in; the paper focus the following tree layers:

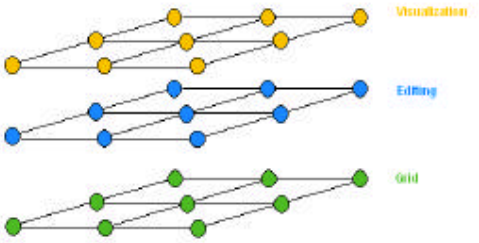


Figure 1 – Distributed TAO using GRID

The **Visualization plane** manages the multimedia flows specifying how knowledge should be transformed to pictorial data. In other words it represents the set of nodes that provides services about the visualization of the multimedia contents. In this paper we extends the concept of knowledge visualization to grid computing and collaboration, by associating Index Cell (IC) to distributed dataflow, and specifying interconnecting links between these Active Index. The visualization plane is not static, but evolves over time as the user explores different learning strategies and levels. In the second layer of the reference model, the **Editing plane**, we bind the Visualization plane to a particular configuration of the software entities: the knowledge creation. In other words we indicates the complete set of nodes attached to the Grid that have the full access to the TAO editing, manipulation and visualization. The Editing plane is expressed independently of any compute resources; it does introduce constraints on the resources required. (e.g. particular time-delay characteristics, processor, QoS for network links and requirement for co-location of components to ensure performance and other criteria to met). Resource constraints are not static and may evolve dynamically as a session progresses (for example new data sources might be brought into play requiring computation to migrate to satisfy performance criteria). The dynamic resource constrains is treated in the next chapter, introducing the concept of "ubiquitous" and "autonomic" Index Cell agents (UA-IC). Finally, the **Grid plane** represents the physical layer that interprets the editing plane specification in terms of particular Grid computing environment. The processes are bound to specific compute resources – but again the binding is not static, rather it evolves as the editing plane changes over time. A resource broker could take the editing specification, the resource requirements and resource availability, and carry out the transformation between editing and physical layers automatically. This layer is implemented using GLOBUS as depicted in the next section. Distributed TAO (D-TAO) represents a new paradigm that introduces concepts of ubiquitous and autonomic Index Cell, to support a layered distributed infrastructure that transform a centralized multimedia dataflow in a distributed one.

4. Ubiquitous Index Cell (U-IC)

The formal specification of the TAO is based on three basic concepts: the Index Cell Base (ICB) that consists of a finite/infinite number of index cells; an active index (IX) that consists of a finite number of Index Cell from ICB; the Index Cell [8], that performs the logic act to manages multimedia contents stored and processed on a computational system locally (prior art). Active Index related to the TAO local computation consists in a time varying collection of index cell

in different states, accepting certain input messages and posting output messages of the output list. In the traditional TAO, synchronization of Active Index could be defined directly at TAO editing time. TAO primitives (e.g. XML based) define the static and dynamic descriptors that tune this parameter, taking in account the multimedia loading latency (mll), multimedia visualization latency (mvl) and the multimedia execution time (met). Distributed TAO (D-TAO) introduces a variant to the traditional IC model. In this case ICB is completely represented using the grid plane metaphor. In other words IC are distributed through the nodes of the Grid environment. The idea behind D-TAO is very simple: define a distributed Tele Action Object environment that controls and share, on top of a Grid, multimedia resources performing the resource editing and visualization, in real-time. In other words D-TAO introduces an overlay network that includes one or more computational systems that cooperate in order to perform both the data gathering and processing of multimedia contents stored and shared on a distributed environment. In this scenario Index Cell, that represents the knowledge system are ubiquitous and distributed over the Grid in order to be consumed dynamically (U-IC). In a distributed environment the synchronization of the ubiquitous IC (U-IC) is function of the network latency. Importunely the network delay is not a predicable parameter, because is dynamic and depending from the network topology. This implies that is not possible define synchronization of the Grid Index Cell using the same criteria also used for the traditional IC. The novel idea is based on the concept of "autonomic" U-IC that permits to self-synchronize the Index Cell using signalling information coming from the Grid environment. In this way we transforms the open chain Index Cell model (traditional) in a close chain system, where feedback information are recovered using the signalling channel defined on top of the grid. Autonomic U-IC could be implemented in a software agent connected to the grid that implements the Index Cell logic (state machine) and synchronize it taking in account the network topology and dataflow distribution.

UA-IC is integrated in each node of the Grid environment. In other word the logic that implements the UA-IC is wrapped in a Grid service and exported in Eclipse environment. Details about this integration are deeply analyzed in the next section.

5. GLOBUS and Eclipse integration

The architecture for Grid Computing is defined in OGSA standard that describes the overall structure and services to be provided in Grid environments [9]**.** Building on concepts and technologies from the Grid and Web Services communities, OGSA architecture defines uniform exposed service semantics (the Grid service) and standard mechanisms for creating, naming, and discovering transient Grid service instances; due to its many benefits, OGSA and Globus [10] [11] are adopted in this paper as a guideline for Grid services [12]. This architecture aims to provide mechanisms automatically adapting Grid network elements to different Grid services and the management of the Grid system itself.

In our dimension, Grid acts as supplier of "Editing" and "Visualization" services; Grid management services are UA-IC agents. In other words the natural evolution of the TAO paradigm is now based on a Service Oriented Architecture (SOA) where multimedia editing and streaming are distributed over the Grid using the "service" approach.

Consumer of Editing and Visualization Services are applications that, implementing the service interface, are able to connect to the Grid and share, in a very transparent way, mechanisms, multimedia contents, knowledge and logics.

In our architecture the consumer is implemented using only one open-source application: Eclipse Callisto. Each different service offered by the platform are "consumed" using a different plug-in. In details the GLOBUS framework acts as distributed communication service that supplies all the possible services required by the e-learning platform; Eclipse act as Graphical User Interface for TAO Editing and Visualization simply using the plug-in concept as "glue " technology. Finally the plug-ins "consumes" the services offered by the Grid platform. Each plug-in is composed by three elements: the eclipse interface, the Grid interface and the logic that implements the UA-IC agent, in order to supply or consume multimedia contents. Next figure shows details about the Eclipse-Globus integration based on the pluggable approach**:**

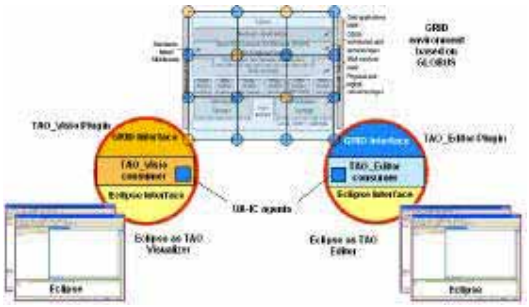


Figure 2: Eclipse-Globus integration

Next section shows details about the implemented prototype.

6. An E-learning simple application based on Grid

In this section we introduces an e-learning adaptive system based on Grid. Lessons are organized through one hypermedia-distributed structure [13]. Such structure remains passive until it does not come visited from the students. Such structure can be rendered active associating the index cells [14] with multimedia documents. The idea is to construct a special document that, if visited from some students, give us notification that they have caught up a sure level of proficiency and therefore the learning materials would have to be modified in order to become more difficult [17]. In the same way, when some students approach to a special document that indicates deficiency, it means that they have found problems and consequently the learning materials would have to be rendered simpler [18]. The special document is ubiquitous, in the sense that knowledge contents are distributed among the nodes of the e-learning network, attached to the same Grid. The model behind the ubiquitous document management is based on three concepts:

Ubiquitous proficiency-level index cell (U-PLIC): This index cell is associated with a specific multimedia document shared on the grid (reachable only from the students well prepared). When it comes primed, the level of proficiency comes increasing of 1. When the level of proficiency has caught up a predefined threshold (as an example 3), the Ubiquitous Index Cell will send to the professor a message that will inform him that a sufficient number of students have caught up this level of proficiency. The U-PLIC will send also messages to some multimedia documents, to the aim of increasing the level of difficulty.

Ubiquitous deficiency-level index cell (U-DLIC): This index cell is associated with a specific multimedia document shared on the grid (reachable only from the students with deficiencies). When it comes primed, the level of deficiency comes increasing of 1. When the level of deficiency has caught up a predefined threshold (as an example 2), the Index Cell will send to the professor a message, than will inform him that a sufficient number of students have caught up this level of deficiency. The U-DLIC will send also messages to some multimedia documents, to the aim to reduce the level of difficulty.

Ubiquitous self-adjustment index cell (U-SAIC): This Index Cell is associated to multimedia documents containing e-learning material. When it receives the message harder, it increases the difficulty of the learning material. In the same way, when U-SAIC receives the message easier, it reduces the difficulty of the learning material to the aim to render them simpler. The instances of Index Cells are associated to multimedia documents distributed over the Grid. Self-adjustment is implemented using feedbacks received on adjacent Grid nodes.

The simple e-learning application [19][20] focus two kinds of lessons distributed over the grid network:

- A more difficult lesson when three students approaches to a multimedia document named Proficiency Document.
- A simpler lesson when two students approaches a special document named Deficiency Document

The distributed multimedia lesson contains five different levels of difficulty.

In this scenario, the hypermedia structure of the e-learning application is composed of:

a) Proficiency Document: it is a special document, which can approach only the students who have exceeded the annexed verifications to the multimedia lessons with profit. The document communicates to the student its result.

b) Deficiency Document: it is a special document, which can approach only the students who have not exceeded the annexed verifications to the lessons with a sufficient result. The document communicates to the student its result.

c) Multimedia Lesson: it is a complex multimedia document, containing the e-learning material. It is divided into five lessons on the same argument with increasing difficulty.

Only a lesson at once must be accessible depending on the proficiency of the virtual class. All these three documents are expressed in terms of TAO [15][16]. As example the TAO hyper-graph could be composed of a macro node, to the whose inside is defined a Ticon, containing a text of congratulations that is shown to the customer, one Icon, containing logo of the department and one Earcon content an audio message of congratulations. The Earcon is tied to the Ticon with a temporal link, in order to synchronize the audio execution with the reading of the text. The Icon is connected to the Ticon using a spatial link that defines the mutual position of the two generalized icons. The described objects represent the hypermedia structure of e-learning application. They constitute a well-defined multimedia system, because it contains the mechanisms that regulate dynamics of the application (see the nature of the UA-IC). In other words the TAO multimedia structure is now complete, because equipped with an autonomic mechanism that react to the user input dynamically. Next figure (3) shows the knowledge network associated to the described multimedia object, distributed on each node of the e-learning Grid.

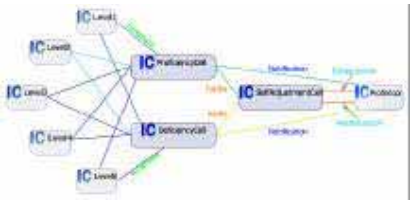


Figure 3. Knowledge structure associated to multimedia objects

The IC Level1-Level5 represent the knowledge structure associated to multimedia documents representing the distributed lessons. Their task are to send an output increment message, with the parameter increase: boolean=TRUE, to the IC Proficiency, if one or more students have obtained a positive result from the test. The IC Proficiency will send a message to the IC Deficiency (Boolean=FALSE), if one or more Students have obtained insufficient score. The IC Professor receives the notification messages sent from both the ubiquitous IC Proficiency (U-ICP) and IC Deficiency (U-ICD) distributed on the e-learning Grid. The U-ICP and U-ICD have the tasks to notify the Teacher that the level of difficulty of the distributed multimedia lesson is, respectively, increased or diminished; and it receives messages of Easier Lesson or Harder Lesson from the Self-Adjustment IC, when the contents distributed are, respectively, too much difficult or too much simple (relating the learning Grid).

Next figure (4) shows a complete view of the proposed e-learning application. In this scenario the knowledge network is distributed over the Grid:

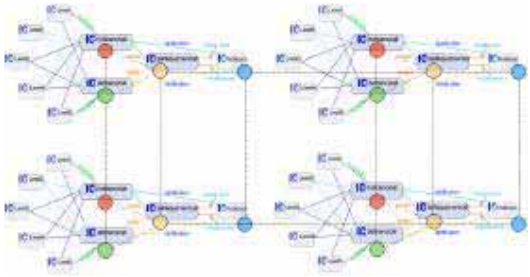


Figure 4. Knowledge structure associated to multimedia objects and distributed over the Grid

Red and green nodes represent the Ubiquitous Proficiency and Deficiency cells. Yellow nodes represent self-adjustment cells. Finally blue nodes represent the professor cells. The distributed organization of Index Cells on the Grid is more powerful: different class of cells communicates and shares information over a Private Learning Network (PLN) where the grid middleware acts as privileged signaling channel. In other words, using this new approach, professor could have in real time information about the average preparation of one or more virtual class (not only a limited group of students); self-adjustment cells could tunes the e-learning grid in a very precious and performed way (e.g using the average proficiency and deficiency information related to more students communities). The knowledge could be distributed and reused on different nodes, creating different network dichotomies, per-faculty, per-courses and per-specializations.

Conclusions and future works

In this paper we have proposed a new architecture that use the basic principles of the Grid computing and apply them to the TAO paradigm, in order to propagate multimedia knowledge over a distributed system.

On the basis of the defined architecture we have implemented a prototype of Grid based ubiquitous Index Cell (U-IC) Agent that validates our thesis. The prototype is realized integrating two open-source technologies: GLOBUS of the GLOBUS Alliance and Eclipse: on the basis of the U-IC implementation the simple application introduced in chapter 6 could be completed.

This demonstrator, developed by the Alcatel Lucent Italia Laboratories and University of Naples Federico II, is now in beta-test beside the Laboratories of the same University.

References

[1]Bonifacio, M.; Bouquet, P.; and Traverso, P. 2002. Enabling distributed knowledge management. Managerial and technological implications. Novatica and Informatik/ Informatique III(1).

[2]Douglas K. Barry. Web Services and Service-Oriented Architectures: The Savvy Manager's Guide. San Francisco, Morgan Kaufmann Publishers, 2003. ISBN 1-55860-906-7

[3]Chang S.K., (1995). Towards a Theory of active index, Journal of Visual Languages and Computing 6 (1): 101-118.

[4]Guercio, A., Arndt, T., Chang S.K., (1998). Visual Tool for a Multimedia IC Development Environment (MICE). Proceedings of the IEEE Symposium on Visual Language, 1998.

[5]Chang S.K., Towards a Theory of Active Index, URL:< http://www.cs.pitt.edu/~chang/365/ic.html >

[6] P. Maresca, T. Arndt, A. Guercio3, P. Donadio, (2003). A Visual Editor for Multimedia document management based on the TAO_XML paradigm.

[7] P. Donadio, "Progettazione ed implementazione di un ambiente per la costruzione, rappresentazione e trasformazione di documenti multimediali basati sul paradigma Tele Action Objects - XML based (TAO_XML)" Laurea degree thesis, Università degli studi di Napoli and Cleveland State University.

[8] Chang H.J., Hou, T.Y., Hsu A., Chang S.K., (1995b). The Management and Application of Tele-action Objects, Journal of ACM Multimedia Systems 3 (5-6): 204-216.

[9]Daniel Minoli - A Network Approach to GRID Computing - Wiley Interscience A John Wiley & Sons,INC.,Publication;

[10] The Globus Alliance http://www.globus.org/;

[11] Foster I. and Kesselman C. "Globus: A Toolkit-Based Grid Architecture". Foster, I. and Kesselman, C. (eds.), The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, 1999. Pages 259-278;

[12] Ian Foster, Carl Kesselman, Jeffrey M. Nick and Steven Tuecke1. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. Open Grid Service Infrastructure Working Group (OGSI-WG) of Global Grid Forum.

[13] Arndt, T., Chang S.K., Guercio, A., Maresca P. (2003). An XML-Based Approach to Multimedia Software Engineering for Disctance Learning, Journal of Distance Education Technologies 1 (1): 40-62.

[14] Maresca P., Arndt, T., Guercio, A., (2003). Unifying Distance Learning resources: The metadata Approach, Journal of Computers 13 (2): 60-76.

[15] Chang H.J., Hou, T.Y., Hsu A., Chang S.K., (1995b). The Management and Application of Tele-action Objects, Journal of ACM Multimedia Systems 3 (5-6): 204-216.

[16] Chang S.K., (1996). Extending Visual Languages for Multimedia, IEEE Multimedia 3 (3): 18-26.

[17] Chang S.K., CS2310 Multimedia software engineering, [documento WWW] URL:< http://www.cs.pitt.edu/~chang/231/231syl.html > accessed on 15th October 2006

[18]Chang S.K., CS2310 Exercise 2, [documento WWW] URL:< http://www.cs.pitt.edu/~chang/231/231ex.html > accessed on 15th October 2006

[19] P.Maresca, S.K. Chang, M. Pesce - Application of Active Index to the Management of E-learning Activities (2005)

[20] Scarfogliero G.M., Sorrentino L. (2006). A Multimedia Knowledge Eclipse Environment (MKEE) Thesis, University of Napoli Federico II.

Teaching with Eclipse through the simulations

Paolo Maresca
DIS Università di Napoli Federico II,
Italy - paolo.maresca@unina.it

Abstract

The ECLIPSE community is made up of developers, and all of them contribute to the improvement of the platform, from all points of view. ECLIPSE is used in schools, universities and for long-life learning, and there are several reasons why it was chosen: it is open source, has many plug-ins, allows collaborations and is free. The use of the eclipse platform in the didactic activities doesn't constitute only the use of an open source software but a philosophy of work that creates a community around a common objective and facilitates the transfer of knowledge between people.
This paper shows as a community of students grows, organizes, transfers knowledge and prepare didactic resources to the purpose to form own knowledge around a didactic objective by using simulations.

1 Introduction

The advent of the new computer technologies in e-learning domain has not changed the way of teaching ! The student always learns from the language.

The simulations are the innovative example of the potentialities of the new computer technologies applied to the e-learning. In all the schools, the language and the learning through it still constitutes the only way to teach and to learn. Although the language allows us to describe and to abstractly represent the realities, many are the limits to its use:

1. ***Motivation***: to learn from the language (written or spoken) it is boring and it asks for strong concentration and a great ability in effecting connections among concepts.
2. ***Superficiality***: the language is able not to be the proper tool to understand in depth the concepts in how much arousing only abstracted reasoning forces to an accumulation of mnemonic knowledge.
3. ***The wealth or the level of knowledge***: To learn through the language needs to know the language. A person that has a limited dictionary cannot easily learn through it and he/she will have some problems and a lazy and limited learning.

A good learning asks for understanding, movement among the concepts also in the practice, connections with preceding experiences, with done experiences.

The problems of the use of the language have become even more important because of the changes that are happening in our society. In fact while in the past the education was reserved to a limited elite of people that knew well the language today the education is addressed to all putting aside from the fact that these people have a complete knowledge of the language. Besides, in the actual society, the people are subject to the long life learning [1,6] and therefore the people have to access the continuous renewal of their knowledges really through the language and and it is not possible that these people improve their own language before acquiring the knowledges that will serve them to improve their own abilities and knowledges in the working field.

The knowledge of a frank language (es. the english language) in a global environment still sets problems. With the advent of internet and the networks it becomes possible to learn and to study through cooperating systems sharing resources, knowledge opening ourself to the cultures and to the experiences. Besides the choice of a common language resolves partly the problem (for a small elite of people) because a lot of people don't know the common language and it reverts in the limitation 3 for which learning in a virtual and global environment (and not in a real classroom) anymore becomes really an obstacle.

The actual state of the things is that we always use a language, sometimes a frank language to learn the novelties and we do it through new technologies.

This however it doesn't improve the learning, it has many costs and the advantages they are not commensurate to the disadvantages.

In this job we want to show that the frank language is not a language but a collection of experiences – named also simulations collections- communicated through an e-learning virtual environment where every experience can be simulated and every thing represents an example of virtual laboratory that can be shared and to constitute the occasion of learning for other people: it is the phenomenon of the collaboratory (collaboration+laboratory). A collaboratory doesn't have limits neither in the simulation phenomenons, neither in

the greatness of the phenomenons. The e-learning virtual environment needs a powerful interface. The choice is fallen on ECLIPSE [2,3,4,5].

2 The Simulation and ECLIPSE

Experience is a strange teacher: first it examines yourself and then it explains the lesson learned. The introduction of the new technologies of the information and the communication (TIC) offer for the first time the possibility to learn on everything (or almost) through to see and to know how to do. The TIC allow to acquire a *know how to do* spendable in the world of the job not only through the oral language. It deals with a to see and to know how to do that don't directly have for object the real things but they have for object visualizations, animations, simulations of the things, inside environments more or less dipped you that they resemble to the real environments.

The student sees the objects, the phenomenons, the mechanisms, the processes, act on the conditions, the variable, the parameters that influence them, and observing the consequences of these manipulations, he understands the nature and the operation of it. Only that all this doesn't happen in the reality but in simulations of the reality.

Despite the new digital technologies offers revolutionary possibility for that concerns the different way of learning (ex. the to see, and to do), the use that has been done up to now has not at all exploited these possibilities but the behaviour has been: *to put old wine in new strokes*, that is to use the traditional channel of oral learning (texts to be read, lessons oral video recorded, discussions on internet, forum, etc) in the new container constituted by the new technologies (for instance a classical environment of and-learning).

Reading a book, watching a picture or a movie or, more rarely, the television the mind works and our emotional life is in movement. But we receive the perception of when we realize that what is outside in relationship to our actions: this is only possible with a computer that simulates hibernated reality !

A simulation is a model of a some aspect of the reality (natural, social, historical, technological, working, etc) expressed in the form of a computer program. When a simulation is dressed again by an appropriate interface that allows the student to see things and to act on them, the simulation becomes an important tool of understanding and learning. In a simulation the language develops entirely a marginal role as channel of communication and learning, while the simulation works because consumer sees the simulated phenomenons and acts on them observing the effects of his/her actions. The fact that then the simulations are the reality it constitutes for the science a novelty. But the simulations are very important because they make to penetrate more deeply in the science and they distribute more quickly the knowledge among those people who use it. In short the simulation becomes once a tool of learning that the simulation has been built a student can learn to know and to understand that part of the reality interacting with simulation; that is on the commands of the computer, and observing as simulation reacts to our actions.

For instance in the fig. 1 are brought a simulation used in our didactic laboratory that allows the student to interact with the not simple environment of eclipse with the purpose to install a perspective that is him useful for the compilation of the languages c/c++. An assignment that if faced through the traditional formative channels (language, reading manuals, resolution problems, consultations forum, reading chat etc) it has middly covered since 3 to the 5 hours. The same simulated assignment allows the student to be operational on the didactic laboratory in less than 30 minutes. The lapel of the medal is the preparation of the same resource and is clear that this is due to some better student of others but the enormous advantage it is secret in the fast transfer of the knowledge from a side (in fact the resource is that more downloaded by the site and by all italian community) and in the middle growth of the knowledge of the community of practice that reaches an objective in way very more rapids that with the use of the spoken language

Figure 1. ECLIPSE CDT Simulation of installation.

The second good motive is that the interface ECLIPSE can be multiform because of its perspectives. This is very important in an environment of simulation in which the virtual didactic laboratory can be changed (also changing the experiences) without changing the interface. In other words all languages can be used. For instance to stay in theme of the fig.1, all the languages that are wanted without modifying the editor that will entertain them and without having problems to recall different compilers. In the fig. 2 are for instance shown the perspective java.

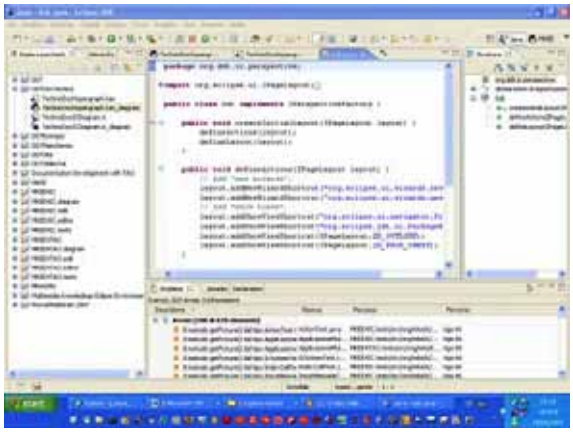


Figure 2. Virtual Java Lab.

In the figure 3 a virtual multimedia lab [7,8,9,10, 11] is shown through which the student interacts also creating complex simulations of real systems as for instance those economic, medical etc.

Finally in the figure 4 and 5 are shown an environment for the construction of documentation of every type to software support, courses and other [12].

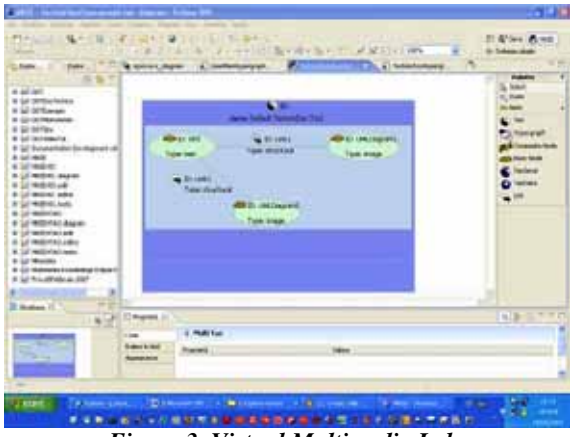


Figure 3. Virtual Multimedia Lab.

They would seem finished the advantages but it is a quarter that is external to the interface ECLIPSE of it but that it is related to the experience accumulated in the use of the same interface and therefore in the use of the simulation and as this experience is appraised and how much quickly this can be transmitted acquired and magnified. In this the practice community ECLIPSE has shown to be a solid and meaningful community of knowledge transfer and the innovation in which the consumers are the carrying one of the innovation that flows through the net from individual to the other.

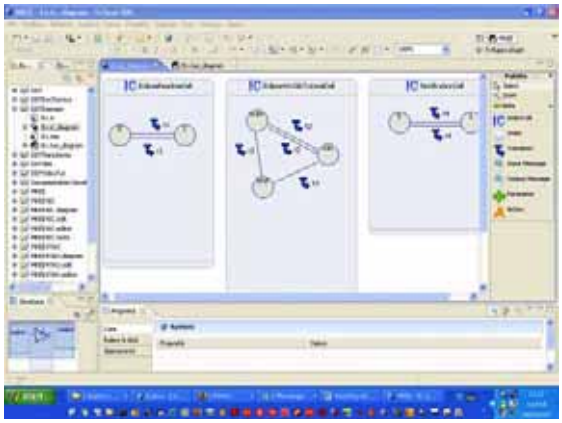


Figura 1 Eclipse Documentation system index cell

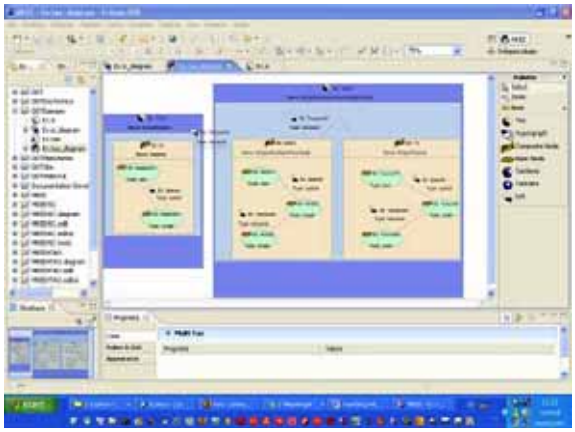


Figura 2 Eclipse documentation system TAO diagram

3 Conclusions

With software open source as eclipse are made many footsteps before but: will it be easy to put the new wine in the new barrels of the new technologies? The answer is that it won't be easy. The simulations to be well done cost and organized communities are not generally prepared to spend money and resources unless same communities are not open. Besides the simulations are innovations that change deeply the traditional disciplines and it is really for this that their penetration will be slow. The simulations as tools of learning require both competences regarding the new technologies that competences regarding the abilities human cognitive and these two competences are rarely found reunited in the same person or team. There are then other more specific obstacles. The simulations are tools of learning in which the language has a marginal aspect while the language still represents a fundamental carrying axle of the universal scholastic arrangement. Undeniable then that the simulations involve a certain dose of automation of

the learning that, some teachers, don't see of good eye because of the marginalization of their role. But it is obvious that with the simulations the role of the teacher changes changing their assignments and their competences. And these are other obstacles. Nevertheless in the university teaching the simulations offer many more advantages because the teachers are formed to change more quickly them of it competences and their assignments.
Some small opening of optimism comes from the open source communities which often produces simulations. Before using it these artifact must be revised and often modified. The Eclipse community can be a good opportunity for the production of qualities simulations for the didactics.

Bibliografy.

1 Bologna pan-European agreement, 1999: http://ec.europa.eu/education/policies/educ/bologna/bologna.pdf.

2 Eclipse community http://www.eclipse.org .

3 Eclipse italian community http://www.dis.unina.it/eclipse

4 Gamma, E.,Beck, K. (2005). Contributing to eclipse: principles, patterns and plug-ins.

5 D'Anjou J., Fairbrother, S., Kehn, D., Kellerman J., McCarthy, P., K (2006). The Java developer's guides to eclipse.

6 Davis, M., Denning, K. (2000). Online learning: frontiers in the creation of learning communities. Proceedings of the Conference: Networked Learning 2000: Innovative Approaches to Lifelong Learning and Higher Education Through the Internet, University of Sheffield.

7 Maresca P. ,Chang S. K., M. Pesce (2006), Application of Active Index to the Management of E-Learning, Rivista della società italiana di e-learning Rivista Si-el, vol. 3, n. 2, 2006, November, pp. 331-341.

8 Maresca P. ,Scarfogliero G.M., Sorrentino M. (2007), Multimedia Knowledge Eclipse Environment: An Eclipse Tool for Multimedia Application Design, in Proceedings of Thirthenth International Conference on Distributed Multimedia Systems (DMS2007) – Distance Education Workshop, San Francisco , USA, September 6 - September 8, 2007.

9 Scarfogliero G.M. (2006). A Multimedia Knowledge Eclipse Environment per le Celle Indice: Progetto ed Implementazione. Ph.D. Thesis, University of Napoli Federico II.

10 Sorrentino L. (2006). A Multimedia Knowledge Eclipse Environment per la gestione del TAOXML: Progettazione ed Implementazione. Ph.D. Thesis, University of Napoli Federico II.

11 Pesce M. (2005). La conoscenza applicata al TAOXML environment. Ph.D. Thesis, University of Napoli Federico II.

12 Marino A. (2006). La progettazione ed il riuso di documentazione in ambiente eclipse utilizzando il multimedia Knowledge Eclipse Environment. Ph.D. Thesis, University of Napoli Federico II.

Improving student's Self-Efficacy using an Adaptive Approach

Rita Francese°, Ignazio Passero°, Giuseppe Scanniello* and Genoveffa Tortora°

° Dipartimento di Matematica e Informatica, Università degli Studi di Salerno,via Ponte don Melillo 1,Fisciano (SA), Italy

* Dipartimento di Matematica e Informatica, University of Basilicata, Viale Dell'Ateneo, Macchia Romana, 85100, Potenza, Italy

francese@unisa.it, ipassero@unisa.it, giuseppe.scanniello@unibas.it, tortora@unisa.it

Abstract

The evaluation of the self-efficacy and its relationship with performance helps students in obtaining an authentic view of their current beliefs and abilities. In this paper we propose a model, named Self-Efficacy Assessment (SEA), that investigates this relationship. Honest students will be prized in the computation of the test's score and adaptive feedback is provided to enhance their self-efficacy. A prototype supporting the proposed model has been developed as a MOODLE plug-in.

1. Introduction

The crisis of the scientific vocations is a phenomenon which involves both Europe and US. The loss of interest in scientific studies risks producing a shortage of scientists. On the other side, often college students arrive in introductory scientific courses with varied levels of confidence and competence. In the specific case of computer science, some of them believe that the discipline is easy and a high percentage of them has not the pre-requisite to success in this kind of studies. As a consequence, they do not invest much preparatory effort in it [1]. Others may have the attitude, but they can have had difficulties with previous studies which had discouraged them.

Also for lifelong learning, to acquire the skills necessary to perform newer and newer activities, human resources have to be able to evaluate their performance accurately and consequently become more skilled at assessing their learning needs. Thus, when the students have to choose their study address or, more generally, when a person has to begin a new activity, a self-evaluation tool could have the potential to help this process. Particular relevance has the measurement of self-efficacy, the student's belief on his capacity of performing a task. The evaluation of the self-efficacy and its relationship with performance helps students in obtaining an authentic view of their current beliefs and abilities [11]. Indeed, it is very important for a student to identify areas in which he or she is relatively effective and areas in which he/she is relatively ineffective, given limits on the time and resources available for self-improvement efforts [10]. These kind of evaluation is natural in face-to-face assessment, where the teacher can verbally highlight to the student his/her potentialities or trouble while (s)he performs a given task. In computer based assessment, the lack of a teacher having this role is particularly relevant and can be overcome introducing self-efficacy assessment [6].

In this paper we propose a model that investigates the relationships between the self-efficacy of a student and its real competence. The mark of a test takes into account both these factors to invite the students to take in the appropriate consideration the questions concerning their confidence. Honest students will be prized in the computation of the test's score. The scoring scheme can be used also as summative evaluation. A prototype supporting the proposed model has been developed as MOODLE plug-in, named Self-Efficacy Assessment (SEA). The prototype enables teachers to choose the feedback typology (textual, numeric, graphical or both) and whether to provide a feedback for each questions or only a summative report. Moreover, the text of the messages corresponding to the various level of confidence can be personalized.

The remainder of the paper is organized as follows: Section 2 presents the related work. The proposed methodology and tool are described in Section 3, while the adaptive feedbacks are described in Section 4. Final remarks and future work conclude the paper.

2. Related work

Relationships between confidence and competence have been largely investigated in medical sector [1][8][12].
A study conducted on biology undergraduate students revealed that they do not always know what they do know [13].
Empirical studies have been also conducted to compare the confidence and the competence in software skill [8]. The results indicated that although students were confident in their ability to perform basic software application tasks, their actual performance level was consistently below their confidence level.
In [6] a confidence-based approach, CBA, for formative and summative assessment has been proposed. Marks have been selected in such a way to provide a great penalty whether the student lies or has a misconception on its confidence on a given questions. CBA is suggested for T/F answers in case of summative exams. Indeed, whether multiple answers are allowed the proposed penalty seems to be too high, because a wrong answer is not always due to guess, but it can be caused by misunderstanding. This is particularly true whether the time available to perform the test is strict. Moreover, this kind of evaluation has to consider the character of a student. Indeed, a student with a low self-estimate can reach a lower score with respect to another one with the same knowledge. Similarly to our approach, immediate feedbacks are provided during formative exercises. Such feedbacks are numerical. In our approach we try to solicit the student to objectively reflect about the reason of errors providing adaptive textual messages.
In [4] fuzzy theory has been adopted for dealing with uncertainty in the assessment of the students and the incomplete knowledge about his Self-Efficacy.

3. The proposed methodology and tool

According to Bandura [2], self-efficacy beliefs for a behaviour is increased by successfully performing the behaviour. It is also influenced by verbal persuasion, which, if realistic, can encourage efforts that are more likely to increase efficacy through success.
The goal of the study is to enable a student to measure his/her awareness considering the following level of knowledge acquisition:

1. **Unconscious incompetence** – I do not know what I don't know.
2. **Conscious incompetence** – I know what I do not know
3. **Unconscious competence** – I do not know what I know. This self-bilief also corresponds to the situation where the student is so practiced that (s)he does not need to know what (s)he knows because (s)he can perform the task without conscious thought.
4. **Conscious competence** – I know what I know

These competence classes are listed in order of danger for the students knowledge.

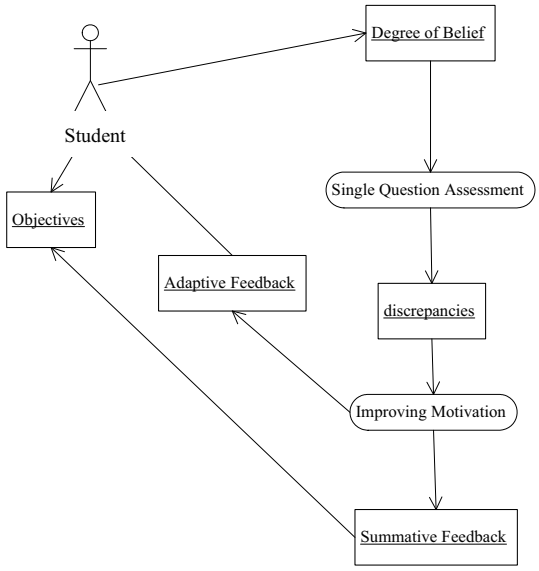


Figure 1. The SEA process

The proposed approach aims at giving appropriate feedback to the students about their cognitive level. It also aims at improving self-efficacy during a self-assessment test composed by several multiple choice questions.
Figure 1 depicts the Self-Efficacy Assessment (SEA) process. A student, following the Bandura's model [3], has a certain degree of beliefs about a given task. A multiple choice question is proposed to evaluate the discrepancies between the student belief and its competence on the task to perform. Discrepancies are then analyzed to improve the student's motivation through the production of an Adaptive Feedback, which takes in account the trends of the student confidence examining the previous answers and his/her performance. When all the questions have been answered, a summative report is also provided.
The Self-Efficacy Assessment environment supporting the proposed methodology has been developed as a MOODLE [7] plug-in. The prototype enables the teacher to choose the feedback typology (textual, numeric, graphical or both) and whether to provide a feedback for each questions or only a summative report. Moreover, the text of the messages corresponding to the various level of confidence can be personalized. The teacher can also decide the weight of the confidence scoring in case o summative assessment of the student.

3.1 The Scoring Schema

The scoring schema we propose is customized by the teacher choosing the number *n* of confidence levels. To not heavy influence the final score by the adoption of confidence levels, differently from [6], we prefer to adopt a scoring scheme based on the Renard Number [9], as depicted in Table 1, where, as an example, we consider two Renard series *R10* and *R20*, three confidence levels and four multiple choice answers.

Table 1. The 3 level confidence SEA scoring schemas

	R 10 Confidence Level		
	Low (1)	Medium (1.25)	High (1.60)
Correct answer (1)	1	1.25	1.60
Wrong answer (-1/3)	- 0.33	- 0.42	- 0.53
	R 20 Confidence Level		
	Low (1)	Medium (1.12)	High (1.25)
Correct answer (1)	1	1.12	1.25
Wrong answer (-1/3)	- 0.33	-0.37	- 0.42

As Table 1 shows, the Renard Numbers enable us to adopt equally spaced factors to better weight the score of the test. This feature provides us with a didactical instrument that investigates the confidence using test scores which are limitedly influenced by the confidence weights.

A stronger penalization is proposed in [6] when a student expresses high confidence on a wrong answer. Differently, we prefer to adopt a lighter scoring schema that on one side judges the level of confidence and on the other side does not alter too much the final score.

Let S_t and S_c the scoring evaluated without and with the confidence level, respectively. To avoid final results to be too reliant to the student nature, such as shy, sure, or boaster, the final score S is weighted with the confidence judgments expressed by the student during the test:

$$S_c = \frac{\sum_{i=1}^{nQ} q_i c_i}{\sum_{i=1}^{nQ} c_i} \quad where \quad q_i = \begin{cases} 1 & for\ correct\ answer \\ \dfrac{-1}{numChoice - 1} & otherwise \end{cases}$$

$$c_i = \begin{cases} 1 & low\ confidence \\ 1{,}25 & medium\ confidence \\ 1{,}60 & high\ confidence \end{cases}$$

$$nQ := number\ \ of\ \ questions$$

$$numChoice := number\ \ of\ \ possible\ \ answers$$

The scoring scheme is devised in such a way to be extremely flexible. To this aim the test has been parameterised to let the teacher finely customize it by introducing the following parameters:

- the *confidence weights* c_i, which are fully customizable in range and starting point. As an example, in our experiment we have chosen the *Renard serie R10* [9] starting from 1 and assigning three confidence levels
- the *influence factor*, *k*, which enables the teacher to decide the weight (s)he wants to assign to the confidence degree of a student. The resulting score *S* is evaluated as follows:

$$S = \frac{100 - k}{100} S_t + \frac{k}{100} S_c$$

- the *time window*, *w*, which sets the size of the sample of recent scores that is compared in average with all the scores already obtained. These statistics are used to provide feedbacks during the test. The *w* parameter enables to give inertia to feedback output avoiding useless and dangerous oscillations.

Figure 2 shows the test results obtainable by all permutations of the three confidence levels *R20* (1, 1.12, 1.25) and the impact on the traditional test score, depicted in each subplot by the horizontal line. Figure 2 depicts the results of a test composed by 3 questions, each of which with four answers. Displayed data are obtained with a *"lazy generation"* of the scores domain, in the sense that each subplot in Figure 2 represents all possible permutation of answers that obtain the same average score. Thus, Figure 2 (a) and (d) depict the application of all possible confidence levels to all wrong answers and all correct ones, respectively. As graphically shown, the adoption of confidence level does not alter the test results in the best and in the worst cases. Differently, the impact on an intermediate score, while motivating students to evaluate correctly themselves, is quite slight. Figure 2 (b) depicts the score fluctuation due to confidence weights in the case of one correct answer and two wrong ones. The test grade is contained in (*0.05*, *0.18)*, while the average is attested to *0.115*. Figure 2 (c) shows that, answering correctly two questions, the final grade may oscillate between *0.49* and *0.62*, while the traditional score is near to *0.56*. It is important to note that the teacher can further customize the impact of the confidence by acting on the *influence factor*, *k*.

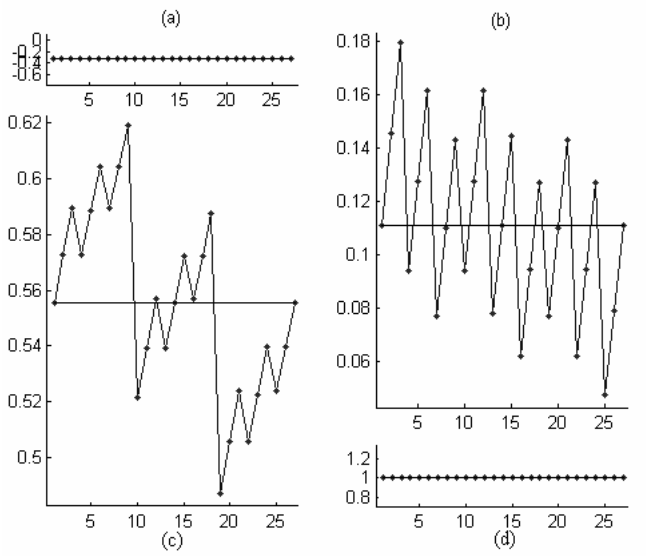


Figure 2. The confidence evaluation of a 3 question test each with 4 possible answers, computed with 3 R 20 confidence levels.

4 Adaptive Feedbacks

The SEA system provides users with different levels of feedback, depending on the teacher options selected when creating the test. The teachers, can also customize the messages provided as feedback to his(her) students. (S)he can also establish if messages are provided in a continuous manner (question by question) or only at the end of the test.

4.1 Question by question feedbacks

After a user has answered a question a summative report is provided which presents several statistics computed on the questions scores:

- the *scoring trend*, which is provided in a graphical way after submitting the answer to question *i* (*i*>*w*). It is computed by comparing the average score obtained on the last *w* questions with the average score computed on previous *i-1* ones.
- the *relative confidence feedback*, which is computed starting after the answer *w,* on the last *w* answers. It is graphically depicted by representing the average score obtained by the student and the *Best* and *Worst* scores obtainable with confidence weights.
- the *confidence trend*, which is computed answer by answer comparing the actual score with the *Best* and *Worst* scores obtainable with confidence weights on the partial test results.
- the *confidence benefit*, which is obtained by comparing the confidence test score S_c and S_t, the classical one. By providing immediate feedback on how much the confidence answers are increasing or decreasing the score, we intend to invite the students to reflect on themselves.

Table 2. An example of confidence scoring

	Normal test	**Confidence test**		
		Answer	*Best*	*Worst*
	C	M	H	L
	W	L	L	H
	C	L	H	L
	W	L	L	H
	W	H	L	H
Score sum	1.01	1.062	2.21	0.416
Average Score	0.202	0.18	0.36	0.06

An example of the evaluation of the confidence scoring is shown in Table 2. To assess the confidence of a student we compare the test score, which is influenced by his/her confidence, with the *Best Confidence Score* (s)he would obtain if (s)he were fully conscious of his/her competence and incompetence, and with the *Worst Confidence Score* representing the opposite case.

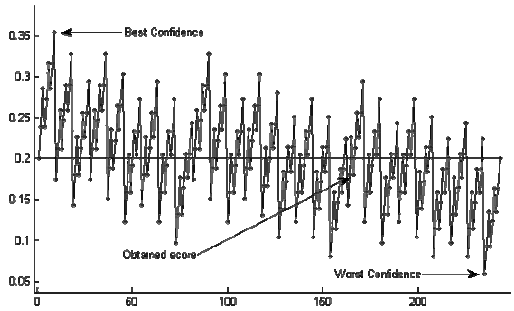


Figure 3. The application of the 3 level of R10 confidence to the test result of Table 2

Figure 3 depicts all scores obtainable with the answers detailed in Table 2, when weighted with all possible confidence judgments. The traditional score is represented by the horizontal line at 0.2. As Figure 3 shows, the *Obtained score* (in our case fixed at 0.18) is, for any other combination of confidence answers, in the range (0.06, 0.36).

Figure 4 shows the feedback window displayed after the submission of the fourth answer. As described in Table 2, the student fails, while expresses a low level of confidence. In particular, the user receives as feedback a graphical representation of his/her confidence degree, as depicted in Figure 4 (a). The relative confidence feedback (33%) is lower than the confidence trend (39%) because the former does not consider the first answer which is correct with medium confidence. The SEA plug-in, while signaling the decreasing of score in Figure 4 (b), provides a positive textual message underlining the user consciousness. The message will become harder after the student misses the last question with a high confidence value. The aim is to stimulate the student attention on the differences existing between its belief on his capacity and the real performances to encourage reflection and the auto evaluation of self-efficacy. Thus, particular attention has to be addressed in this kind of message: a wrong formulation can decrease the student self-estimate.

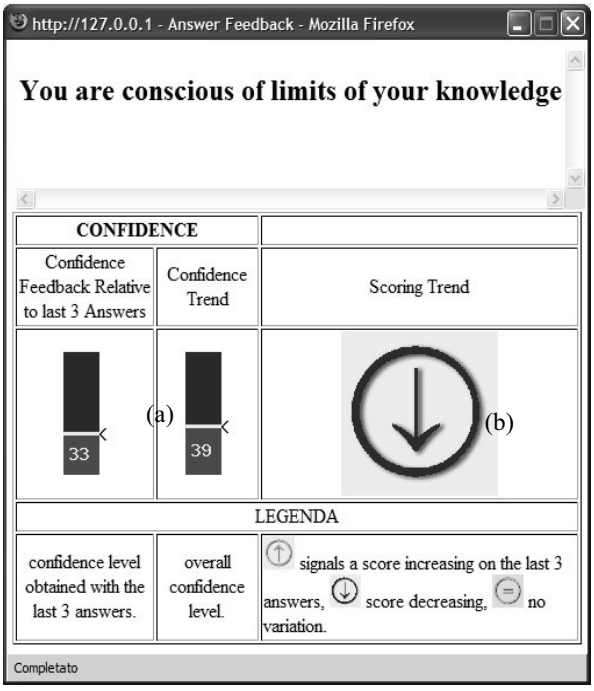


Figure 4. An example of feedback representing the user confidence (consciousness of limits)

4.2 Final feedback

At the end of the test the final feedback provided to the students has to be accurately designed. Indeed, the whole confidence revealed by the student during the test has to be summarized in a message.

Particular attention has to be devoted in the two most problematic cases: the unconscious incompetence and unconscious competence. While the latter is less dangerous, the former can seriously misunderstand his/her need of an appropriate corrective learning process.

Table 3. Confidence level classification

	Confidence ranges		
Scores ranges	Low	Medium	High
[-n/3, n/3]	conscious incompetence K_1 and C_1	Semi-conscious Incompetence K_1 but C_2	unconscious incompetence K_1 but C_3
[n/3, 2n/3]	medium consciousness, medium competence, K_2 but C_1	High consciousness, medium competence K_2 and C_2	K_2 but C_3 medium consciousness, medium competence
[2n/3, n]	unconscious competence K_3 but C_1	Semi-conscious, competence K_3 but C_2	conscious competence K_3 and C_3

In Table 3 a classification of the relationship occurring between confidence and competence is shown together with an example of adaptive feedback. To each cell of the table we associate a message K_i Ж C_j, where the messages K_i and C_j are described in Table 4 and Ж is *"and"* if i=j , *"but"* otherwise. The text of these messages is reported in Table 4.

These textual feedbacks are provided for each subtopic group of questions to signal to the student troubles in knowledge in that specific subtopic.

Table 4. Final feedback examples

Id message	**Message text**
K_1	You have to hard improve your competence
K_2	You have a little competence
K_3	You are very competent
C_1	You have to reflect a lot on your competence
C_2	You are quite conscious of your competence
C_3	You are very conscious of your competence

4. Conclusion and future work

In this paper we have presented the SEA approach, aiming at improving the self-efficacy of a student. Respect to a traditional self-assessment system, when a student assesses his/her competence on the proposed system first of answering to a question (s)he has to examine himself on his competence degree. An immediate feedback puts in evidence the discrepancies between his/her self-belief and his/her real competence. The final score takes in account the confidence expressed, in such a way to induce the student to be honest and to reflexive.

The proposed assessment methodology has been implemented as a MOODLE plug-in. The system will be experimented during the ARCHIMEDE project, an on going research project aiming at promoting the scientific culture in the secondary school, while experimenting self-assessment methodologies. The SEA prototype runs on a dual Quad-Core Xeon server, which will host a thousand of student's accounts from twenty secondary schools.

During the initial phase of this project an assessment test has been proposed using the SEA plug-in. A secondary school class composed by 18 students answered the survey presented in Table 5 concerning the SEA functionalities. Before attending the test the students were introduced to the SEA environment and its objectives. No time was fixed to accomplish the test.

Table 5. Survey questionnaire

QID	Question	%					
		E	D	C	B	A	n. a.
Q1	The objectives of the self-efficacy assessment test are clear to me	0	0	22%	56%	22%	0
Q2	The task I had to perform was perfectly clear to me	0	0	11%	78%	11%	0
Q3	It is important to have an individual feedback	0	0	33%	22%	45%	0
Q4	The feedback messages are clear	0	11%	33%	23%	23%	0
Q5	I believe that messages provided by the system improve my self-efficacy	0	5%	33%	56%	5%	0
Q6	I believe that reflecting on my self-efficacy improve my attention	0	22%	33%	34%	11%	0
Q7	I believe that the tool help me in the identification of the areas in which I am relatively effective and areas in which I am relatively ineffective	0	0	33%	23%	33%	11%

As Table 5 revealed, the results for the SEA functionalities are quite encouraging and the most important achievement concerns the help provided by the tool in the identification of effective/ineffective areas (Q7). All the students consider useful individual feedbacks and 94% of them positively consider the impact of the feedback messages on their self-efficacy.

Future work are devoted to further encourage students in avoiding to guess the answer and in better assessing their self-efficacy using adaptive assessment.

Acknowledgments

The authors would like to thank Pietro Aurilio who implemented the MOODLE plug-in supporting the proposed approach.

References

[1]. Bandura, A. (1982). Self-efficacy mechanism in human agency. American Psychologist, 37(2) 122-147.

[2]. Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.

[3]. Bandura, A. Self- Efficacy - The exercise of control. New York: Freeman.(1997).

[4]. Bica F., Verdin R., Vicaril R. M., Towards Cognitive Modeling of Students' Self-Efficacy. In Proceedings of the Sixth International Conference on Advanced Learning Technologies (ICALT'06).

[5]. C Carlisle, Confidence and competence in skills acquisition. Blackwell Science Ltd MEDICAL EDUCATION 2000;34:886-887.

[6]. A.R. Gardner-Medwin and M. Gahan, Formative and Summative Confidence-Based Assessment. Proc. 7th International Computer-Aided Assessment Conference, Loughborough, UK, July 2003, pp. 147-155.

[7]. MOODLE, www.moodle.org.

[8]. P. J. Morgan and D. Cleave-Hogg, Comparison between medical students' experience, confidence and competence. Blackwell Science Ltd MEDICAL EDUCATION 2002;36:534–539.

[9]. Renard Numbers ISO 3:1973 http://www.iso.org/

[10]. Regehr G, Hodges B, Tiberius R, Lofchy J. Measuring self-assessment skills: an innovative relative ranking model. Acad Med 1996;71 Suppl:S52±4.

[11]. S. M. Smith. Software Skills Acquisition: Confidence vs. Competence. Information Technology, Learning, and Performance Journal, Vol. 22, No. 2, Fall 2004.

[12]. J. Stewart, C. O'Halloran, J. R. Barton, S.J Singleton, P. Harrigan and J. Spencer, Clarifying the concepts of confidence and competence to produce appropriate self-evaluation measurement scales. Blackwell Science Ltd MEDICAL EDUCATION 2000;34:903-909.

[13]. S. Vickery and H. Cooper. Confidence or Competence? – auditing information literacy skills of biology undergraduate students.

Planning and Prototyping of a Telecare System for Music Performance Health

Yu-Huei Su[1],Mei-Ju Su[2],Yu-Yuan Chang[1],JJ Luh[5],Chao-Cheng Lin[6],
Min-Chun Liao[1] Heng-Shun Chen[3,4]*

Abstract

Musical performance requires delicate mental work and good physical skill. While musical performance anxiety not only affects stage accomplishment but also undermines confidence. Besides, musical performance injury brings pain of muscles and bones, even forcing some musicians to give up their performance career. To elevate the healthcare awareness of music performers, this study will construct Telecare system for music performer's health, integrating e-learning and e-training systems, to prevent music performer's anxiety and injury. This study includes establishing an interdisciplinary team of distance mental and physical care, constructing the knowledge structure, devising evaluation tools, developing improvement strategies, and planning online e-learning platform with user communities. The expected result is a Telecare System for Music Performance Health, consisting of healthcare database, digital on-line testing, and two interactive self-care systems, Guided imagination- music performer's anxiety relief system and Stretch and warm up exercise- music performer's injury prevention system. To evaluate the effectiveness of the system, psychological and physiological measurement before and during musical performance can be transmitted to the system, interpreted by the professionals, and the feedback will be provided to the users. By this way Telecare system could offer music performers a comprehensive prevention and self-learning guidance.

Keywords: music performance anxiety, music performance injury, e-learning

[1]Department of Music, National HsinChu University of Education; [2]Graduate Institute of Electronics Engineering, National Taiwan University; [3]Department of Family Medicine;[4]Department of Medical Informatics, National Taiwan University College of Medicine; [5]School and Graduate Institute of Physical Therapy, National Taiwan University College of Medicine;[6]Department of Psychiatry, National Taiwan University Hospital.
Correspondence to: Heng-Shun Chen ,Department of Medical Informatics, National Taiwan University College of Medicine,1 Jen-Ai Road, Taipei ,Taiwan
E-mail: chenhs@ntu.edu.tw

A musical performance requires delicate mental function and good physical skill. However, performance anxiety and injury limit the development of performance ability. The former will not only influence the performance, but also hurt the performer's self-esteem. The latter will cause physical pain due to the improper posture and over tension of the muscle. Performance anxiety and injury are inevitable for music performers. Hence, it is an urgent issue for music performers to improve their mental and physical healthcare literacy.

According to investigations, nearly 60% of the performers feel performance anxiety affecting their quality of performance and life (Kemenade, 1995；Stephenson & Quarrier, 2005)[1,2], and 83.1% of the music major students face performance anxiety (Miller & Chesky, 2004)[3]. Performance injury is also a frequent problem during music learning. Most are due to inappropriate use of strength and gestures, which may lead to injury of the muscle, bone and nerve system in the long run. While less severe cases are suffering from lasting pain and being able to perform only less effort-taking works. Severe sufferers may be forced to give up their performing career. According to investigations, nearly 76% of the performers regard that performance medical problems affect their quality of performance (Fishbein, Middlestadt, Ottati, Straus & Ellis, 1998；Brandfonbrener & Lederman, 2002)[4,5]. The research in Taiwan indicates the rate of performance injury is very high among music major students. 97.9 percent of them have experienced profession-related injury. The remarkable issue is that when injury happens, 60.4% of the students tend to handle by themselves or to ask for help from their teachers. Only 21.6% of the students go to the doctors (Pon & Sun, 1994)[6].

Current internet resources of performing art medicine mainly provide information of conferences, articles, group discussion, and links of medical services. Online learning for improvement of performance anxiety and injury is rare, most of which are only ads of paying DVD about exercise practice therapy. Due to the widespread usage of the internet, our country has developed an excellent platform for distance learning of music performance healthcare literacy. Based on the above, the purpose of this study is to apply the concept of digital learning to music performer's mental and physical care, combining case studies of musical performance, basic and clinical researches of psychiatry and physical therapy, knowledge management of informatics and technology education, and support from electrical engineering to have an interdisciplinary study to achieve the purpose of preventive education.

Methods

This study, with the trend of e-learning with internet platform, uses the online- training to promote music performer's healthcare literacy to achieve the goal of prevention. Methods are as following:

1. Establishing an interdisciplinary team of music performance long-distance health care

Interdisciplinary cooperation is the main contribution to this study. Members of this team include performing and music educational experts in music department, psychotherapy and informatics experts in hospital, scholars of electrical measurement equipment of electrical engineering, and multimedia experts .

2. Developing a knowledge structure of performance health literacy

Performance healthcare literacy comprises mental and physical aspects. The mental part includes the definition of performance anxiety, the character with performance anxiety, social anxiety, self-evaluation, the relation between performing situations, and the strategies to deal with performance anxiety, etc. The physical part includes the right practice gestures, some common symptoms like muscle and bone pain, warm up activities before practice, the proper rest, the evaluation and healing of performance injury.

3. Developing the evaluation tools for performance anxiety and injury

This study adopts the online-survey approach. The survey of performance anxiety will adopt "Music Performance Anxiety Inventory for Adolescent, MPAI-A" (Osborne, & Kenny, 2005)[7] and "Kenny Music Performance Anxiety Inventory, K-MPAI" (Kenny, Davis, & Oates, 2004)[8]. The evaluation of performance injury will adopt "Disabilities of the Arm, Shoulder, and Hand Outcome Measure, DASH" (Hudak, Amadio & Bombardier, 1996)[9] which can effectively evaluate the forelimb barriers of the testee and their effect on life. In DASH, if there is any item checked intermediately difficult by surveyors, they will further fill out "Nordic Musculoskeletal Questionnaire, NMQ" (Kuorinka, Jonsson, Kilborm, Vinterberg, & Biering-Sorensen, 1987)[10]. This survey investigates frequent discomfort of muscle and bone. Those above surveys are all in English and need to be translated into Chinese and tested their effectiveness.

4. Developing improving strategies

The improvement strategies for performance anxiety refer to the four steps of anxiety pressure management and health improvement (Li Ming-bin, 2005)[11]: (A) Trace the origin of anxiety; (B) Train and imagine how to deal with anxiety; (C) Evaluate cognition and improve the capability of positive thinking; (D) Practice and evaluate the result. Anxiety improvement tools refer to guiding imagination, integrating relaxation practice and cognition

therapy to guide participants to think positively and reduce performance anxiety through cognition reestablishment. The improvement strategies, adopted from "The Prevention of Health Injury Brochure of Computer Technicians" and "Brochure of Promotion of Health", are instructed by physical therapists and put into practice, doing 15-20 min stretch exercise under the guidance of teacher before every performance.

5. Designing digital learning platform and physiological signal measurement lab

Through convenience of the Internet, the long distance digital platform of music performance healthcare literacy offers performers and the public a more easy, convenient and efficient access to music performance healthcare literacy and online learning. But the security of online data plays a more important role. This system will adopt hardware firewalls to separate inner database from the public to avoid illegal data stealing while legal users can access information they need through firewalls and web servers.

This platform adopts the design structure of web base. Users can utilize browsers and connect to the platform of music performance healthcare literacy anytime, anywhere. In the design of database, the multimedia database is separated from general databases so that multimedia information can be presented independently. In this way the efficiency can be improved and expansion possibility can be increased. Program development will adopt object-oriented programming and all objects can be used repeatedly. This platform can support the file formats of text, graph, image, sounds, DVDs, and auxiliary teaching software.

Besides, in the physiological signal measurement labs, the sensors including an clinical thermometer, ECG machine, an EMG machine and an Oximeter are integrated to measure the ECG, EMG, heartbeat, finger temperature and skin impulse. Through a collective interface, these physiological signals will be uploaded to e-health care center and will be read by the professionals of mental hygiene and physical therapy.

6. Developing learning communities

After the construction of this platform, four learning communities will be established afterwards, including portal websites of professional performers, music teachers, music classes and music major students, and the general public to provide different information for professional necessities and experience sharing . This platform will also build up online-learning, set up mail system, and deliver mails through auto-transfer, informing four social groups of new learning units and online updates. Moreover, the audio and video function of online performance injury prevention will be set up to utilize guiding imagination strategies to direct users to deal with anxiety and to provide appropriate gestures, rest and workout advice to prevent performance anxiety. As for learning evaluation, online survey for learners is the main focus to investigate their anxiety and injury knowledge and behavior improvement. Finally, the platform build up transfer information, providing database, linking websites, to provide related information for online searching, like information of experts and medical centers for anxiety and injury in the county.

Expected Result

1. The e-learning structure of music performance health literacy

This system can provide a healthcare database about music related performance. It includes digitalization of self-evaluation of performance anxiety and injury, online learning and community. The structure is as the fig 1:

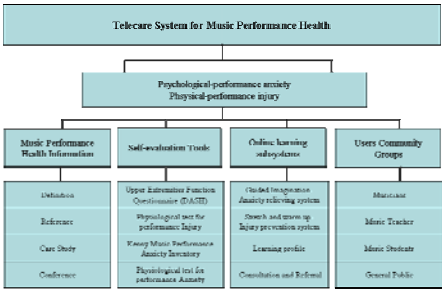


Fig 1. Structure of a telecare system for music performance health

2. Designing an interactive learning system to prevent performance anxiety and injury

This research constructs the online interactive learning system of guided imagination-performance anxiety improvement training. This training is of cognition behavior model. We can not only analyze the results by "Music Performance Anxiety Inventory for Adolescents, MPAI-A "(Osborne, & Kenny, 2005)[] and "Kenny Music Performance Anxiety Inventory, K-MPAI"(Kenny, Davis, & Oates, 2004)[], but also examine the change of anxiety index by some psychological equipment, such as heartbeat, finger temperature, and skin electric signal sensors. As for the online interactive learning system "stretch warm up exercise for injury prevention", it's to instruct the performers stretch warming up exercise. An examining lab should be built in the music department classroom to monitor the performers' psychological signals, like electric signals of muscle groups on the hands, Electro Cardio Grams, images of action. The signals will be sent to medical centers to decide the appropriate timing for rest, and the messages will be sent back to the performers by email so that the performers will know more about prevention of injury. To establish long distance healthcare mode, the physiological signal output terminal are in the

performing classrooms and the input terminal are in medical centers. And the signals are to be read and analyzed by the professionals.

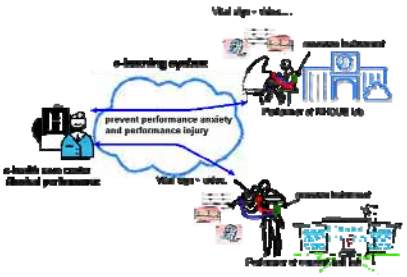


Fig 2. Interactive leaning system- prevention of performance injury and anxiety

Discussion

It's essential for every music performer to learn to deal with performance anxiety and prevent performance injury. According to IFPAM (International Foundation for Performing Arts Medicine), Performing Arts Medicine was not taken seriously until the last two decades. Currently the related courses are few in Taiwan and the access for consulting is not popular. Therefore, when such problem happens, only a few performers ask for help from the professionals. Some even turn to folk therapy. They are afraid of being tagged and choose to hide their problems, thus make their situation worse. Severe sufferers are even forced to give up their performing career. Even if they go to the doctors, they have difficulties to follow the healing process due to their insufficient time and tight schedule. Hence, a long-distance educational system of music performance health literacy, taking the advantage of the internet, can offer music performers an online learning way of self-care.

Current websites of performing art medicine mainly provide information of conferences, periodical articles, courses, art performance, references, article issue, association connection, and on-line shopping. In the circumstances of lacking resources of on-line learning of performance anxiety and injury prevention, this system will enable performers to seek professional support and knowledge in the digital learning platform and provide performers with a digital way of self-learning and prevention.

References

[1]Kemenade, J.F., Son.M.J.& Heesch, N.C.(1995). Performance anxiety among professional musicians in symphonic orchestras: a self-report study. Psychological reports, 77(2):555-62.

[2]Stephenson , H. & Quarrier , F.(2005).Anxiety Sensitivity and Performance Anxiety in College Music Students. Medical Problems of Performing Artists,20 (3),119-125.

[3] Miller, S. R. & Chesky, K.(2004).The Multidimensional Anxiety Theory: An Assessment of and Relationships between Intensity and Direction of Cognitive Anxiety, Somatic Anxiety, and Self-Confidence over Multiple Performance Requirements among College Music Majors, Medical Problems of Performing Artists,19(1),12-20.

[4]Fishbein, M., Middlestadt, S. E., Ottati, V., Straus, S., & Ellis,A.(1998). Medical problems among ICSOM musicians: Overview of a national survey. Medical Problems of Performing Artists, 3, 1-6.

[5] Brandfonbrener, A.G. & Lederman, R. J.(2002). Performing Arts Medicine. in The New Handbook of Research on Music Teaching and Learning: a Project of Music Educators of National Conference, Part Ⅲ: Neuroscience, Medicine, and Music, (eds. R. Colwell & C. Richardson), pp.1009-22. New York: Oxford University Press.

[6]Pon hui-wen, Sun pi-chang, Chuang hwa-on, Shu dou-chang (1994). The Survey of Colledge Music Majors Professional Injury - Preliminary Report. Medical Journal of China Physiotherapy for Restoration from Sickness 22(1), 105-109.

[7]Osborne, M. S.& Kenny, D. T(2005) .Development and Validation of a Music Performance Anxiety Inventory for Gifted Adolescent Musicians. Journal of Anxiety Disorders,19,725-751.

[8]Kenny D.T., Davis P.J.,& Oates J.(2004). Music Performance Anxiety and Occupational Stress amongst Opera Chorus Artists and their Relationship with State and Trait Anxiety and Perfectionism, Journal of Anxiety Disorders, 18,757-777.

[9]Hudak, P.L., Amadio, P.C.& Bombardier, C. The Upper Extremity Collaborative Group (UECG)(1996). Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). Am J Ind Med., 29, 602–608.

[10]Kuorinka, I., Jonsson, B., Kilborm, A., Vinterberg, H. & Biering-Sorensen, F.(1987).Standardised Nordic Questionnaire for the analysis of musculoskeletal symptoms. Applied Ergonomics, 18 (3), 233-237.

[11]Li ming-bin, 2005. Pressure Life - Emotion Management and Health Improvement. Taipei: Health Culture

Beyond Learning Management Systems in Lifelong Learning

Giovanni Adorni, Serena Battigelli, and Mauro Coccoli

DIST, Department of Computer, Communication, and Systems Science
University of Genova, Via OperaPia, 13 – 16145 Genova, Italy

Abstract: *During the last years Internet has been changing the way people work as well as the way people work is changing the Internet. New communication paradigm, new tools, and new technology have grown-up driven by the users. The network is moving fast in the direction of a social network where people collaborate by sharing their knowledge and experience, their thoughts, their materials, their documents, their pictures and videos, their presentations, etc.. The leaders of this revolution are the basic concepts underlying the semantic web in which content plays the key role. Distributed and shared bookmarking systems, forum, wiki, blog, and push technologies such as feed are used rather then search engines so that browsing and searching are not any more the main activities that users perform on the web. In this new networked social working environment learning management systems too must be able to adapt their functionality and standards to the emerging ones. The learning process could strongly benefit of going outside of traditional learning environments and teaching materials must be considered in the terms of semantic web content. This is to be considered the road-map to the development of novel learning services in the actual social network scenario.*

Introduction

In the last years, the number of the Internet users has been growing fast. In the present situation in which new technology has made web applications more and more appealing and easy to use, users also are continuously evolving their network culture, thus their approach to the Internet and to the web contents, applications, and services is changing. In the actual scenario web surfers are customary visitors and they spend a long time over the Internet while performing several different activities for both their work and their fun. It is very common that people are gathering and consuming information over the Internet; also it is getting very common that people supply some information. In this respect two mainstreams are becoming the key to the future Internet society: social networking and knowledge sharing. Every one should keep into account these two aspects while designing Internet application and services.

The shape of the information on the web also is changing. It is evolving from a "page-centered" configuration in which the documents and the content of the page are the most important things towards a relationship-centered observation of the pages in which the content is still important but linking also has a pre-eminent role. From a behavioural point of view, the approach of the users to the web is changing: it is kind of a social evolution attesting that a new degree of maturity has been reached by the people working and living in the net. They are moving from having a walk-on part to having a lead role. Such new protagonists can be easily identified looking at the communities of customers of the same service or e-shop, the "bloggers" whose number is increasing day by day, the communities of file sharing such as personal pictures or home made videos, and so on.

In the past years, thanks to the use of LMS, both learners and teachers have become more familiar with the use of ICTs and web technologies and in the field of e-learning too something is moving and learning services as well as teaching strategies are moving towards the use of a variety of new services. As it happens in e-business and e-commerce, systemic changes induced by the emerging technologies are also changing e-learning and the relationships between learners and teachers, who makes lectures and who is attending. In this respect one can easily make a separation among formal e-learning and informal e-learning. Successful implementations of informal e-learning can be seen in self-teaching and lifelong learning processes made by one-self for his/her own desire of knowledge or for improving professional

skills. In the informal learning process, people spontaneously group them-selves in discussion boards where they can freely exchange information and they can benefit of each others' experience. Most of the times, traditional Learning Management Systems are not ready to cope with such new needs and thus new services must be thought. E-learning environments must change their style and include new tools for promoting knowledge circulation and sharing and for guaranteeing a better level of communication among people, while hiding the supervision functionality of control over the users. The idea of integrating different systems with different functionalities pushes learning management systems in a new direction, for them to become open learning content management and sharing platforms. Interoperability is strongly needed as well as a bunch of common languages and standards, interfaces and bridges to other systems. Teaching strategy and methodology also must be dynamic especially when related to topics that are rapidly evolving (e.g., technological and scientific). References also are continuously changing as well as teaching materials that must be guaranteed to be up to date.

After this short introduction dealing with the technological scenario, the paper is organized as follows. In the first section the difference between formal and informal learning will be quickly depicted. Then the case of lifelong learning activity is presented as a typical application of non formal learning concepts. Social constructivism is introduced as the didactical strategy to be implemented in the considered case of lifelong learning. Semantic web concepts are introduced as the humus in which the previously described concepts can be set. An overview of application and services aware with the semantic web and fully in line with the social networking ideas are listed, which could be the tools and technologies available in the next generation Learning Content Management Systems. Conclusions follow.

Formal and Informal Learning

When working in an educational institution or when people are requested to show some degree or a certificate there is the need of following a formal learning path which may include lectures or distance education and a final examination. On the other hand, in every day's life and in organisations, learning activity also may happen outside formal training programs.

It can be interesting to be able to distinguish between what is part of a learning program and what is not. A categorization of different kind of learning is reported [1].

Formal learning includes the hierarchically structured school system that runs from primary school through the university and organized school-like programs created in business for technical and professional training.

Informal learning describes a lifelong process whereby individuals acquire attitudes, values, skills and knowledge from daily experience and the educative influences and resources in his or her environment, from family and neighbours, from work and play, from the market place, the library and the mass media.

Intentional learning is the process whereby an individual aims to learn something and goes about achieving that objective.

Accidental learning happens when in everyday activities an individual learns something that he or she had not intended or expected.

Lifelong learning

Together with the shift towards this new model of knowledge proposed within the knowledge society, the continuous evolution of information and communication technologies as well as the fast varying conditions of work, personal skills and knowledge continuously need to be enhanced. The use of e-learning systems is also one of the strategic objectives of the EU/IST Work Programme, with the specific goal of "improving the efficiency and cost-effectiveness of learning, for individuals and organizations, independent of time, place and pace" [2]. In this respect, novel methodologies and techniques are needed as well as infrastructure and devices for this goal to be reached. Such an achievement can facilitate the sharing of knowledge between individuals and in organisations too.

The basic idea in lifelong learning is that "Every time is the right time for learning something new", a philosophy that has taken root in a whole host of different organizations. Lifelong learning is attitudinal; that one can and should be open to new ideas, decisions, skills or behaviours. In the actual scenario people are provided with ever seen before learning opportunities at all ages and in numerous contexts: at work, at home and through leisure activities too, not just through formal channels such as school and educational institutions.

Lifelong education is a form of pedagogy often accomplished through distance learning or e-learning, continuing education, home schooling or correspondence courses. It also includes postgraduate programs for those who want to improve their qualification, bring their skills up to date or retrain for a new line of work. Internal corporate training has similar goals, with the concept of lifelong learning used by organisations to promote a more dynamic employee base, better able to react in an agile manner to a rapidly changing climate. In later life, especially in retirement, continued learning takes diverse forms, crossing traditional academic bounds and including recreational activities. One of the reasons why lifelong education has become so important is the acceleration of scientific and technological progress. Despite the increased duration of primary, secondary and university education (14-18 years depending on the country), the knowledge and skills acquired there are usually not sufficient for a professional career spanning three or four decades [3]. In UK and elsewhere, the "University of the Third Age" (U3A) provides an example of the almost spontaneous emergence of autonomous learning groups accessing the expertise of their own members in the pursuit of knowledge and shared experience. No prior qualifications and no subsequent certificates feature in this approach to learning for its own sake and, as participants testify, engagement in this type of learning in later life can indeed 'prolong active life'.

Social Constructivism

Thanks to the increasing use of Learning Management Systems and technologies, relationships among students and teachers are changing and didactical strategies too are going in the direction of implementing the paradigm of the social constructivism [4]. According to the point of view of social constructivism the process of knowledge construction which is carried on by the students must be influenced and driven by the personal experience and attitude [5]; it is only marginally supervised by teachers and/or tutors and unpredictable exogenous factors play a key role. In such a scenario students and teachers are co-protagonist in the learning process while people outside of the group (a classroom or a virtual classroom) can also participate to the knowledge construction by interacting and exchanging ideas and experiences in a virtuous social network system. Students and teachers when working together in an e-learning environment also participate of a social activity that involves other students, administrators, friends, and participants in various forms of activity. This takes into account the social nature of both the local processes in collaborative learning and in the discussion of wider social collaboration in a given subject. The language becomes the vehicle by which ideas are considered, shared and developed. The exchange and building of thoughts and ideas which takes place in the course of the interaction, conversation, activity and social context plays a fundamental role in the development of an individual [6] that is the concept at the basis of social constructivism. Other researchers [7] have tried to find out whether the "mind" is located in the head or in social action. Their conclusion is that both perspectives should be used together, as they are both useful. In fact what is seen from one perspective as the reasoning of a collection of individuals mutually adapting to each other's actions can be seen in another as the norms and practices of a classroom community [8]. We can exploit the social-networking capabilities to this aim.

Some research show how the metaphors of “acquisition” and “participation” of learning interrelate and interact in synergistic ways [9]. In some examples, the social entity as a learner (consider a football team, a business or a family), is compared with the learning of an individual in a social setting, and three main types of relations are identified: individual learning can be less or more socially-mediated learning; individuals can participate in the learning of a collective, sometimes with what is learned distributed throughout the collective more than in the mind of any one individual; individuals and social aspects of learning in both of these senses can interact over time to strengthen one another in a “reciprocal spiral relationship”.

Teaching strategies based on the social constructivism paradigm include teaching in contexts that might be personally meaningful to students, negotiating taken-as-shared meanings with students, class discussion, small-group collaboration, and valuing meaningful activity over correct answers [10]. In such teaching strategies motivation is considered to be a significant fact for learning: learners are motivated by rewards provided by the knowledge community; anyway, due to the fact that knowledge is constructed actively by the learner, learning depends on the learner's internal drive to promote the learning process itself. Results can be appreciated through the observation of the construction process of the wide community of intellectual publishers: liberal quoting of

each other's ideas, combining, arguing, extending and recombining them in order to construct our social and cultural understanding of thought, understanding and ultimately human nature.
Viewed in this way, teaching becomes the establishment and maintenance of a language and a means of communication between the teacher and students, as well as between students. Simply presenting material, giving out problems, and accepting answers back is not a refined enough process of communication for efficient learning.

Semantic web

The present and future standard for the web are depicted in the semantic web specifications released by the w3 consortium. According to the w3c, *"the Semantic Web is a web of data"*. Semantic web specifications are aimed to provide common formats for integration and combination of data drawn from diverse sources, and to find a language for recording how the data relates to real world objects. The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners [11]. It is based on the Resource Description Framework (RDF) [12] and on the use of specific ontology outlined in the Web Ontology Language (OWL) [13].
Given the actual trends of growth of the semantic web (both in content and applications) it can be considered that a certain level of maturity has been reached and the web is a real Distributed Knowledge Repository and it is rapidly evolving in the direction of a truly Distributed Knowledge Management System. Much effort is also requested to contributors in clustering and categorizing their works so that indexing can be made easy and can also be automated. Moreover, so much information needs efficient systems for it to be retrieved and circulated.
In particular it appears to be interesting the use of metadata search engine for knowledge management in learning communities [14, 15].
A semantic web based learning management system should act as a collector of information for a group of users that have the common task of acquiring some specific knowledge or skills; also it should provide facilities for communication and information exchange within the above community as well as between different communities; it should employ the Semantic Web technology for advanced communication and information exchange facilities

Application and services

The way web surfers go and find whatever they are looking for over the Internet has been changing together with the network itself while it has been evolving towards a multi-user shared knowledge management environment. The old fashioned "bookmark" is almost unused and it has given way to the more profitable, flexible, and portable on-line bookmark while social bookmarking services are growing more and more. Users do not save links to their favourite sites and web pages on their PC or laptop or PDA but they store their personal bookmarks somewhere on the web, in a personal space, so that they can retrieve them anytime, anywhere, even when they do not have access to their own devices. Personal web space is not any more a private room but information inside it can be shared with other people on the web by registering to specific social bookmarking services whose aim is circulating information, knowledge, and relationships among people with common interests.
Nowadays the web is the first place to look at, when seeking some information about something. Maybe someone somewhere has already had our same problem and maybe he/she has already found the way to solve that problem. Even when the problem is learning something new, the web (more than a library) is the fastest possible source for having some information (at least some preliminary information) about a topic of interest.
When learning activity is performed on the web by means of e-learning systems and techniques using web resources appears natural and in many cases it is encouraged to make it become part of the teaching/learning activity carried on by the students and teachers within a virtual classroom.
The habit to the use of the web shows its influence when preparing teaching materials too. A rich bibliography is at the basis, and then a selection of web sites too is presented as support to the learning process with the aim of being the up to date state of the art in that topic. Should some change in the state of the art occur, both minor modifications or new significant knowledge has been acquired in the meantime, a book can have a second edition printed in a couple of weeks while the web resource is immediately modified and the web reference can still remain the same.

Having a selection of web pages that you may want to visit at a fixed time-interval (daily, weekly or monthly, to check for update or for the presence of new contents) can take a very long time. The time spent for reading all the pages in a given list can be even longer if you consider that a time is needed for "context switching" when jumping from one page to another one, with different contents, colours, styles and layout. Searching and browsing and checking for updates are time consuming jobs and the use of feed can help in this. By using Syndication services it is possible to have the news we are interested in redirected to our PC, in a specific desktop application. All the information sources selected can be aggregated into one only container showing all of them in a single page sorted by topic or author or date in a clear, text-centred representation.

Keywords for the change to take place are feed, blogging, social bookmarking, wiki, knowledge sharing, on-line collaboration. Some implementations already exist and have been developed [16] based on the use of RSS and this is a first step towards the integration of the world outside into e-learning platforms and systems.

RSS is acronym for Really Simple Syndication and indeed it is a format for syndicating news and the content of news-like sites or personal weblogs. Yet it can be used for a variety of applications since it just works when some change occurs on a web page (e.g., the "recent changes" page of a wiki, a changelog of CVS check-ins, a new post on a personal weblog). When an RSS feed is claimed, then an RSS-aware program can show that change has happened. Typical applications are called news aggregators or more in general feed aggregators. Such a program can help people to keep up to date with all their favourite information sources [17].

Blog (blend of the words web and log, Web log) is an easy to use web publication tool, based on the architecture of a Content Management System. [18]. In a few words a blog is just a user-generated website where entries are made in journal style and displayed in a reverse chronological order. Moreover blogs give readers the possibility of commenting blog entries. A typical blog combines text, images, and links to other blogs, web pages, and other media related to its topic. One of the reasons of the great success of blogs is the possibility for readers to leave comments in an interactive format so that they become part of a wider network of social media.

Social bookmarking is not a new idea but it is getting very popular now. The Social bookmarking sites are just places to store, classify, share and search links with other people on the web.

Wiki is in Ward's original description: "*The simplest online database that could possibly work*". Wiki is a piece of server software that allows users to freely create and edit Web page content using any Web browser. Wiki supports hyperlinks and has simple text syntax for creating new pages and cross-links between internal pages on the fly. The use of wiki is unusual among group communication mechanisms in that it allows the organization of contributions to be edited in addition to the content itself. Like many simple concepts, "open editing" has some profound and subtle effects on Wiki usage. Allowing everyday users to create and edit any page in a Web site is exciting in that it encourages democratic use of the Web and promotes content composition by non-technical users [19].

Conclusions

Guidelines can be outlined based on the observation of how lifelong learners manage their knowledge and the available web tools and applications. Free and unconstrained learning is not applicable in the framework of educational institutions, yet it can give good ideas to depict the road map for the design and development of next generation learning management systems.

Future development can consider the integration of podcasting and audio/video streaming instead of traditional text based documents for passing information and the use of social connection within the whole learning/teaching process. Workgroups and discussion boards such as forums are not enough any more for people has developed new social networking ability. Knowledge sharing and the collaborative knowledge construction are the key points and they can be achieved by means of already available technologies.

Bibliography

[1] Conner, M. L. (2004) "Learn More Now: 10 Simple Steps to Learning Better, Smarter, and Faster", Wiley

[2] Cardinali, F. (2006) "Innovating eLearning and Mobile Learning Technologies for Europe's Future Educational Challenges, Theory and Case Studies" in Lecture Notes in Computer Science, Innovative Approaches for Learning and Knowledge Sharing, Proc. of the First European Conference on Technology Enhanced Learning, EC-TEL 2006 October 1-4, 2006, Crete, Greece

[3] Field, J. (2006) "Lifelong Learning and the New Educational Order" Trentham Books

[4] Tobin, K. & Tippins, D (1993) "Constructivism as a Referent for Teaching and Learning" in: K. Tobin (Ed) The Practice of Constructivism in Science Education, pp 3-21, Lawrence-Erlbaum, Hillsdale, NJ

[5] Gergen, K.J. (1995) "Social Construction and the Educational Process" in L.P. Steffe & J.Gale (Eds) Constructivism in education (pp 17-39). Hillsdale, New Jersey: Lawrence Erlbaum

[6] Vygotsky, L. S. (1978). "Mind in society" Cambridge, MA: Harvard University Press

[7] Cobb, P. (1994) "Where is the mind? Constructivist and Socio-cultural Perspectives on Mathematical Development" in Educational Researcher, 23(7), pp 13-20

[8] Cobb, P. (1998) "Analyzing the mathematical learning of the classroom community: the case of statistical data analysis" in Proc. of the 22nd Conf. of the International Group for the Psychology of Mathematics Education, pp 33-48, University of Stellenbosch, South Africa

[9] Salomon, G. and Perkins, D. (1998) "Individual and Social Aspects of Learning" in P. Pearson and A. Iran-Nejad (Eds) Review of Research in Education 23, pp 1-24, American Educational Research Association, Washington, DC

[10] Wood, T., Cobb, P. & Yackel, E. (1995) "Reflections on learning and teaching mathematics in elementary school" in L. P. Steffe & J.Gale (Eds) Constructivism in education (pp 401-422). Hillsdale, New Jersey: Lawrence Erlbaum

[11] W3C Semantic Web Activity, available on-line at the URL http://www.w3.org/2001/sw

[12] W3C Resource Description Framework (RDF), available on-line at the URL http://www.w3.org/RDF

[13] Mc Guinness, D.L., van Harmelen, F. (2004) "OWL Web Ontology language overview" w3c recommendation, available on-line at the URL http://www.w3.org/TR/owl-features

[14] Yang, J, (2004) "Metadata search engine for knowledge management in learning communities" in Proc. of the Int. Conf. on Computers in Education, pp. 787-792, Melbourne, Australia

[15] Moon, G.S. (2001) "Design and Implementation of Metadata Management System for WWW Courseware", in Proc. of the 31st Conference on Frontiers in Education, Volume: 2, 2001, P160 -164

[16] Zhao, C., Wan, L., and Yu, Y. (2006) "Construction of a Distributed Learning Resource Management System Based on RSS Technology" in Proc. of the 36th ASEE/IEEE Frontiers in Education Conference, October 28-31, 2006, San Diego, CA

[17] Powers, S. (2005) "What are syndication feeds", O'Reilly Media Inc.

[18] Kline, D., Burstein, D. (2005) "Blog!: How the Newest Media Revolution is Changing Politics, Business, and Culture", Squibnocket Partners, L.L.C.

[19] What is a Wiki (2002) Available on-line at the URL http://wiki.org/wiki.cgi?WhatIsWiki

Developing a Personalized Health Promotion Weight Management System

Heng-Shuen Chen, Lan-Hsin Hung, I-Wen Chang, Sheng-Hui Hung, Shiow-Li Hwang

Heng-Shuen Chen
Department of Medical Informatics
National Taiwan University College of Medicine
Taipei, Taiwan
chenhs@ntu.edu.tw

***Abstract*— Recently, we have developed a Personalized Health Promotion Weight Management System (PHPWMS). This system is intended to coordinate the "six e-learning website plan" originally developed by the Ministry of Education (MOE) in Taiwan. The system provides electronic nutritional knowledge, on-line food exchange database access and the ability to record daily energy intake to the individuals who want to lose weight. Individuals can easily understand the relationship between dietary planning, exercise, and weight control through PHPWMS. We expect that overweight individuals can have a healthy dietary intake and lifestyle through this personalized weight management system, and gradually lose weight successfully and safely.**

Keywords - Personalized; health promotion, weight-management; e-learning

I. INTRODUCTION

Along with the growing world economy, many countries tend to experience increasing chronic diseases in younger populations, and found obesity as an important index of many of these diseases [1]. Therefore, educating people to adhere to an adequate and balanced diet as set forth in the dietary guide is important to promoting national health. Furthermore, personal computers are becoming more widely used in the general population, and therefore an electronic personalized health promotion weight management program will be an effective way to help people get information for a healthy diet and promote good weight management. Since 1993, the MOE of Taiwan has constructed six e-learning websites, including life, health, nature, science, arts, and history to promote high quality digital learning content. PHPWMS belongs to the health category of the six e-learning websites (as shown in Figure 1).

Research has shown that nutrition, lifestyle and exercise can be successful keys to losing weight [3, 4]. Diet and lifestyle are risk factors in obesity and will influence the results of a weight-management program [5]. Changes in dietary habits will be an efficient way to lose weight [6]. Therefore, by providing personalized information rather than neutral messages, the user will benefit by keeping with the weight-management plan and consequently have better results in the program as a whole [7]. The weight-loss E-learning program, through integration with courses at various junior high schools, has shown good results [8, 9]. However, regaining weight after the regimen is a significant problem after weight loss, so it is necessary to monitor the weight maintenance to see the efficacy of weight-loss e-learning program.

The purpose of this study is to develop a personalized health promotion weight management system by providing electronic nutrition learning lectures, creating a graphical food exchange database, and an online calorie calculating function. Furthermore, the system intends to help the user to establish accurate weight-loss concepts and to gain good dietary behavior to achieve their weight-loss goal and prevent the reoccurrence of weight-gain.

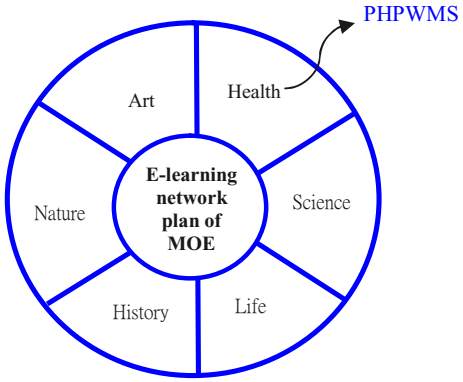


Figure 1. The six e-learning website as promoted by the Ministry of Education in Taiwan.

II. METHOD

We proposed the following strategies to enhance the personalized health promotion and weight management functions.

A. Establish the personal database

1)The recommendation of personal dietary intake

Weight-management members should enter their personal information, including gender, height, weight and physical activity into the database. The PHPWMS will calculate the individual's ideal body weight and recommended daily energy intake according to the dietary reference intakes (DRIs) of DOH in Taiwan. In addition, the personal DRIs can be used as a reference to create a weight-management plan.

2) *Personal diet and exercise record*

According to the food exchange table, PHPWMS classifier food into six categories including: grains; fish, beans, and eggs; dairy products; vegetables; fruits; and oils, sugars and salts. The image of seventy-three common food was presented on the personal recording system with the unit per exchange (e.g. one exchange unit of milk is one 240mL bottle of milk). In the recording process, the weight-management member was guided step by step to accurately record the food intake. This process can further help weight-management individual to understand the relationship between food and energy (Figure 2). Once the individual understands the concept of food exchange units and their consequences in the recording process, it will be helpful for the weight-management member in dietary self-management in the future. Moreover, in reference to the recommended daily food guide, weight-management members can arrange their personal menus by the online food recording system.

Beside the dietary control, exercise is another way to help lose weight. PHPWMS divides activity into five categories based on the exercise pyramid. There are fifty four exercise regimens in the exercise database. The weight-management member can calculate the energy expended during exercise by inputting activity types, exercise durations and the heart rate. This function is very helpful for weight-management member to get enough exercise (Figure 3).

3) Set the personal weight-loss goal

Rapid weight-loss is not healthy. The better weight-loss rate is reducing the subject's food consumption by 500-1000 kilocalorie daily or 0.5-1kg per week. However, personal cooperation and compliance are the determining factors. After setting the weight-loss goal, PHPWMS will calculate the recommended daily dietary intake and exercise. Though diet control and exercise are healthy ways to lose weight, the effects of the weight-loss regimen will not be prominent at first. Thus, PHPWMS provides flexibility for the weight-controller function to adjust the prearranged weight-loss rate to reduce the discomfort and to prevent loss of heart from the patients.

4)Customized Exercise Selection

Depending on eleven different physiological and health conditions, PHPWMS can provide the appropriate exercise list in order to avoid injury from the exercises.

Figure 2. By selecting food type and inputting the food exchange, PHPWMS can show the energy contained by food.

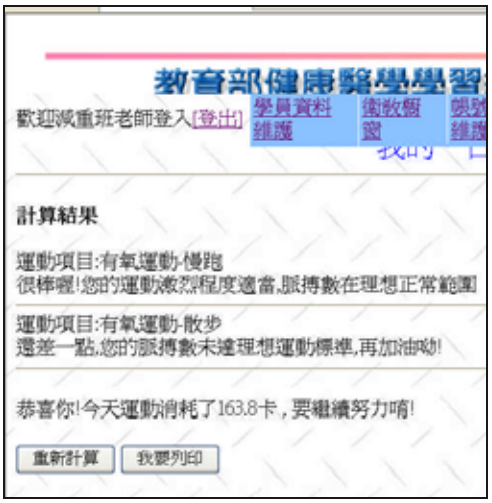


Figure 3. PHPWMS can calculate the energy consumed by exercise.

5) Establishing the Personal Weight-Management Database

PHPWMS also provides weight-management curves using available weight data. This helps to establish observations of long-term patterns in the subject. PHPWMS provides a drawing function to produce weight-variance curves such as body mass index (BMI) curves and waist to hip ratio curves to monitor the weight-loss progress.

B The user-oriented interface

1) Really Simple Syndication (RSS): PHPWMS consist of

the RSS function to provide news, discussion message and a few weight-control electronic courses to the subscribed members. Because of its ease of use, RSS can remind the weight-controller to record the daily weight-control data.

2) Message Board: The message board is designed for weight-management members to easily leave personal questions or messages to their Health Promotion managers. It can reduce the frustration during the weight-loss period and such group support is critical for the ensuring the success of the weight-management plan. For educational purposes, the Health Promotion managers can use the message board to share weight-management teaching experiences with others.

3) Online Availability of the Electronic Course: The convenience of having the weight-management courses available online for the weight-management member will significantly increase the cooperation of the subject. This convenience is further compounded by having the courses in popular formats such as MS Word, Flash, or graphical documents.

4) PHPWMS Customization for Various Demographic Groups: For example, in the children's version, the simple concept of a balanced diet can be taught through many interesting games. In the student version, the importance of exercise can be emphasized through demonstrations about exercising in a healthy lifestyle. Interactions with parents are also included in this version to encourage the weight-management plan's application in the entire household. In the general population version, a large quantity of knowledge about health and cooking skill are contained.

5) Evaluation and Ranking Boards: The evaluation and ranking board are designed for members to score their knowledge of health promotion and weight control in order to evaluate its usefulness in weight-management. The feedback will be a reference for dietitians to improve PHPWMS and to select a more comfortable personalized weight-management plan in the future.

III. DISCUSSION

1. According to the tenants of the daily diet guide, "it's proper to have balanced food kinds in one day". It is very convenient for individuals to get a reference to next meal by the food record, so individuals can perform adequate weight management through using PHPWMS.

2. The system focuses on the total energy intake in order to implement the weight control plan. However, for the middle-aged adult with a chronic disease, the regulation of specific nutrients such as cholesterol and sodium should be emphasized. In the future, we can integrate the 「Taiwan Food composition nutrient database」 to the system. There are twenty four nutrients of each food in this database. Therefore we can not only monitor the intake of total energy but also the specific nutrients for each individual. This will be helpful for a dietitian to provide suggestions as early as possible to those who would take extreme weight-loss plans while consuming only specific foods. Furthermore, we can set up the personal drug database in the PHPWMS to be a clinical nutrition consult assistant system.

3. Although obesity is prevalent in modern societies, malnutrition—especially in the more-elderly populations—is also a chronic problem that needs to be addressed. PHPWMS can also be employed to insure that these individuals obtain a well-balanced diet. This adaptation will utilize the PHPWMS's ability to present nutritional information on various food items that can then be compared to the dietary guide.

4. In regards to the diverse characteristics of various food groups, PHPWMS has a few new functions. Other than merely selecting the food item and exchange, PHPWMS has the ability to monitor the energy consumed. The regulation of 'Nutrition Labeling for Packaged Food' will be in place by 1st January 2008. This will ensure that packaged foods will provide nutrition labeling- including contents of energy, proteins, fats, carbohydrates and salts presented at a conspicuous place on the outer package or container of the product. Due to the prevalence of eating outside the home, this regulation will protect consumer by guaranteeing the right to know their diets and will be convenient for the weight-management controllers to record energy consumption. Moreover, we hope the government will promote the labeling of nutrients at restaurants in the future. That will not only help individuals to do weight management, but also stimulate restaurants to provide a healthy diet choice.

IV. CONCLUSION

PHPWMS is a system that effectively integrates weight-management members and health promotion managers using an electronic interface (Figure 4).

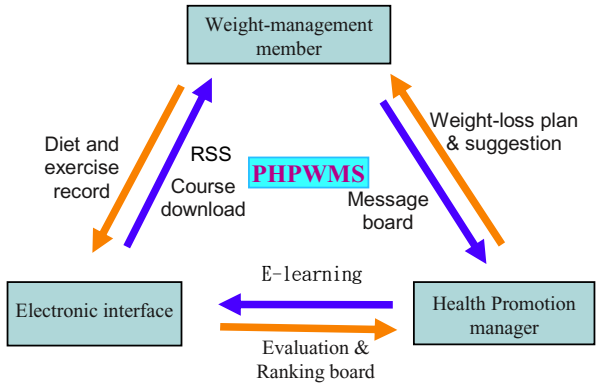


Figure 4. PHPWMS is an integrated system to connect among the dietitian, weight-management member and electronic interface.

The PHPWMS consists of the following components:

- Providing accurate electronic weight-management knowledge by a professional dietitian.
- Diet and exercise recording databases to help individuals to understand food exchange skills, change their dietary behaviors, reduce weight gain and to eventually lose weight successfully.
- Message board and evaluation functions are used to provide feedback to the dietitian in order to improve the weight-management system.

We expect to develop PHPWMS as a personal digital assistant in the future to enhance the health self-management of individual to increase their personal quality of life.

REFERENCES

[1] Thomas, Non Eleri. Baker, Julien S. Davies, Bruce. Established and Recently Identified Coronary Heart Disease Risk Factors in Young People: The Influence of Physical Activity and Physical Fitness. Sports Medicine; 2003, Vol. 33 Issue 9, p633-650, 18p

[2] Geise, T 1; Cooper, L 1; Fiutem, J 1; Dell'Olio, J 1; Turgiss, J 1; McInnis, K FACSM 1; Rippe, J FACSM 1 Identifying Individualized Goals, Perceived Barriers, and Readiness for Behavior Change in Obese Adults. Medicine & Science in Sports & Exercise. 34(5) Supplement 1:S63, May 2002.

[3] Cooper, L; Geise, T; Fiutem, J; Dell'Olio, J; Turgiss, J; McInnis, K FACSM; Rippe, J FASCM. Perceived Barrier to Weight Loss in Obese Adults. Medicine & Science in Sports & Exercise. 34(5) Supplement 1:S62, May 2002.

[4] Chiu-Ming Hu, Heng-Shuen Chen, Chih-Ying Huang, Sheng-Hui Hung, Jing-Ing Yu, Li-Jie Huang, Wei-Yi Fan, Chien-Yeh Hsu Developing a Health Education E-Learning System for Weight-loss Program. e-Health Networking, Applications and Services, 2006. HEALTHCOM 2006. 8th International conference . pp. 24- 27.

[5] Sullivan, S.Cloninger, C. R. Przybeck, T. R. Klein, S. Personality characteristics in obesity and relationship with successful weight loss. International Journal of Obesity; Apr2007, Vol. 31 Issue 4, pp669-674.

[6] W Anderson L Grant L Gotthelf & L T P Stifler. International Journal of Obesity. 2007, No.31, pp488 - 493.

[7] Moreno, Roxana; Mayer, Richard E. Engaging Students in Active Learning: The Case for Personalized Multimedia Messages. Journal of Educational Psychology. Vol.92 No.4,:pp.724-733, December 2000.

[8] Sheng-Hui Hung, Heng-Shuen Chen, Chien-Yeh Hsu, Hsiao-Lien Chen, Ging-Ying Lin, Wen-Yin Chang, Shiow-Li Hwang. The Effectiveness of a Weight-loss E-learning Program for Obese Junior High School Students. e-Health Networking, Applications and Services, 2006. HEALTHCOM 2006. 8th International Conference . pp. 50- 54.

[9] J Reese, Shelly.Going Online to Take Pounds Off. Prevention; Dec2003, Vol. 55 Issue 12, pp149-151.

Virtual Community systems: a pragmatical approach to integrate e-learning and information systems

Luigi Colazzo, Andrea Molinari

Università di Trento, Dipartimento di Informatica e Studi Aziendali
via Inama 5 38100 Trento tel. 0461 882144 eMail colazzo@cs.unitn.it amolinar@cs.unitn.it

In this paper we present our experience in integrating a in-house e-learning system into a Virtual community system that support the activities of a Faculty, not only from the didactical point of view, but also from a "information system" perspective. This experience involved designing, implementing and maintaining a totally new system, the "Online Communities" project. Although many LMSs exist on the market, also free, there are many reasons that lead us to rewrite from scratch it on totally new foundations. In late 2003, the old LMS has been replaced by a new one based on a different approach, specifically a Virtual Community System (VCS) called "Online Communities". The new concepts implemented into the system help a traditional Academic institution to change the interaction paradigm with its users, from an approach based on "physical" relationships to one based on virtualization - as far as possible - of these relationships. This virtualization has been created through the mediation of ICTs, and in particular with the help of a virtual community system oriented not only to the typical e-learning aspects, but extended to all possible communication forms among the different actors that play a role in the academic community. More specifically, the system integrates the traditional e-learning services with the ordinary functions offered by an information system, like managing core-business transactions, checking access to the system, supporting decisions, reporting etc.

Introduction

Learning management systems (LMS) [A'herran 2001], [Hall 2001], [McMahon & Luca 2001], [Hanna et al 2000] are software applications, normally based on Web technologies, used to plan, create, and possibly assess the learning processes of an educational institution. Typically, teachers find in a LMS a set of functions to create and deliver content, monitor student participation, and assess their performance. To students, a LMS can provide with the possibility to download, see and use the educational material, to use interactive tools such as discussion forums, chat, video conferencing etc. LMS are therefore full-featured systems, normally complex to use, but rarely they adapt to the specific needs of a Faculty, due to their general mechanism of managing specific issues.

For this reason, personalization is usually a difficult and costly task, due to the lack of personalization tools or to the complexity of the system to be personalized. Moreover, a lot of information present in a LMS are extremely useful for administrative tasks: for example, functions required to teachers that could be easily performed in a centralized way by the dean's secretariat, or reports (normally managed by secretaries) derived directly by the daily teachers' activities. Many other bureaucratic-specific activities, like the drawing up of the exams offered by the Faculty, have significant influences on the rest of the didactics, because the assignment of a teacher to a course will derive directly from this list. Also the relationship students-administrative offices could be facilitated, for example for the verification of exams, program of studies, plan of the lessons etc.

In the late '90s, the Faculty of Economics of the University of Trento decided to set up a Learning Management System (LMS) to be used for educational activities in a blended modality. One of the crucial starting points was, as many others around the world, the "make or buy" decision, a typical dilemma for those who are in front of a new project that will intensively use software products and that have internal potential to build it on their own. The paper presents our experience in going in two different directions, respect to the mainstream that most of educational institutions follow today. First, we decided to replace the LMS approach, typically based on the concept of "course", with a highly pervasive "information system" approach. On top of this idea, we based our design on a more "community" vision of the educational institution, seen like a set of Virtual Communities of many different users with different roles inside different communities [Parchoma, G. 2005], all based on a broader idea on "information management system", rather then "Learning management system". Second, we decided to "Make" the system starting from an existing LMS we already built. The availability of the source code of our system was not the only reason we decided to "make" it. Due to the concept of "virtual communities" we wanted to implement, and due to the fact that we strongly believe in a deep connection among learning/teaching activities, community activities and Faculty/Campus Information Systems, we were not able to find satisfying systems around the Free/Off-the-shelf markets.

Both of them sound a little bit like "swimming against the stream" of a generalized choice in academic institutions. We believe that the e-services offered to users are one of the main factors that could distinguish educational institutions and their formative offer, and having a personalized, open-source, internally developed solution is a crucial competitive factor. In our opinion, the quality of an academic institution is strongly related to the quality of its educational processes. These, in turn, are based

on the complex relationship between professional competencies of researchers/teachers and used teaching methods. In the good and in the bad, the real processes occurring in an academic institution are based on the quality of this relationship.

Studying in a high-quality educational institution means to learn from an intellectual environment where the knowledge transfer (education) and creation (research) are interwoven, perceptible and "internalizable" by students. In marketing terminology, this could be considered as the "brand" of an academic institution. If we aspire to be seen as an educational institution with a high-quality teaching, we better try to reify our best practises into software processes and systems, instead of importing standard, pre-digested, homologated educational processes determined by software products that will be used by thousands of other institutions.

In this paper, we will try to justify why the shift from the traditional paradigms included in a LMS towards a Virtual Community system makes a learning environment more profitable and performant for its users, like our experience is demonstrating. We believe there are many interesting aspects following this paradigm shift we are presenting. We will concentrate on two elements. First, we will discuss the extension of the traditional e-learning approach towards a concept of virtual community. Here we include issues related not only to educational aspects, but also to the solution of the multitude of communication and interaction problems that users have while interacting with the academic institution. A second aspect regards the approach used in the design, construction and evaluation of the system. The project has been structured and implemented under the steady supervision of end-users, i.e. the actors of the system.

1. Constructing a Learning Management System

The first experience in using ICTs in learning context for our University was done late in 1999 by the Faculty of Economics, developing a software system call "Online-courses" [Colazzo et al., 2002]. The first version of the system was based on some assumptions of the authors that the following successful experimentations demonstrated to be valid.

1) User-centred design. The system is built with and around users' requirements, using agile and evolutionary prototyping software techniques. The software developing cycle and its architectural implications are derived from the needs of the actors involved in the process. We spent many hours and many resources in going around teachers' rooms and students' classes, to look at the needs and the usage people did of the system. The system, therefore, was not a software architect's conception; it has been substantially driven by users' suggestions.
2) Teaching methods vary depending on the disciplinary domain of the courses and on the specific user's preferences. In our opinion, a software system should not overwhelm the way people act by simplifying a complex relationship like the training process.
3) The usage of a LMS should not be mandatory. In a real situation, there are people not attracted by e-learning technologies, and that will therefore avoid using the system. These subjects could not be labelled as "bad teachers".
4) Every constraint on the nature/type of learning objects used inside the system is needlessly restrictive and counter-productive. This does not mean that learning objects built using standards (like SCORM) are not usable in the system, but we cannot imagine that this is a necessary condition. Imposing such condition means to move users away.
5) There must be no filters between a teacher and his/her students. A teacher should not need for an "intermediary" in order to interact with his/her students through ICTs. This "welfare" teaching model, where an intermediary helps teachers to produce educational material has failed in our University as probably in others. In our opinion, the reason is very simple and has many reasons. First, there is no enough time and money to pay huge numbers of experts that help all the teachers in a Faculty, and it is not possible to find few people so prepared to be able to help teachers in domains that range from fuzzy logic to labour psychology, from telecommunication to economic geography (just to cite some areas in our Faculty). Second, it is impossible to find an expert in the field of the teacher able to comprehend and transmit the topic like the teacher herself. Third, every teacher is able to teach with her own method, and will probably have problems in recognizing herself in something produced by others. Probably this is the reason why many of us look at learning objects produced by other colleagues, but in the end we produce their own teaching material, because of their need to "feel" it as a personal opera, customized on own needs. Based on this assumption, the software should behave like an amplifier of the communication space between teachers, students, tutors, and assistants. It will happen for sure that some colleagues will use the system under its capabilities, but this is largely better than nothing (and it's explicitly allowed, see point 3). We are aware that these ontological prejudices are arguable [Herrington et al., 2005]; in fact, they do not coincide with those of many other software implementations, but our experience presented in this paper has verified them all as being successful starting points.

The system, named "Online Courses", started its services in late 1999. It was a traditional Learning management system built around the concept of "course. The majority of its functions were devoted to the two main actors of every educational process, i.e., the student and the teacher. In fig. 1 the central

classes of the system with their relationships are show in UML notation

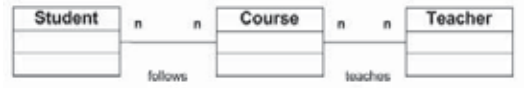


Figure 1 Core classes in "Online Courses"

The functionalities were all included in the services available inside one course. The most important services were the following:

- Synchronous services: chat, remote application control, videoconferencing, FTP
- Asynchronous services: bulletin board, forum, mail
- Informative services: contact information, course program, course organizative information, timetable, recommended books, bibliography, exam modalities, course syllabus
- upload /download services

This system has been mainly used by the Faculty of Economics, but during these years it has been extended to courses of other Faculties, with the cited approach "help yourself if you like it". In table 1, the main highlights of this experience are listed

Online courses present since 1999	670
Online courses activated 2004/05	218
People enrolled in the system	11.939
Number of total qualified hits since the full operational activity of the system (2001-2004)	1.273.333

Table 1: summary of "online courses" usage

2. Moving towards a Virtual Community System

The successful experience of the system "Online Courses" showed a number of questions that stimulated us towards new approaches. The first issue regarded the systemic nature of a learning system. If we use such a software system in order to improve the training processes of an educational institution, we can observe that this system is strictly connected with other sub-systems of the general information system of the institution itself. An obvious relationship is with the students' secretariat, where typically we have all data regarding students' careers, financial information regards fees, exams, etc.

Another trivial relationship regards the information system that helps the Dean and her staff to manage their activities, from simple to complex decisions (timetable, course assignments, exam and bachelor commissions, study plan formulation, lists of courses offered etc.). This second option clearly regards the governance of academic institutions, and specifically didactic activities. An effective system for managing teaching activities mainly needs data streams from outside the institution, but at the same time needs to be feed with a variety of information from its internal educational processes.

The possibility of having this information is intrinsically connected to the availability of some form of computerized learning management system. Without a computer-mediated communication universe, it's hard to extract data from the phenomenon going on in the field: the only possibilities are to extract some of them (not all of them) from the records of the students' secretary or from questionnaire, and this means to have a relevant delay in the availability of the information. Besides, decisions involving process changes (for example, course assignments, study plans etc.) are assumed corporately in different discussion tables (Degree course council, Faculty council, Campus-wide council). Most of the ordinary activities are articulated in diffused and capillary discussion moments, that sometimes overpass the official schema and meetings: are we really sure that we, the teachers, are "teaching" in a proper sense only when we are in the classroom? [Chang 2003] [Michailidou 2002] [Ma 1999] Under this perspective, the abstract concept of "course", around which many LMSs are built, seems to be an unnatural restriction.

Two of the most relevant problems in the university education are the drop-out rate and, on the other side, the prosecution of studies beyond the terms. Usually, the reasons behind these phenomenons are only partially known, and most of the times involve subjective motivations. Students leaving the studies should be considered as a defeat for the educational model of an academic institution. A greater attention given to students with difficulties could decrease the phenomenon, but also in this case, every possible intervention will be useless or inappropriate if not supported by data from the information system. Finally, Italian academic institutions slowly start to comprehend the advantages of maintaining relationship channels with their own ex-students. The more evident advantages regard the possibility of creating a sort of "fidelity program" with the past students, in such a way that they can find chances for post-lauream professional studies. This is an interesting possibility also from a business perspective, and a way to provide Universities with a deeper territory and market relationship. For academic institutions, this is a necessity that can face the growing request of high-level education, generated by productive processes of companies that are involved in frequently changing scenarios, where personnel's competences and skills are crucial for successful activities. Electronic communication is a good candidate for these kinds of services that Universities could supply to the business world, also through their own past students.

In conclusion, the necessity of combining the systemic nature of a learning system with the collective nature of decision-making processes in an academic institution, and the necessity to supply technological tools to extend educational institutions' policies, led us to reconsider the nature and the intrinsic architecture of the Learning Management System used by our Faculty. As the foundation of this process of revision of the general architecture of the system, we set the idea of "virtual community" as

the cornerstone of the system. For a review of the most recent experiences in Online/Virtual communities, see [Havelock, 2004]

3. The Virtual Community system (VCS) "Online Communities"

In the history of the sociological thought, the label "community" has been used in different ways, to point out a variety of social phenomenons. Nevertheless, right in the difficulty found for the definition of the concept of "community", we find the first aspect that unifies the "real" and the "virtual" concepts . Usually, in the sociological literature, the concept of «community» is considered in relationship with the concept of «society», and the separation line of the two social structures is marked by the degrees of freedom of the individuals. In fact, in a community it is possible to observe a greater quantity of ties (for intensity and extension) than in a society. In return, individuals get cohesion, safety and solidarity. Virtual communities [Beamish 1995], [Jones 1995, 1997], [Lévy 1994, 1995] [Rheingold 1993], [Turkle 1995], are something else respect to the concept of community as in the classics. Their use in a formative environment is further different. For example, the absence of anonymity, a typical aspect of these communication spaces, produces substantial effects on the dynamics of virtual relations. These aspects will not be further discussed, but it's important for us to stress that the software was built in order to virtualize real processes that happen in real communities. The traditional model implemented inside LMS (like "Online Courses" itself) demonstrated some limitations and contradictions. The majority of these systems, in fact, assign fixed roles to subjects participating to educational processes.

These roles are normally included in a vision of the training process that we could define "transfer" model: the teacher (and to a lesser extend, the tutor) owns the knowledge, and this is transferred to students via a sequence of lectures. The student learns from references or books while guided by the teacher's lecture. For courses where the "learning by doing" is important, the course is joined with some lab activities, but the overall model remains the same.

This model, once reified in the software, represents a crystallisation of the original model, only apparently more participatory (students have to do some actions in the virtual), but substantially it enforces the power of one side (teachers, in fact, can use the software with higher privileges than students). Consequently, many non-traditional learning and teaching approaches (like Problem based learning, Cooperative learning, Learning by projects, etc.), instead of being facilitated, are obstructed by the software itself. Moreover, if we consider technical courses, where more practical skills must be acquired and demonstrated, very often students have to develop a project. This could be an individual work, but more frequently it is a joint effort among students of a group. It would be more productive to let students work together also using cooperative tools that allow them to interact among themselves and with teachers / tutors, but this kind of collaboration "freedom" (with all the administrative problems behind the scenes) is very often extraneous to a typical e-learning system, like our old "Online Courses". "Online Courses" got immediately the appreciation of two classes of students: working-students (in our Faculty, approximately 50%) and students living outside our town. Students belonging to these two categories are those that most frequently leave the study degree, and after the release of the system we recorded a decrease of this drop-out effect, though there is no clear evidence about this relation.

In order to intervene on these situations, we could proceed in two directions. The first consists of creating an alerting application available for the students' secretary. When the student is at a crisis in his/her career, the frequency to the exams decreases like (very often) the results in the exams. It would not be difficult to highlight these situations and to point out them to university / faculty authorities, thus avoiding that this important information stays hidden into administrative meanders. Nevertheless, this way of acting highlights the problem when it's too late, when the student's problems have already produced their effects.

A second solution could be easily implemented using electronic communication, by creating direct communication mechanisms between students and teachers, thus avoiding delays and further difficulties for the students. Probably both of the solutions must work together, but what is clear to us was that the limited nature of the "course" concept was not suitable for this scope. Similarly, another typical problem of academic institutions regards the ability to intercept the request for lifelong learning that is rising from the market. This request could find adequate responses in computer-mediated communication, but in order to do this, universities must activate stable, affordable and effective communication channels with the outside world. One of these sources of information could be immediately available: direct relationship with its ex-students. In the italian tradition, this communication channel has been used seldomly, or has been interpreted in the wrong way (recreation, ceremony, bureaucracy etc.). In other countries, the relationship between students and their university is managed through the associations of ex-students, and this channel is not only an informative pipe, but also an economic flywheel.

We saw that normally this relationship is maintained within the first years after the bachelor, and that normally this is a relationship with the professor that supervised the student's thesis. The virtual space created between the student and the supervisor should enter totally in a learning system context, but normally an LMS avoid past students to use the system, once they got their bachelor. The creation of a virtual communication space between those that are studying and working now inside an educational

institution, and those who have worked and studied inside the same institution, and that are now employed in the productive business can represent a non-trivial advantage for lifelong learning programs. Nevertheless, also in this case the traditional "course" approaches and spaces are evidently inadequate.

These are only some examples for representing our decision to shift towards a more flexible, structured, wide definition of a communication space between people acting inside an educational institution going beyond the limitations of the approach based on the idea of "course". In fig. 2 we quickly sketch the central part of the schema adopted for the realization of "Online Communities": at present, the system is based on more than 300 classes, but all the main concepts are managed around these concepts: People; Role; Community; Right; Duty and Permissions.

Online Communities is exited from the experimentation phase in June 2005, after a long test phase and now is used in extensive way at the Faculty of Economics (an average of about 1.000 unique accesses per day). We summarize its main characteristics:

- better, clearer, easier roles/duties/rights management thanks to the creation of a community container, with people enrolled in the community with specific roles, duties and rights
- decentralized administration to the responsible of the community, that can in turn create other communities in which deciding the permission mechanism for their users.
- possibility of managing people having different roles in different communities
- further complexity in managing relationships among communities, instead of being limited with a "degree-course" two level taxonomy. For example, traversing hierarchies as in the case of teachers that have people in their courses from different faculties or institutions, or adding intermediate level of hierarchies also when a hierarchy has been created.
- possibility of creating communities between persons not related with educational activities (like secretaries, research groups, students' associations, recreation groups, post-lauream associations, etc.) with specific permissions on the community.
- extensibility of the hierarchy and the relationships between communities to whatever level and degree of complexity we desire
- extensibility of the system to new services that are instantly activated to the respective communities simply by choosing permissions for groups, single participants, entire branches of communities etc. A typical example regards mobile services, like SMSs or shared agenda and appointments on mobile phones.
- more flexible management of hierarchical relationships between communities, that have different relationships respect to "courses"
- time management inside the community: the visibility of events of the community could be seen from a hierarchical point of view, from a roles/rights/duties point of view, from a device point of view (computer, mobile phones, PDA ….) [Colazzo et al.,2004]
- more flexible and easy administrative tasks, thanks to community hierarchies and relationships

4. The deployment of "Online Communities"

The development of "Online Communities" started in 2003 from a discussion regarding the limits of "Online Courses" system. During 2002 the project team designed the new architecture, and in early 2004 the new system was ready to be tested on the field. The primary objective was the implementation of all the functionalities provided by the old system, but the necessity of submitting the system to a long experimentation, due to its novel approach, has emerged immediately. The experimentation phase took 4 bimesters, at the beginning with some courses where students were asked to contribute to the project. This initial step was fundamental to orient the development of the system, and to capture users' need and difficulties with the new approach based on the concept of community. A second, wider test was performed during the 2004/2005 academic year. In figure 3 the monthly accesses from November 2004 to July 2005. Since first bimester 2005/2006, Online Communities is operative for all the courses in the Faculty. In the right part of figure 3, this information is presented.

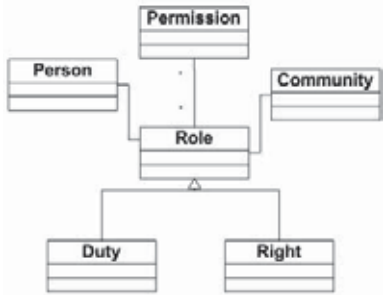
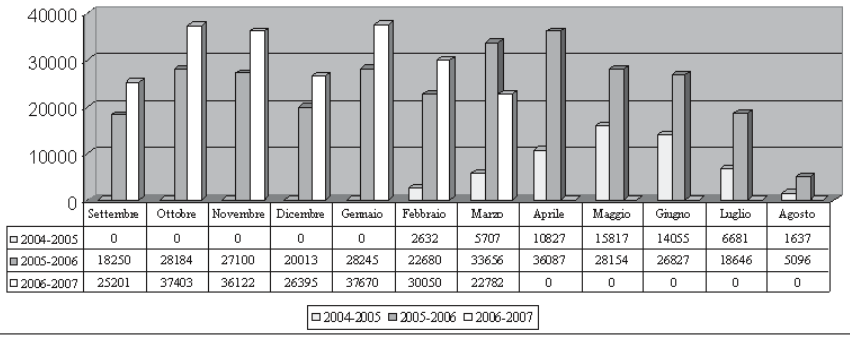


Figure 2. Core classes of "Online Communities

Conclusions

Facing teaching and learning problems using ICT tools in academic settings emphasizes the systemic nature of these processes. The governance of these systems assumes a crucial role. Academic institutions, during one thousand years since their birth, have built governance methods and practises based on the conscious participation of innovational and educational processes' actors, on a diffused network of decisional moments, on unwritten though not less important and binding traditions. These are complex governance forms that cannot be simplified beyond a certain limit. The risk of an excessive simplification in the name of efficiency (for example, eliminating many of decisional moments in order to create a more reactive system) is real and present. We believe that software systems in academic institutions should be built assuming that complexity and personalization are fundamental requisites, not

simply burdens or problems to be ignored.
We would like to express our gratitude to all the people that are working with us in this project: Francesco Conte, Riccardo Borsato, Mario Ossi and all the other people that, with their contributes, helped us in creating a successful experience.



	Settembre	Ottobre	Novembre	Dicembre	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio	Agosto
2004-2005	0	0	0	0	0	2632	5707	10827	15817	14055	6681	1637
2005-2006	18250	28184	27100	20013	28245	22680	33656	36087	28154	26827	18646	5096
2006-2007	25201	37403	36122	26395	37670	30050	22782	0	0	0	0	0

Figure 3 "OLCom access rate 15 Feb. 2004 - 22 Mar. 2007

References

A'herran A., (2001) *Integrating a course delivery platform with information, student management and administrative systems*, in Proc. EDMedia 2001, Tampere, Finland, June 25-30 2001

Beamish A.,(1995) Communities on-line, tesi di dottorato di Anne Beamish, testo disponibile all'indirizzo http://loohooloo.mit.edu/anneb/cn-thesis/index.html, 1995

Chang, Yung-ta (2003). *Learners as Designers: A Case Study on the Virtual Learning Environment*. Honolulu, Hawaii, USA: AACE.

Colazzo L., Grasso R., Molinari A. (2002) Corsi on line un sistema per distribuire materiali didattici in una Facoltà Universitaria, Atti del convegno "Didamatica 2002", Napoli 14/15 Febbraio 2002

Colazzo L., Molinari A. (2004) *The Management of Chrono - Referenced Events in Virtual Communities*, in proc ED-MEDIA 2004-World Conference on Educational Multimedia, Hypermedia and Telecommunications, Vol. 2004, Issue. 1, 2004, pp. 117-124

Havelock, Bruce (2004). Online community and professional learning in education: Research-based keys to sustainability. AACE Journal, 12(1), 56-84.

Hall B, (2001) *Learning Management Systems. How to Chose the Right System for your Organisation*, Brandon Hall, 2001

Hanna, D. E., Glowacki-Dudka, M. & Conceicao-Runlee, C. (2000). *147 Practical tips for teaching online groups: Essentials of Web-based education*. Madison, WI: Atwood Publishing

Herrington, Jan, Reeves, Thomas C & Oliver, Ron (2005). *Online Learning as Information Delivery: Digital Myopia*. Journal of Interactive Learning Research, 16(4), 353-367.

Jones S. G., (1995), *Cybersociety*, Sage, London, 1995.

Jones S. G., (1997), *Virtual Culture*, Sage, London, 1997.

Lévy P., La Découverte, Paris, 1994 (trad. It. *L'intelligenza collettiva. Per un'antropologia del cyberspazio*, Feltrinelli, Milano, 1996).

Lévy P., (1995), *Qu'est-ce que le virtuel?*, La Découverte, Paris, 1995 (trad. It. Il virtuale, Raffaello Cortina, Milano, 1997).

Ma, Chia-Ling, O'Krinsky, Evelyn, Smith, Janice, Bump, Wren & Kayfes, Marlane (1999). *Virtual Teams in the Classroom*, Society for Information Technology and Teacher Education International Conference (SITE) 1999, AACE.

McMahon M., Luca J, (2001) *Courseware Management Tools and Customised Web Pages: Rationale, Comparisons and Evaluation*, Proc. EDMedia 2001, Tampere, Finland, June 25-30 2001

Michailidou, Anna (2002). *ELearn: A collaborative educational virtual environment*. World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (ELEARN), Montreal, Canada: AACE (2002)

Parchoma, G. (2005). *Roles and Relationships in Virtual Environments: A Model for Adult Distance Educators Extrapolated from Leadership in Experiences in Virtual Organizations*. International Journal on E-Learning, 4(4), 463-487.

Rheingold H.(1993), *The Virtual Community*, 1993

Turkle S.(1995), *Life on Screen. Identity in the Age of Internet*, New York, Simon & Shuster

Logging and Analyzing Learner Behavior in Web-Based E-testing

Gennaro Costagliola, Vittorio Fuccella, Massimiliano Giordano and Giuseppe Polese

Dipartimento di Matematica e Informatica, Università di Salerno
Via Ponte don Melillo, 84084 Fisciano (SA), Italy
{gcostagliola, vfuccella, mgiordano, gpolese}@unisa.it

Abstract

Several experiments about learners behavior during structured exam tests based on multiple choice questions have been carried out so far. Most of these experiments were performed by exploiting the think out loud method: the learners were informed of the experiment and had to modify their behavior in order to allow the experimenters to record information about their habits during the test. In this paper we describe a system which logs the interactions of the learner with the e-testing system interface during the test. Our system allows us to record information about their habits during on-line tests without informing them of the experiment and, consequently, without asking them to modify their behavior, thus obtaining more realistic data. In order to demonstrate the effectiveness of our system, we describe how it can be used for replicating on-line several experiments performed for traditional papery testing.

1. Introduction

E-testing systems are more and more widely adopted in academic environments combined with other assessment means. Through these systems, tests composed of several question types can be presented to the learners in order to assess their knowledge. *Multiple Choice* question type is extremely popular, since, among other advantages, a large number of its outcomes can be easily corrected automatically. Those items are composed of a *stem* and a list of *options*. The *stem* is the text that states the question. The only correct answer is called the *key*, whilst the incorrect answers are called *distractors* [13].

Among the disadvantages of *structured* tests, a low acceptance of the exam type by the learners is rather frequently noticed: many learners are afraid of not being able to best express their capacity, due to the characteristic of *multiple choice* questions of being *closed*.

In order to teach to the learners how to better perform on *structured* tests, several experiments aimed at tracking learners behavior have been carried out in the past using the think out loud method: the learners were informed of the experiment and had to speak during the test to explain what they were thinking, while an operator was storing their words using a tape recorder. This technique can result quite invasive, since it requires that the learners must modify their behavior in order to record the information to analyze.

In this paper we discuss a complete system that allows us to record and to analyze information about learners habits during on-line tests without informing them of the experiment and, consequently, without asking them to modify their behavior thus obtaining more realistic results.

The proposed system is web based and adopts the *AJAX* [10] technology in order to capture all of the learners interactions with the *e-testing* system interface (running in the Web browser). The system is composed of a logging framework which can be instantiated in any e-testing systems and of a stand-alone application which analyzes the obtained logs in order to extract information from them and to produce statistical analysis.

Finally in order to demonstrate the effectiveness of our system, it has been used for repeating several experiments performed in the past. Our experiments have been carried out in the ambit of a university course at the *Dipartimento di Matematica e Informatica* of the *University of Salerno.* The grade obtained on the tests has concurred to determine the final grade of the course exam.

The rest of the paper is organized as follows: In section 2 the logging *framework* and its integration in *eWorkbook* is presented. In section 3, we discuss the techniques employed and the results obtained through the experiments. Several final remarks and a brief discussion on future work conclude the paper.

2. The Logging *framework*

The purpose of the Logging *framework* is to gather all of the learner actions during the browsing of the Web pages of

the test and to store raw information in a set of log files in *XML* format.

The *framework* is composed of a server-side and a client-side module. The client-side module is responsible for being aware of the behavior of the learner while he/she is browsing the test pages. The server-side module receives the data from the client and creates and stores log files on the disk.

Despite the required interactivity level, due to the availability of *AJAX*, it has been possible to implement the client-side module of our *framework* without developing plug-in or external modules for Web browsers. *JavaScript* has been used on the client-side to capture learner interactions and the text-based communication between the client and the server has been implemented through *AJAX* method calls. The client-side scripts are added to the *e-testing* system pages with a light effort by the programmer.

The events captured by the *framework* are the following:

- Actions undertaken on the browser window (open, close, resize, load, unload);
- Actions undertaken in the browser client area (*key* pressing, scrolling, mouse movements and clicks);

The *event* data is gathered on the browser and sent to the server at regular intervals. It is worth noting that the *event* capture does not prevent other scripts present in the page to run properly.

The server-side module has been implemented as a Java servlet which receives the data from the client and prepares an *XML Document* in memory. At the end of the test *session* the *XML Document* is written to the disk. The logger can be instantiated and then enabled through the configuration.

The information *Model* used for the log data is quite simple and is shown in figure 1. The information is organized per learner test *session*. At this level, the username (if available), the IP of the learner and *session* identifier are logged as well as the agent information (browser type, version and operating system). Inside a *session*, a list of *event* elements is present. The data about the user interactions are the following:

- Event type;
- HTML source object involved in the *event* (if present);
- Timing information (timestamp of the *event*);
- Mouse information (pressed button, coordinates);
- More information specific of the *event*. I.e. for a response type *event* (a response given to a question), the question and option identifiers and the indication whether the response was right or wrong are recorded.

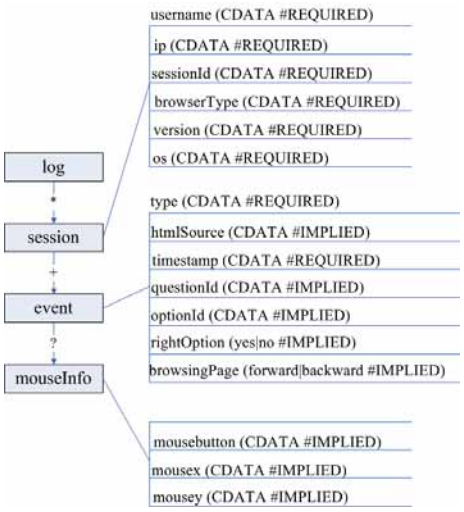


Figure 1. The information *Model* for log data

An important concern in logging is log size. If an experiment is done involving a large set of learners and the test is composed of many questions, log files can reach big sizes. A configuration system, including the following configuration settings has been conceived in order to reduce log sizes:

- List of events to capture;
- Sub-set of attributes to store in the log for each *event*;
- Sections of the Web pages (divs or table cells) to monitor as sources of the events;
- Time interval between two data transmissions from the client to the server;
- Sensitivity for mouse movements (short movements are not captured).

The configuration is read by the server-side module but affects the generation of the *JavaScript* modules running on the client-side. The architecture of the *framework* is graphically represented in figure 2.

On the user machine, everything can be done in the web Browser. The *JavaScript* modules for *event* capturing, dynamically generated on the server according to the configuration settings, are downloaded and run in the browser interpreter. Data is sent to the server through an *AJAX* request. On the server-side, a module called *RequestHandler* receives the data and sends it to a module called *LoggerHandler*, which organizes the *XML Document* in memory and flushes it to the disk every time a learner test *session* finishes.

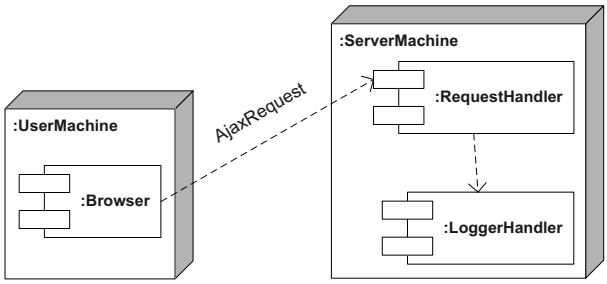


Figure 2. The Logging *framework* Architecture

2.1. The *eWorkbook* System and Instantiation of the *framework*

The above described *framework* has been instantiated in an existing Web-based *e-testing* system, named *eWorkbook* [4], which is used for evaluating learner's knowledge by creating (the tutor) and taking (the learner) on-line tests based on *multiple choice* question types.

eWorkbook is a Web-based *e-testing* system that can be used for evaluating learner's knowledge by creating (the tutor) and taking (the learner) on-line tests based on multiple choice question types. The questions are kept in a *hierarchical* repository. The tests are composed of one or more sections. There are two kinds of sections: static and dynamic. The difference between them is in the way they allow question selection: for a static section, the questions are chosen by the tutor. For a dynamic section, some selection parameters must be specified, such as the difficulty, leaving the system to choose the questions randomly whenever a learner takes a test. In this way, it is possible with *eWorkbook* to make a test with banks of items of different difficulties, thus balancing test difficulty, in order to better assess a heterogeneous set of students. *eWorkbook* adopts the classical *three-tier* architecture of the most common *J2EE* Web-applications. The *Jakarta Struts framework* has been used to support the *Model 2* design paradigm, a variation of the classic *Model View Controller (MVC)* approach. In our design choice, Struts works with JSP, for the View, while it interacts with *Hibernate*[7], a powerful framework for object/relational persistence and query service for Java, for the Model. The application is fully accessible with a Web Browser. No browser plug-in installations are needed, since its pages are composed of standard *HTML* and *ECMAScript*[5] code.

The integration of the server-side component in the *eWorkbook* system has been rather simple: the JAR (Java ARchive) file containing the *framework* classes has been imported as a library in the system. A modification to the system's deployment descriptor has been necessary in order to deploy the server-side module (servlet) which receives the events from the client.

The integration of the client-side component of the *framework*, composed of several *JavaScript* files, has been slightly more complicated, due to the structure of the *eWorkbook* interface: the test is launched in a child browser window of the main system Web page. This window displays a timer to inform the learner of the remaining time to complete the test and contains the controls to flow among the questions (*forward* and *backward* buttons) and the button to submit the test. The *stem* and the form containing the *options* are loaded in an *iframe* window present in the centre of the page.

3. Log Analysis

The system has been experimented by *using* it across a test *session* in a university course: *eWorkbook* has been used to administer on-line tests to learners. The learners were not informed of the experiment; they just knew that the grade obtained on the tests concurred to determine the final grade of the course exam.

The test, containing a set of 25 items to complete in a maximum time of 25 minutes, was administered to 80 learners, who took the test concurrently in the same lab. The logger was enabled and an approximately 4Mb sized *XML* log file has been obtained. The logging activity produced no visible system performance degrading. The next subsection shows the data extraction phase, while the subsequent sub-sections describe the experiments performed *using* log data.

3.1. Data Extraction

The objective in this phase is to obtain information at a higher level from the raw data of the log. For the sake of future extensions the analyzer does not directly produce the data matrix but acquires data from the log and populates a data *Model* that will be later translated in a tabular format.

The data *Model* is organized in learner test sessions. Each *session* is a sequence of page visits (each page contains a question). The following information are relevant for a page visit:

- Duration of the visit;
- Presence and duration of inactivity time intervals (no interactions) during the visit;
- The sequence of the responses given by the learner during the visit;
- Estimation of the times spent by the learner in evaluating the *stem* and each of the *options* for the question.

A precise estimation could only be possible through an eye tracking logging. Nevertheless, an eye tracking study [3] shows that there is a significant correlation between the eye movements and the mouse movements: tracking the trajectory drawn by the mouse pointer could be useful for obtaining the probable trajectory of users eyes and, in this case, the option the learner is evaluating. In the light of the above argument, we have equipped our logger *framework* with the capability of registering mouse movements, in particular, it can register the option on which the mouse pointer is moving.

In our *Model*, the learner activity is organized, for each test *session*, in a set of *ItemView*, *Response* and *Observation* objects. Each *ItemView* represents the visit of the learner at the web page containing a given question. Usually, during a test *session*, there are at least one *ItemView* per item, but in some cases the learner can skip or revise one or more questions.

Each *ItemView* embodies a set of *Observations* made by the learner during the evaluation of the item. With the term evaluation we mean the mouse activities of the learner on the web page containing the *stem* and the set of *options*. For each observation we store the nature of user interactions (i.e. the kind of mouse movements) and the duration of that observation. Lastly, in each *ItemView*, the learner can give one or more responses, and for each response we store its correctness and timestamp.

Figure 3 shows the *static* structure of the analyzer, which is written in Java and uses several *XML* libraries.

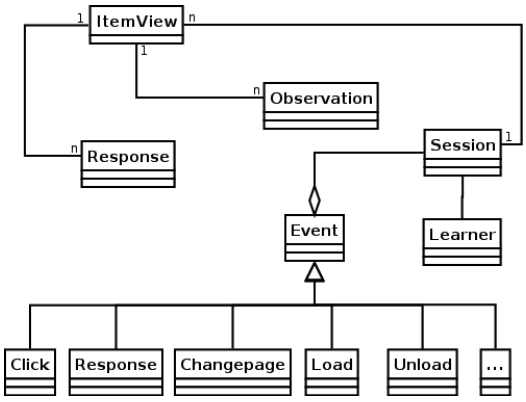


Figure 3. Structure of the Log Analyzer

3.2. Experiment I: Test and Item Completion Times

The time needed by the learners to complete the test has been subject of studies and experiments, mainly by measuring its correlation to the final result obtained on the test. Paul and Rosenkoetter [11] found no significant correlation between the above measures, while in a previous experiment Johnston [8] had demonstrated that the variability in the results of early and late finishers is bigger than that in the results of learners which complete the test in a time close to average.

Having more detailed data, we have repeated Paul and Rosenkoetters experiment, extending its domain to the time spent by the learners on the evaluation of the single items, correlating these times with the correctness of the responses given for them. The data matrix has been obtained considering the following measures:

- *ItemId*: A number used by the *e-testing* system to uniquely identify a question item;
- *AvgTime*: The average time needed by the learners to evaluate the item, in milliseconds.
- *Avg Score*: The average score obtained by the learners on the question, where a value of 1 has been assigned to a correct response and a value of 0 to an incorrect response.
- *Corr*: The correlation between the time needed by the learners to evaluate the question and the score obtained on it.

ItemId	Avg Time	Avg Score	Corr
5738	10665	0,78	0,14
5750	10945	0,73	-0,07
5762	14939	0,82	0,28
5774	26367	0,82	-0,22
....	...	...	...

Table 1. An extract of Data Matrix for Experiment I

As established for the whole test and its result, we can conclude that no correlation exists between the time spent on the single items and the scores obtained on them: a longer time spent evaluating the question does not imply a better result on it. We can evict this by observing the *Corr* column in the table, where a similar number of positive and negative values are present.

3.3. Experiment II: Response Changing

When executing *structured* tests based on *multiple choice* question types, is it better for learners to trust in their first impression or going back for evaluating the items again and, eventually, for changing the given answers? Furthermore, are the strongest or the weakest learners those who change idea more frequently?

Many experiments have focused on answer changing in *multiple choice* item based tests. Best [2] tried to answer the former of the above questions, following the example of other researchers, and his results confirmed the previous ones: *wrong-to-right* changes are more frequent than *right-to-wrong* ones. The answer to the latter question is still uncertain, since several experiment carried out in the past [1, 2] gave contrasting results.

UserName	R2W	W2R	NOC	Score
achpap	0	1	2	13
adomig	0	0	0	13
alepet	0	1	1	14
andast	0	1	1	20
.....	...	...	...	...
MEAN R2W			0,91	
MEAN W2R			1,36	
MEAN NOC			3,46	
CORR (NOC, SCORE)			0,11	

Table 2. an extract of data Matrix for Experiment II

We repeated the experiments *using* our system. The following measures have been reported in the data matrix:

- *UserName*: The name used by the *e-testing* system to uniquely identify a learner;
- *R2W*: Number or *right-to-wrong* changes made by the learner during a test *session*;
- *W2R*: Number of *wrong-to-right* changes made by the learner during a test *session*;
- *NOC*: The total number of changes made by learner during a test *session*. This number also includes wrong-to-wrong changes not considered in previous measures.
- *Score*: The number of questions correctly answered by the learner during the test *session*.

In table 2 we summarize some of the results obtained during our experiments. Even in our experiment, *wrong-to-right* changes are slightly more frequent than *right-to-wrong* ones, as shown by the average values for these measures. Furthermore, there is some correlation between the total number of changes performed by the learner during the test and the score he/she obtained on the test, meaning that changes are more frequently made by strong learners.

It is worth noting that through the analysis of a classical Web application log it would not have been possible to obtain such detailed data: due to the scarce interactivity of *HTML* documents, only a part of the given responses would have been submitted to the server and registered.

3.4. Experiment III: Learners' Behavior

A more comprehensive experiment [9] focuses on the strategies used by several learners with different abilities ("A", "C" and "F" students) to complete the tests. In particular, the frequency of several habits has been recorded and then correlated to the learners' final mark on the test. These are:

1. The habit of giving a rash answer to the question as soon as a plausible option is detected, often neglecting to consider further *options*;
2. The habit of anticipating the answer before *reading* the *options* provided;
3. The habit of eliminating incorrect *options* by stating the reasons they are inappropriate;
4. The habit of initially skipping the questions whose answer is more uncertain, in order to evaluate them subsequently, etc.;

Only the measure in point 4 is easily inferred from our log data. The measures in point 2 and 3 can only be obtained with the "think out loud" method, since no Web-based interaction can be appropriate for inferring them. As for the measure in point 1, we can obtain an approximate estimation of it by *using* our log data, as explained in section 4.1. Nevertheless, the calculation of this measure is left for future extension.

The measures considered in our experiment are the following:

- UserName: The name used by the *e-testing* system to uniquely identify a learner;
- Skipped: The total number of questions initially skipped by the learner during a test *session*.
- Score: The number of questions correctly answered by the learner during the test *session*.

We have divided the learners participating in the experiment in three *groups* according to their score. The first group contains 27 learners with a low score; the second group 27 learners with a medium score; the third group 26 learners with a high score. The latter group has a lower average number of skipped items. This suggests that strong learners are less likely to skip questions than weak ones. Nevertheless, the strategy of skipping at least one item gives good results, since the learners who did not skip any item have a lower average final score (15,67 / 25) than those learners who skip at least one item (16,58 / 25).

Score	Mean	SD
low	5,11	4,96
medium	5,26	6,38
high	3,92	4,65
0 Skipped Mean Score = 15,67		
more than 1 Skipped Mean Score = 16,58		

Table 3. Results of experiment III

4. Conclusion

In this paper we have presented a system for capturing and analyzing the behavior of the learners during *e-testing* sessions based on *structured* tests. The system is composed of a logging *framework* which can be instantiated in *e-testing* systems and of a stand-alone application which analyzes the obtained logs in order to extract information from them.

In order to demonstrate the effectiveness of the system, we have replicated on-line several experiments performed in the past on the habits of the learners during the tests. No significantly different results from those experiments have been obtained in our tries, but our main aim was not that of contradicting or confirming the theories maintained by psychologist: as computer scientists interested in e-learning, our research is mainly aimed at taking advantage from the new computing technologies in order to provide the psychologists with effective tools for enhancing their experiments. As shown in this paper, *using* our system, some experiments can be exactly repeated on-line, some can be enhanced with the availability of more information, others can only be partially and approximately repeated. In the case of our system, we are confident that, for some kinds of experiments, more realistic data can be gathered, since learners are not informed of the experiments and are not forced to modify their behavior.

As for future work, we foresee to use information visualization techniques in order to integrate the tabular data obtained in the analysis phase. Our aim is to obtain a tool for charting the whole learner test *session*. The availability of such a tool could help tutors in detecting some learners' habits (i.e. cheating) not directly manifest in the data but which could *catch our eyes* if visualized.

References

[1] J.A. Bath, Answer-changing Behavior on objective examinations., The Journal of Educational Research, 1967, 61, 105-107.

[2] J.B. Best, Item difficulty and answer changing, Teaching of Psychology, 1979, 6, 228-240.

[3] M. C. Chen, J. R. Anderson, M. H. Sohn Moore, What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing, In CHI '01 extended abstracts on Human factors in comp. syst., 2001.

[4] G. Costagliola, F. Ferrucci,V. Fuccella, F. Gioviale, A Web Based Tool for Assessment and Self-Assessment, Proceedings of ITRE '04, London, UK, 2004, pp. 131-135.

[5] ECMAScript, Standard ECMA-262, ECMAScript Language Specification, http://www.ecma-international.org/publications/files/ECMA-ST/Ecma-262.pdf, 2007.

[6] K.A. Edwards,C. Marshall, First impressions on tests: Some new findings, Teaching of Psychology, 1977, 4,193-195.

[7] Hibernate, http://www.hibernate.org, 2007.

[8] Johnston, J.J. Exam Taking speed and grades. Teaching of Psychology, 1977, 4, 148-149

[9] McClain, L.: Behavior during examinations: A comparison of "A", "C" and "F" students. Teaching of Psychology, Vol.10, No 2 April 1983

[10] Murray G., Asynchronous JavaScript Technology and XML (AJAX) With Java 2 Platform, Enterprise Edition http://java.sun.com/developer/technicalArticles/J2EE/AJAX/

[11] C.A. Paul, J.S. Rosenkoetter, The relationship between the time taken to complete an examiniaton and the test score received, Teaching of Psychology, 1980, 7, 108-109.

[12] S.E. Rindler, The Effect of Skipping over more difficult items on time limited tests: Implications for test validity, Educational and Psychological Measurement, 1980, 40, 989-998.

[13] K. Woodford, P. Bancroft, Multiple Choice Items Not Considered Harmful, in Proceedings of the 7th Australian conference on Computing education,(2005), Newcastle, Australia, pp. 109-116

Building a Multimedia e-Learning Model for Preventing Club Drug Abuse

Wei-Yi Fan, Graduate Institute of Medical Informatics, Taipei Medical University, Taiwan
m110094008@tmu.edu.tw
Wenn-Chieh Tsai, Graduate Institute of Medical Informatics, Taipei Medical University, Taiwan
m110092001@tmu.edu.tw
Yu-Chen Lee, Graduate Institute of Medical Informatics, Taipei Medical University, Taiwan
m110093007@tmu.edu.tw
Chih-Ying Huang, Graduate Institute of Medical Informatics, Taipei Medical University, Taiwan
cyhuang_4@tmu.edu.tw
Chiu-Ming Hu, Graduate Institute of Medical Informatics, Taipei Medical University, Taiwan
terry@tmu.edu.tw
Chien-Yeh Hsu, Graduate Institute of Medical Informatics, Taipei Medical University, Taiwan
cyhsu@tmu.edu.tw

ABSTRACT

According to the statistical analysis, the rate of seizing Club Drugs has been increased these years. Thus, it is very important to find ways for preventing drug abuse. In this study, we combined drug abuse knowledge, situated learning theory, and multimedia methods to implement an internet-based learning model for preventing abuse of club drugs. An interactive scenario simulation e-learning model was created, which is based on a very distinctive design called Virtual Drug Abuse Scenario Simulation (VDASS).The idea of virtual scenario simulation was adopted in this study to create the material and activities for the on-line e-learning system. The content was advised and contributed by many experts who have practical medicine experiences. Incorporated with multimedia knowledge database, the system management platform was established to provide the information about drug abuse and the education resources for medicine development, and become a portal web page for the education and information retrieval of drug abuse.

The evaluation results show that the outcomes (test scores) of the post-test is better than that of the pre-test. Besides learning effects evaluation, the result of the satisfied questionnaire of system shows that the overall satisfaction is 76.96%. Therefore, the results support our hypothesis that the effect of applying VDASS e-learning model is suitable and feasible. In the future, more evaluation can be done for different user groups such as the general populations. Furthermore, searching for the full-text and related external web pages will be constructed in this system.

Key Words: *Drug Abuse, Multimedia, E-Learning, Internet, Scenario Simulation*

INTRODUCTION

Knowledge is power. Education in advance is more important than compensation after drug abuse really happen. According to the statistical analysis, the rate of seizing club drugs has been increased these years in Taiwan. It is very important to find ways for preventing drug abuse. From the point of view of harm reduction, traditional propaganda and education methods become more and more inefficient and ineffective. Since the emergence of the internet, even though there are some web pages provide information which related to drug abuse, most of them are based on passive learning model. Therefore, providing interesting and accurate anti-drug abuse information via this multimedia e-learning model is the most important objective to prevent drug abuse and increase public health belief.

PURPOSE

The purpose of this study is to combine club drug abuse information, situated learning, and digital technology to develop internet-based and game-based e-learning models for preventing abuse of club drugs. People can learn anti-drug abuse knowledge and increase their health belief from this e-learning platform by themselves anytime and anywhere.

LITERATURE REVIEW

Nightclubs, bars, dances, and pubs attract many adolescents, university students, and young adults. Although many of the young people visiting these places only want to get together with friends, talk, and dance, some may seek to heighten their mood via substance abuse, including the use of tobacco, alcohol, and drugs [1]. According to the statistics for seized narcotics drugs

and controlled drugs from 1997 to 2005, there is an obvious grow up of club drugs in Taiwan. Club drugs abuse include tobacco, alcohol, Ecstasy (MDMA), GHB, GBL (transfer to GHB in the body), Ketamine, Flunitrazepam (Rohypnol, FM2), (meth-) amphetamine, and LSD (acid). Therefore, how to design a well drug abuse education program is the most important issue for preventing abuse of club drugs [2].

According to the "The Goals of Drugs Education in Schools and Support Systems for Drug Abuse Prevention" of Taiwan Bureau of Controlled Drugs, the goals of schools drug education are as follows: enhance students' knowledge of proper drug use and drug abuse, recognize the influence of drugs on physical and mental health, nurture attitudes towards proper drug use and the avoidance of drug abuse, nurture or assist students in tackling the problem of drug abuse, enhance students' ability to cope with the pressures of schoolwork, thus avoiding the choice of drugs as an escape method, nurture students' ability to maintain a calm attitude when dealing with problems, nurture problem solving and ways to resist peer pressure to take drugs, and understand organizations providing drug abuse prevention, guidance and information and their functions [3].

As a matter of course, school teachers should pay close attention to students' words and deeds language and behavior. Through the use of various kinds of literature and audio visual material, students should be supplied with the relevant information about drugs and the drug abuse harm to their body.

When a student is found to have an unusual expression and abnormal manner, the situation should be investigated further in coordination with a guidance councilor. The parents shall be kept in close contact the student come under extra supervision from the time the discovery is made. Where the student already exhibits signs of drug abuse behavior, the local health authority shall be contacted and assisted in rehabilitation treatment. Moral support shall be given to the student and a report made to police authorities to investigate the source of the drugs [4, 5]. Therefore, education in advance is more important than compensation after drug abuse really happen.

In order to provide an interesting and accurate anti-drug abuse information platform, besides traditional anti-drug abuse knowledge, combining situated learning is the most important way to enhance users' learning desire. Situated learning was originally proposed as a model of instruction to bridge the gap between learning and use of knowledge for no experience learners [6]. Key components of the model are apprenticeship, collaboration, reflection, coaching, multiple practice, and articulation of learned skills. It has been used most widely to help children and youth learn reading, writing, and mathematics [7]. It has been also applied to course design in undergraduate and graduate education, and to a limited extent in adult continuing education and distance education [8]. There is plenty literature examples of situated learning design in continuing education. One report [7] described a program in which community pharmacists were taught how to clinically assess patients' needs and provide patient guidance through a combination of simulated problem scenarios in a workshop format and authentic experiences in the workplace. More recently, Casey [9] described the application of a situated learning model to the design of a multimedia continuing education program for meteorologists. Furthermore, there are many related examples of situated learning design in specialized fields in Taiwan. Such as e-learning supporting system for medical education based on digital game [10]. They developed a game-based e-learning supporting system for medical education on a special topic of Japanese encephalitis virus (JEV). The learning contents including epidemiology, the virus pathogenic, symptoms and sequelae, circulation way, virus construction and the course of duplicating in host's cell were presented by means of integrating the characteristics of e-learning and on-line games. According to the results of literature examples, it is a good way to attract learners by integrates specialized knowledge and situated e-learning.

METHODS

In this study, we combined anti-drug abuse knowledge, situated learning theory and multimedia methods to implement internet-based and game-based e-learning models for preventing abuse of club drugs.

The idea of sharable content object reference and virtual scenario simulation were adopted in this study to create the material and activities for the on-line e-learning system [11, 12]. The anti-drug abuse knowledge and real social events were advised and contributed by many experts who have practical medicine experiences. Incorporated with multimedia knowledge database, the anti-drug abuse knowledge system was established to provide the information about drug abuse and the education resources for medicine development, and become a portal web page for the education and information retrieval of drug abuse.

The web-based anti-drug abuse knowledge system was constructed by Apache server, XOOPS package software, MySQL database, Flex and PHP language. (http://vdass.tmu.edu.tw)

Figure 1. Web-based anti-drug abuse knowledge system

In this study, we created a portal web page named "The education and prevention of drug abuse", which contains several functions including knowledge for experts/general population, the interactive e-learning scenario simulation -Virtual Drug Abuse Scenario Simulation, on-line e-learning resources, self-evaluations, user management, system management and learning / discussion community. There are 222 drug abuse related documents contained in Anti-drug abuse knowledge base which include MDMA, Ketamine, Flunitrazepam, g-hydroxybutyrate, Amphetamine and other drugs. Users can retrieve and download related drug abuse literatures, news and bookmarks from this knowledge base.

The interactive scenario simulation e-learning was created and which is based on a very distinctive design called Virtual Drug Abuse Scenario Simulation (VDASS). This e-learning model simulated many scenarios of drug abuse including the feeling, symptoms, behaviors, side effects, sequela, and the ways of preventing or solving those problems. Multimedia materials and learning activities were implemented on a knowledge management platform. The learners were expected to be more interested in and had higher motive on the learning model. The idea of this study was to build a learning system for users to reduce harm of drug abuse.

In this study, we have successfully built a game-based virtual drug abuse scenario simulation for MDMA, Amphetamine, Ketamine, Flunitrazepam (FM2), and g-hydroxybutyrate [13, 14, and 15]. We apply situated learning theory [16, 17] to develop and simulate club drug abuse scenario. Learners could increase their anti-drug abuse knowledge and health belief after they contact the virtual reality drug abuse scenario simulation, law punishment and sequela in VDASS.

In virtual drug abuse scenario simulation, learners will get in to about 20 drug abuse scenarios such as campus, work place, pub, KTV, etc. Each place in different environments has its own individual drug abuse events. Figure 2-3 are scenario examples which were extracted from VDASS.

Figure 2 . Scenario example: Coming across the drug seller

In this game-based learning model, learners will encounter different drug abuse simulation events and learn related anti-drug abuse knowledge by situated learning.Figure 2 shows learners meet with drug seller in pub. Learners can chat with drug seller and ask some club drugs' features.

Besides club drug abuse scenario simulation, there are 10 anti-drug abuse classes such as basic drug abuse introduction, Ketamine (Figure 3), Flunitrazepam, GHB, Amphetamine, how to abstain from drug abusing, and related anti-drug abuse organization information, etc. These medicine classes are provided by pharmacist experts from school of pharmacy, college of medicine, Taipei Medical University.

Figure 3 . Scenario example: Anti-drug abuse class

RESULTS

The VDASS system was evaluated in two aspects. One is the perceptions of the pharmaceutical knowledge and the other is the effectiveness of the learning model. Several times of expert meeting were holds to confirm the correctness of pharmaceutical knowledge. The evaluation of learning effects includes the comparison between pre- and post-test of applying VDASS e-learning method. In order to prevent the bias of learning effects, we design pre- and post- test using the same drug abuse questions to keep the equal degree of difficulty. The satisfaction score was obtained using questionnaire to test the suitability and feasibility of VDASS.

We recruited 45 participants (college students) from Taipei Medical University in Taipei, via posters and announcements in classes. In the beginning, participants are required to take the drug abuse knowledge pre-test to understand each participant's anti-drug abuse knowledge. After the pre-test, participants started to surf the portal web page of "The education and prevention of drug abuse". During self learning period, participants can retrieve related drug abuse knowledge and experience the interactive e-learning scenario simulation – VDASS. After 1 hour self learning, participants are demanded to take the post-test. There are total 10 drug abuse questions in the test.

Table 1. The statistical data of the number of students who answer correct in test

Question	Pre-test	Pro-test	Progress(%)
1	32	37	11.4
2	22	37	34.1
3	16	20	9.1
4	18	30	27.3
5	14	19	11.4
6	17	26	20.5
7	39	42	6.8
8	37	39	4.5
9	26	23	-6.8
10	22	32	22.7

Between pre- and post-test, the statistical and analysis difference are in the following tables.

Table 2. Paired sample statistic

Test	Mean	n	Variation
Pre-Test	5.52	44	3.33
Pro-Test	6.93	44	2.62

Table 3. Paired two tailed t-test

	Difference of means	t	Degree of freedom	P-value
Pre-test – Post-test	-1.41	-4.28	43	0.0001 (two tails)

The evaluation results show that the outcomes (test scores) of the post-test is better than the pre-test. According to the values of t = - 4.28 and p=0.00005<0.01 from paired t-test, we can determine that there is a significant improvement in test scores after the studying of VDASS.

Besides learning effects evaluation, according to the statistic of the questionnaire, the overall satisfaction is 76.96%. Therefore, the results support our hypothesis that the effect of applying VDASS e-learning model is suitability and feasibility.

In the future, more evaluation can be done for different user groups such as the general populations. The development procedures used for creating VDASS can be adopted for building other interactive e-learning platforms. Furthermore, searching for the full-text and related external web pages will be constructed in this system

REFERENCES

1. Bureau of Controlled Drugs, Department of Health, Retrieved Aug. 29, 2004 from http://www.nbcd.gov.tw/prop/prop_2-02.asp
2. Bureau of Controlled Drugs, Department of Health, , Retrieved Oct. 20, 2005 from http://www.nbcd.gov.tw
3. Epidemiology and Education Division ,Bureau of Controlled Drugs, Department of Health, , Retrieved Oct. 20, 2005 from http://www.nbcd.gov.tw/oldweb/english/prop/Goals.asp
4. Cline R.J.W., Haynes K.M: (2001). Consumer health information seeking on the Internet: the state of the art. Health Education Research, 16(6): 671-92.
5. Thiagarajan: S. Evaluation of simulations and games: a comprehensive procedure and a case history. Health Education Monographs. 5 suppl 1:64-73, 1977.
6. Brown, J.S., Collins, A. and Duguid, P., "Situated cognition and the culture of learning," Educ. Res., 18, 32-42(1989).
7. Farmer, J.A., Buckmaster, A. and LeGrand, B., "Cognitive apprenticeship: Implications for continuing professional education," New Directions for Adult and Continuing Education, 55, 245-253(1992).
8. Pharmacist perceptions of situated learning in an Internet-based continuing education program American Journal of Pharmaceutical Education, Winter 2000 by Wright, Dale E
9. Casey, C., "Incorporating cognitive apprenticeship in multi-media," Educ. Tech. Res, Dev., 44, 71-84(1996).
10. Chiu-Ming Hu, E-learning Supporting System for Medical Education Based on Digital Game, The Journal of Taiwan Association for Medical Informatics, 15(2): 55-69, 2006 Jun.
11. Bartfay E, Bartfay WJ: Promoting health through fun and games. [Clinical Trial. Journal Article. Randomized Controlled Trial] Nurse Educator., 19(5):7-8, 1994 Sep-Oct.
12. Czuchry M, Sia TL, Dansereau DF, Dees SM: Downward Spiral: a pedagogical game depicting the dangers of substance abuse. [Journal Article] Journal of Drug Education. 27(4):373-87, 1997.
13. National Institute on Drug Abuse , Retrieved Oct. 20, 2005 from http://www.amstat.org/sections/srms/Proceedings/papers/1991_053.pdf
14. National Institute on Drug Abuse , Retrieved Oct. 20, 2005 from www.nida.nih.gov
15. Pharmacist perceptions of situated learning in an Internet-based continuing education program American Journal of Pharmaceutical Education, Winter 2000 by Wright, Dale E
16. Situated Learning, J. Lave Retrieved Oct 15, 2006 from http://tip.psychology.org/lave.html
17. UNESCO - Learning Without Frontiers, Constructing Open Learning Communities to Inspire a Changing World, A Simulation Game, Retrieved Aug. 29, 2003 from

Proceedings
International Workshop on
Visual Languages Computing
(VLC 2007)

Co-Editors

Gennaro Costagliola, University of Salerno, Italy
Giuliana Vitiello, University of Salerno, Italy
Monica Sebillo, University of Salerno, Italy

Incorporating Negation into Visual Logics: A Case Study Using Euler Diagrams

Gem Stapleton
Visual Modelling Group
University of Brighton, Brighton, UK
g.e.stapleton@brighton.ac.uk

Judith Masthoff
University of Aberdeen, Aberdeen, UK
jmasthoff@csd.abdn.ac.uk

Abstract

Many diagrammatic logics based on Euler diagrams have been defined with the aim of making precise reasoning easier for people. In such logics, it is unusual to find the negation operator ($\neg$) included. This is likely to impact the usability of the logics when users try to make statements that are naturally phrased using the $\neg$ operator. Furthermore, if one wishes to use semantic tableaux methods for the purposes of establishing entailment then including $\neg$ is essential. Thus, there are good reasons for extending existing diagrammatic logics to include $\neg$ explicitly. In this paper, we take Euler diagrams and extend the notation to include the $\neg$ operator, as well as $\vee$ and $\wedge$. Various expressiveness results for the logic are established. We present a sound and complete set of reasoning rules for the logic, drawing parallels with existing completeness proof strategies and highlighting differences that arise due to including negation.

1 Introduction

Recently, a variety of diagrammatic logics have been developed with an aim of making reasoning tasks easier for people. Many such logics are based on either Euler or Venn diagrams; see, for example [2, 4, 5, 9, 11, 14]. Whilst these logics mostly include some logical connectives (usually $\vee$ and occasionally $\wedge$), none of the sound and complete logics developed have included both $\vee$ and $\wedge$ along with $\neg$. Indeed, none of the systems just referenced include $\neg$; those in [4, 15] do not incorporate logical connectives at all. In some cases extending such logics to include $\neg$ (or $\wedge$) does not lead to an increase in expressive power. Even when expressiveness is not increased, it is likely that excluding $\neg$ and $\wedge$ will impact usability because to make a negated statement without using $\neg$, for example, one has to find an equivalent non-negated statement. Furthermore, if we wish to utilize semantic tableaux methods for establishing semantic entailment (as in [7]) then incorporating explicit negation is essential.

We conjecture that this typical exclusion of explicit negation is to simplify the systems from a formal perspective; for example, fewer reasoning rules will need to be defined which makes the proof of soundness easier to derive. Moreover, the completeness strategies employed are also more straightforward when $\neg$ is not included. Whilst $\{\vee, \neg\}$ and $\{\wedge, \neg\}$ are both minimal sets of operators, our own experience suggests that all three operators are useful when making logical statements. Thus it is important for us to have a full understanding of the effects that including all three operators have on expressiveness and on the completeness proof strategies available to us.

As a case study, we extend Euler diagrams to include $\neg$, $\vee$ and $\wedge$ and highlight various effects that this extension brings with it. Euler diagrams are an ideal choice for such a case study, since all of the logics referenced above extend this language. The strategy we take to prove completeness is likely to adapt to other visual logics that are extended to include all of these connectives because the completeness proof strategies used when $\neg$ is not included in these systems are generally similar.

2 Euler Diagrams: Syntax and Semantics

Here, we will briefly overview the syntax and semantics of Euler diagrams; for a formalization of the syntax see [12] and for the semantics see [13], adapting to the Euler diagram case.

There are two *unitary* Euler diagrams in figure 1. The diagram d_1 contains three *contours*; these are the closed curves labelled A, B and C. Contours represent sets and their spatial relationship is used to make statements about containment and disjointness of sets. So, in d_1 the contours assert that the sets A and B are disjoint because the contours do not overlap in any way; similarly $A \cap C = \emptyset$. The placement of C inside B expresses $C \subseteq B$. Shading is used to assert the emptiness of sets. Thus, d_1 also expresses $B - C = \emptyset$. The contours give rise to *regions* in the plane. A region is a set of *zones*, which are maximal sets of points

in the plane that can be described as being inside certain contours and outside the rest of the contours. The diagram d_1 contains four zones of which one is shaded and can be described as inside B but outside both A and C.

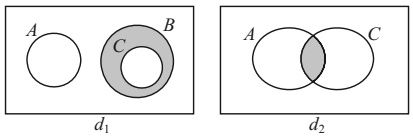


Figure 1. Two Euler diagrams.

Formally, we define a zone to be an ordered pair, $(inside, outside)$, which forms a two-way partition of the contour label set; the shaded zone in d_1 is $(\{B\}, \{A, C\})$. We can now talk about zones that can be described by such a pair but which do not correspond to regions present in a diagram. For example $(\{A, B\}, \{C\})$ is a zone which is not present in d_1 (no zone is inside both A and B but outside C). Such zones are said to be *missing* from the diagram; d_1, therefore, has four missing zones whereas d_2 has no missing zones. An Euler diagram with no missing zones is called a *Venn diagram*.

Informally, a *zonal region* can be thought of as a region that becomes a zone when contours are removed. In figure 1, the two zones inside A in d_2 become a zone when C is removed; in total d_2 has nine zonal regions. We define zonal regions in a similar way to zones: a zonal region in a diagram d is a set of zones that are described by a containing label set P and an excluding label set Q; more precisely, given any such disjoint sets P and Q, the set of zones $\{(inside, outside) : P \subseteq inside \wedge Q \subseteq outside\}$ is a zonal region. Whilst the two sets P and Q are disjoint, they need not form a partition of d's label set (unlike the sets used to describe a zone).

Unitary diagrams form the building blocks of *compound* diagrams. If d_1 and d_2 are Euler diagrams then so are $\neg d_1$, $(d_1 \vee d_2)$ and $(d_1 \wedge d_2)$. For simplicity, we adopt the usual convention of omitting brackets where no ambiguity arises. At the drawn diagram level, we denote $\neg d_1$ by $\overline{d_1}$, $(d_1 \vee d_2)$ by $\boxed{d_1 - d_2}$ and $(d_1 \wedge d_2)$ by $\boxed{d_1\, d_2}$ (i.e. by juxtaposition), where the rectangles act as brackets.

At the semantic level, an *interpretation* is a universal set, U, together with an assignment of a subset of U to each contour (strictly, to contour labels) which is extended to interpret zones and regions. A zone, (a, b), represents the set $\bigcap_{l \in a} set(l) \cap \bigcap_{l \in b} (U - set(l))$ where $set(l)$ is the set assigned to label l. A region in a unitary Euler diagram, d, represents the set which is the union of the sets represented by the region's constituent zones. Briefly, we say that an interpretation is a *model* for d if all of the zones which are shaded in d or missing from d represent the empty set. The semantics extend to compound diagrams in the obvious way. A diagram, d_2, is a *logical consequence* of diagram d_1, denoted $d_1 \vDash d_2$, if all of the models for d_1 are also models for d_2.

3 Expressiveness

In this section, we will provide a series of results concerning the expressiveness of the Euler diagram system. Our first observation is that including $\neg$ in the Euler diagram case impacts expressiveness.

Theorem 3.1. *Removing $\neg$ from the Euler diagram logic reduces expressiveness.*

Proof (Sketch) Every satisfiable non-negated diagram is satisfied by the empty model (i.e. $U = \emptyset$). There are negated diagrams that have only non-empty models. □

Our second observation is that including $\wedge$ does not impact expressiveness, regardless of whether $\neg$ is included.

Theorem 3.2. *Removing $\neg$ and $\wedge$ from the Euler diagram logic is equivalent in terms of expressiveness to removing only $\neg$ from the logic.*

To prove the above theorem, we can assume, without loss of generality, that our diagrams are in disjunctive normal form. We illustrate the proof strategy by the example in figure 2. The diagram $d_1 \wedge d_2$ can be replaced by the unitary diagram d_5. First, we make the contour label sets of d_1 and d_2 identical, giving d_3 and d_4 respectively. It is then easy to see how $d_3 \wedge d_4$ gives rise to d_5: take the union of the shaded zones in d_3 and d_4 to give the shaded zones in d_5; in general, the shaded zones in d_5 will be the union of the shaded and missing zones. This process can be used to turn the conjunction of any pair of unitary diagrams into a single unitary diagram.

We cannot, however, necessarily replace the disjunction of two unitary diagrams by a single unitary diagram. For example, taking d_3 and d_4, figure 2, in disjunction (as opposed to the conjunction displayed in the figure) provides a compound diagram that has no unitary counterpart, justified as follows. The diagram $d_3 \vee d_4$ has only models where $A - B = \emptyset$ or $C - A = \emptyset$. Thus, a model for $d_3 \vee d_4$ can allow $A - B$ to be non-empty, provided $C - A$ is empty. Therefore, any unitary diagram, d, semantically equivalent to $d_3 \vee d_4$ must have a model where $A - B$ is non-empty. Similarly, d must also allow $C - A$ to be non-empty. It can be shown that any such d also has a model where both $A - B$ and $C - A$ are non-empty, which is necessarily not a model for $d_3 \vee d_4$, showing that no unitary diagram is semantically equivalent to $d_3 \vee d_4$.

Next, we observe that a negated unitary diagram is never equivalent to a unitary diagram.

Theorem 3.3. *Let d_1 be a unitary diagram. Then $\neg d_1$ is not semantically equivalent to any unitary diagram.*

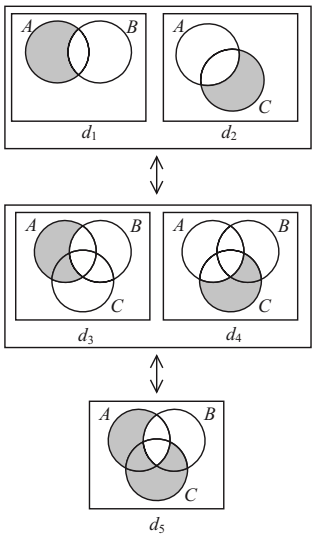


Figure 2. Removing conjunction.

Proof (Sketch) Every unitary diagram is satisfied by the empty model (i.e $U = \emptyset$). The negation of d_1, therefore, is not satisfied by the empty model. Thus $\neg d_1$ is not semantically equivalent to any unitary diagram. □

This result can be extended to show that $\neg d_1$ is not equivalent to any compound diagram that does not include negation. In fact, given any diagram, D, that involves negation, there only exists a semantically equivalent diagram that does not involve negation when the standard propositional logic equivalences (such as involution and De Morgan's Laws) can be used to remove the negation from D.

We can establish the expressiveness of the Euler diagram logic by comparing with other logics. First, we observe that the diagram d_2 in figure 1 is equivalent to the Monadic First Order Logic (MFOL) sentence $\forall x \neg(A(x) \wedge C(x))$. In MFOL all of the predicate symbols are unary (they correspond to contour labels). It is easy to translate Euler diagrams into MFOL sentences. It is also possible, but not so easy, to convert MFOL sentences into Euler diagrams. As a straightforward example, the MFOL sentence $\exists x\, A(x) \wedge \forall x\, B(x)$ is equivalent to the diagram in figure 3.

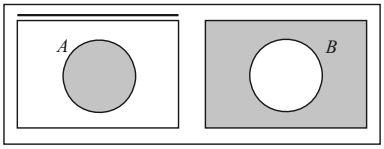


Figure 3. Comparing with MFOL.

Theorem 3.4. *Euler diagrams are equivalent in expressive power to MFOL.*

It follows that Euler diagrams are less expressive than spider diagrams, which are equivalent to MFOL with equality [13].

In seminal work, Shin introduced a variant of Venn-Peirce diagrams [8] called Venn-II [9] which extends Venn diagrams by including $\otimes$-sequences to express the non-emptiness of a set. For example, figure 4 shows a Venn-II diagram that expresses either $A - B \neq \emptyset$ and $A \cap B = \emptyset$ or $C \neq \emptyset$.

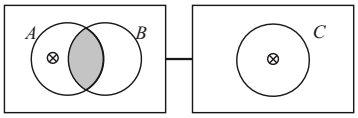


Figure 4. A Venn-II diagram.

In Venn-II, the only explicit logical connective is disjunction; there are no explicit negation or conjunction operators. Shin proved that Venn-II is also equivalent in expressive power to MFOL. As a consequence, if we extend Venn-II to include the negation and conjunction operators then we do not increase expressiveness.

Whilst Euler diagrams and Venn-II are equivalent in expressive power, there are advantages of developing a sound and complete logic for Euler diagrams over Venn-II extended to include $\neg$ and $\wedge$: many diagrammatic logics extend Euler diagrams but not Venn-II. For example, constraint diagrams [6] extend Euler diagrams and include all the connectives considered here. Furthermore, unlike Venn-II, $\neg$ and $\wedge$ are not both syntactic sugar in the constraint diagram case (see [3] for a formalization).

4 Reasoning Rules

In [12], it was found that rules which make 'medium level' changes to diagram syntax are generally good choices if we want to automatically find proofs efficiently. Moreover, it is sensible, from an automated theorem proving perspective, to define rules that change few different types of pieces of syntax because this helps us to define good search strategies; for example, a rule that applies to a compound diagram, changing its unitary parts and its tree structure changes many pieces of syntax. The rules presented here have been defined following both of these philosophies.

4.1 Information Preserving Rules

Our first rule allows us to change the contours in a diagram. For example, in figure 5, we can add a contour to d_1 giving d_2. The two diagrams are semantically equivalent, with neither providing any information about the set

D. We can also remove D from d_2 to yield d_1. We will write $d_1 \rightarrow d_2$ to mean that d_1 can be replaced by d_2 by the application of a reasoning rule.

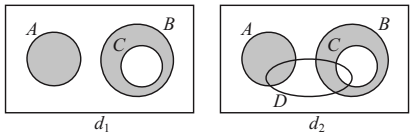


Figure 5. Changing contours.

Rule 1. Let d_1 and d_2 be unitary diagrams such that all of the contours in d_1 are in d_2 and d_2 has exactly one more contour, l. If we can add l to d_1 in such a way that l splits all of the zones in d_1 into two zones and the shading does not change to give d_2 then we say $d_1 \rightarrow d_2$ by applying the **add a contour** rule.

Rule 2. Let d_1 and d_2 be unitary diagrams such that $d_1 \rightarrow d_2$ by applying the add a contour rule. Then $d_2 \rightarrow d_1$ by applying the **remove a contour** rule.

We know that shaded regions represent the empty set, as do missing regions. The diagram d_1 in figure 5 asserts that $A \cap B = \emptyset$ because the two zones inside both A and B are missing. We can replace d_1 by d_3 in figure 6, using the *add a shaded zonal region rule*. We can also replace d_1 by d_4 using a *remove a shaded zonal region rule*.

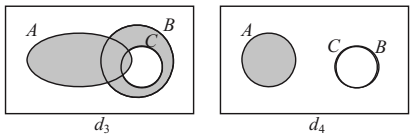


Figure 6. Changing zonal regions.

Rule 3. Let d_1 and d_2 be unitary diagrams such that the only difference between them is that d_2 contains an entirely shaded zonal region that is completely missing from d_1. Then $d_1 \rightarrow d_2$ by applying the **add a shaded zonal region** rule.

Rule 4. Let d_1 and d_2 be unitary diagrams such that $d_1 \rightarrow d_2$ by applying the add a shaded zonal region rule. Then $d_2 \rightarrow d_1$ by applying the **remove a shaded zonal region** rule.

In addition to the unitary rules introduced above, there are rules that operate on compound diagrams. In figure 7, d_1 asserts that $C - A = \emptyset$. Therefore, we can add shading to d_2 in the region corresponding to $C - A$ to give d_3; $d_1 \wedge d_2$ is semantically equivalent to $d_1 \wedge d_3$.

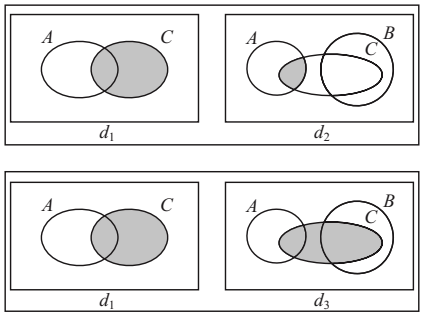


Figure 7. Changing shading.

Rule 5. Let d_1 and d_2 be unitary diagrams such that there exists a zonal region, zr, with the properties that

1. none of the zones in zr are missing from either d_1 or d_2 and
2. zr is completely shaded in d_1 and
3. zr is completely non-shaded in d_2

Let d_3 be identical to d_2 except that all zones in zr are shaded in d_3. Then $d_1 \wedge d_2 \rightarrow d_1 \wedge d_3$ by applying the **add shading to a zonal region** rule.

Rule 6. Let d_3 and d_4 be diagrams such that $d_3 \rightarrow d_4$ by applying the add shading to a zonal region rule. Then $d_3 \rightarrow d_4$ by applying the **remove shading from a zonal region** rule.

The remaining rules in this section operate on the structure of compound diagrams and they have analogies in propositional logic. Firstly, we observe that our explicit representation of true is a unitary diagram, d, that contains no contours and no shading; in other words $d = \Box$. Therefore, the diagram $\neg\Box$ is a contradiction. Throughout the remainder of this section, let d_1, d_2 and d_3 be diagrams. We write $d_1 \leftrightarrow d_2$ to denote $d_1 \rightarrow d_2$ and $d_2 \rightarrow d_1$.

Rule 7. Identity Law $d_1 \wedge \Box \leftrightarrow d_1$.

Rule 8. Complement Laws $d_1 \vee \neg d_1 \leftrightarrow \Box$, and $d_1 \wedge \neg d_1 \leftrightarrow \neg\Box$

Rule 9. De Morgan's Laws $\neg(d_1 \vee d_2) \leftrightarrow \neg d_1 \wedge \neg d_2$, and $\neg(d_1 \wedge d_2) \leftrightarrow \neg d_1 \vee \neg d_2$.

Rule 10. Involution $\neg\neg d_1 \leftrightarrow d_1$.

Rule 11. Distributivity $d_1 \wedge (d_2 \vee d_3) \leftrightarrow (d_1 \wedge d_2) \vee (d_1 \wedge d_3)$ and $d_1 \vee (d_2 \wedge d_3) \leftrightarrow (d_1 \vee d_2) \wedge (d_1 \vee d_3)$.

Rule 12. Idempotency $d_1 \leftrightarrow d_1 \wedge d_1$ and $d_1 \leftrightarrow d_1 \vee d_1$.

For simplicity, we assume commutativity and associativity. Furthermore, we assume that we can apply the above rules to any part of a compound diagram; for example, if d_1 can be replaced by d_2 by applying some rule then $(d_1 \wedge d_3) \vee \neg d_1 \vee d_4$ can be replaced by $(d_1 \wedge d_3) \vee \neg d_2 \vee d_4$.

4.2 Information weakening rules

Next, we introduce some rules that weaken the informational content of a diagram. These information weakening rules can only be applied inside an even number of negation signs. For example, if d_1 can be replaced by d_2 by applying one of the weakening rules then only the second occurrence of d_1 can be replaced by d_2 in the compound diagram $\neg d_1 \vee \neg(d_3 \wedge \neg d_1) \vee \neg(d_1 \vee d_4)$.

Rule 13. Inconsistency $\neg\Box \rightarrow d_1$.

Rule 14. Connecting a diagram $d_1 \rightarrow d_1 \vee d_2$.

Rule 15. Removing a diagram $d_1 \wedge d_2 \rightarrow d_1$.

Information weakening rules cannot be applied inside an odd number of negation signs because such inferences are not necessarily sound. As an illustration, in figure 8, the diagram $d_1 \vee d_2 \vee d_3$ can be obtained from $d_1 \vee d_2$ by applying the connecting a diagram rule. An interpretation in which $A = D = \emptyset$, $B = \{1\}$ and $C = \{2\}$ is a model for $\neg(d_1 \vee d_2)$ but not a model for $\neg(d_1 \vee d_2 \vee d_3)$.

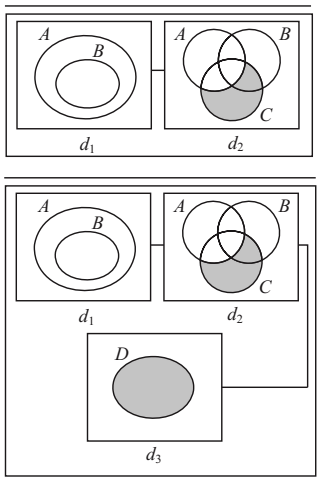


Figure 8. Incorrectly applying rules.

To conclude this section, we define obtainability.

Definition 4.1. *We say d_2 is **obtainable** from d_1, denoted $d_1 \vdash d_2$, if there exists a finite set of diagrams, $\{d_3, ..., d_n\}$, such that $d_1 \rightarrow d_3 \rightarrow ... \rightarrow d_n \rightarrow d_2$. If $d_1 \vdash d_2$ and $d_2 \vdash d_2$ then d_1 and d_2 are **syntactically equivalent**, denoted $d_1 \equiv_{\vdash} d_2$.*

5 Soundness and Completeness

To prove that the system is sound, an induction argument is used. The base case is trivial. For the inductive step it is sufficient to prove each rule is sound (that is, applying the rule maintains or enlarges the model set).

Theorem 5.1. ***Soundness** If $d_1 \vdash d_2$ then $d_1 \vDash d_2$.*

The system is also complete. We give a reasonably thorough account of the completeness proof, omitting the details of the stages which have analogies in the *spider diagram* logic [5]. For spider diagrams, given $d_1 \vDash d_2$, the high level strategy is to convert both d_1 and d_2 into a disjunction of unitary diagrams each of which has some given set of labels, L, and set of zones, Z, without changing their semantics, and then reason about these disjunctions. For this Euler diagram system it is not possible to transform some diagrams into such disjunctions because of the presence of negation. Thus we need to modify the completeness proof strategy and it is likely that our new strategy will extend to other systems that include negation.

Our approach is also to begin by introducing contours to each unitary part of d_1 and d_2 until all unitary parts have the same label set, L say, giving d_1^L and d_2^L respectively. Next, we convert each unitary part of both d_1^L and d_2^L into Venn diagrams (so there are no missing zones) using the add a shaded zonal region rule, giving d_1^Z and d_2^Z respectively. So, each unitary part of both d_1^Z and d_2^Z has zone set $Z = \{(a, L - a) : a \subseteq L\}$. We now deviate from the spider diagram strategy.

Our aim is to convert d_1^Z into disjunctive normal form and d_2^Z into conjunctive normal form where, in both cases, each unitary part contains at most one shaded zone. From now on, unless stated otherwise we will assume, without loss of generality, that all unitary diagrams have zone set Z and, therefore, have no missing zones.

Definition 5.1. *Let d be a unitary diagram with no missing zones and at most one shaded zone. Then d and $\neg d$ are called **literals**.*

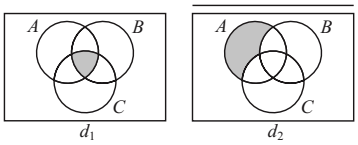


Figure 9. Two literals.

A literal, d, with a shaded zone z, gives us a single piece of information: in the *positive* case d tells us that z represents the empty set and in the *negative* case d tells us that z does not represent the empty set. For example, the positive

literal d_1 in figure 9 asserts that $A \cap B \cap C = \emptyset$ whereas the negative literal d_2 expresses that $A - (B \cup C) \neq \emptyset$. A literal with no shaded zones is either universally valid (the positive case) or a contradiction (the negative case).

Definition 5.2. *A diagram of the form*

$$\bigwedge_{1\leq x\leq p} \bigvee_{1\leq y\leq q} d_{x,y},$$

*where each $d_{x,y}$ is a literal, is in **literal conjunctive normal form** (LCNF). A diagram of the form*

$$\bigvee_{1\leq i\leq m} \bigwedge_{1\leq j\leq n} c_{i,j},$$

*where each $c_{i,j}$ is a literal, is in **literal disjunctive normal form** (LDNF).*

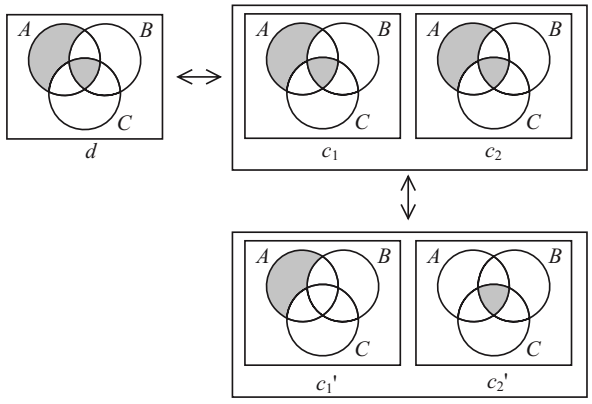


Figure 10. Obtaining LCNF.

Lemma 5.1. *Every unitary diagram, d, is syntactically equivalent to a diagram in LCNF.*

Proof If d has no shaded zones or exactly one shaded zone then d is in LCNF. Otherwise, enumerate the shaded zones in d, giving $\{z_1, ..., z_n\}$, and apply the idempotency rule to d in order to obtain $\bigwedge_{1\leq i\leq n} c_i$ where each c_i equals d (see figure 10). For each c_i apply the remove shading from a zonal region rule, removing shading from all of the shaded zones except z_i, giving c_i'. Clearly $d \equiv_{\vdash} \bigwedge_{1\leq i\leq n} c_i'$. □

Theorem 5.2. *Every diagram, d, is syntactically equivalent to a diagram in LCNF.*

Proof (Sketch) Replace each unitary part of d by a conjunction of literals (this can be done by lemma 5.1). Then use the information preserving rules that have analogies in propositional logic to convert to LCNF. □

Theorem 5.3. *Every diagram, d, is syntactically equivalent to a diagram in LDNF.*

The proof is similar to the LCNF case.

For the next step in our completeness argument, we replace d_1^Z by a syntactically equivalent diagram in LDNF, say

$$d_1^{LD} = \bigvee_{1\leq i\leq m} \bigwedge_{1\leq j\leq n} c_{i,j}.$$

The diagram d_2^Z is replaced by a syntactically equivalent diagram in LCNF, say

$$d_2^{LC} = \bigwedge_{1\leq x\leq p} \bigvee_{1\leq y\leq q} d_{x,y}.$$

We observe that $d_1 \equiv_{\vdash} d_1^{LD}$ and $d_2 \equiv_{\vdash} d_2^{LC}$. Therefore, if $d_1^{LD} \vdash d_2^{LC}$ then $d_1 \vdash d_2$. Hence, if we can show $d_1^{LD} \vdash d_2^{LC}$ then we have proved that the system is complete. In order to show $d_1^{LD} \vdash d_2^{LC}$, we further observe that it is sufficient to prove, for each i $(1 \leq i \leq m)$

$$\bigwedge_{1\leq j\leq n} c_{i,j} \vdash d_2^{LC}.$$

Since, when showing completeness, we start by assuming $d_1 \vDash d_2$, we know

$$\bigwedge_{1\leq j\leq n} c_{i,j} \vDash d_2^{LC},$$

so we can deduce that, for each x $(1 \leq x \leq p)$,

$$\bigwedge_{1\leq j\leq n} c_{i,j} \vDash \bigvee_{1\leq y\leq q} d_{x,y}.$$

We prove

$$\bigwedge_{1\leq j\leq n} c_{i,j} \vdash \bigvee_{1\leq y\leq q} d_{x,y}$$

from which it follows that the system is complete.

We will first establish some syntactic relationships between $\bigwedge_{1\leq j\leq n} c_{i,j}$ and $\bigvee_{1\leq y\leq q} d_{x,y}$ when $\bigwedge_{1\leq j\leq n} c_{i,j}$ is satisfiable. Recall that $\square$ is a unitary diagram containing no contours or shading. The notation $\square^L$ will be used to denote a positive literal with label set L, zone set Z and no shading. We extend this notation, and denote a positive literal with label set L, zone set Z and shaded zone z by $\square_z^L$.

Theorem 5.4. *Let $C = \bigwedge_{1\leq j\leq n} c_j$ be a satisfiable conjunction of literals and let $D = \bigvee_{1\leq y\leq q} d_y$ be a disjunction of literals. If $C \vDash D$ then either*

(1) there exists j and y such that $c_j = d_y$ or

(2) there exists x and y such that $d_x = \neg d_y$ or

(3) the diagram $\square^L$ occurs in D.

Proof We proceed by contradiction and assume that none of the above three conditions hold. Reduce D to a disjunction of literals in which $\neg\square^L$ never occurs, using remove a contour (removing all of the contours in L from $\neg\square^L$ to give $\neg\square$) and inconsistency (replacing $\neg\square$ by any literal in D that contains a shaded zone), to give D'; this can be done because (3) is false and, because C is satisfiable, D is satisfiable. For clarity, the diagram D' is a disjunction of literals, each with a shaded zone. Moreover, $D \vDash D'$ so $C \vDash D'$. Thus, showing that $C \nvDash D'$ will give us a contradiction.

We will construct a model, m, for C that is not a model for D'. We take the universal set to be

$$U = \{z \in Z : \square_z^L \text{ is in } D' \text{ or } \neg\square_z^L \text{ is in } C\}$$

and each zone represents the empty set if z is not in U, otherwise z represents itself.

First, we show that m satisfies C. Let c be a literal in C and suppose c is positive. For m to satisfy c it must be that any zone which is shaded in c (of which there is at most one) represents the empty set. If c has no shaded zones then m trivially satisfies c. Otherwise, $c = \square_z^L$, for some $z \in Z$, and z does not represent the empty set if and only if $\square_z^L$ is in D' or $\neg\square_z^L$ is in C (by the definition of m). In the former case c occurs in both C and D' which cannot happen (condition (1) above is false). The latter case also cannot happen since C is satisfiable. Hence m satisfies c.

Alternatively, c is negative. If c has no shaded zones then $c = \neg\square^L$, which is unsatisfiable, contradicting the assumption that C is satisfiable. Therefore c has exactly one shaded zone, z say, and $c = \neg\square_z^L$. Then the only way m can fail to be a model for c is if z represents the empty set. But, by the definition of m, $z \in U$ so z represents itself in m and not the empty set. Hence m satisfies c. In any case, m is a model for each c in C, so m is a model for C.

Now, we show that m is not a model for D'. To do so, we show that m is not a model for any d' in D'. Let d' be a literal in D' and suppose that d' is positive. Then $d' = \square_z^L$ for some z and we want to show that $z \in U$ (i.e. z does not represent the empty set); this follows trivially from the definition of U, so m is not a model for d'. Alternatively, d' is negative and of the form $\neg\square_z^L$ for some z; $\neg\square_z^L$ asserts the non-emptiness of z so, for m not to satisfy $\neg\square_z^L$ we must show that $z \notin U$. Since (1) is false, $\neg\square_z^L$ is not in C. Furthermore, we know that, for each unitary diagram, d, if d appears in D' then $\neg d$ does not appear in D' because (2) is false. Therefore, $z \notin U$, so z represents the empty set and m does not satisfy d'. Hence m is not a model for any literal in D', so m is not a model for D'.

Thus, we have reached a contradiction, since $C \vDash D \vDash D'$ implies $C \vDash D'$ but we have shown that $C \nvDash D'$. Therefore at least one of the three conditions stated in the theorem holds. □

Using theorem 5.4, we are able to show that $C \vdash D$.

Theorem 5.5. *Let $C = \bigwedge_{1 \le j \le n} c_j$ be a satisfiable conjunction of literals and let $D = \bigvee_{1 \le y \le q} d_y$ be a disjunction of literals. If $C \vDash D$ then $C \vdash D$.*

Proof We break the proof down into three cases, corresponding to those in the statement of theorem 5.4.

(1) *There exists j and y such that $c_j = d_y$.* Choose such a c_j and d_y. Remove diagrams from C until only $c_j = d_y$ remains, using the removing a diagram rule. Next, connect a diagram to $c_j = d_y$ yielding D, using the connecting a diagram rule. Hence $C \vdash D$.

(2) *There exists x and y such that $d_x = \neg d_y$.* In this case, take C and use the identity law to obtain $C \wedge \square$. Using the removing a diagram rule, obtain $\square$, to which we apply the complement law to give $d_x \vee d_y$. From this, use the connecting a diagram rule (if necessary) to obtain D.

(3) *The diagram $\square^L$ occurs in D.* Apply the identity law and removing a diagram rule to obtain $\square$ from C. Next, add contours to $\square$ to give $\square^L$ (which is in D). Finally, use the connecting a diagram rule (if necessary) to obtain D.

In each case, we have shown that $C \vdash D$. □

To conclude our completeness proof, we now show that, if C is unsatisfiable, then $C \vdash D$. To do so, we prove that $C \vdash \neg\square$ after identifying some syntactic conditions on C.

Lemma 5.2. *Let $C = \bigwedge_{1 \le j \le n} c_j$ be an unsatisfiable conjunction of literals. Then either C contains $\neg\square^L$ or C contains both $\neg\square_z^L$, and $\square_z^L$ for some zone z.*

Proof Suppose that C does not contain $\neg\square^L$. Define an interpretation, I, where

$$U = Z - \{z' : \square_{z'}^L \text{ is in } C\}$$

and every zone in U maps to itself in I and all of the other zones map to the empty set. Obviously, I satisfies all of the positive literals in C so, since C is unsatisfiable, there exists a negative literal in C, say c, with shaded zone z, which is not satisfied by I; $c = \neg\square_z^L$. Since I does not satisfy $\neg\square_z^L$, z represents the empty set and, therefore, must be shaded in some positive literal, namely $\square_z^L$, in C as required. □

Lemma 5.3. *Let $C = \bigwedge_{1 \le j \le n} c_j$ be an unsatisfiable conjunction of literals. Then $C \vdash \neg\square$.*

Proof We consider two cases, corresponding to those in the statement of lemma 5.2.

1. *C contains $\neg\Box^L$.* Remove all of the contours from $\neg\Box^L$, to give $\neg\Box$ using the remove contour rule. Replace $\neg\Box$ by a conjunction of literals so that, in the result, each unitary part appears both positively and negatively using the inconsistency rule. Replace each pair of literals, $d \wedge \neg d$, by $\neg\Box$ using the complement laws (this may also require idempotency). Finally, use idempotency to obtain $\neg\Box$.

2. *C contains $\neg\Box_z^L$ and $\Box_z^L$ for some zone z.* Replace $\Box_z^L \wedge \neg\Box_z^L$ by $\neg\Box$, using the complement law and then proceed as the first case.

Therefore $C \vdash \neg\Box$. □

Theorem 5.6. *Let $C = \bigwedge_{1\leq j\leq n} c_j$ be an unsatisfiable conjunction of literals and let $D = \bigvee_{1\leq y\leq q} d_y$ be a disjunction of literals. Then $C \vdash D$.*

Proof By lemma 5.3, replace C with $\neg\Box$ then use inconsistency to obtain D. □
Hence the system is complete.

Theorem 5.7. ***Completeness*** *If $d_1 \vDash d_2$ then $d_1 \vdash d_2$.*

6 Conclusion

In this paper, we have presented a sound and complete Euler diagram logic that includes operators $\neg$, $\wedge$ and $\vee$. Typically, logics that extend Euler diagrams include $\vee$ and, sometimes, $\wedge$ but not $\neg$. Including $\neg$ impacts the strategy used to prove completeness. Typically, the Euler diagram based logics that do not include $\neg$ rely on being able to reduce d_1 and d_2, where $d_1 \vDash d_2$, to a disjunction of unitary diagrams as part of the completeness proof. Reducing to disjunctions of unitary diagrams, say $\bigvee c_i$ and $\bigvee d_j$ allows one then to prove completeness by simply showing that each c_i implies one of the d_j's, for example. See [5, 9, 10, 11] for logics whose completeness proofs use this type of strategy.

When negation is included, obtaining such disjunctions is not necessarily possible, as in this Euler diagram case. It is likely that our strategy, to reduce d_1 to literal disjunctive normal form and d_2 to literal conjunctive normal form, will extend to other Euler diagram based logics where constructive completeness proofs exist (i.e. in decidable languages). At the very least, we anticipate that the strategy will extend to the systems presented in [1, 4, 5, 9, 10, 11, 15].

Acknowledgement Gem Stapleton is supported by a Leverhulme Trust Early Career Fellowship and by UK EPSRC grant EP/E011160/1 for the Visualization with Euler Diagrams project.

References

[1] L. Choudhury and M. K. Chakraborty. On extending Venn diagrams by augmenting names of individuals. In *Proceedings of 3rd International Conference on the Theory and Application of Diagrams*, volume 2980 of *LNAI*, pages 142–146. Springer-Verlag, March 2004.

[2] A. Fish and J. Flower. Investigating reasoning with constraint diagrams. In *Visual Language and Formal Methods 2004*, volume 127 of *ENTCS*, pages 53–69, Rome, Italy, 2005. Elsevier.

[3] A. Fish, J. Flower, and J. Howse. The semantics of augmented constraint diagrams. *Journal of Visual Languages and Computing*, 16:541–573, 2005.

[4] E. Hammer. *Logic and Visual Information*. CSLI Publications, 1995.

[5] J. Howse, G. Stapleton, and J. Taylor. Spider diagrams. *LMS Journal of Computation and Mathematics*, 8:145–194, 2005.

[6] S. Kent. Constraint diagrams: Visualizing invariants in object oriented modelling. In *Proceedings of OOPSLA97*, pages 327–341. ACM Press, October 1997.

[7] O. Patrascoiu, S. Thompson, and P. Rodgers. Tableaux for diagrammatic reasoning. In *Proceedings of the Eleventh International Conference on Distributed Multimedia Systems, International Workshop on Visual Languages and Computing*, pages 279–286. Knowledge Systems Institute, September 2005.

[8] C. Peirce. *Collected Papers*, volume 4. Harvard University Press, 1933.

[9] S.-J. Shin. *The Logical Status of Diagrams*. Cambridge University Press, 1994.

[10] G. Stapleton, J. Howse, and J. Taylor. A constraint diagram reasoning system. In *Proceedings of International Conference on Visual Languages and Computing*, pages 263–270. Knowledge Systems Insitute, 2003.

[11] G. Stapleton, J. Howse, and J. Taylor. A decidable constraint diagram reasoning system. *Journal of Logic and Computation*, 15(6):975–1008, December 2005.

[12] G. Stapleton, J. Masthoff, J. Flower, A. Fish, and J. Southern. Automated theorem proving in Euler diagrams systems. *Accepted for Journal of Automated Reasoning*, to appear 2007.

[13] G. Stapleton, S. Thompson, J. Howse, and J. Taylor. The expressiveness of spider diagrams. *Journal of Logic and Computation*, 14(6):857–880, December 2004.

[14] N. Swoboda and G. Allwein. Using DAG transformations to verify Euler/Venn homogeneous and Euler/Venn FOL heterogeneous rules of inference. *Journal on Software and System Modeling*, 3(2):136–149, 2004.

[15] N. Swoboda and G. Allwein. Heterogeneous reasoning with Euler/Venn diagrams containing named constants and FOL. In *Proceedings of Euler Diagrams 2004*, volume 134 of *ENTCS*. Elsevier Science, 2005.

On the Descriptional Complexity of a Diagrammatic Notation

Aidan Delaney and Gem Stapleton
Visual Modelling Group,
University of Brighton, UK
{a.j.delaney,g.e.stapleton}@brighton.ac.uk

Abstract

Spider diagrams are a widely studied, visual logic that are able to make statements about relationships between sets and their cardinalities. Various meta-level results for spider diagrams have been established, including their soundness, completeness and expressiveness. In order to further enhance our understanding of spider diagrams, we can compare them with other languages; in the case of this paper we consider star-free regular languages. We establish relationships between various fragments of the spider diagram language and certain well-known subclasses of the star-free regular class. Utilising these relationships, given any spider diagram, we provide an upper-bound on the state complexity of minimal deterministic finite automata corresponding to that spider diagram. We further demonstrate cases where this bound is tight.

1 Introduction

It is widely recognised that diagrams play an important role in various areas, including visualizing information and reasoning about that information. They are often useful for conveying (sometimes complex) information in accessible and intuitive ways. This is one reason behind the widening perception of the importance of diagrams. Traditionally in mathematics and logic, diagrams have been excluded from formal proof techniques and were considered only as a heuristic aid. Whilst some people have held the view that diagrams *cannot be formalised*, so as to be permitted when reasoning formally, it has been shown that this view is incorrect: Shin devised a sound and complete diagrammatic logic [19]. Her work is widely regarded as a seminal piece, overturning the view that diagrams could not yield a formal reasoning system. Thus, diagrams are now being recognised as a valuable tool that can be exploited in a logical setting (see [9] for an extensive discussion on the importance of diagrams in numerous reasoning contexts).

Since the work of Shin, many other diagrammatic logics have emerged. One such logic is the widely studied language of spider diagrams (see, for example [3, 7, 8, 12, 20]). With regard to applications of spider diagrams, they have been used to assist with the task of identifying component failures in safety critical hardware designs [1]. They have also been (implicitly) used for displaying the results of database queries [21], representing non-hierarchical computer file systems [2], in a visual semantic web editing environment [10, 24] and for viewing clusters which contain concepts from multiple ontologies [5]. In all of these application areas, there are other languages that could be used instead. It is, therefore, useful if we can compare spider diagrams to other languages. If we can show that, for example, spider diagrams offer significant descriptional savings over another descriptional system then practitioners may choose to represent their specifications in this more succinct form.

This paper shows that spider diagrams can be used to define languages from a particular subset of the star-free regular languages. Sections 2, 3 and 4 consider preliminaries of spider diagrams, star-free regular languages and descriptional complexity respectively. Section 5 identifies various relationships between spider diagrams and star-free regular languages and section 6 investigates descriptional complexity by providing upper bounds on the size of a deterministic finite automata accepting the language generated by a spider diagram.

2 Spider Diagrams

This section will provide a brief overview of the spider diagram syntax presented in [8]. In figure 1 the spider diagram d_1 contains two labelled *contours*, A and B. Contours are simple closed curves. The diagram also contains three minimal regions, called *zones*. There is one zone inside A, another inside B and the

other zone is outside both A and B. Each zone can be described by a two-way partition of the contour label set. The zone inside A can be described as inside A but outside B and contains two *spiders*; spiders are trees whose vertices, called *feet*, are placed in zones (in d_1, the spiders each consist of a single vertex). Spider diagrams can also contain *shading*, as in d_2 (which contains three spiders and four zones).

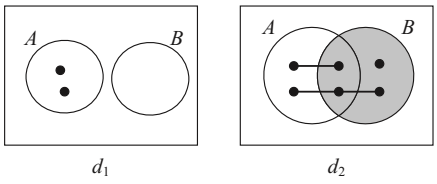


Figure 1. Two spider diagrams.

The syntax is defined at an abstract level. The contour labels in spider diagrams are selected from a finite set $\mathcal{L}$. A **zone** is defined to be a pair, (in, out), of finite disjoint subsets of $\mathcal{L}$. The set in contains the labels of the contours that the zone is inside whereas out contains the labels of the contours that the zone is outside. The set of all zones is denoted $\mathcal{Z}$. To describe the spiders in a diagram, it is sufficient to say how many spiders are placed in each region. Thus, the abstract definition of a spider diagram will specify the labels used, the zones, the shaded zones and use a set of spider identifiers to describe the spiders.

Definition 2.1. *A **unitary spider diagram**, d, is a quadruple $\langle L, Z, ShZ, SI\rangle$ where*

$L = L(d) \subseteq \mathcal{L}$ *is a set of contour labels,*

$Z = Z(d) \subseteq \{(a, L-a) : a \subseteq L\}$ *is a set of zones,*

$ShZ \subseteq Z(d)$ *is a set of shaded zones,*

$SI = SI(d) \subset \mathbb{Z}^+ \times (\mathbb{P}Z - \{\emptyset\})$ *is a finite set of spider identifiers such that for all* $(n_1, r_1), (n_2, r_2) \in SI(d)$ $(r_1 = r_2 \implies n_1 = n_2)$.

*The symbol $\perp$ is also a unitary spider diagram. We define $L(\perp) = Z(\perp) = ShZ(\perp) = SI(\perp) = \emptyset$. If d_1 and d_2 are spider diagrams then $(d_1 \vee d_2)$, $(d_1 \wedge d_2)$ and $\neg d_1$ are **compound spider diagrams**. Given a unitary diagram, d, a zone (a, b) is said to be **missing** if it is in the set $\{(a, L-a) : a \subseteq L\} - Z(d)$. If d has no missing zones then d is in **Venn form**. The set of **spiders** in d is defined to be*

$$S(d) = \{(i, r) : (n, r) \in SI(d) \wedge 1 \leq i \leq n\}.$$

*For spider (i, r), each zone in r is a **foot** of (i, r).*

So S(d) can be thought of as a bag of spiders generated by the identifiers.

By convention, we employ a lower-case d with or without subscripts to denote a unitary spider diagram. An upper case D with or without subscripts will denote an arbitrary spider diagram. The usual convention of omitting brackets where no ambiguity arises is adopted.

Our attention now turns to the semantics. Spider diagrams make statements about sets (represented by contours) and their cardinalities (by using spiders and shading). In figure 1, d_1 expresses that A and B are disjoint, because there are no points interior to both of the contours. Spiders assert the existence of elements, so d_1 specifies that there are (at least) two elements in A. The spiders in d_2 assert that there are at least three elements, one of which is in A, another is in $A \cup B$ and the third is in $B - A$. Shading is used to place upper bounds on set cardinality. For example, d_2 expresses that the set $B - A$ contains at most two elements, $A \cap B$ contains at most two elements and B contains at most three elements.

The semantics of spider diagrams are model-based. An *interpretation* is a universal set, U, together with an assignment of a subset of U to each contour (strictly, to contour labels) which is extended to interpret zones and regions. A zone, (a, b), represents the set $\bigcap_{l \in a} set(l) \cap \bigcap_{l \in b} (U - set(l))$ where $set(l)$ is the set assigned to constant label l. A set of zones, Z, represents the set which is the union of the sets represented by Z's constituent zones. An interpretation is a **model** for unitary diagram $d(\neq \perp)$ whenever

1. all of the zones which are missing represent the empty set,
2. all of the regions represent sets whose cardinality is at least the number of spiders placed entirely within that region and
3. all of the entirely shaded regions represent sets whose cardinality is at most the number of spiders with a foot in that region.

If $d = \perp$ then the interpretation is not a model for d (i.e. $\perp$ is a contradiction). The definition of a model extends to compound diagrams in the obvious (inductive) manner. The semantics are formalised in [20].

The operators $\neg$ and $\wedge$ are syntactic sugar in the spider diagram language, captured by the following theorem.

Theorem 2.1. *The language of spider diagrams is equivalent in expressive power to the fragment in which the only operator is $\vee$.*

3 Star-Free Regular Languges

The class of star-free regular languages is the set of languages which are a subset of the free monoid over an alphabet Σ including the finite languages which are closed under finite boolean operations and catenation (catenation is the product operation of the monoid) [13]. The Straubing-Thérin hierarchy describes some well studied subclasses of star-free regular languages. The more well known subclasses being the class of shuffle-ideal languages and piecewise testable languages.

A deep result by Schützenberger characterises star-free languages as those which have an aperiodic syntactic monoid [18]. Alternative characterisations of subsets of star-free languages are given later based on the shuffle product of languages. The shuffle product of two languages L_1, L_2 denoted $L_1 \sqcup L_2$ informally takes all words from L_1 and intersperses letters from all words in L_2. More formally, the words in $L_1 \sqcup L_2$ are precisely those of the form $w_0 w_1 w_2 \dots w_n$ where, for some subset $I = \{p_1, p_2, \dots, p_m\}$ of $\{1, \dots, n\}$

1. $w_{p_1} w_{p_2} \dots w_{p_m} \in L_1$ where $p_i < p_{i+1}$
2. $w_{q_1} w_{q_2} \dots w_{q_{n-m}} \in L_2$ where $\{q_1, \dots, q_{n-m}\} = \{1, \dots, n\} - I$ and $q_i < q_{i+1}$.

Considering the shuffle product and boolean operations of $\cup, \cap$ and $\subset$ gives us an immediate insight into the Straubing-Thérin catenation hierarchy. Level 0 is the set of languages $\{\Sigma^*, \emptyset\}$. Level 1/2 is the well known shuffle ideal set, which is the polynomial closure of Level 0. An alternative characterisation is that languages of catenation level 1/2 are of the form $k \sqcup \Sigma^*$ where k is a finite set of words. Level 1 is defined as the boolean closure of 1/2. This hierarchy has been extended by Pin to consider varieties of languages [14].

Spider diagrams are a monadic first order logic with equality (MOFLe). Therefore we are interested in the relationship between logics and star-free regular languages. The main body of literature discussing this relationship [6, 14, 15, 17, 22] assumes the existence of an order relation $<$ adjunct to the standard monadic first order operators of $\neg, \vee, \wedge, \iff$, the quantifiers $\exists$ and $\forall$ and predicates of the form $P_a(x)$ which states that the letter a is at positive position x in word w. Intuitively, if we do not have an order relation, $<$, as in MOFLe then any language corresponding to a formula will be closed under permutation. In other words, languages of the form, for example, $\Sigma^* A \Sigma^* B$ (A comes before B in every word) do not correspond to languages arising from formulae in MOFLe. Consequently spider diagrams do not contain facilities for ordering elements.

4 Descriptional Complexity

Descriptional complexity is concerned with the economy of representation offered by descriptional systems [11]. Descriptional systems consist of a set of finite descriptors which describe each instance of a language in a class of a languages [4]. For example, finite automata are a descriptional system which describe languages in the class of regular languages (REG). Other well-known descriptional systems include pushdown automata, linear bounded automata, Turing machines and spider diagrams.

Economy of representation is computed through the use of some metric. In the case of automata as descriptional systems, common metrics include the length of the automaton description or the number of states in the automaton. Other typical metrics include the number of productions in a grammar system or a measure of non-determinism in a system. Rabin and Scott's [16] well known result on the state complexity of deterministic and non-deterministic finite automata neatly illustrates the utility of such metrics: there exists a non-deterministic finite automaton with n states where the minimal deterministic automaton accepting the same language contains 2^n states. By convention, Rabin and Scott's result can be stated as (taken from [4])

$$\mathrm{nFA}_n \xrightarrow{states} \mathrm{dFA}_{\leq 2^n}$$

To draw conclusions on the descriptional complexity of spider diagrams we require a metric, which establishes their complexity, and a descriptional system for comparison. We have chosen the number of spiders in a diagram as the metric. This measure is independent of the actual layout of the diagram and is directly linked with diagram semantics. We establish relationships between spider diagrams and a descriptional system for regular languages. More precisely, we establish a relationship between the number of spiders in a spider diagram and the number of states in the minimal deterministic finite automaton which accepts the same language, thus comparing their descriptional complexity.

5 Relationships Between Spider Diagrams and Star-Free REG

The finite models for a spider diagram can be considered as giving rise to a star-free REG language. For example, in figure 2, the diagram d_1 has a model with universal set $U = \{1, 2\}$, with A representing the set $\{1\}$ and B representing $\{2\}$. So, in this model, there is exactly one element in $(\{A\}, \{B\})$ and exactly one

element in $(\{B\},\{A\})$ but the model does not specify any order on these elements. We can think of the zones as being the letters in a language; with this alphabet, the word $(\{A\},\{B\})(\{B\},\{A\})$ indicates that there is an element in $(\{A\},\{B\})$ which comes before an element in $(\{B\},\{A\})$; the word $(\{B\},\{A\})(\{A\},\{B\})$ indicates the existence of the same two elements, but in the opposite order. Consequently, the information provided by the two words matches the information provided by the model. The set of words in a language corresponding to d_1 contains precisely the words that arise from the finite models for d_1 in this way. We observe that isomorphic models give rise to the same words. In what follows, we formalise the *language of a spider diagram* and establish various relationships between spider diagrams and some subsets of star-free REG.

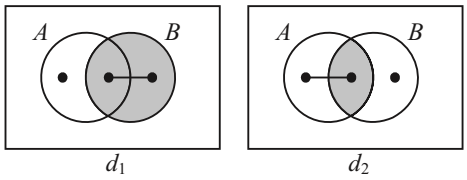


Figure 2. Words for a diagram.

To begin, we note that it can be shown that every spider diagram can be transformed into a syntactically equivalent diagram (see [8] for a set of sound and complete reasoning rules) involving only disjunctions of unitary diagrams, $D = d_1 \vee ... \vee d_n$ where each d_i has no missing zones and the same zone set. We assume, without loss of generality, that each d_i contains all of the labels in $\mathcal{L}$. Such a disjunction, D, is said to be in **disjunctive Venn form**; we call this fragment of the spider diagram language **disjunctive-SD** denoted $\mathcal{DSD}$. There is an algorithm that translates arbitrary diagrams into disjunctive Venn form (see [8] for the conversion of diagrams into *associated zone diagrams* and adapt the details in the obvious way).

To identify a star-free REG language for an arbitrary diagram, D, we can first convert D into disjunctive Venn form and then identify such a language for this disjunction. Thus, we need only identify a star-free REG language for each diagram in disjunctive Venn form. From this point forward all diagrams will be assumed to be part of the disjunctive fragment $\mathcal{DSD}$. The approach we adopt is to first identify a language for each unitary diagram and then extend to the whole of $\mathcal{DSD}$.

For example, d_2 in figure 2 can be translated into the star-free REG language that has alphabet $Z(d_2)$ and words that contain the letters (breaking into cases determined by the two-footed spider):

1. $(\{A\},\{B\})$ (for the two-footed spider) and $(\{B\},\{A\})$ (for the one-footed spider) but not $(\{A,B\},\emptyset)$ (because of the shading), or

2. $(\{A,B\},\emptyset)$ (for the two-footed spider) and $(\{B\},\{A\})$ (for the one-footed spider) but not more than one occurrence of $(\{A,B\},\emptyset)$ (because of the shading).

In general, the alphabet over which our star-free REG language is constructed is

$$\Sigma = \mathcal{Z}_{\mathcal{L}} = \{(a,b) \in \mathcal{Z} : a \cup b = \mathcal{L}\}$$

and the language generated by Σ is $\Sigma^* = \{\text{finite words over } \mathcal{Z}_{\mathcal{L}}\}$, including the empty word λ. A letter of the alphabet is simply a zone, $(\{in_1,...,in_n\},\{out_1,...,out_m\})$; informally, we may instead write $in_1...in_n\overline{out_1}...\overline{out_m}$. As a notational convenience we may also write letters of a word within square brackets, thus "$[AB][\overline{AB}]$" is a two letter word containing AB and $\overline{AB}$. We define, for unitary diagram d, $\Gamma(d) = Z(d) - ShZ(d)$ (recall that $Z(d) = \mathcal{Z}_{\mathcal{L}}$), so $\Gamma(d) \subseteq \Sigma$. Given an arbitrary diagram, D, some words in Σ^* correspond to the meaning of the diagram and the rest do not[1].

Definition 5.1. *Let w be a word in Σ^* and d $(\neq\perp)$ be a unitary diagram. The bag (or multi-set) of letters of which w consists is denoted $bag(w)$. The word w* ***conforms****, to d if and only if there exists an injection, $f\colon S(d) \to bag(w)$ satisfying*

1. *$f(s)$ is a foot of s,*

2. *f is bijective when the image is restricted to the maximal sub-bag of w whose elements are shaded zones in d.*

For $d = \perp$, no words in Σ^* conform to d.

So, w conforms to unitary diagram $d \neq \perp$ provided, for each spider, s, in d,

1. each spider in d gives rise to a letter in w by way of selecting a foot,

2. for each shaded zone, z, the number of occurrences of z in w is precisely the number of spiders whose selected foot is z.

[1] There are two special cases: when D is universally valid, all of the words in Σ^* correspond to the meaning of D and when D is a contradiction none of the words in Σ^* correspond to D.

To illustrate, in figure 2, the following are examples of words that conform to d_2:

$$[A\overline{B}][B\overline{A}],\ [B\overline{A}][A\overline{B}],\ [AB][B\overline{A}],$$
$$[B\overline{A}][A\overline{B}][AB],\ [B\overline{A}][A\overline{B}][\overline{AB}],\ [B\overline{A}][AB][A\overline{B}].$$

The word $[AB][AB][B\overline{A}]$ does not conform to d_2 because the letter $[AB]$ occurs twice (d_2 asserts that the zone $(\{A, B\}, \emptyset)$ contains at most one element).

Definition 5.2. *Let d be a unitary diagram. The* ***language of d*** *is the set of words in Σ^* that conform to d denoted $\mathscr{L}(d)$. Let $D = d_1 \vee ... \vee d_n$ be a spider diagram. The* ***language of D*** *is $\bigcup_{1 \leq i \leq n} \mathscr{L}(d_i)$.*

Definition 5.3. *Let d be a unitary diagram. We define $k(d)$ to be the set of* ***words generated by S(d)****:*

$k(d) = \{w \in \Sigma^* : w$ conforms to d where f is bijective$\}$

(taking f as in definition 5.1). For $d = \perp$, $k(d) = \emptyset$.

Lemma 5.1. *For unitary diagram d, $\mathscr{L}(d) = k(d) \sqcup \Gamma^*$ where $\Gamma = \Gamma(d)$.*

Proof. $\mathscr{L}(d)$ is of the form $\Gamma^* a_1 \Gamma^* a_2 \Gamma^* \dots \Gamma^* a_{|S(d)|} \Gamma^*$ where $a_1 a_2 \dots a_{|S(d)|} \in k(d)$. □

Corollary 5.1. *Let d be a unitary diagram such that $ShZ(d) = \emptyset$. Then $k(d) \sqcup \Gamma(d)^*$ is a shuffle-ideal.*

Proof. In this case, $\Gamma(d)^* = \Sigma^*$. □

Let $\mathcal{DSD}_{\mathcal{NS}}$ be the class of spider diagrams in $\mathcal{DSD}$ that do not contain shading.

Theorem 5.1. *Let $D = d_1 \vee ... \vee d_n \in \mathcal{DSD}_{\mathcal{NS}}$. Then $\mathscr{L}(D)$ is a shuffle-ideal.*

Proof. $\mathscr{L}(D) = \bigcup_{1 \leq i \leq n} \mathscr{L}(d_i)$ is a shuffle-ideal because the class of shuffle-ideals is closed under union. □

Theorem 5.2. *The class $\bigcup_{D \in \mathcal{DSD}_{NS}} \{\mathscr{L}(D)\}$ is a strict subclass of the class of shuffle-ideal languages.*

Proof. By theorem 5.2, $\bigcup_{D \in \mathcal{DSD}_{NS}} \{\mathscr{L}(D)\}$ is a class of shuffle-ideals. For strictness, we start by observing that, for each unitary diagram $d \in \mathcal{DSD}_{\mathcal{NS}}$, the language of d is closed under permutation of words. Therefore, the language of any diagram in $\mathcal{DSD}_{NS}$ is closed under permutation of words. However, the shuffle-ideal $l = \{z_1 z_2\} \sqcup \Sigma^*$, where $z_1, z_2 \in Z(d)$ are two distinct zones, is not closed under permutation of words since, for example, the word $z_1 z_2$ is in l but $z_2 z_1$ is not in l. Consequently, l is not the language of any diagram in $\mathcal{DSD}_{NS}$. □

There are compound diagrams whose languages are not equal to the language of any unitary diagram, captured by the following lemma.

Lemma 5.2.

$$\bigcup_{\{d \in \mathcal{DSD}: d \text{ is unitary}\}} \{\mathscr{L}(d)\} \subset \bigcup_{D \in \mathcal{DSD}} \{\mathscr{L}(D)\}.$$

Proof. (Sketch) Taking $D = d_1 \vee d_2$, where d_1 and d_2 are shown in figure 3, provides an example of a compound diagram whose language, $\mathscr{L}(D)$, is not the language for any unitary diagram. □

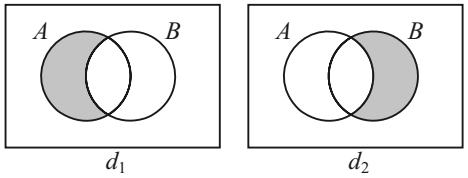


Figure 3. Classes of languages.

Theorem 5.3. *The class $\bigcup_{D \in \mathcal{DSD}} \{\mathscr{L}(D)\}$ is a strict subset of languages of catenation order 3/2.*

Proof. (Sketch) From [14] we know that sets of level 3/2 are finite unions of sets of the form $\Gamma_0^* a_1 \Gamma_1^* a_2 \Gamma_2^* \dots a_j \Gamma_j^*$ where $a_1, a_2, \dots, a_j \in \Sigma$ and $\Gamma_0, \Gamma_1, \dots, \Gamma_j \subseteq \Sigma$. From lemma 5.1 and lemma 5.2 we can see that $w \in \mathscr{L}(D)$ is of a similar, but restricted, form where $\Gamma_0 = \Gamma_1 = \dots = \Gamma_j \subseteq \Sigma$. □

Corollary 5.2. *The set of shuffle-ideal languages is not a subset of $\bigcup_{D \in \mathcal{DSD}} \{\mathscr{L}(D)\}$.*

Proof. From theorem 5.2 we know that there exists l such that $l \notin \bigcup_{D \in \mathcal{DSD}_{NS}} \{\mathscr{L}(D)\}$ but is a shuffle-ideal. The argument extends to show that $l \notin \bigcup_{D \in \mathcal{DSD}} \{\mathscr{L}(D)\}$. □

Corollary 5.3. *The class $\bigcup_{\{d \in \mathcal{DSD}: d \text{ is unitary}\}} \{\mathscr{L}(d)\}$ is not a subset of the class of shuffle-ideal languages.*

Proof. Let d be a unitary diagram in $\mathcal{DSD}$ such that $ShZ(d) \neq \emptyset$. Then $\Gamma(d) \subset \Sigma$. A simple application of the `TestShuffle` [6] to the automaton accepting $\mathscr{L}(d) = k(d) \sqcup \Gamma(d)^*$ verifies that $\bigcup_{D \in \mathcal{DSD}} \{\mathscr{L}(D)\}$ contains languages outside the class of shuffle ideal languages. □

The spider diagram in figure 4 illustrates the relationships between various classes of languages, established in this section, where DSD denotes the class $\bigcup_{D \in \mathcal{DSD}} \{\mathscr{L}(D)\}$ and DSD_{NS} denotes $\bigcup_{D \in \mathcal{DSD}_{NS}} \{\mathscr{L}(D)\}$.

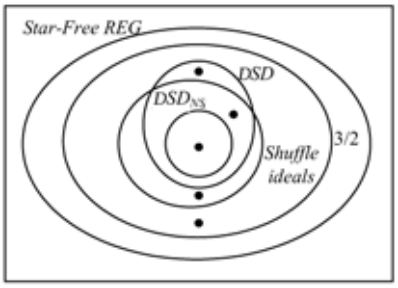


Figure 4. Relationships between languages.

6 The Descriptional Complexity of Spider Diagrams

For any given spider diagram, a regular language arises from the finite models, as demonstrated in the previous section. We now provide upper bounds on the number of states required in the minimal deterministic finite automaton accepting the language that arises from these models of a spider diagram, and show that our upper bound is exact in some cases. Given a unitary diagram d we will establish an upper bound on $|k(d)|$ in terms of the number of spiders in d. When $|k(d)|$ meets this upper bound, we establish the exact minimal number of states required in a finite state machine.

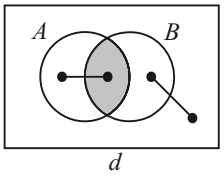


Figure 5. Large $k(d)$.

As an illustration, d in figure 5 has

$$k(d) = \{[A\overline{B}][B\overline{A}], [B\overline{A}][A\overline{B}], [A\overline{B}][\overline{AB}], [\overline{AB}][A\overline{B}], [AB][B\overline{A}], [B\overline{A}][AB], [AB][\overline{AB}], [\overline{AB}][AB]\}.$$

Given that there are two spiders, s_1 and s_2 say, in d, each with two feet, the largest number of words that can be in $k(d)$ is 8: suppose a word, w, has first letter which arises from s_1 and second letter from s_2, then there are $2 \times 2 = 4$ possible choices for w; since $k(d)$ is closed under permutation of words, every permutation of w must also be in $k(d)$, giving $4 \times 2 = 8$ words in total. Since the two spiders in d are placed in disjoint regions, every such permutation gives rise to a distinct word. Hence there are 8 words in $k(d)$.

Lemma 6.1. *Given a unitary diagram d*

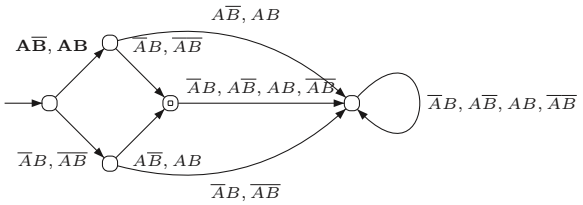
$$|k(d)| \leq |S(d)|! \times \prod_{(i,r) \in S(d)} |r|.$$

Given a unitary diagram d with set of words, $k(d)$, satisfying $|k(d)| = |S(d)|! \times \prod_{(i,r) \in S(d)} |r|$, we can construct a minimal finite automaton accepting precisely the words in $k(d)$. If d is entirely shaded, then this automaton accepts the language of d (since, in such a case, $\mathscr{L}(d) = k(d)$). Alternatively, the automaton can be trivially modified (adding loops, removing edges and possibly removing the sink state) to accept $\mathscr{L}(d)$. The number of states remains the same or reduces by 1 (the sink state is necessary only when at least one zone is shaded or $d = \perp$). For example, considering d in figure 5, a minimal finite automaton accepting precisely the words in $k(d)$ is $\mathcal{A}(k(d))$ in figure 6(a). To convert $\mathcal{A}(k(d))$ to a finite state automaton accepting $\mathscr{L}(d)$, we add loops and remove edges, enlarging the set of words accepted, in the obvious way (see figure 6(b)). Given any unitary diagram d we define $\mathcal{A}(k(d))$ to be the minimal automaton accepting precicely the words in $\mathscr{L}(k(d))$.

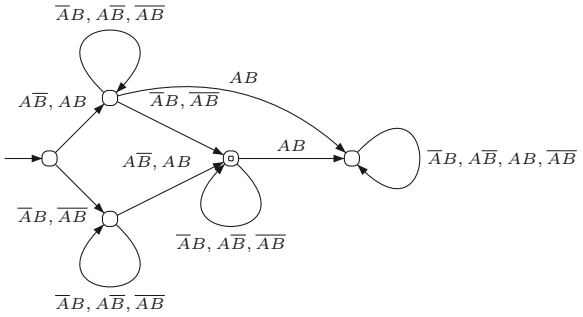
The bold typeface transition label in figure 6(a) represents two transitions: a transition from the start state to a state on the letter $A\overline{B}$ and another transition between the same two states on the letter AB. Each tranistion occurs on a letter that arises from a unique spider foot and is labeled with the appropriate letter. We allow the commonly accepted transition label shorthand in figure 6 by defining the function $l(s_i), s_i = (i, r) \in S(d)$ which returns a label consisting of the zones in r.

Theorem 6.1. *Let d be a unitary diagram with $|k(d)| = |S(d)|! \times \prod_{(i,r) \in S(d)} |r|$. The minimal complete automaton accepting precisely the words in $k(d)$ has $2^{|S(d)|} + 1$ states.*

Proof. (Sketch)We prove this by induction over the size of $S(d)$. In the base case, seen in figure 7(a), we accept all one-letter words corresponding to the single spider s in $S(d), |S(d)| = 1$. The cases where $|S(d)| = 2$ and $|S(d)| = 3$ are presented in figure 7. We assume that



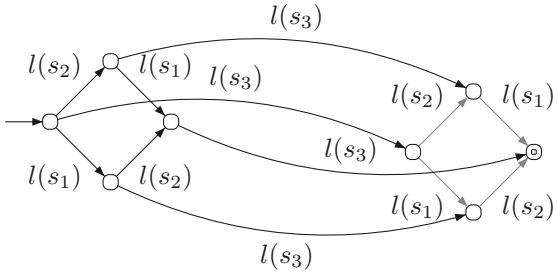
(a) Automaton accepting $k(d)$



(b) Automaton accepting $\mathscr{L}(d)$

Figure 6. Constructing a finite state machine.

(a) $k(d)$ where $S(d) = \{s_1\}$ (b) $k(d)$ where $S(d) = \{s_1, s_2\}$



(c) $k(d)$ where $S(d) = \{s_1, s_2, s_3\}$

Figure 7. Constructing an automaton accepting *k(d)*.

where $|S(d)| = n$ the size of the automaton is 2^n. We may construct the automaton for the $n+1$ case by duplicating the automaton for the n case and "shifting" the duplicated structure by one letter. The inductive step is visible in the examples 7(a) through 7(c). To complete the automaton we add a sink state q_{sink} as the target of all missing transitions. The proof of minimality is omitted. □

Theorem 6.2. *Let d be a unitary diagram such that $|k(d)| = |S(d)|! \times \prod_{(i,r)\in S(d)} |r|$. The minimal automaton accepting $\mathscr{L}(d) = k(d) \sqcup \Gamma(d)^*$ has exatly $2^{|S(d)|} + 1$ states provided $\Gamma(d) \subset \Sigma$. Alternatively the minimal automaton has $2^{|S(d)|}$ (when $\Gamma(d) = \Sigma$).*

Proof. By theorem 6.1 $\mathcal{A}(k(d))$ has $2^{|S(d)|} + 1$ states. Intuitively we remove all transitions from $\mathcal{A}(k(d))$ either sourced on or targetted at q_{sink}. We add transitions from each state s of $\mathcal{A}(k(d))$ as follows. Let σ be a letter not labelling any transition sourced on s. If σ is not a shaded zone, label a transition from s to itself with σ. Otherwise, label a transition from s to q_{sink} with σ. □

This construction process is demonstrated in figure 6.

Corollary 6.1. *Let d be a unitary diagram. The minimal automaton accepting $\mathscr{L}(d)$ has at most $2^{|S(d)|} + 1$ states:*

$$d_n \xrightarrow{spiders,states} \mathrm{dFA}_{\leq 2^n+1}$$

Theorem 6.3. *Let $D = d_1 \vee \ldots \vee d_m$ be a spider diagram. The minimal automaton accepting $\mathscr{L}(D)$ has at most $2^{\sum_{i=1}^{m} |S(d_i)|} + 1$ states.*

$$D_{\sum_{i=1}^{m} |S(d_i)|} \xrightarrow{spiders,states} \mathrm{dFA}_{\leq 2^{\sum_{i=1}^{m} |S(d_i)|}+1}$$

Proof. By the well known upper-bound [23] for the union of m deterministic finite state machines $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$ is of order $O(\prod_{i=1}^{m}(|\mathcal{A}_i|))$. This gives us an upper bound for the size of $\mathcal{A}(\mathscr{L}(D))$ where, by corollary 6.1, $\mathcal{A}_1$ accepts $\mathscr{L}(d_1)$, $\mathcal{A}_2$ accepts $\mathscr{L}(d_2)\ldots$ and $\mathcal{A}_m$ accepts $\mathscr{L}(d_m)$. □

7 Conclusion

We have established relationships between spider diagrams and various subsets of star-free REG (see figure 4 for a summary). These results show that it is possible to use spider diagrams to visualise languages drawn from these various subclasses.

We have also demonstrated economy of representation offered by spider diagrams over deterministic finite state machines: in some cases a unitary spider diagram containing n spiders is equivalent to a finite state machine containing $2^n + 1$ states, for example. The in-

tuition behind the economy of representation that spider diagrams bring over deterministic finite state machines is that, to not specify an order on letters using deterministic FAs many states are required (ensuring one path exists from the start state to the final state for each ordering of the letters) whereas spider diagrams achieve this naturally. Deterministic FAs are more expressive than spider diagrams (partly due to being able to specify an ordering) but their restrictive syntax forces us to use many states when we do not wish to specify any ordering.

Acknowledgement Gem Stapleton is supported by a Leverhulme Trust Early Career Fellowship and by UK EPSRC grant EP/E011160/1 for the Visualization with Euler Diagrams project.

References

[1] R. Clark. Failure mode modular de-composition using spider diagrams. In *Proc. of Euler Diagrams 2004*, volume 134 of *Electronic Notes in Theoretical Computer Science*, pages 19–31, 2005.

[2] R. DeChiara, U. Erra, and V. Scarano. VennFS: A Venn diagram file manager. In *Proc. of Information Visualisation*, pages 120–126. IEEE, 2003.

[3] J. Flower, J. Masthoff, and G. Stapleton. Generating readable proofs: A heuristic approach to theorem proving with spider diagrams. In *Proc. of 3rd Intl. Conf. on the Theory and Application of Diagrams*, volume 2980 of *LNAI*, pages 166–181. Springer, 2004.

[4] J. Goldstine, M. Kappes, C. M. Kintala, H. Leung, A. Malcher, and D. Wotschke. Descriptional complexity of machines with limited resources. *J. of Universal Computer Science*, 2002.

[5] P. Hayes, T. Eskridge, R. Saavedra, T. Reichherzer, M. Mehrotra, and D. Bobrovnikoff. Collaborative knowledge capture in ontologies. In *Proc. 3rd Intl. Conf. on Knowledge Capture*, pages 99–106, 2005.

[6] P.-C. Héam. On shuffle ideals. *Theoretical Informatics Applications*, 36:359–384, 2002.

[7] J. Howse, F. Molina, J. Taylor, S. Kent, and J. Gil. Spider diagrams: A diagrammatic reasoning system. *J. of Visual Languages and Computing*, 12(3):299–324, 2001.

[8] J. Howse, G. Stapleton, and J. Taylor. Spider diagrams. *LMS J. of Computation and Mathematics*, 8:145–194, 2005.

[9] J. Larkin and H. Simon. Why a diagram is (sometimes) worth ten thousand words. *J. of Cognitive Science*, 11:65–99, 1987.

[10] J. Lovdahl. *Towards a Visual Editing Environment for the Languages of the Semantic Web*. PhD thesis, Linkoping University, 2002.

[11] A. Meyer and M. Fischer. Economy of description by automata, grammars, and formal systems. In *Proc. 12th Annual Symposium on Switching and Automata Theory*, pages 188–191, 1971.

[12] O. Patrascoiu, S. Thompson, and P. Rodgers. Tableaux for diagrammatic reasoning. In *Proc. 2005 Intl. Workshop on Visual Languages and Computing*, pages 279–286. Knowledge Systems Institute, 2005.

[13] J.-E. Pin. Finite semigroups and recognizable languages: an introduction. In *NATO Advanced Study Institute: Semigroups, Formal Languages and Groups*, pages 1–32. Kluwer Academic Publishers, 1995.

[14] J.-E. Pin. *Syntactic semigroups*, pages 679–746. Springer-Verlag New York, Inc., 1997.

[15] J.-E. Pin and P. Weil. Polynomial closure and unambiguous product. *Theoretical Computer Science*, 30:1–39, 1997.

[16] M. O. Rabin and D. Scott. Finite automata and their decision problems. *IBM J. of Research Development*, pages 114–125, 1959.

[17] A. Rabinovich. Star free expressions over the reals. *Theor. Comput. Sci.*, 233(1–2):233–245, 2000.

[18] M. Schtzenberger. On finite monoids having only trivial subgroups. *Information and Control*, pages 190–194, 1965.

[19] S.-J. Shin. *The Logical Status of Diagrams*. Cambridge University Press, 1994.

[20] G. Stapleton, S. Thompson, J. Howse, and J. Taylor. The expressiveness of spider diagrams. *J. of Logic and Computation*, 14(6):857–880, 2004.

[21] J. Thièvre, M. Viaud, and A. Verroust-Blondet. Using Euler diagrams in traditional library environments. In *Euler Diagrams 2004*, volume 134 of *ENTCS*, pages 189–202. ENTCS, 2005.

[22] W. Thomas. *Languages, automata, and logic*, pages 389–455. Springer-Verlag New York, Inc., 1997.

[23] S. Yu. State complexity of regular languages. In *IWD-CAGRS: Proc. Intl. Workshop on Descriptional Complexity of Automata, Grammars and Related Structures*, 1999.

[24] Y. Zhao and J. Lövdahl. A reuse based method of developing the ontology for e-procurement. In *Proc. Nordic Confernce on Web Services*, pages 101–112, 2003.

Usability evaluation of interactive visual applications: a quantitative approach

Rosanna Cassino, Maurizio Tucci
Dipartimento di Matematica e Informatica,
Università di Salerno (Italy)
{rcassino, mtucci}@unisa.it

Abstract

Most of the existing techniques for the evaluation of interactive visual applications must be manually performed and iteratively repeated at each refinement step after a rapid prototyping of the system. This paper introduces a usability evaluation approach of graphical environments that tends to realize an automatic evaluation process to achieve quantitatively measurable parameters. The process analyzes the source code of a visual application, the design model of its interface, the graphical aspects and the interaction mechanisms implemented in each frame, panel and/or page of the system, to produce a report of the quantitative evaluation of heuristic factors. In particular, we consider interactive applications developed in a markup language or in Java and a subset of the Nielsen heuristics as usability metrics: completeness, correctness, aesthetic and minimalist design, error prevention/user control and freedom, consistency - recognition rather than recall, visibility of the system status. As a case study, we describe the application of the proposed evaluation approach in the analysis of the interface of a no-profit foundation web site.

1. Introduction

The evaluation of interactive visual applications is traditionally performed by means of usability testing involving a rapid prototyping of the system and an evaluation based on the use of the product by end users or specialists. Popular techniques to detect usability deficiencies are based upon testing the developed graphical environment by executing test cases and checking the execution results [7]. Heuristic evaluation, heuristic estimation, cognitive walkthrough, pluralistic walkthrough, feature inspection, consistency inspection, standards inspection, formal usability inspection, are standard methods for usability inspection [10]. In general, the evaluation of usability heuristics involves a group of end users. Usual evaluation methods include cooperative work [19], questionnaires and usability test. These costly and time-consuming activities often produce results that are considerably depending on the measure acquisition, on the punctual usability objective definition, on the type and the number of the performed tasks, on the data to retrieve, on the performance evaluation standards [18]. Most of the previous techniques must be manually performed and iteratively repeated at each refinement step.

In this perspective, we propose a usability evaluation approach for interactive visual applications that tends to overcome the problems of the traditional testing techniques above. To achieve a more automatic evaluation of quantitatively measurable parameters, the process analyzes the source code of a visual application, the design model of its interface, the graphical aspects and the interaction mechanisms implemented in each frame, panel or page of the application, to produce an evaluation report based on the measured values. We consider interactive application developed in a markup language or in Java.

Usability metrics and automatic or semi-automatic processes to evaluate them, aid and improve the analysis of visual applications reducing the evaluation costs.

Layout based metrics can reduce the manual work of designers by automatically generating initial layouts and analyzing selected aspects of a potential design [14]. In addition, user interface events can also be used to automatically extract usability information during test case executions [5]. An event gives information about the interactions that an user is performing by an external device (mouse, keyboard, etc…) and acting on particular components of the visual environment (buttons, labels, menus, input fields, etc…). Such events represent a very effective source of information regarding application usage and usability, and they can be easily captured and specified in terms of source object, target object and attributes that are modified as a result of an interaction.

In this work, we consider a subset of the Nielsen heuristics [8] and we describe an usability evaluation process of interactive visual applications to achieve quantitative measures of these properties based on the analysis of the source code of the application.

As a case study, we analyzed the interface of a no-profit foundation web site [17]. The evaluation results are used to perform refinements of the site in terms of several usability properties. To show the effectiveness of the approach, we compared our analysis values to a manually usability test performed involving a group of the tester users. The assessment of the two methods shows that similar results are obtained, that is the same basic weaknesses have been detected, so our method can reduce evaluation costs without loss of effectiveness.

The paper is organized as follow. Section 2 presents some of the existing approaches to the usability evaluation of software applications. In section 3 we detail the proposed technique to achieve quantitative usability measures of a

visual environment. Section 4 presents a case study related to the analysis of a web site. Section 5 contains some concluding remarks.

2. Usability evaluation techniques

Much work has been done to find practical usability properties that can be effectively checked and measured to evaluate the usability of interactive applications.
The analysis of user interface events can be used to extract usability information of an application. A survey of computer aided techniques for extracting usability information deriving from UI events can be found in [5]. Nevertheless, the usefulness of these techniques is rather limited if a "subjective" analysis must be performed in any case by the human observation in order to evaluate each particular usability issue under investigation.
Quantitative measure of usability is one of the main problems of the standards in the context of usability of software quality. Quantitative evaluations are performed, for example, by the GOMS (Goals, Operator, Methods, and Selection rules) approach and the keystroke level model (KLM) [2]. GOMS can be used to predict how long it will take to learn a certain task. With GOMS, the total time can be predicted by associating time with each operator. With these quantitative predictions, GOMS can be applied for example in a comparison between two systems. The GOMS model also has its limitations. Preece et al. [12] suggest that GOMS can only really model computer-based tasks that involve a small set of highly routine data-entry type inputs. The model is not appropriate if errors occur.
A quantitative approach to evaluate the compliance of the mobile phone UI can be found in [6]. This and other approaches [4] are related to a specific domain and cannot be generalized: each project has its specific features, and the usability methods should be selected and tailored on the base of the specific context of the project.
Quantitative measures to evaluate the design, the communication, the functionality, the contents, the management, the accessibility and the usability properties of web sites are described in [11]. The usability is defined in terms of effectiveness, efficiency and satisfaction sub-properties, and then evaluated using questionnaires. In fact, questionnaire for user interaction satisfaction (QUIS) [3] are a simple and economic tool to obtain quantitative measure of usability by the user's subjective rating of the human computer interface. Nevertheless, the approach is sensible to subjectivity of data, user experience, etc….
In the proposed approach we have identified a subset of the Nielsen heuristics: completeness, correctness, aesthetic and minimalist design, error prevention/user control and freedom, consistency - recognition rather than recall, visibility of the system status; and we propose computable measures to evaluate each property and its related sub-property. These heuristics are well suited for usability evaluation of graphical user interfaces, and a detailed definition of metrics is available to practically measure them.

3. Quantitative evaluation of user interfaces

In this section, we describe an algorithmic process to achieve quantitative measures of a subset of the Nielsen usability properties [8], based on the analysis of the source code of the application. If an application is implemented in a markup language (i.e., HTML, XML, VRML, SVG, etc…) a general tag processing allows to identify the components of the pages and the related attributes, whereas the "href" tags allow to identify the interaction mechanisms. Similarly, if the application is implemented in an object-oriented language (i.e., Java) an examination of the classes allows to identify the containers and the widgets, an analysis of the methods allows to define the interface-flow diagram which describes the interaction processes. The delegation model of the event management, defined by the listener objects that wait the event generation from the source objects, allows to performs checks and evaluations of the tasks retrieving quantitative measure of usability.
For each considered usability property, we define the *calculated* value as the arithmetic average of the values of the related sub-properties; and the *weighted* value as the weighted average of the weighted values of its sub-properties. The weighted value is calculated on the base of the weight which the evaluator wants to associate to each characteristics. We assume that weights range from 0 to 1: 0 if the sub-property is not relevant to the evaluation, 1 if it is extremely important.
From the analysis of the source code of an application it is possible to recover the modelling of the interactive visual application, for example in terms of an interface – flow diagram where the nodes represent the scenes and the external components of the application; the edges represents the interaction paths that characterize the system. Then during the design analysis it is possible to evaluate the *completeness*, the *correctness*, the *aesthetic and minimalist design*, the *error prevention* and then, the *user control and freedom* usability properties.
The *completeness* can be evaluated in terms of the *interactions incompleteness* identifiable in the interface – flow diagram as the number of the edges outgoing from a source node and that don't have the corresponding target node. Then, if we label with CM the calculated value of the completeness and CMp the number of the dangling references in the application, results: CM = CMp. If CMw is the weighted value of the *completeness* property, W_{CM} is the weight of this property, CMpw the weighted value of the interactions incompleteness and W_{CMp} the weight of this sup-property, results: CMpw = CMp * W_{CMp} and then, CMw = CMpw/W_{CM}.
Lesser is CMw, better is the usability level for this property.
The *correctness* can be evaluated in terms of the *frames or pages not achievable* and corresponding to the number of the nodes not reachable by any edge outgoing from any sourcing scene. Then, if CR is the calculated value of the correctness and CRp the number of the scenes inside of the application but not achievable, results: CR = CRp. If CRw is the weighted value of the *correctness* property, W_{CR} is the weight of this property, CRpw the weighted

value of the interactions incompleteness sub-property and W_{CRp} the weight of this sup-property, results: $CRpw = CRp * W_{CRp}$ and then, $CRw = CRpw/W_{CR}$. Lesser is CRw, better is the usability level for this property. In the case of web site evaluation, there are countless web pages potentially achievable from any page of the site. Nevertheless, the related links could not be inserted for design choices and not for implementation mistakes.

The *aesthetic and minimalist design* can be evaluated in terms of the complexity of the actions to perform a task, which might correspond to the length of the interaction paths necessary to achieve a scene or an external component. In the design diagram this can be represented by the number of the edges from a "start" source node (the index page in the case of a web site, the main frame o the start panel in a general application) to a target node. Then, the quantifiable measure of this property is described by the maximum nesting level of a node. Formally, if AE is the calculated value of the aesthetic and minimalist design, NL the calculated value of the complexity of the interaction paths and NL(i) is the nesting level of a node corresponding to the lesser number of the tasks (edges in the diagram) to achieve the node "i" from the start node, result: $NL= max \{NL(i), 0<i\leq N$, N is the number of the nodes in the diagram} and then, AE = NL. If AEw is the weighted value of the *aesthetic and minimalist design* property, W_{AE} is the weight of this property, NLw the weighted value of the complexity of the interaction paths and W_{NL} the weight of this sup-property, results: $NLw = NL * W_{NL}$ and $AEw = NLw/W_{AE}$. Lesser is AEw, greater is the usability level for this property.

The *error prevention, user control and freedom* property corresponds to the number of the tasks to back return and the number of the tasks to return at the start scene. In the diagram design model, the number of the tasks to back return can be evaluated in terms of the number of the edges from a node to the previous. Then, if TR is the calculated value of this sub-property, W_{TR} is its weight, N the number of the nodes in the application diagram, if $TR\approx N$ better is the usability level for this sub-property (that means that for each scene there is a task to back return). The weighted value, TRw, is: $TRw = TR * W_{TR}$.

Similarly, the number of the tasks to return at the start scene (resp., to the home page in the case of a web site) can be evaluated in terms of the number of the edges from a node to the first (i.e., the initial scene). Then, if TS is the calculated value of this sub-property, W_{TS} is its weight, the weighted value, TSw, is: $TSw = TS * W_{TS}$. Greater is TS, better is the usability level for this sub-property. Moreover, the *error prevention, user control and freedom* property can be evaluated in terms of the number of the tasks to exit from the application, the check of the number of the "Undo"/"Redo" tasks and the "Dialog" components inside to each scene of the application. That sub-properties can be analyzed during the *implementation analysis level*. As a matter of fact, during the analysis of the source code of the graphical and interactive aspects of the application, the number of the tasks to exit from the application can be evaluated in terms of the *exit* instructions implemented. In the case of web site it is impossible to evaluate this sup-properties because the "site navigation" finishes by changing the url address in the browser or closing the same browser. Then, in the analysis of a stand alone application, if TE is the calculated value of the *exit* instructions (at first TE = 1, since we assume that exist at least one possibility to exit from the application), W_{TE} is the weight of this sub-property, the weighted value, TEw, is: $TEw = TE * W_{TE}$. Greater is TE, better is the usability level for this sub-property. On the other hand, the check of the undo/redo tasks can be calculated in terms of the number of the classes that define "undo" and "redo" methods in an objects language, or the number of the tags that define "undo" and "redo" functions in a markup language. Then, if BU is the calculated value of the "undo" actions and BR is the calculated value of the "redo" actions and BUR the sum of the undo and redo tasks in the application, W_{BUR} is the weight of this sub-property, result: BUR = BU + BR. If BU = BR and the value if great, better is the usability level for this sub-property. The weighted value of the check of "Undo"/"Redo" actions, BURw, is: $BURw = BUR * W_{BUR}$. The check of the presence of the dialog components in the examined application can be calculated in terms of the number of the classes/tags that define a "*dialog frame*". Then, if BD is the calculated value of the check of the presence of the dialog frames sub-property, W_{BD} the related weight, the weighted value of the check of the presence of dialog frames sup-property, BDw is: $BDw = BD* W_{BD}$. Greater is BDw, better is the usability level for this sub-property. Therefore, the calculated value, ER, for the *error prevention, user control and freedom* usability property is the arithmetic average of the calculated value of its sub-properties (TR, TS, TE, BUR, BD), that is:

$$ER = (TR+TS+TE+BUR+BD)/5$$

The weighted value, ERw, is calculated as the weighted average of the weighted value of its sub-properties:

$$ERw = (TRw+TSw+TEw+BURw+BDw)/W_{ER}$$

Greater is ERw better, is the usability level for this property.

During the interface implementation analysis, the *consistency*, defined also as *recognition rather than recall* usability property can be evaluated in terms of the check of the location and of the font of the similar elementary objects in several scenes, the check of the expressivity of the dynamic objects labels and then, in terms of the check of the presence of a text linked to each icon which identifies a component. Then, if CS is the calculated value of the consistency, LC the number of the similar objects with different location inside of the several scenes, W_{LC} the weight of this sub-property, the weighted value of this sub-property, LCw, is: $LCw = LC * W_{LC.}$ Lesser is LCw, better is the usability level for this sub-property. Similarly, if FT is the number of the similar objects with the same font inside of the different scenes, W_{FT} the weight of this sub-property, the weighted value of this sub-property, FTw, is: $FTw = FT * W_{FT}$. Lesser is FTw, better is the usability level for this sub-property.

The check of the expressivity of the label which identifies a dynamic object, can be evaluated by the match between

the description (title) of the source object and the description of the corresponding listener object in the objects language (resp., the match between the "href" tags name and the corresponding anchor in a markup language). Then, if EX is the value of this sub-property calculated in terms of the number of the similar label, W_{EX} its weight, the weighted value of this sub-property, EXw is: $EXw = EX * W_{EX}$. Greater is EXw, better is the usability level for this sub-property. The check of the presence of a description linked to any icon which identifies an elementary object, can be evaluated analysing the constructor of each button or label inside of each scene of the application. In a markup language for each component defined by an image, for example, we should have: <img title="Enter"… src="Immages/ enter.gif">. Then, if DS is the value of this sub-property calculated as the number of the elements defined by an image and a text, W_{DS} the related weight, the weighted value of this sub-property, DSw, is: $DSw = DS * W_{DS}$. Greater is DSw, better is the usability level for this sub-property. Therefore, the calculated value, CS, for the *consistency* usability property is the arithmetic average of the calculated value of its sub-properties (LC, FT, EX, DS), that is: CS = (LC + FT + EX + DS)/4

The weighted value, CSw, is calculated as the weighted average of the weighted value of its sub-properties:

$$CSw = (LCw+FTw+EXw+DSw)/W_{CS}$$

Greater is CSw, better is the usability level for this property.

The *visibility of the system status* usability property can be evaluated in terms of the check of the change of the form of the mouse pointer as a result of an interaction, the check of the number of the error messages and the check of the presence of the progress indicators linked to the definition of a long task. Then, if MP is the value of the first sub-property, calculated in terms of the number of mouse pointer changing, W_{MP} the related weight, the weighted value of this sub-property, MPw is: $MPw = MP * W_{MP}$. Greater is MPw, better is the usability level for this sub-property. In the case of the web site interface analysis, this sub-property is generally always respected, since the browser, in presence of a link, change the mouse pointer (i.e., from an arrow to a hand) automatically.

The check of the presence of the error messages in the examined application, can be calculated in terms of the number of the classes (resp., tags) that define an "*error frame*" corresponding to each interaction mechanism implemented. Then, if EM is the calculated value of the number of the "*error frames*", W_{EM} is the weight of this sub-property the weighted value of the check of the presence of error messages sup-property, EMw, is: $EMw = EM * W_{EM}$. Greater is EMw, better is the usability level for this sub-property. Similarly, the check of the presence of the progress indicators linked to the complexes tasks (expressed in terms of the number of the instructions inserted) can be calculated in terms of the number of the *progress bar* objects corresponding to each interaction mechanism implemented. Then, if PI is the calculated value of the number of the *progress bar*, W_{PI} is the weight of this sub-property corresponding to the *source objects* (resp., *href tag*), the weighted value of the check of the presence of progress indicator sup-property, PIw is: $PIw = PI* W_{PI}$. Greater is PIw better, is the usability level for this sub-property. Therefore, the calculated value, VS, for the *visibility of the system status* usability property is the arithmetic average of the calculated value of its sub-properties (MP, EM, PI), that is:

$$VS = (MP + EM + PI)/3$$

The weighted value, VSw, is calculated as the weighted average of the weighted value of its sub-properties:

$$VSw = (MPw+EMw+PIw)/W_{VS}$$

Greater is VSw better, is the usability level for this property.

Since the *standards respect* usability property depends from the particular interest domain, we have not improved this heuristic in the present work, sending back to further works, related to the evaluation of specific interactive visual applications, the quantitative analysis of that metric.

In the next section we describe a case study related to the analysis of the web site of a no-profit foundation performed by the proposed evaluation approach.

4. The usability evaluation of a web site

To show the effectiveness of the presented approach, we examined a real application and we compared the evaluation results with a manual test performed by a usability test involving a group of users. In particular, we have formulated informal queries for each usability property to evaluate. The users have answered assigning a vote in a range of values (i.e., from 0 to 5). We have underlined similarity relationships between the values calculated by the proposed analysis methodology and the users tester votes. Then, we have analyzed the web site of a no-profit foundation [17] and, on the base of the obtained values, the application is redefined. In fact, the use of the first release of the web site has highlighted several accessibility and usability problems.

It is out from the purpose of this work to describe the evaluation of the accessibility of the site and the related adopted solutions.

As described in the previous section, we have evaluated the "calculated values" of the usability properties related to the completeness, the correctness, the aesthetic and minimalist design, the error prevention, the user control and freedom and the visibility of the system status.

The obtained values are shown in the Table 1. For each sub-property we have chosen a weight in the range from 0 to 1. The weight of the related usability property is the sum of the several "sub-weights". The calculated value is the arithmetic average of the value of the related sub-properties. The weighted value of each usability property is the weighted average of the weighted value of its sub-properties.

The graphic in figure 2 makes visible the usability level of each property. The evaluation report shows a medium level of completeness and correctness, caused, respectively, by the presence of a link not connected to any web page and by a "forum page" not achievable.

USABILITY PROPERTIES	CALCULATED VALUE	WEIGHT	WEIGHTED VALUE
Completeness	1	1	1
Interactions incompleteness	1	1	1
Correctness	1	0,5	1
Frame/Pages not achievable	1	0,5	0,5
Aesthetic and minimalist design	3	1	3
Complexity of the interaction paths	3	1	3
Error prevention, user control and freedom	11	2,5	4,4
Tasks to back return	4	1	4
Tasks to return at the start scene	7	1	7
Tasks to exit from the application	0	0	0
Check of "Undo"/ "Redo" buttons	0	0	0
Check of the presence of the "Dialog" frames	0	0,5	0
Consistency/ Recognition rather than recall	14	2,3	3,43
Check of the location of the similar object in several scenes	4	0,5	2
Check of the font of the similar object in several scenes	4	0,5	2
Check of the expressivity of the label of the dynamic objects	3	0,3	0,9
Check of the presence of a description label linked to any icon of an object	3	1	3
Visibility of the system status	8	1,5	5
Check of the pointer properties	0	0	0
Check of the presence of the error messages	1	0,5	0,5
Check of the presence of the progress indicator	7	1	7

Table 1. Usability values of the first release of the site.

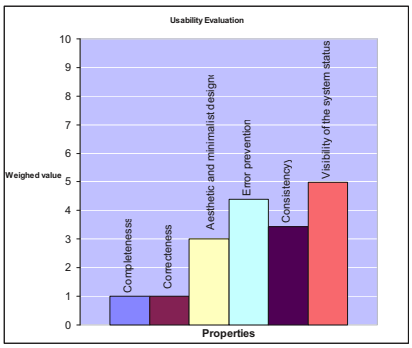


Figure 2. The graphic of the first evaluation.

A good level of aesthetic and minimalist design is guaranteed by a lower level of nesting of the pages achievable from the home page (maximum three steps). Similarly, the error prevention and user control is obtained thanks to a sufficient number of buttons to back and to the home page return inside of the web pages.

The medium level of the consistency property is caused by the presence of similar objects in two several web pages that have different font and location, and also by the presence of a button to achieve the "contacts page" represented by an icon not connected to any textual description. Nevertheless a good visibility of the system status is obtained thanks to the presence of a number of error messages and a progress bar in the browser to show the loading level of a page.

On the base of the assessments of the obtained values, a new release of the web site was realized. The graphic in figure 3 makes visible the improvement performed in the new release of the site. As visible, the correction of the problems has decreased the value of sub-properties related to the completeness and the correctness and has increased the values related to the errors prevention, the user control and freedom, the consistency, improving the corresponding usability level for these properties.

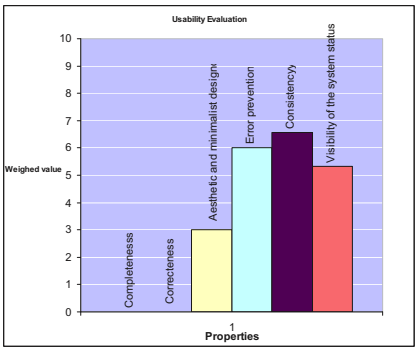


Figure 3. The graphic after the new evaluation.

To show the effectiveness of the approach, the obtained results are compared to a manual test performed by tester users. In particular, we have formulated informal queries for each usability property under evaluation, and the results were submitted to a pool of domain-expert users, generic users, and students. A total of 14 users were involved to drive the tests. The users have answered assigning a vote in a range of values (i.e., from 0 to 5). Table 2 shows an example of questionnaire developed to test the first release of the web site.

USABILITY PROPERTIES	QUESTIONS	VALUE
Completeness	Are the functions to carry out the transactions adequate?	4
Correctness	Are the pages easy to achieve?	4
Aesthetic and minimalist design	Is the graphic of the web site user-friendly?	2
Error prevention, user control and freedom	Is the errors management adequate?	3,5
Consistency/ Recognition rather than recall	Do the pages, the colour of the objects and the characters of the text make to understand the contents easily?	3
Visibility of the system status	Is it possible to check the state of loading of a page or the execution of a task?	4

Table 2. The subjective evaluation of the web site.

The graphic in figure 4 makes visible the results of the evaluation performed by the questionnaire. The results of

the evaluation questionnaire of the reimplemented site shows that it is possible to define similarity relationships between the values calculated by the proposed approach and the users tester votes. As visible, the values of the evaluation of the completeness and the correctness are inversely proportional to the votes of the users. This is because, for example, a high number of dangling references corresponds to a lower vote from the user related to the accessibility of any page. The evaluation of aesthetic and minimalist design, error prevention, consistency and visibility of the system status correspond almost precisely. Slight differences appear due to the subjective nature of user evaluation, as the answers to questionnaires are marginally influenced by how the questions are formulated.

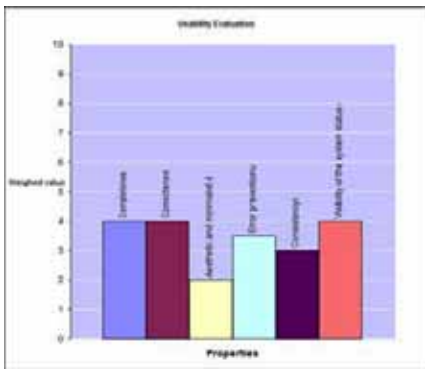


Figure 4. The graphic of the questionnaire results.

5. Conclusion

In this paper, we have presented a process to obtain quantitative measures of usability to analyze interactive visual applications. The proposed approach is centred on the evaluation of a subset of the Nielsen heuristics desirable in a graphical user interfaces [8].

As a case study, we have applied the proposed approach to analyze the web site of a no-profit foundation. The measures automatically calculated agree with the values obtained by an heuristic evaluation based on the results of a questionnaire proposed to a group of end users. Nevertheless, the proposed approach reduces the evaluation costs and the need of expert or tester users.

We expect that the technique runs for other types of interface. In this perspective, further work will be done to extend and to show the usefulness of the described usability evaluation process in a wider range of desktop and mobile applications. Then, further improvements can be achieved by adding usability checks during the development process of an application to prevent usability problems before the software is released and to manage usability issues and controls based on a formal specifications of the visual language corresponding to the graphical application.

References

[1].Len Bass, and Bonnie E. John, "Linking usability to software architecture patterns through general scenarios", *Journal of Systems and Software* – vol. 66, Issue 3 , 15 June 2003, Pages 187-197.

[2].JOHN, B. E. AND KIERAS, D. E. 1996b. "Using GOMS for user interface design and evaluation: which technique?", *ACMTrans. on Comput.–Hum. Interaction* 3, 4.

[3].Chin, J. P., Diehl, V. A. & Norman, K. L. (1988), "Development of an instrument for measuring user satisfaction of the human computer interface". *Proceedings of CHI '88*, 213-218.

[4].Jeffrey W. Dawson, Ping Chen, Yanshen Zhu, "Usability study of the virtual test bed and distributed simulation", *Proceedings of the 2005 Winter Simulation Conference M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.*

[5].Hilbert, D.M. and D.F. Redmiles, "Extracting usability information from user interface events", *ACM Computing Surveys* (CSUR). 2000. p. 384 - 421.

[6].Timo Jokela , Jussi Koivumaa , Jani Pirkola, Petri Salminen and Niina Kantola, "Methods for quantitative usability requirements: a case study on the development of the user interface of a mobile phone", *Personal and Ubiquitous Computing Publisher - Springer London* - Volume 10, Number 6 /October, 2006.

[7].A. Memon, I. Banerjee, A. Nagarajan, "GUI Ripping: Reverse Engineering of Graphical User Interfaces for Testing, *Proceedings of the 10th Working Conference on Reverse Engineering* (WCRE'03).

[8].Nielsen, J. (1994b). "Heuristic evaluation", in Nielsen J., and Mack R.L. (Eds.), *Usability Inspection Methods*, John Wiley & Sons, New York, NY.

[9].J. Nielsen, "Iterative User Interface Design", published in *IEEE Computer* Vol. 26, No. 11 (November 1993), pp. 32-41.

[10].Jakob Nielsen and Robert L. Mack, "Usability Inspection Methods" - *John Wiley & Sons*, New York, NY, 1994.

[11].R. Polillo, "Il check-up dei siti web – Valutare la qualità per migliorarla", *Edizioni Apogeo*, 2004.

[12].Preece J, Rogers Y, Sharp H (2002), "Interaction design". Beyond *Human–Computer Interaction*, Wiley, New York.

[13].Pyla, P., Pérez-Quiñones, M., Arthur, J., & Hartson, R. (2003) "Towards a Model-Based Framework for Integrating Usability and Software Engineering Life Cycles", *SE & HCI Communities Workshop* in INTERACT 2003.

[14].Sears, A. (1995), "AIDE: A Step Toward Metric-Based Interface Development Tools", *Proceedings of the ACM Symposium on User Interface and Software Technology* (UIST 1995), 101-110.

[15].Melanie Tory, Torsten Moller, "Evaluating Visualizations: Do Expert Reviews Work?", *IEEE Computer Graphics and Application*, September/October 2005 (Vol. 25, No. 5) pp. 8-11.

[16].www.cs.umd.edu/projects/hcil/

[17].www.fondazioneorlando.org

[18].http://www.useit.com/alertbox

[19].http://web.tiscalinet.it/userware/

A Comparison between 2D and 3D Iconic Visual Arrangements for Representing Topological Relationships

Enzo Del Fatto, Luca Paolino, Monica Sebillo, Genoveffa Tortora, Giuliana Vitiello
Dipartimento di Matematica e Informatica - Università di Salerno
Via Ponte don Melillo, 84084 Fisciano (SA) Italy
{vdelfat, lpaolino, msebillo, tortora, gvitiello}@unisa.it
+39 089 963324

ABSTRACT
In this paper, we describe the result of an experiment we conduct at our Department meant to understand which visual representation between two-dimensional or three-dimensional icon arrangements is better for stimulating the user mental model in order to recognize topological relationships. The experiment has involved two groups of 10 subjects to which we have submitted some tasks representing common topological relationships. The experiment consisted in two sessions During the first session, for each task we asked users to assign a score indicating how much the task may be represented by each arrangement, then we calculated the existing correlation between the approaches. Resulting values show that using three-dimensional icons there are no significant improvements in efficacy with respect to bi-dimensional iconic arrangements. However, the second session of the experiment revealed that the degree of user's satisfaction may be positively influenced by the use of three-dimensional icons.

Keywords
3D icons, 2D icons, topological relationships, Pearson correlation index.

1. INTRODUCTION

One of the fundamental concepts necessary for the analysis of data in a Geographic Information System (GIS) or an Image Management System is a formal understanding of how geometric objects are spatially related in order to provide users with a description of the data contents. This is commonly achieved thanks to a special class of relationships, named *Topological*, which is a particular subset of geometric relations, invariant with respect to transformations such as translation, rotation, and scaling. Typical examples of this relationships are: *Equals*, *Disjoint*, *Touches*, *Within*, *Overlaps*, *Contains*, *Crosses* and *Intersects* as described by the OpenGeospatial Consortium [7], which is a set of relationships providing a qualitative rather than quantitative description of content.

The same principle of describing the content of images by means of topological relationships is also used for interpreting examples of data the users are looking for according to the Query-by-Example[10] methodology. That is to say, user visually composes some examples of what s/he needs, the composition is successively translated into a set of objects and spatial relationships, which are then compared with the target images in the databases/GIS, also expressed as objects and relationships.

In the last two decades, this kind of approach has been largely investigated and several proposals have been made, which allow users to compose examples by means of (two or three-dimensional) iconic representations (see, e.g., [9], [8], [1], the [6], [2], [4] and [5]).

Disregarding the natural attractiveness of three-dimensional visualization and the objective necessity to have this kind of representation for depicting some directional and metrical relationships [8], we wished to verify whether there are substantial improvements when recognizing or composing 2D topological relationships by using environments which manage three-dimensional icons instead of two-dimensional ones.

In this paper, we report on a cognitive experiment we conducted at our Department, in order to understand which visual representation between two-dimensional or three-dimensional icon arrangements is better for stimulating the user mental model when recognizing topological relationships.

The paper is organized as follows. In Section 2, we recall some basic concepts related to the manipulation of spatial data. In Section 3, we present the experiment and, in Section 4 we describe the results we gained from the experiment.

2. Basic Concepts

In this section, we take a look at the basic concepts relative to the research we present in this paper.

First of all, we recall Egenhofer's 9-intersection model[3]. In this model, the topological relations between two simple spatial entities A and B are transformed into point set topology problems, in terms of the intersections of A's interior and boundary with B's interior and boundary. It means that spatial relationships are calculated by the matrix where column are indexed with the interior, exterior and boundary of the first entity the, row are indexed with the interior, exterior and boundary of the second entity and each cell contains a Boolean value indicating whether the row attribute intersects the column attribute.

In Figure 1, we applied Egenhofer's model to identify how many and which different topological arrangements exist between two objects represented by cubes. We identified six different arrangements which are labeled *Op1-Op6*.

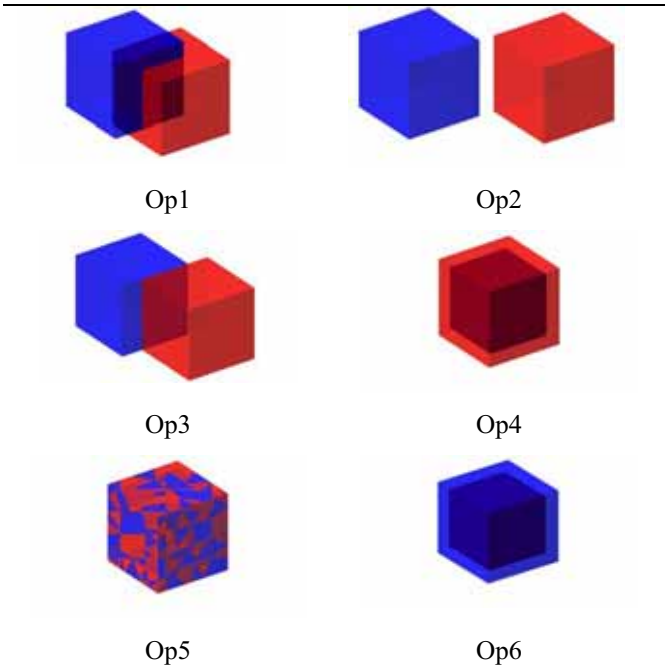


Figure 1. Visual compositions by 3D icons, as generated by Egenhofer's 9-intersection matrix

The same approach has been used to identify the possible arrangements between a pair of 2D icons. Also in this case, we recognized six different arrangements, as depicted in Figure 2.

The set of topological relationships most commonly used in a GIS is part of the standard specification issued by the OpenGeospatial consortium [7] and includes *Equals*, *Disjoint*, *Touches*, *Within*, *Overlaps*, *Contains*, *Crosses* and *Intersects*. Given two spatial entities g1 and g2, therefore:

- *Disjoint* is TRUE if the intersection of g1 and g2 is the empty set.
- *Touches* is TRUE if the only points in common between g1 and g2 lie in the union of the boundaries of g1 and g2.
- *Within* is TRUE if g1 is completely contained in g2.
- *Overlaps* is TRUE if the intersection of g1 and g2 results in a value of the same dimension as g1 and g2 that is different from both g1 and g2.
- *Crosses* is TRUE if the intersection of g1 and g2 results in a value whose dimension is less than the maximum dimension of g1 and g2 and the intersection value includes points interior to both g1 and g2, and the intersection value is not equal to either g1 or g2.
- *Intersect* is a convenience predicate, namely it is TRUE if the intersection of g1 and g2 is not empty.
- *Contains* is a convenience predicate, that is it is TRUE if g2 is completely contained in g1.

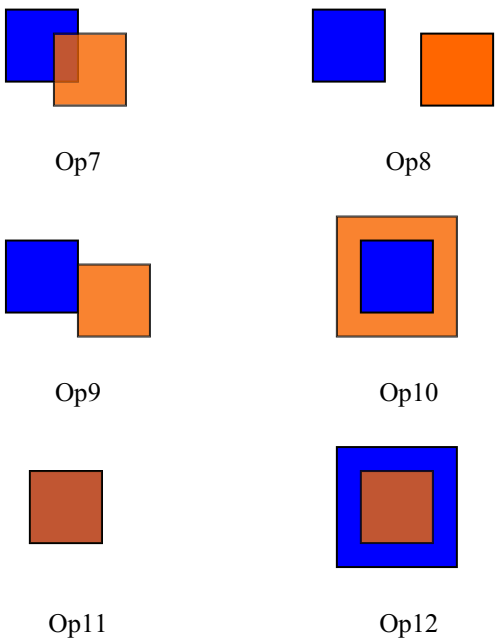


Figure 2. Visual compositions by 2D icons, as generated by Egenhofer's 9-intersection matrix

3. USER STUDY

The 3D and 2D icons arrangements described in the previous section have been used in a usability survey aiming to verify effectiveness of those representations for posing topological queries.

The aim of the survey was twofold. On one hand, we wanted to find out a satisfactory matching between visual compositions (*Op1-Op12* above) and topological relationships in order to define an interpretative model as precise as possible. Such work represents an important prerequisite for improving the quality of the interaction in visual query systems where visual compositions represent geometric objects and their relations.

On the other hand, we wanted to perform a comparative usability study evaluating the 3D composition approach against the traditional 2D approach.

3.1 Apparatus and procedure

A sample of twenty users was involved in the evaluation. We selected twenty participants including students and researchers of the Department of Matematica e Informatica of the University of Salerno. The subjects were males and females between the ages of 20 and 50; all of them were usual PC users and were familiar with GIS technology and the notion of topological relationship. We randomly divided subjects into two groups (named A and B, in the following) of ten people each. The experiment was conducted separately with groups A and B, in a quiet classroom after courses.

We adopted a between-group evaluation process and split the experiment into two sessions:

- In the first session, group A operated on 3D icons only, while group B managed 2D icons only, and were asked to assign resemblance scores to the considered visual compositions.
- In the second session we both groups were asked to simulate the given scenarios performing 3D query composition tasks and as well as the 2D counterparts.

Within each group, in order to avoid biases, tasks and conditions were submitted in different order to participants. A lap-top computer was used for displaying the visual compositions, which were built using a custom application implemented for the survey. In order to record the interaction on the screen and the subjects' comments, we used the Camtasia Studio™ software by Techsmith ™. The survey team was present during testing and included a test assistant and an observer/note-taker. We wrote a script that the test assistant read to each participant in order to provide a common background..

3.2 The Tasks

The two sessions of the experiment were carried out on two different kinds of tasks.

All the tasks were related to a realistic situation where the participant was asked to play the role of a tourist who had to identify some locations of a point of interest in a map. Then, the following scenarios were considered, which involved the most common topological relationships:

S1. Equals. You have two layers Bars and Tobacconist. You are a tourist visiting the city of Rome, you want to know which bars are also tobacconist.

S2. Within. You have two layers, Bars and District. You want to know which bars are inside the district of Parioli.

S3. Crosses. You have two layers, Rivers and Roads. You want to know which roads pass through the Tevere river.

S4. Disjoint. You have two layers, Bars and Districts. You want to know which bars are outside the district of Parioli in Rome.

S5. Touches. You have just the District layer. You want to know which districts confine with the district of Parioli.

S6. Intersect. You have two layers, Rivers and Districts. You want to know which district is partially covered by a river.

S7. Crosses. You have two layers, Bars and Museums. You want to know which districts contains some museums.

During Session 1, subjects from Group A were asked to consider, for each scenario, all possible 3D visual compositions of icons (see Fig.1) and associate with them *resemblance scores* indicating their degree of appropriateness with respect to the topological relationship expressed by the considered scenario.

In a separate room, subjects from Group B, were asked to associate with each, a set of *resemblance scores* indicating the appropriateness of bi-dimensional arrangements, as depicted in Fig 2.

The scale for the resemblance parameter was set from 0 to 10, where 0 was associated to the value *it does not correspond to the topological relationship* and 10 was associated to the value *it fully corresponds to the relationship*.

In Session 2, subjects from Group A were asked to consider scenarios S1-S7 and to perform the required tasks using 3D arrangements, while subjects from Group B performed 2D composition tasks. The goal of this session was to perform a comparative evaluation of the degree of user's satisfaction with respect to the 3D and 2D composition approaches. This was achieved by combining a 'think-aloud' technique with the submission of a user questionnaire. The think-aloud technique consisted in encouraging subjects' comments during their query composition tasks, while making appropriate annotations. Upon task completion, the subjects were invited to complete a questionnaire, meant to estimate users' satisfaction.

4. Evaluation Results

In this section, we describe results of our experimental study.

Session 1. Resemblance Evaluation

The results of the first session are shown by means of the bar diagrams described from Figure 3 to 9. Results are grouped for relationship, that is, given a topological relationship, the *resemblance scores* of Op1-Op12 for that relationship are put together and represented by a bar chart in order to understand which 2D and 3D representations best fit that relationship. Percentage represented in the bar chars are further divided into two sets:

1. a set of red bars showing which 3D visual composition best fits the relationship considering Op1-Op6, and finally,
2. a set of yellow bars showing which 2D visual composition corresponds to the relationship.

Percentages are calculated by summing, for each task, the resemblance values submitted by all the subjects; the sum is

then normalized, assigning 100% to the overall sum of all the resemblance values assigned by the subjects for all the relationships.

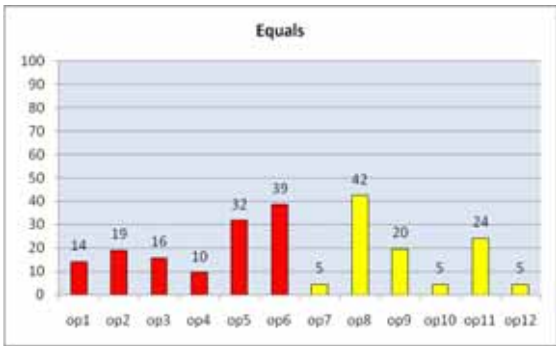


Figure 3. These results indicate the correspondence degree between the Equals task and the operations as depicted in Figures 1 and 2.

As for the *Equals* relationship, the predominant values of the two sets are related to the *Op5* and *Op8* operators, whose values are 32 and 42, respectively. It means that people who arrange 3D icons as shown in *Op5* for representing examples of the *Equals* are only the 32-percent of cases. A more interesting value concerns with the *Op8* operator. However, the resemblance trend of operations in *Op1-Op6* is flatter than the trend of the set composed by operations in *Op7-Op12*. From a mathematical point of view, it highlights the fact that the standard deviation of the first set (8) is lower than the standard deviation of the second set (15), namely, that the *Equals* operation may be easily confused by any operation belonging to the first set, while just with *Op8* and *Op11* of the second set.

In order to measure the correlation between three-dimensional and bi-dimensional visual arrangements we used the index of correlation of Pearson

The Pearson's correlation index of two statistic variables X and Y is defined as the ratio between the covariance and the product of X, Y standard deviations.

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

where the numerator is the covariance, while the denominator is the product of the standard deviations.

The *r* value may range between -1 and +1.

For the value equals to 1 the direct correlation is maximum, for the value equals to-1 is maximum the inverse correlation and for values equals to 0 data are not correlated.

As for the *Equals* relationship, the correlation index between bi-dimensional visual representations and three-dimensional visual representations is 0.59.

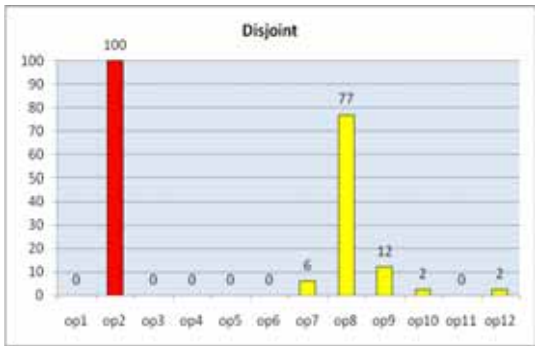


Figure 4. These results indicate the correspondence degree between the Disjoint task and the operations as depict in Figures 1 and 2.

Results concerning with the *Disjoint* relationship as summarized in Figure 4. In this case, 2D and 3D representations are very similar, in fact, two important values, 100 and 77, may be easily identified by the *Op2* and *Op8* operations. With regards to the 2D representation, *Op2* is the only operation which may be associated with the *Disjoint* relationship, it means that this kind of arrangement will be always interpreted in the same way. As for 3D arrangements, the *Disjoint* relationship is not fully associated with just one operation, and some other associations have been quite relevant, as an example the Op9 operation will be indicated as the *Disjoint* relationship by the 12-percent of users.

In this case, the correlation index of Pearson is higher with respect to the Equals relationship. This is to say, 0.99.

The diagram shown in Figure 5 highlights that Op3 and Op9 are the most important operations associated with the *Touches* relationship.

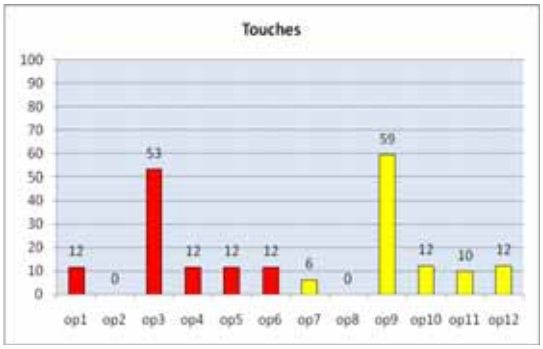


Figure 5. These results indicate the correspondence degree between the Touches task and the operations as depict in Figures 1 and 2.

In particular, *Op3* is the most relevant among the 3D arrangement while *Op9* is the most relevant for the 2D. Other significant values exist for both the arrangements but no one can be effectively compared with the diagram picks. Anyway, the sum of choices different from the picks is close 40-percent, namely four person of ten make a different choice with respect to the principal.

The Pearson's index value, in this case, is 0.99, similar to the Disjoint relationship.

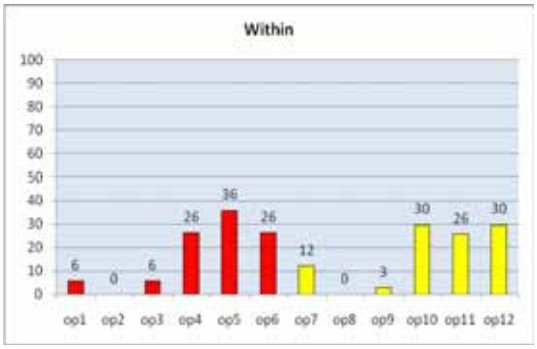


Figure 6. These results indicate the correspondence degree between the Within task and the operations as depict in Figures 1 and 2.

A very dispersed distribution exist for the *Within* relationship. As a matter of fact, three associations may be considered as correct interpretations for each group, namely *Op4*, *Op5* and *Op6* for the 3D arrangements and *Op10*, *Op11* and *Op12* for the 2D. It means that both the visual representations are not able to provide with an arrangement for univocally determining this kind of relationship.

By measuring the correlation between 3D and 2D arrangements also in this case we have found out that corresponding operation are very similar. It causes that the Pearson index is about 0.91.

A very interesting result concerns with the *Crosses* relationship. In fact, in this case, the 3D visual representation allows users to better univocally identify a specific operation than the 2D iconic representation. As a matter of fact, 3D arrangements has a pick for the *Op1* operation whereas all the 2D operations are very close to their average (17). It is even highlighted by the fact that the standard deviation of 2D is just 3.6 whereas 16.63 is for 3D.

Anyway, even though there are so many differences between the two visual arrangements, the Pearson's index is high enough, that is 0.61, which implies a sufficient correlation.

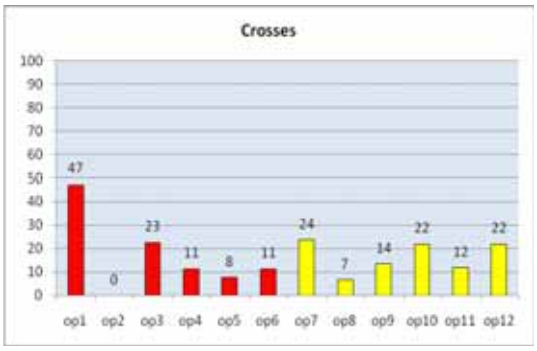


Figure 7. These results indicate the correspondence degree between the Crosses task and the operations as depict in Figures 1 and 2.

The *Contains* relationship is interpreted similarly to the *Within* relationship, probably because the first one is specular with respect to the second one, so the same discussions are valid. The same arrangement distribution also implies a similar Pearson index, namely 0.97.

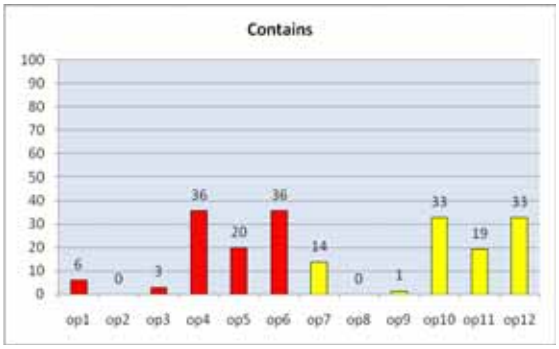


Figure 8. These results indicate the correspondence degree between the Contains task and the operations as depict in Figures 1 and 2.

As for the *Intersect* relationship, it is possible to note that more or less all the operations except *Op2* and *Op8* are quite close. It means that there is no one operation which can be used to represent the meaning of the *Intersect* relationship. However, this result is quite obvious because for definition the *Intersect* is the opposite of the *Disjoint* relationship which has been univocally identified by the *Op2* and *Op8* operations and now are set to 0.

Finally, we measured the correlation between these pair of arrangements. As with previous cases, the mental model of users does not recognized particular differences between 2dD and 3D, and therefore, the index of Pearson is higher enough, namely 0.87.

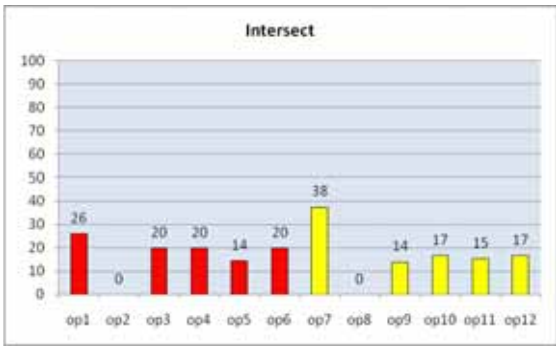


Figure 9. These results indicate the correspondence degree between the Equals task and the operations as depict in Figures 1 and 2.

Session 2.User's Satisfaction

In order to estimate the overall user satisfaction with the 3D composition approach, a think-aloud technique was adopted. During task performance, subjects were observed by an evaluator, who encouraged comments and took annotations. Moreover, upon completion of all that tasks, each subject was asked to fill-up a questionnaire, which aimed at measuring the degree of preference of one approach with respect to the other.

For the sake of brevity we do not report the details of this part of the experiment. However, the collected results revealed that user's degree of satisfaction is positively influenced by adoption of a 3D composition approach. Since user's satisfaction is to be considered a crucial usability factor, together with efficiency and efficacy, we may conclude that the adoption of three-dimensional icons in the composition of spatial queries better supports user's understanding and deployment of the given system.

5. Conclusion

In this paper, we presented a comparison between 2D and 3D icon visual arrangements for representing topological relationships as defined by the OpenGeospatial Consortium in [7]. The result of this experiment has proved that there are no substantial differences in the subject mental model for expressing this kind of relationships. Of seven comparisons just two may be considered partially correlated because their Pearson's correlation index are 0.59 and 0.61, the other ones are higher than 0.90, which implies a very strong correlation. In the following table, we summarize the result correlation values for each relationship.

The final row shows the mean of the correlation indexes. This value gives a summarizing value which highlight the correspondence between the two visual arrangements.

Table 1 Pearson's correlation indexes

Relationship	Correlation Index of Pearson
Equals	0.59
Disjoint	0.99
Touches	0.99
Within	0.91
Crosses	0.61
Contains	0.97
Intersect	0.87
Mean	0.85

6. REFERENCES

1. D. Calcinelli, M. Mainguenaud, Cigales, a Visual Query Language for a Geographical Information System: the User Interface, Journal of Visual Languages and Computing, 5(2) (1994) 113-132.
2. A. Del Bimbo, M. Campanai, and P. Nesi, "A Three-Dimensional Iconic Environment for Image Databases Querying," IEEE Trans. Soft. Eng., vol. 19(10), 1993.
3. M. Egenhofer and J. Herring Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases Technical Report, Department of Surveying Engineering, University of Maine, 1990.
4. V.N. Gudivada, "Spatial Knowledge Representation and Retreival in 3-D Image Databases," Proc. Int'l Conf. Multimedia and Computing Systems, May 1995.
5. J. Hildebrandt and K. Tang, "A Two and Three Dimensional Ship Database Application," Proc. Workshop Spatial Reasoning, Aug. 1993.
6. Y.C. Lee, F.L. Chin, An Iconic Query Language for Topological Relationships in GIS, International Journal of GIS 9(1) (1995) 25-46.
7. OpenGIS, Simple Features Specification for SQL. Version 1.1.0 part I, part II, 2005, Open GIS Consortium Inc. http://www.opengeospatial.org/specs/?page=specs.
8. L. Paolino, F. Pittarello, M. Sebillo, G. Tortora, G. Vitiello, WebMGISQL - A 3D Visual Environment for GIS Querying, in: Proceedings of International Conference on Visual Languages and Computing (VLC'03), Miami, Florida, USA, 2003, 294-299.
9. M. Sebillo, G. Tortora, G. Vitiello, The Metaphor GIS Query Language, Journal of Visual Languages and Computing 11(4) (2000) 439-454.
10. M. Zloof. *Query by Example*. In Proceedings of American Federation of Information Processing Societies AFIPS, 44, 1975.

A Visual Application Generator for the Chronobot

Shi-Kuo Chang, Mehmud Abliz, Yaw Gyamfi and Chang Liu
Department of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260 USA
{chang, anti, mehmud, changliu}@cs.pitt.edu

Abstract: *We describe the design of a visual application generator, which creates customized interfaces that can be easily modified and deployed. With this visual application generator, the user can specify the requirements using an easy-to-understand iconic language. Unlike most other visual programming languages the icons are allowed to have multiple meanings. This unique feature makes our tool easy to use by the end user. The visual application generator for the chronobot is described in detail, including the customized syntactic/semantic rules to translate visual sentences into the application program.*

1. Introduction

In this paper we describe a visual application generator designed based on the visual language theory. There are several advantages: First, using a visual language to describe the problem for the application generator is more concrete, direct, and convenient than using text. Second, users can express their ideas more easily and freely in the 2D space available. Third, our tool provides a reconfigurable user interface which makes full use of commercially available code generation toolkits to provide a general purpose visual application generator. Fourth, the tool enhances flexibility by supporting dynamic addition/removal of icons (the building blocks of visual sentences) while using the same parser in the core. Last but not least, the icons can have multiple meanings, which are automatically disambiguated by the parser so that the tool is easy to use by the end user.

We try to make this tool a general purpose visual application generator that can be customized and retargeted for different application domains. However, the target application of this tool is the chronobot project.

A chronobot is a device for storing and borrowing time [4]. Using the chronobot one can borrow time from someone and return time to the same person or someone else. It is a convenient device for managing time. The underlying premise of the chronobot is that there is a way to exchange time and knowledge. For example one spends time to acquire knowledge and later uses this knowledge to save time. A group of people can also find some means to exchange time and knowledge. Thus the chronobot is a device to facilitate the exchange and management of time and knowledge for the benefits of a community.

The chronobot was implemented as a web service and applied to e-learning (for learners to exchange tutoring time and course-related knowledge), job placement (for job seekers to find temporary/permanent jobs and exchange job seeking knowledge) and senior citizen care (for health care providers and family members to exchange health care service and knowledge). For the quick deployment of the chronobot service it is desirable to have a chronobot application program generator. An application generator is a software program that generates application programs from intuitive descriptions of the problems to be solved, rather than by traditional programming methods. It enables users to quickly create a program of their own with minimal effort and little programming knowledge. With an application generator, a user only needs to specify the steps required for his or her program, with very little or even no codes to be written. In what follows we describe the user interface (Section 2), the visual language (Section 3), the visual application

generator (Section 4) and an experimental system (Section 5) in detail.

2. The User Interface

The visual application generator's user interface is shown in Figure 1. It provides the users with a panel and buttons with icons. Using this user can compose iconic sentences. The iconic sentence is composed in a 2D space using *Hor*, *Ver*, and *Overlay* operators on two icons or two sentences or on icon and sentence. Since there is no grid to restrict the programmer, he/she is free to place icons anywhere on the screen. The onus is on the programmer/user to specify the groupings. For instance, if the user wants to write *icon1 hor icon2*, the user must utilize the hor-group operator provided in the editor. The operators are color coded for convenience: the horizontal operator is colored in blue, the vertical operator in yellow and the overlay operator in red.

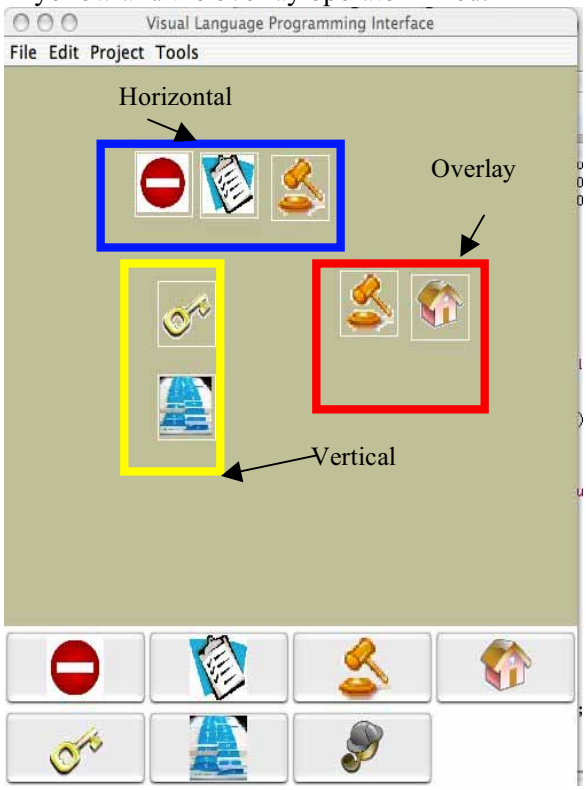


Figure 1. The user interface.

After the user has finished composing the iconic sentences, the parser parses the icons (tokens) to check for syntax and semantics. It returns the groups of icons that form distinct sentences or the icon that is responsible for an error if any. The tokens are the output from the visual editor. Embedded in the visual sentence parser is a grammar definition for the iconic language. The visual sentence is parsed by the parser, which creates a simple parse tree and the code generator takes on from there.

3. The Visual Language

A visual language is a set of visual sentences. Each visual sentence is a spatial composition of object icons and/or operation icons. It represents a complex conceptual entity or a sequence of operations. An icon I is represented in the following form: $I = (IM, PM)$, where IM and PM are the physical image and the set of semantic meanings or the attributes of icon I respectively. Each icon can have a single physical image and multiple attributes. If the corresponding attributes of two icons are matching based on some definite operations, a new meaning will be generated as the result of combination. Thus, depending on their attributes, several icons could be combined by operators to generate a visual sentence.

The general domain-independent spatial icon operators are Ver (for vertical composition), Hor (for horizontal composition), Ovl (for overlay composition) and Con (for conceptual composition). The first three are implicitly represented in the spatial relation between the two icons, while the last one is explicitly represented as a connected line and often used to connect two concepts.

In our work we replace the last conceptual composition operator with a group operator Grp to represent the group composition. The usage of this Grp operation is similar as the usage of bracket in mathematical expressions. If we specify that some icons are in one group, these icons will be parsed first. Thus, the group operation allows us to make several icons into a group with high priority. This group will be regarded as a new entity in the rest of the parsing process.

The icons may have multiple meanings. Such ambiguous icons are disambiguated when the icons are combined with other icons by the

operators. For example the icon "hammer" means either "bid" or "place a bid" depending on its position. If the icon "hammer" is used as the first operand of the vertical, overlay or horizontal operator its meaning is "bid". If it is used as the second operand its meaning is to "place a bid". In general icons can be disambiguated depending on what other icons are present, what operators are applied to them, and their positions as operands. The user can freely enter a visual sentence and the parser will generate the translated sentence and present the sentence, or the code, to the user for verification purpose. After a while the user will gain an intuitive understanding on how to use the icons to compose visual sentences.

3.1. Syntax

In order to design the visual interface of the application generator we define a context-free grammar, which could be customized for different applicant areas such as Chronobot, Chinese characters, etc. The grammar is defined as: $G = (N, X, OP, s, R)$

where N is the set of non-terminals, X is the set of the terminals (or icons), OP is the set of spatial relation operators, s is the start symbol, and R is the set of production rules [A].

Based on the above definition, we further define a context-free grammar. To describe the grammar directly, we used the pair of bracket symbols *[...]* to represent the *Grp* operation. We also defined the priorities for the four operations in decreasing order: *Grp* > *Ovl* > *Ver* > *Hor*.

N (non-terminal set) := {Sentence, Term, Factor, Group}

X (terminal set) := {icon}

OP (operator set) := {Ver, Hor, Ovl, [ ]}

s (start symbol) := Sentence

R (production rules set) := {

Sentence := Sentence Hor Term | Term

Term := Term Ver Factor | Factor

Factor := icon Ovl Group | Group

Group := icon | [Sentence]

}

In this grammar, we could group the icons in sequence together by using the operation *[ ]* first, and then we could take this group as a new icon which can be arranged with other icons to build up valid sentences with any complexity.

3.2. Semantics

We introduce the semantic rules for the above grammar in two parts. Firstly, the general attributes of terminals and non-terminals are described. Secondly, the semantic rules for computation are described based on the definitions of attributes. To simplify the descriptions, the terminals and non-terminals will be both called as *objects*.

3.2.1. Attributes of terminals and non-terminals

There are two types of attributes in this grammar: unitary ones and binary ones.

A unitary attribute represents the meaning of current *object* when it is used alone without any syntactic operation. At most time, it represents the direct meaning, such as "*what it looks like"* and etc. Thus, we used *.meaning* as the unitary attribute.

The binary attribute is to represent the meaning of current *object* when it is in a syntactic operation. Since each syntactic operator has only two operands, the binary attribute of an *object* must be able to distinguish the different locations where the *object* is in. For each *object*, it has two binary attributes op_1 and op_2 related to one syntactic operation. op_i means the *object* is the *i-th* operand in the current operation *op*.

3.2.2. Semantic rules

We are interested in three types of semantic operators: likeness, location and conceptual joint. Likeness means that what the icon looks like, which is related only to a single icon. Location means that the different locations of icons represent different meanings. Location operator is corresponding to the overlay operation of two icons. Conceptual joint operation means that linear combination of two icons represents the concept. This operation is represented by vertical or horizontal operations.

The general descriptions for the above three semantic operations are as listed in the Table 1:

Semantic Operation	Likeness	Location	Concept Joint
General Description	Get the direct meaning of current *object*	1. get location of I_1, then can get the attribute *.ovl1* 2. get location of I_2, then can get the attribute *.ovl2* 3. new attributes = I_1.*ovl 1* + I_2.*ovl 2*	1. get location of I_1, then get the attribute *.hor1/.ver1* 2. get location of I_1, then get the attribute *.hor2/.ver2* 3. new attributes = I_1.*hor1/ver1* + I_2.*hor2/ver2*

Table 1. General Description of Semantic Operations.

Thus, we define the semantic rules as in Table 2, the symbol *.attributes* means the set of all attributes to the current *object*.

Seq.	Syntactic rules	Semantic rules
1	Sentence := Sentence *Hor* Term	Create a new *object* R; Sentence.*attributes* = Sentence.*hor1* + Term. *.hor2*; Return R;
2	Sentence := Term	Create a new object R; Sentence.*attributes* = Term.*attributes;* Return R;
3	Term := Term *Ver* Factor	Create a new object R; Term.*attribues* = Term.*ver_1* + Factor.*ver_2* Return R;
4	Term := Factor	Create a new object R; Term.*attribues* = Factor.*attribues* Return R;
5	Factor := Group *Ovl* Group	Create a new object R; Factor.*attribues* = Group.*ovl_1* + Group.*ovl_2* Return R;
6	Factor := Group	Create a new object R; Factor.*attribues* = Group.*attribues* Return R;
7	Group := icon	Create a new object R; Group.*attribues* = *icon.attribues* Return R;
8	Group := [Sentence]	Create a new object R; Group.*attribues* = *Sentence.attribues* Return R;

Table 2. Semantic Rules.

4. The Chronobot Application Generator

In this section we will take the Chronobot as an example to illustrate how the grammar and semantic rules work for a particular application. These rules are generated from the grammar described in Section 3, with some special characteristic for the Chronobot application. In other words, the grammar we defined above is general to many applications. But for each specific application, we can customize the icon definitions and semantic rules.

4.1. The Visual Language

4.1.1. Syntax

In the Chronobot application, there are two types of functionalities. The simple one describes only one user action, while the other one describes more than one user actions. In other words, the second functionality, we call it complex action, is composed of several simple functionalities. Thus, we will describe the grammar for complex functionalities based on the description of simple ones.

We are interested in three major functionalities of Chronobot application: add function (which adds something into others), list function (lists something required by users) and delete function (delete or remove some existing entity). We are also interested in the functionalities which could be represented by only one icon in the visual language.

Derived from the generic grammar shown in Section 3, we defined a specific grammar for the Chronobot application as follows. The valid

visual sentence should be simple or complex functionalities. We will use the symbol CSimple to represent the simple ones, and CComplex for the complex ones. Also, the terms as *Add_X* in the following grammar represent the add functionality, and the same for *List_X*, *Delete_X* and etc. The operations are the same as them in the generic grammar. In order to differentiate from the generic definition, we use the terms begin with the capital letter C for the Chronobot application. We divided the grammar into three parts:

(i) Description for valid visual sentence:
CSentence := CSimple | CComplex

(ii) Description for Simple functionalities:
CSimple := Add_X | List_X | Delete_X | X
Add_X:= Add_icon Hor X
List_X := List_icon Hor X
Delete_X := Delete_icon Hor X
X := CIcon | CIcon Hor X

Here, the term *Add_icon, List_icon* and *Delete_icon* represent the icons corresponding to the actions as their names shown.

(iii) Description for Complex functionalities:
CComplex := CSimple Ver CFactor
CFactor := CSimple
CFactor := CSimple Ovl CGroup
CFactor := CGroup
CGroup := [CSentence]
CGroup := CSimple

4.1.2. Semantics

Before we describe the semantic rules for the grammar defined in section 4.1.1, we will first introduce two types of predicates. There are two similar predicates in each type. They are used to detect if the current object in the grammar rule has the corresponding attributes. The predicates are:

Overlay Predicates is to detect if the current object *cobj* has the attributes *ovl_1* or *ovl_2*. If it has, then the predicate will return a true value, otherwise it will return false.

Vertical Predicates: is to detect if the current object *cobj* has the attributes *ver_1* or *ver_2*. If it has, then the predicate will return a true value, otherwise it will return false.

Based on the above two types of predicates, we will define the semantic rules for the grammar in Section 4.1.1 as shown in Table 3. We also divide them into three parts as before.: (i) description for valid visual sentence such as Rule 1 below, (ii) description for simple functionalities such as Rule 3 below, and (iii) description for complex functionalities such as Rule 12 below. The following table gives details for the three examples of semantic rules.

Seq.	Syntactic rules	Semantic rules
1	CSentence := CSimple	Create a new *object* R for sentence; CSentence.*attributes* = CSimple.*attributes* Return R;
3	CSimple := Add_X	Create a new object R for CSimple; CSimple.*attributes* = Add_X.*attributes;* Return R;
12	CComplex := CSimple *Ver* CFactor	Create a new object R for CComplex; If ((Ver_1(CSimple) == True) && (Ver_2(CFactor) == True)) then CComplex.*attribues* = *CSimple.ver_1* + *C*Factor.*ver_2* Return R;

Table 3. Three examples of semantic rules.

4.2. An Example

To demonstrate the expressiveness of the grammar, we give a specific example, which is related to the Chronobot application [C]. There are three related icons in this example:
Icon_room  , Icon_list  , Icon_hammer  .

The semantic attributes for each icon is as following (the symbol "/" means that there is no corresponding semantic meaning for the attribute by far, but may be used for error detection in the future):

Icon_room		Icon_list		Icon_hammer	
{		{		{	
*Hor*1 :		*Hor*1 :	List	*Hor*1 :	Bid
*Hor*2 :		*Hor*2 :	/	*Hor*2 :	Place a Bid
	Room				
*Ver*1 :		*Ver*1 :	/	*Ver*1 :	Bid
*Ver*2 :		*Ver*2 :	/	*Ver*2 :	Place a Bid
*Ovl*1 :		*Ovl*1 :	/	*Ovl*1 :	Bid
*Ovl*2 :		*Ovl*2 :	/	*Ovl*2 :	Place a bid
Meaning		*Meaning*	/	*Meaning*	/
}		}		}	

Table 4. Semantic attributes for the icons.

Thus, the syntactic structure for the generated sentence "List all the bid rooms I placed a bid" could be represented as:

(Icon_list ***Hor*** *Icon_hammer* ***Hor*** *Icon_room)* ***Ovl*** *Icon_hammer*.

Figure 2 shows the parse tree for this syntactic structure, where each node represents a terminal or non-terminal and each arc represents an operation between nodes.

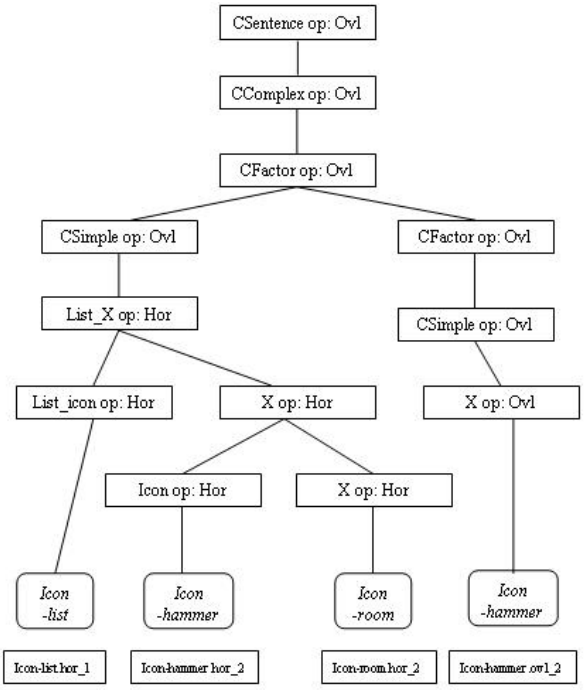


Figure 2. Syntactic Structure for the Example Sentence.

The above parse tree is what will be generated by the VLParser. A subsequent traversal of the parse tree will relay the results of the parser to the code generator. In the case of the previous example, the output from the parse tree traversal will be:

Icon-list.hor_1 Icon-hammer_hor_2 Icon-room.hor_2 Icon-hammer.ovl_2.

Invoking the names of the corresponding attributes, the sentence will be read as ***list bid_room room if bid=true***. This sentence can be read as list all bid rooms for which the condition place bid is true. The visual sentence is shown in Figure 3.

Figure 3 shows a sample screen dump. Icons within the same semantic unit are grouped under one color. Currently it is the user responsibility to specify the groups. This was found to be the easiest way of developing the interface since parsing is done in multiple directions. The only other alternative was to have the system try all acceptable ordering, which would obviously not be efficient if the number of icons in the sentence are beyond a certain number and the sentence employs multiple operators. For instance, the coding of the above example will be as follows: [1 HOR 2 HOR 3] OVL [2] where '1' refers to the first icon "list", '2' the second icon "hammer" and '3' the third icon "room".

Figure 3. Visual sentence (upper box) and screen dump for code (lower left box).

5. An Experimental System

The experimental system comprises the visual language editor, the parser and the application generator. The experimental system serves as a proof-of-concept prototype. The visual editor takes in the user/programmer's input, converts it to a string of tokens which is then passed to the parser and subsequently the code generator. In our example, the code generator is a web-page generator that builds dynamic pages with "jsp."

In this section we describe the screen dumps of the examples shown in Section 4. The first example is *list bid rooms*. It consists of three icons: *icon_list*, *icon_hammer* and *icon_room* grouped by the 'horizontal' operator. The second example is *bid and remove bid*. This example uses two distinct icons; *icon_hammer* (for bid) and *icon_delete* (for delete). The sentence for this example is *icon_hammer ver [icon_delete hor icon_hammer]*, where the 'ver' operator stands for conjunction. To form this sentence with the visual editor, the user/programmer selects the appropriate icons, arranges them in the correct order and envelops them in the required operator from the 'Tools' menu. In this case *icon_delete* and *icon_hammer* will be placed in a 'Horizontal' operator which for the following group: *[icon_delete hor icon_hammer] (1)*. The next step is to enclose (1) and the stand-alone *icon_hammer* in the 'Vertical' operator, signifying a conjunction, resulting in *[icon_hammer ver [icon_delete hor icon_hammer]]*.

The following figures show the Visual editor's screen dumps. Figure 4(a) shows the screen dump for the process of creating the icon groups for *[icon_list hor icon_hammer hor icon_room]* while Figure 4(b) shows the screen dump for *[[icon_hammer] ver [icon_delete hor icon_hammer]]*. In the code window superimposed on the editor, the resulting token string is represented as a combination of numbers. These numbers serve as indices to the icon definition structure array for the parser. This was found to be the easiest implementation. The parser uses the index to retrieve the corresponding icon definition from the array.

Figure 4(a)

Figure 4(b)

The parser and the code generator work in the background. After the visual coding phase, the result (in the superimposed code window) is fed to the parser which produces an output of the form: *bid and delete bid*, for the sentence in Figure 4(b).

6. Discussion and Future Work

There are many visual programming approaches [2, 3, 8]. Our work is more closely related to the following three: Chimera, which is a visual programming system for the end user based upon the concept of program-by-examples [6]; Viper, which is a general-purpose visual

programming environment based on the data-flow model with a fully programmable interpretive command language to define the flow graph operators [1]; and Hi-Visual, which is a iconic programming system based upon icons [5]. The major difference between these approaches and ours is that in our approach the same combination of icons can have several different meanings, and the meanings are determined by the parser. In other words, in our approach ambiguity in the visual expression is regarded as a virtue rather than a vice. It simplifies the user's task of specifying the application program. In this regard our work is influenced by Minspeak, which is a visual language for augmentative communication using multiple-meaning icons [7]. The difference is that Minspeak is a linear visual language while our visual language is two-dimensional.

When new functions are to be added to an application, the user can enrich the icons by assigning them new meanings. For example the icon "hammer" together with the icon "user" can be assigned a new meaning "select a bid". The user interface can be extended to allow user's enrichment of icons. The user can assign multiple meanings to an icon in the enrichment process. However at the end such ambiguity will have to be resolved by the user and a well-defined application program is generated. This unique feature makes our tool easy to use by the end user.

An open research topic is whether the user can be assisted in the systematic introduction of new meanings to an icon, for example the first operand always signifying one thing, and the second operand signifying another. The explicit specification of such rules will lead to a "universal icon algebra". However one can also argue against such an approach because it will ultimately restrict the creative use of ambiguity.

References:

1. Massimo Bernini and Mauro Mosconi, "VIPERS: a Data Flow Visual Programming Environment based on the Tcl Language, *Proceedings of the workshop on Advanced visual interfaces, 1994.*

2. Margaret M. Burnett, "Visual Programming", *Encyclopedia of Electrical and Electronics Erngineering pp 275-283 , 1999.*

3. Margaret M. Burnett, Maria J. Baker, Carisa Bohus, Paul Carison, Sherry Yang, Pieter Van Zee, "Scaling up visual programming languages", *IEEE Volume 28 Number 3: March 1995, pp. 45-54.*

4. Shi-Kuo Chang "A Chronobot for Time and Knowledge Exchange in E-Learning", Journal of Distance Education Technologies JDET, 2005.

5. Masahito Hirakawa, Minoru Tanaka, Tadao Ichikawa , "An Iconic Programming System, HI-VISUAL," *IEEE Transactions on Software Engineering Vol. 16 No. 10 Oct. 1990.*

6. David Kurlander, "Chimera: Example-Based Graphical Editing. In Allen Cypher, editor, Watch What I Do: Programming by Demonstration," *270-290. MIT Press, Cambridge, MA*, 1993.

7. *http://www.minspeak.com,* 2006.

8. N. C. Shu, "Visual programming: Perspectives and approaches*," IBM Systems Journal, 28(4):525–547*, 1989.

Co-Evolution of Users and Interactive Systems in the Web

Maria Francesca Costabile[1], Rosa Lanzilotti[1], Andrea Marcante[2],
Piero Mussio[2], Loredana Parasiliti Provenza[2], Antonio Piccinno[1]
[1]*Università di Bari, Dip.Informatica, via Orabona 4*
70125 Bari, Italy, {costabile, lanzilotti, piccinno}@di.uniba.it
[2]*Università di Milano, DICO, via Comelico 39/41*
20135 Milano, Italy, {marcante, mussio, parasiliti}@dico.unimi.it

Abstract

This paper addresses the issue of co-evolution of users and systems in the web era and discusses an approach to designing web applications that support this co-evolution. The approach is illustrated through the design of a web portal for accessing, managing and annotating documents about the Valchiavenna territorial district, in Northern Italy. Users may perform End-User Development activities, so that the collaborative creation of knowledge and tailoring of tools to manage this knowledge are possible.

1. Introduction

It is widely acknowledged that during the usage of a system, both users and systems evolve [1], [2], [5], [11]. Nielsen says that "using the system changes the users, and as they change they will use the system in new ways" [11]. For Norman "the individual is a moving target" [12]. As a consequence, a design of an interactive system may be good today but no longer appropriate tomorrow. Once people gain proficiency in system usage, they would like to use the system in different ways and need different interfaces than those they required when they were novice users. The system must also evolve to cope with these new needs, thus *co-evolution of users and systems* occurs. The traditional approach is that software engineers take care of system evolution, but this takes a lot of time and other resources. Our goal is to allow users to modify and evolve the system according to their changing needs and to the task to be performed.

We have proposed a model of Interaction and Co-Evolution processes (ICE model) that is used as a unifying framework to help to identify the causes of interaction difficulties affecting software systems [5]. This model is at the base of an approach to designing novel interactive systems capable of supporting people activities and fostering users and systems co-evolution, leading to a profound re-examination of the way interactive systems are designed. Software engineers and Human-Computer Interaction (HCI) experts have the possibility of collaborating with representatives of users to evolve the system.

End-User Development (EUD) is a new research area that aims at empowering people to create and/or modify software artefacts [4] [15]. In order to support the co-evolution of users and systems, we should develop systems that permit EUD activities (called EUD systems hereafter), so that the collaborative creation of knowledge and tailoring of tools to access and manage this knowledge are possible.

In this paper, we focus on EUD systems accessible via the web. The web was actually developed by Berners-Lee to provide researchers with an environment through which they could collaborate. But those researchers were physicists and were familiar with computers, so they were able to understand the same system language. Nowadays, web systems are used by very different users, but with the same wish to collaborate among themselves to overcome spatial and temporal distances. Even in the same application domain, the users are very different due to various reasons: a) different culture, skills, specific abilities (physical and/or cognitive) and tasks; b) different roles in performing work activities; c) different context of activity and geographical dispersion.

Because of the diversity of users, the challenge is to ensure the universal access and universal usability of interactive systems. The slogan "one fits all" cannot be applied to the user interface since it is well known that different users experience many difficulties when they interact with an interface not tailored to them. In the last few years, we have developed a design methodology that addresses this challenge by providing different communities of users with software environments that they may access and manipulate by exploiting their own system of signs [6].

In this paper we show how we applied this methodology to the design of the "Valchiavenna" portal. This is a web portal for accessing, managing and annotating documents pertaining to the territorial district of Valchiavenna (Chiavenna valley), in Northern Italy. The documents refer to tourism,

history, monuments, etc. They can be manipulated by users through various tools available in the portal.

The paper is organized as follows. Section 2 presents the web portal. Section 3 briefly reports the model of the Interaction and Co-Evolution processes (ICE model) occurring between users and interactive systems. Section 4 briefly describes the portal design and shows how co-evolution of users and systems is supported. Finally, Section 5 concludes the paper.

2. Annotating documents through the web

We describe herein the prototype of a web portal for accessing, managing and annotating documents concerning the territorial district of Valchiavenna. The portal provides all tools needed for retrieval, management and updating, through electronic annotations, of historical and tourist documents. An electronic annotation is a multimedia note that provides a comment or an explanation to a document or to a part of it [9]. The Valchiavenna portal implements an annotation tool to support single or cooperative reasoning on document content. Users can manage two types of electronic annotations: *specialized annotations*, whose content is certified by experts in a particular discipline; and *personal annotations,* being the stream of electronic annotations (in general, private) expressing users comments or observations on a specific document. The Valchiavenna portal is designed and developed to permit its evolution in order to satisfy the evolving user needs with respect to the portal behaviour and the knowledge accumulation. The portal is equipped with tools that allow users to enrich the knowledge base managed by the portal. This offers users a great possibility, but there needs to be a guarantee that the knowledge base will remain consistent and socially acceptable [11].

The portal addresses different communities of users:

- *visitors*, whose goal is to access the documents they are interested in and to record their comments about a specific resource of interest through personal annotations, for personal or public use;
- *publishers*, whose task is to control knowledge base updating in order to guarantee the quality of the knowledge publicly available through the portal;
- *experts*, i.e., specialists in different fields like history, geology, etc., whose main task is to generate new contents relative to their own expertise, guaranteeing their consistency and correctness.

Systems like the Valchiavenna portal are designed to support the evolution of knowledge in terms of accumulation and sharing of information (through annotations). Moreover, use of the system changes the way users access and manage information and creates new needs; the system itself has to evolve to face new users needs. As we will see in Section 4, the annotation tool we have developed permits knowledge enrichment, but it also supports evolution of the system, because it allows users to point out problems or needs in the system usage and hence to fix them.

3. The ICE model

In our work with end users, we found that several usability problems depend very much on the rigidity of interactive systems. Users want systems to be able to cope with the changes occurring in their activities and/or in their organizational context and to support the evolution of the knowledge shared in their application domain. The evolution of the organizational context and activities results in the need for new interaction tools and procedures, which must fit the user culture and be consistent, from the user point of view, with the pre-existing tools. The system must be able to support users in steering this co-evolution process. However, the knowledge made available by the system also evolves as a result of the user activities. This evolution often requires the development of new knowledge management tools and/or the re-organization of the knowledge base. Again, the system should support users in steering this knowledge co-evolution process. These requirements dictate a profound revision of the way interactive systems are designed, identifying new theoretical foundations on which new approaches could be based. A model-based approach to the design of usable systems must then consider two processes occurring in a working environment augmented by the support of interactive systems: the human computer interaction process, i.e., the interactive use of the system occurring to perform activities in the application domain; the co-evolution process occurring during the use of an interactive system.

The ICE (Interaction and Co-Evolution) model we propose encompasses both interaction and co-evolution processes, serving to identify the causes of interaction difficulties affecting software systems and to support co-evolution. The interaction process is modelled as a cyclic process in which the user and the interactive system communicate by materializing and interpreting a sequence of messages (the images on the screen in visual interaction) at successive points in time – $i(t_0)$, $i(t_1)$, ..., $i(t_n)$ – (see Figure 1). These messages are subject to two interpretations: one made by the user, depending on her/his role in the task, as well as on her/his culture, experience, and skills, and the second one internal to the system, associating the

image with a computational meaning, as determined by the programs implemented in the system (see Figure 1). The emphasis given in this model to these two interpretations addresses a problem that arises in HCI, that is the *communication gap* between users and designers. The interpretation performed by the system reflects the designers' understanding of the task at hand, implemented in the programs that control the machine. Designers develop the interactive system and focus primarily on the computational and management aspects, rather than on the users' problems, so that the interaction language is often too general and machine-oriented rather than situation and user oriented. In this way, users are forced to break the continuity of their reasoning in order to translate and express their problems and solutions in the computerized language.

Some preliminary aspects of co-evolution were investigated in [10], actually described as co-adaptation of user and system. Preliminary models were presented in [3] and [2]. Carroll and Rosson focused on a task/artefact loop and Bourguin et al. stated more explicitly it by observing that co-evolution also involves organizational contexts and technological changes. More specifically, the task/artefact loop in [3] implicitly referred to the co-evolution process by considering that software artefacts are produced to support some user tasks; however, such artefacts suggest new possible tasks so that, to support these new tasks, new artefacts must be created. This task-artefact cycle is denoted as "cycle 1" in the ICE model in Figure 1, indicating a first co-evolution cycle. Technology advances offer computer scientists new possibilities of improving interactive systems once they are already in use: this leads to new interaction possibilities that might change end users working habits. For example, recently improved voice technology allows software engineers to add voice commands to their interactive systems and this might provide end users with a more easy and natural way to use the system. On the other hand, the user social and organizational context evolves over time, requiring new tasks and/or different ways of performing tasks. Therefore, technology and social and organizational contexts repeatedly affect each other: this is represented in our ICE model by a second co-evolution cycle, denoted as "cycle 2" in Figure 1.

Software engineers are required to produce the tools to support the interaction and co-evolution processes. In other words, they must not only produce interactive systems supporting user activities, but also the tools that permit the system to evolve with user evolution.

4. Designing for Co-Evolution

A methodology for designing novel interactive systems capable of supporting people activities has already been presented in [4]. We illustrate here how it is exploited in the design of the Valchiavenna portal and show its capability of supporting the co-evolution process.

In this methodology, software environments are designed by analogy with artisan workshops. Artisan workshops are small establishments where artisans, such as blacksmiths and joiners, manipulate raw materials in order to manufacture their artefacts. At each step of their activity, traditional artisans can extract from a repository the tools necessary for the current activity and set aside/put back those no longer needed. In this way, each artisan adapts the environment to her/his needs and has available all and only those tools needed in the specific situation. By analogy, a software environment is designed as a *virtual workshop*, in which the end user finds a set of (virtual) tools whose shape, behaviour and management are familiar to her/him. Such an environment allows end users to carry out their activities, adapt the environment and also create new

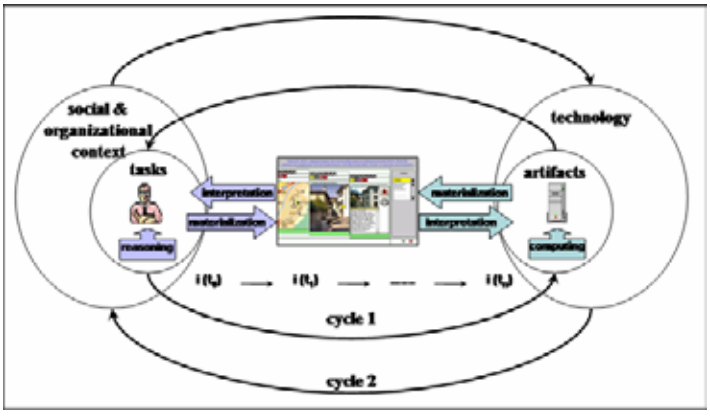


Figure 1. The Interaction and Co-Evolution (ICE) model.

tools without any textual program code, but using direct manipulation activities and high-level visual languages tailored to their needs. Moreover, end users get the feeling of simply manipulating the objects of interest in a way similar to what they might do in the real world. Obviously, while traditional artisans shape real supplies, end users shape software artefacts. For this reason we call these environments *Software Shaping Workshops* (SSWs) [4].

The SSW methodology provides each end user sub-community with a workshop, called an *application workshop*, which supports them in their daily work. An application workshop is customized to users' culture, background and skills, and can possibly be tailored by the users themselves, also by creating new artefacts [4]. Application workshops are not directly created (and successively evolved) by software engineers, but their design, development and modification are carried out, with a participatory approach [14], by a multi-disciplinary team that, besides software engineers, includes domain experts, as representatives of end users, and HCI experts. Each member of the design team uses a particular type of workshop, called *system workshop,* customized to her/his culture, background and skills, in order to carry out the design, development and evolution of other workshops.

According to the SSW methodology an interactive system is not conceived as a monolithic piece of software but as a network of SSWs. In general, a network is organized in levels:

- *Meta-design level* (the top level). Software engineers use a system workshop to create customized workshops to be used by other experts in the design team and to participate themselves in the design, implementation, and validation activities.
- *Design level*. HCI experts and end users (as experts of the application domain) cooperate in design, implementation, and validation activities. A design member belonging to the community X participates in the design using a system workshop created by the software engineers and customized to the needs, culture and skills of community X; the various experts design application workshops and also tailor their own system workshops.
- *Use level*. End users of the different communities cooperate to perform a task in their work practice; end users belonging to the community X participate in task achievement using an application workshop customized to their needs, culture, and skills.

Both meta-design and design levels include all the system workshops that support the design team in performing the activity of participatory design. Such system workshops can be considered User Interface Development Environments (UIDEs) [13]. The novel idea is that these environments are very much oriented to the application domain and have specific functionalities, so that they are easy for experts who are not computer scientists to use. Domain experts and HCI experts perform design activities (creating, evaluating and modifying workshops) through direct manipulation, without needing to know any programming language.

The Valchiavenna portal has been developed according to the SSW methodology as a network of software shaping workshops (see Figure 2). The visitor workshop, W-Visitor, at usage level, is the application workshop for accessing and managing tourism documents from a shared knowledge base, navigating among their linked resources and generating personal annotations. The design level includes system

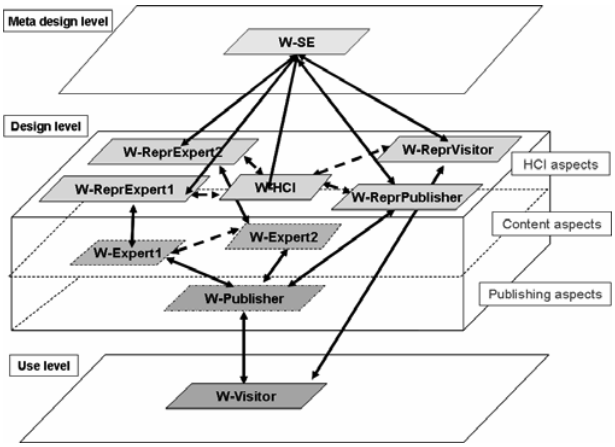


Figure 2. The SSW network for the Valchiavenna portal.

workshops for customizing and adapting the other workshops in accordance with the evolving knowledge and user needs. They are organized on three different layers according to the aspects they consider in the design: at the lower layer, the publisher workshop, W-Publisher, is devoted to content publishing and validation aspects: it permits the publisher to control the knowledge base evolution and to certify the consistency and quality of its contents by validating specialized annotations as socially acceptable; at the middle layer, expert workshops (W-Expert1 and W-Expert2 in Figure 2) are each devoted to specialists in a specific field (geology, history, ecc.), allowing them to access the shared knowledge base and to enrich it by designing and evolving new certified contents; at the top layer, system workshops (W-HCI, W-ReprExpert1, W-ReprExpert2, W-ReprPublisher and W-ReprVisitor), devoted to HCI experts and representatives of end users, permit them to design and evolve the other workshops in the lower part of the network. Note that publisher and expert workshops are system workshops, since they enable not only the addition of new contents but also the design of adequate tools for accessing and using them. Finally, the meta-design level includes the system workshop for software engineers, W-SE, which allows them to generate and maintain all the SSWs in the network.

In order to foster the co-evolution process, communication paths exist among application and system workshops (arrows in Figure 2). In this way, once the overall interactive system (all workshops of a network) is in use, the design team can have feedback of end user activities, new usages of the system and new procedures induced by the evolving organization. The design team can receives end user complaints and suggestions about the workshops they interact with as system annotations, i.e., electronic annotations the user associates with the system (and/or the interface) itself to denote a usability problem or a functionality to be evolved [9].. Based on these annotations, the design team updates (evolves) one or more application workshops, so that they will better fit the new user requirements.

Figure 3 shows a screenshot of W-Visitor: it illustrates how the visitor "Andrea" can access various information regarding the map on the left. The spot on the map indicates that some information is stored in the knowledge base. Around the spot there are triangular widgets that, together with the spot, form the star operator. The visitor accesses this information through the star operator, thus seeing the specialized annotations associated to that point in the map. In Figure 3, the spot indicates the building "Palazzo Pretorio" in the Chiavenna town. By acting on the triangular widgets, the user has visualized the panels in Figure 3 that show two different pictures of the building and a textual description shown in the overlapping window at the bottom left.

The visitor Andrea can also add his own note using the annotation bar at the top of each panel. The yellow emoticon on the building in the second panel from the left indicates that a personal annotation, created by Andrea, has been added. This annotation has been created through an annotation window shown in the centre of Figure 3. Andrea can then decide to submit it to the publisher in order to make it public and available to the whole community.

Figure 4 shows a screenshot of W-Publisher. As we can see, the organization of the window is similar to the application workshop, but the toolbar on top of the left panel is different since it is tailored to a specific community of users. A publisher can enrich the knowledge base by associating specialized annotations with a document and by creating or updating the star

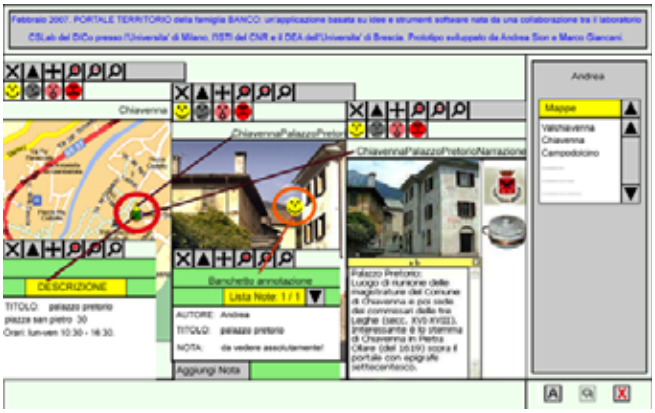


Figure 3. Screenshot of the W-Visitor workshop showing how the user accesses a document, and the associated annotations.

operator to permit access to the new or modified contents. As an example, in Figure 4 the publisher is adding another widget to the existing star operator to provide visitors with new information.

Moreover, to support the co-evolution, visitors and publishers can make a system annotation (by clicking on the button "A" at the bottom right in figures 3 and 4) to require the design team to evolve their own workshop according to their evolved needs.

5. Conclusions

End User Development responds to a paradigm shift in interactive system usage. Nowadays, users need and demand to become knowledge producers rather than just consumers [7] [8]. To be effective, the process of knowledge production and evolution must be tuned to the user culture and not driven by the designer culture. User culture, however, is not a static body, but rapidly evolves under the evolution of ITC technologies, which in turn evolves due to the user culture evolution. Moreover, web systems are accessed by a variety of users sharing some common knowledge and requirements, but belonging to different cultural communities. As a consequence, the way interactive systems are designed needs to be profoundly revised.

This paper addresses these issues, firstly illustrating the ICE model, which frames the HCI process in a co-evolution scenario. Next, the SSW methodology for interactive system design is outlined by describing the design of a web portal. The SSW methodology starts from the ICE model to arrive at the design of interactive systems personalized to the user culture, which are tailorable by the users to better fit their current needs, and capable of supporting the co-evolution process. An interactive system resulting from the application of the SSW methodology consists of a network of environments, the software shaping workshops, which support end users in their activities and permit them to evolve their knowledge but also to steer the evolution of the system structure. To users, the system no longer appears as a static product but as a collection of services which allow them to carry out their activities, create new knowledge and collaborate with designers to achieve a co-evolution of the system.

Figure 4. A screenshot of the W-Publisher workshop.

6. Acknowledgments

This work is supported by the Italian MIUR and by EU and Regione Puglia under grant DIPIS.

7. References

[1] Arondi, S., Baroni, P., Fogli, D., Mussio, P.: "Supporting co-evolution of users and systems by the recognition of Interaction Patterns", *Proc. AVI 2002*, ACM, 2002, pp. 177-189.

[2] Bourguin, G., Derycke, A., Tarby, J.C., "Beyond the Interface: Co-evolution inside Interactive Systems - A Proposal Founded on Activity Theory", *Proc. IHM-HCI 2001*, Springer Verlag, 2001, pp. 297- 310.

[3] Carroll, J.M., Rosson, M.B., "Deliberated Evolution: Stalking the View Matcher in design space", *Human-Computer Interaction*, 6 (3 and 4), 1992, pp. 281-318

[4] Costabile, M.F., Fogli, D., Mussio, P., Piccinno, A., "End-User Development: the Software Shaping Workshop Approach", in Lieberman, H., Paternò, F., Wulf, V. (Eds), *End-User Development*, Springer, The Netherlands, Dordrecht, 2006, pp. 183-205.

[5] Costabile M.F., Fogli D., Marcante A., Piccinno A., "Supporting Interaction and Co-evolution of Users and Systems", *Proc. AVI 2006*, ACM, 2006, pp. 143-150.

[6] De Souza, C.S., Barbosa, S.D.J., "A Semiotic Framing for End-User Development", in H. Lieberman, F. Paternò, V. Wulf (Eds), *End-User Development*, Springer, The Netherlands, Dordrecht, 2006, pp. 401-426.

[7] Fischer, G., Giaccardi, E., Ye, Y., Sutcliffe, A.G., Mehandjiev, N. "Meta-Design: a Manifesto For End user Development", *Communications of the ACM*, 47(9), 2004, pp. 33-37.

[8] Fischer, G., Giaccardi, E., "Meta-Design: A Framework for the Future of End-User Development", in Lieberman, H., Paternò, F., Wulf, V. (Eds), *End-User Development*, Springer, The Netherlands, Dordrecht, 2006, pp. 427-457.

[9] Fogli, D., Fresta, G., Mussio, P., "On Electronic Annotations and its implementation", *Proc. AVI 2004*, ACM, 2004, pp. 98-102.

[10] Mackay, W.E., "Triggers and Barriers to Customizing Software", *Proc. CHI'90*, ACM, 1991, pp.153-160.

[11] Nielsen, J., *Usability Engineering*, Academic Press, San Diego, CA, 1993.

[12] Norman, D. A., "Human-centered design considered harmful", *Interactions*, 12(4), 2005, pp. 14-19.

[13] Preece, J., *Human-Computer Interaction*, Addison-Wesley, Wokingham, UK, 1994.

[14] Schuler, D., Namioka, A. (Eds.), *Participatory Design - Principles and Practices*, Lawrence Erlbaum Associates, Hillsday, NJ, 1993.

[15] Sutcliffe, A., Mehandjiev, N., (guest Eds.), "End-User Development," *Communications of the ACM*, 47(9), pp. 31-32.

A Method for the Development of Web GIS

S. Di Martino[2], F. Ferrucci[2], A. Ginige[1], L. Paolino[2], M. Sebillo[2], G. Vitiello[2], G. Avagliano[2]
1 University of Western Sydney, Australia
a.ginige@uws.edu.au
2 University of Salerno, Via Ponte Don Melillo, I-84084 Fisciano (SA), Italy
{sdimartino, fferrucci, lpaolino, msebillo, gvitiello}@unisa.it

Abstract

The increasing popularity of the Internet has led to the development of Web applications, known as Web GIS, intended for dissemination and manipulation of spatial knowledge. In this paper we present a method for the development of Web GIS based on the WebML approach. In particular, we adapted the WebML development process to this specific domain, by defining the required steps and actors and by extending the visual formalism to capture specific GIS concepts. The resulting method consists of two orthogonal perspectives: Geodata and Metadata Conceptual Model and Hypertext and Web Mapping Model. In particular, the Spatial E-R model by Calkins has been exploited to design the conceptual schema of spatial data. While, the WebML Hypertext Model has been extended to include new notations specifically tailored for GIS concepts and tasks.

1. Introduction

Geographic Information Systems (GIS) are a kind of software system specifically suited to deal with data and attributes spatially referenced [8, 9]. Compared with traditional information systems, GIS provide the unique features to support spatial queries/analyses of the data. They turn out to be fundamental in a great variety of contexts, ranging from urban planning to marketing, route planning, natural disaster prevention/handling, archaeology, and so on. In the past few years, the increasing popularity of the Internet has greatly affected the activity of GIS researchers and laboratories, leading to the development of Web applications, known as *Web GIS*, intended for dissemination and manipulation of spatial knowledge in specific domains. [11].

However, one problem with GIS technology is that developers are often faced with the inherent complexity of data and are forced to become familiar with the related concepts. This is particularly true for Web GIS, since software environments supporting the development of these applications are very hard to use and require specialized skills [10]. Moreover, with the increased time-to-market pressure, it is no longer possible to deal with low-level issues, and create Web GIS applications from scratch. Thus, there is a growing need for tools and methodologies that allow us to rapidly develop this kind of Web applications and to rapidly modify them to meet the ever-changing business needs [6]. Moreover, there is the need of suitable visual models allowing the different actors involved in the development process to focus the attention on specific and relevant aspects of the applications, to improve communication and to foster reuse.

To develop traditional Web applications, many solutions are currently available. Among them, *WebML* (*Web Modeling Language*) is a high-level, formal visual language specifically conceived to design data-intensive Web applications [2, 3, 12]. It provides suitable models, a development process and a CASE tool (*WebRATIO*) [4] to build applications that rely mainly on data management and movements. WebML has been applied in many different contexts, from industry to academia, with positive results. Web GIS can be considered as a particular class of data-intensive Web applications, since they are mainly devoted to handle (spatial) information to and from the user. Thus, in this paper we present an adaptation of the WebML methodology, to deal with the Web GIS context. In particular, we adapted the WebML development process to this specific domain, by defining the required steps and actors and by extending the visual formalism, to include new notations specifically tailored for GIS concepts. As current work we are extending the CASE tool WebRATIO to provide potential Web GIS designer with a Rapid Application Development environment, where he/she can design in a visual fashion the application, that will be automatically generated by the tool.

The remainder of the paper is organized as follows: in Section 2 we illustrate the proposed development process, while in Section 3 we describe the proposed modeling concepts and notations which extend WebML. In Section 4 we present an example of application, while some remarks and future work conclude the paper.

2. A WebML-based development process for Web GIS

WebML is a modeling language suited to support users in designing data-intensive Web applications [2, 3, 12]. To this aim, it provides a set of visual notations to model the structure, data, and navigational aspects of a Web application. Indeed, the design of a Web application is based on two orthogonal perspectives: data and navigation. The former is described in terms of relevant entities and relationships, by means of the E-R model. The latter is modelled by one or more hypertexts which describe content organization within pages, and how they are linked together to constitute a hypertext. WebML is also accompanied by a development process and a CASE tool.

Web GIS applications are a special case of data-intensive Web systems, meant to deal with the complex, expensive, and voluminous geographic data and share them across several users for different business goals. Geographic information is usually distributed across different layers, which a Web GIS user should be able to handle separately or in overlay modality. Thus, besides common Web navigation and composition tasks, far more complex functionalities are needed in Web GIS for visualization and content management. As a matter of fact, with respect to traditional Web applications, this kind of systems require special focus on the following crucial characteristics:

- Spatial data may be acquired from different sources and stored in different formats, for all of which the user should be offered direct support.
- Interface Usability. This should be gained through: a) ease of use, which means little or no training required; b) interaction robustness and efficiency, especially in terms of response time needed by the system to provide the results of a spatial query.

Therefore, the development of Web GIS applications usually involves people with different expertise, ranging from GIS specialists, to Web engineers, to Web interface designers, to interface usability professionals etc. The interaction among those experts as well as their roles and responsibilities should then be specified within a disciplined development process which should take into account the specific characteristics of Web GIS applications. To this aim, we propose a Web GIS development process based on the one of WebML and which relies on an extension of the WebML modelling language, where suitable visual metaphors have been introduced for geographic information management and visualization. In the next subsection we will outline the proposed development process by only describing the phases that are more affected by the specific nature of Web GIS applications. In subsection 2.2 we highlight some actors that play a relevant role in this development process.

2.1 The development phases

The development of Web GIS applications is based on the macro-steps depicted in Figure 1. The activities which are specifically tailored for Web GIS mainly involve requirements specification, geodata and metadata design, hypertext and Web mapping design, user interface design, and Web interface usability evaluation.

For Web GIS development, requirements elicitation is aimed at gaining two critical pieces of information:

- The list of GIS functions that will be needed in the target Web application,
- A master list of geographic data and metadata.

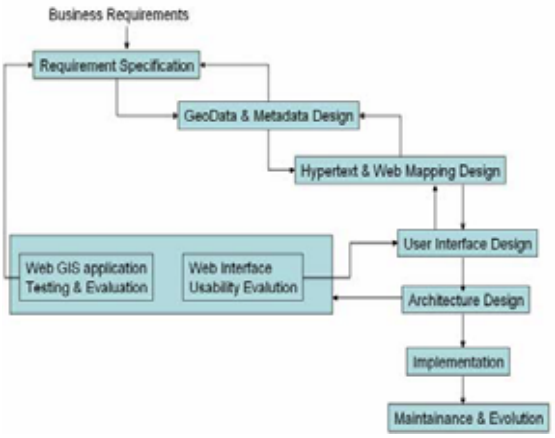


Figure 1: The phases of the development process

In analogy to traditional requirements engineering process, representative subjects of the target user community are interviewed in order to specify a set of GIS application descriptions, a list of important data, and a description of management processes. As a result, the information needed to plan the development of the Web GIS application will be grouped into the following categories:

GIS Functions required - These will include standard operations, such as map navigation, fruition of the associated descriptive data, spatial selection, querying, and spatial analysis.

Data needed in the GIS database – Answering the "where is what?" question represents the main goal of the analysis of the mini-world of interest, which is aimed to provide designers with information about objects and fields featuring the involved domain.

Data maintenance procedures - By looking at the application stakeholders and at typical domain work flow and processes, responsibility for data creation, updates and maintenance will become apparent.

Once requirements are elicited and formally specified, the next step is the design of geodata and metadata conceptual schemas. To this aim, conventional models, such as E-R, have been extended to provide designers with the ability of describing the complex nature of geographic data, where the two components, descriptive and spatial, should be analyzed and managed in a joint manner (e.g. [1, 7]). In our proposed methodology, we suggest the adoption of the Spatial E-R model [1]. The expected output from this geo-data design task consists of a set of logical schemas, which organize data in a relational way and associate them with layers, according to their meaning and the underlying data structure, which can be raster- or vector-model based.

The third step of the proposed methodology extends the hypertext design top-down process, by taking into account composition and navigation requirements specific to Web GIS applications. These entail a number of functionalities meant to dynamically navigate a geographic map, that are commonly known has *Web mapping*. Thus, in the top level coarse design activity, a special focus will be put on the role played by geographic data elements producing a high level specification of the corresponding site views. Typical of this description will be a default map area, whose contents will informally be expressed in terms of entities and relationships of the underlying spatial schema. At detailed design level, composition and navigation will be modelled by exploiting newly defined WebML GIS units besides the set of canonical configurations of pages and units.

User interface design is another critical activity, devoted to guarantee usability of the resulting application, in terms of efficiency, efficacy, and user satisfaction. At this stage, interface designer's expertise is employed to make crucial decisions, aimed at gaining central usability factors, such as ease of use and robustness for each of the designed interaction tasks. When dealing with Web GIS interfaces, usability also results from the attention given by the interface designer to the completeness, relevance and status of available geographic information at a Web page. Then, the Web GIS application is tested and evaluated against the specified user requirements, providing possible refinements. As a matter of fact, requirements as well as functional specifications are hard to identify when starting to develop a Web GIS application, since they may significantly change during the development process. This is related to the intrinsic nature of Web GIS applications, for which requirement engineering has to be considered an evolving activity to be performed throughout the development process. As for the Web interface usability evaluation, both expert evaluators and users will be involved in testing activities, meant to address issues related to the completeness, relevance and status of available geographic information at a Web page, as well as issues more closely related to human-computer interaction and user satisfaction.

As in any iterative development process, the results of testing and evaluation activities are subsequently used for possible refinements.

2.2 Actors

The definition of key actors in a development process allows us to clearly break down the tasks, distributing the responsibility for each of them on the basis of specific skills. This allows us also to modularize and to parallelize the development process. As in the case of the WebML development process different figures can be identified for the proposed development process, namely application analyst, data architect, application architect, graphic designer, tester, etc. The nature of Web GIS applications affect essentially the figures of data architect, application architect and graphic designer. In particular:

- The *Data Architect* is responsible to manage all the aspects about data, thus carrying out the *Geodata and metadata design* task. In particular he/she has to design the geographical database and to make available in suitably modelled repositories both vector and raster layers.
- The *Graphic Designer* is responsible to develop the layouts of the various (sets of) pages composing the Web site. He/she should also deal with usability aspects, not only by guaranteeing a uniform disposition of widgets within the pages, and in general a suitable and consistent *look and feel*, but also by identifying appropriate metaphors for Web mapping. He/she creates and manages a repository of corresponding HTML/XSL templates and performs the *Web interface usability evaluation*, exploiting the results of such

activities to make adequate improvements to the designed interfaces.

- The *Application Architect* is responsible to design the application, by exploiting suitable visual models. To this aim, for each class of intended users he/she also chooses the set of layers (from the previously defined repository) and of functionalities to provide.

3. An extension of WebML for Web GIS

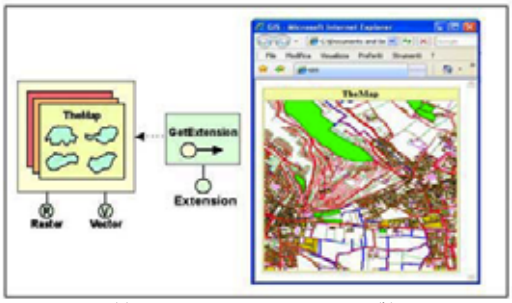
Graphical and diagrammatic representations play a central role in the field of software and Web engineering [5]. They provide powerful aids to describe and understand complex systems, by means of abstractions and different views. Moreover, they help to identify and describe design patterns that can be reused in different applications, thus avoiding to create from scratch. These represent the main motivations for the success of WebML and for our proposal to adapt the WebML approach to develop Web GIS. However, Web GIS present many specific characteristics, making them different from traditional data-intensive Web applications. In particular, from a design point of view, relevant differences are related to the data model and to the navigational model. As for the first aspect, we suggest to adopt the *Spatial E-R* model proposed in [1] to describe spatial entities and relationships. In agreement with this model, simple entities describe spatial "things" which have attributes and are spatially related to each other. Each set of geodata is described as a *spatial entity* characterized by a set of attributes, a geometry and a couple of coordinates. A similar extension has also been defined for relationships to model entities which are spatially related.

As for the navigation model, it is worth noting that besides the common interactions a user performs when navigating and executing a web application, we have to take into account the actions directly performed on a map, when carrying out web mapping or spatial query tasks. Such actions usually produce data, which are not presented in a new web page but are rather rendered onto the same map. Thus, we have extended WebML with new units able to capture the particular characteristics of this kind of applications and to model specific Web GIS interaction tasks, as described in the following.

The *MultiMap* Unit is a visual metaphor for a map viewer, i.e. a graphic component within a Web browser, able to render both vector and raster data arranged in layers. It is characterized by the following properties:

- Name: the name chosen by the *Application Architect* for the *MultiMap* Unit.
- Raster Sources: the reference to raster entities. This source is optional, because a Web GIS application may not require raster layers.
- Raster Selector [optional]: predicate useful to determine which objects belonging to the raster entities share in the unit content.
- Vector Sources: the reference to vector entities. This source is optional, because a Web GIS application may not require vector layers.
- Vector Selector [optional]: predicate useful to determine which objects from the vector entities share the unit content.

Moreover, in order to correctly render source data, the *MultiMap* Unit requires a global parameter, named *Extension*, which contains coordinates of the upper-left and lower-right corners of the presented area. Every time this parameter is modified by some external operations, the area rendered in the *MultiMap* Unit is updated accordingly.

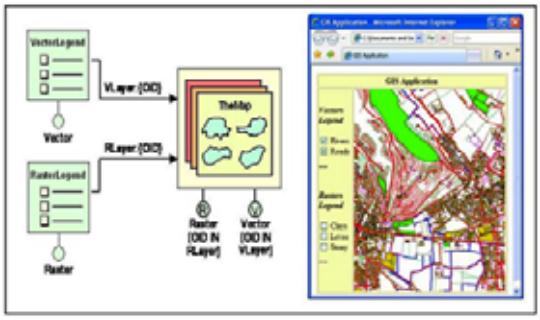


(a) (b)

Figure 2: Graphic notation of an instance of a *MultiMap* Unit (a) and its representation in a Web browser (b)

Figure 2(a) depicts the visual notation of a *MultiMap* Unit instance, namely *TheMap*, as well as the associated global parameter. The resulting representation in a Web browser is shown in Figure 2(b). Here, both raster and vector layers are displayed, as determined by invoking both source entities, while no selectors are set, thus loading all the instances of the underlying data. Differently, when a Web GIS application is expected to allow the activation and display of layers from a Table of Contents on users' demand, a different modelling is required as shown in Figure 3(a). It contains a *MultiMap* Unit to render the layers, and two *Multi-choice Index* Units, to represent the Table of Contents which is made up of two different legends. This scenario also requires selectors associated with the *MultiMap* Unit sources, which are used to select data in agreement with parameters set through the two instances of the *Multi-choice Index* Units of WebML, namely *RasterLegend* and *VectorLegend*. The resulting application on the Web

browser has a checkbox, as shown in Figure 3(b), by which users may activate a layer and visualize the corresponding data on the map.



(a) (b)

Figure 3: Modeling the layer selection associated with a map

For Web GIS, the spatial selection is one of the most common modalities of interaction between the user and the application. It can be used either to select an area of interest to magnify/reduce, or to select a set of geo-data on which a basic functionality may be applied. In order to model the associated geometry acquisition, we have introduced the *Geometry Entry* Unit, which may be used to represent interaction metaphors, typical of a Web GIS environment. This unit is characterized by a name and the geometry type on which the interaction is based, which can be a point, a multipoint, a line, a polyline, a rectangle, a polygon and a drag. Figure 4(a) shows the graphic notation of an instance of the *Geometry Entry* Unit, which refers to a rectangle selection, namely the *Rectangle Entry* unit. The corresponding Web representation is depicted in Figure 4(b).

The *Zoom In (Out)* units refer to the conventional zoom in and zoom out operations. Both take three input values: the global parameter with the current extension of the map, the global parameter with a magnification/diminishing factor and the instance *Point* of *Geometry Entry* Unit, which is used to identify the area on which the operation should be applied. As a result, if the operation succeeds, it sets a new value for the *Extension* global parameter, according to the zoom factor. This arrangement is graphically depicted in Figure 5 where an OK link leads to a new value for the *Extension*, while a KO redirects towards a generic error page.

(a) (b)

Figure 4: Graphic notation of a *Geometry Entry* Unit (a) and its representation in a Web browser (b)

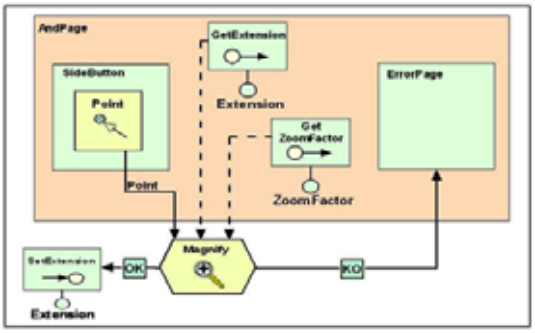


Figure 5: An example of *ZoomIn* Unit

The *Pan* unit represents the pan operation that shifts the focus of a map toward a specified direction. This unit takes three input values, namely the two points of the map where the user presses and releases the mouse button, and the global parameter with the current extension of the map. If the operation succeeds, a new value for the *Extension* global parameter is set, i.e. the new area of the map which is visualized. In Figure 6(a), visual symbols involved in a pan operation are sketched.

For Web GIS, a query processing usually outputs a set of geodata which satisfy user's criteria. In order to display such geodata onto a map, they are organized in a new temporary layer, which may be overlaid to the existing ones. Users may use them to perform other basic operations and derive new geographic information. In order to describe such a functionality through a WebML-like approach, we introduced the *CreateOverlay* Unit, which can be exploited to model scenarios where users pose queries and apply further spatial operations to the corresponding output in a cascade-like fashion. The *CreateOverlay* Unit requires two parameters, namely a spatial filter and a layer on which the filter should be applied.

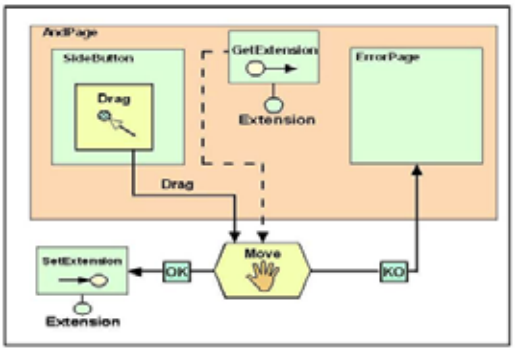


Figure 6(a): An example of the *Pan* operation

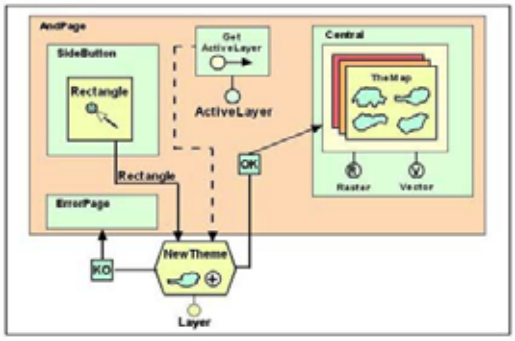


Figure 6(b): the *CreateOverlay* operation

In Figure 6(b), an instance of the *CreateOverlay* unit is depicted, namely *NewTheme*, where the spatial filter is obtained by a *Geometry Entry* Unit which returns a rectangle, and the layer is the active one. In case the operation succeeds, the OK link indicates that the OID of the newly created layer will be forwarded to the *MultiMap* unit, which will eventually render it.

4. Conclusions

In the present paper we described a methodology for Web GIS applications based on WebML approaches. We adapted the development process advocated for data-intensive Web applications to the context of Web GIS. Indeed, it relies on the iterative and incremental approaches that turn out to be suitable for Web GIS applications both for the need to quickly deploy them on the market and for the evolving nature of their requirements. The WebML methodology seems to be especially adequate also because allows us to concentrate on different concepts: data and presentation. In particular, we proposed to exploit the Spatial E-R model to define spatial data and we introduced some visual notations to model relevant interaction and navigation operations that are specific for Web GIS. In order to make more effective the proposed approach we would like to extend the WebRatio CASE tool to allow for the design of Web GIS applications and their automatic generation.

5. References

1. Calkins, Hugh W., Entity Relationship Modeling of Spatial Data for Geographic Information Systems, International Journal of Geographical Information Systems, January 1996.
2. S. Ceri, P. Fraternali, and A. Bongio, "Web Modeling Language (WebML): a Modeling Language for Designing Web Sites," Computer Networks: The Int. J. of Computer and Telecommunications Networking, V.33 n.1-6, pp.137-157, June 2000.
3. S. Ceri, P. Fraternali, A. Bongio, M. Brambilla, S. Comai, M. Matera. Designing Data-Intensive Web Applications. Morgan-Kaufmann Publishers. 2002.
4. S. Ceri, P. Fraternali, M. Matera. WebML Application Frameworks, a Conceptual Tool for Enhancing Design Reuse. Procs. "Web Engineering", Hong Kong, 2001.
5. F. Ferrucci, G. Tortora, G. Vitiello. "Exploiting Visual Languages in Software Engineering" - Handbook of Software Engineering and Knowledge Engineering, 2001, ISBN: 981-02-4973-X
6. X. Liang, A. Ginige "Smart business object a new approach to model business objects for Web applications", Procs. 1st International Conference on Software and Data Technologies, Setúbal, Portugal, Sept. 11-14, 2006. ISBN 972-8865-69-4, pp. 30-39.
7. L. Yu, L. Baojia, T. Dashi, "UML-G: An Object-Oriented Modeling Language for Geographical Information Applications".
8. P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind, Geographic Information Systems and Science, Wiley, 2001, 472 p.
9. P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind, Geographical Information System Principles – Second Edition 2001. Wiley.
10. L. Paolino, M. Sebillo, G. Tortora, G. Vitiello "VIEW-GIS – An Environment Supporting Web GIS Developers", Procs 12th International Workshop on Visual Languages and Computing (VLC), Grand Canyon USA, August 2006, pp. 141-146.
11. ZR. Peng, MH.Tsou, Internet GIS, Distributed GIServices for the Internet and wireless networks, 2003 Wiley.
12. WebML site. Available at www.Webml.org

Planet Diggum: Multi-point Touch and Gesture Control of Gaming Environments

Paul J. Diefenbach, William Muto, Sara Colucci, Chester Cunanan, Justin Dobies, James Grow, Matthew Smith
RePlay Lab and Digital Media Program
Antoinette Westphal College of Media Arts and Design
Drexel University
Philadelphia PA 19104, USA
{pjdief, wlm24, sec29, csc29, jmd84, jrg45, mas97}@drexel.edu

Abstract

While devices such as the Wii and Playstation's Eye (formerly EyeToy) have expanded user interface possibilities in the gaming world, there is still a traditional disconnect between the user and the physical screen image. The advance of gesture computing and the new availability of multi-point touch screens permit a novel mechanism and universal vocabulary for direct interaction with a virtual world. This paper discusses ***Planet Diggum****, a multi-user god-game which uses a novel combination of finger and hand-stroke gestures. The goal of this project is to create a testbed kiosk where multiple users can interact with the system without need of training or formal introduction.*

1. Introduction

The recent popularity of the Nintendo Wii highlights the importance of the man-machine interface over graphics fidelity. As gaming systems have grown in complexity, so to have their user interfaces.

Game interfaces have gone through several different iterations since their inception. The traditional game pad – the "Pong" dial and single-button joystick – has been replaced by increasingly more sophisticated controllers: from a controller with a 4-directional pad and two buttons, to multiple analog sticks and a plethora of buttons, to tilt sensors that can detect any directional movements. Other input devices include touch screens, mice, and programmable keyboards that allow for macros to be bound to a single key.

Single-point touch screens have had many uses since their invention, such as the widespread utilization as point-of-sale systems in the food industry or for information kiosks. They have wide acceptance in the gaming industry as stand alone systems in bars and casinos, mainly as instruments for gambling. Today, touch screens have become more integrated into mobile platforms as the GameBoy DS for gaming and into multi-function devices including PDAs and smartphones. Unlike earlier systems, gestures are replacing the typical button-click interaction [1].

Multi-point touch screens and gesture tracking offer new opportunities to create highly intuitive and user-friendly systems for even the novice user. While Oviat argues against the possibility of designing an idealized "everyperson information kiosk" using even multi-modal input [2], this project is attempting to create a testbed kiosk where multiple users can interact with a complex game system without need of specialized training. Planet Diggum is a god-game in the tradition of the DS game, Nintendogs, in which multiple users can interact with individual "Diggum" creatures and their world via natural finger and hand gestures.

The remainder of this paper is structured as follows: Section 2 discusses gesture-based computing and its growing role in user interface design, as well as examining the multi-point touch screen as a UI platform. Section 3 describes the goals of this project and the desired user experience of our testbed game application, and Section 4 discusses the design and implementation framework used. A summary and a description of future work are given in Section 5.

2. Related Work

In order to reinforce the intuitive nature of the interface, we are incorporating gestures to perform

various actions on our diggums and the world in which they live. This project leverages prior work in general gesture-based computing, its application in gaming, and the recent advances in multi-point touch systems.

2.1. Gesture-based Computing

Gesture-based computing has greatly evolved since its early implementations over half a century ago. Sutherland's Sketchpad Graphical Communication System (1963) utilized a light pen to send position information, using physical switches and knobs to change modes and control rotation and magnification [3]. Once a diagram was drawn using this system, it could be transferred to paper using a PACE plotter. Sutherland's system could not only draw lines, arcs, and regular geometry, but it could also fasten corners of geometry, group symbols, and instance groups.

Jones and Katyl of IBM introduced GRAPHPAK (1974), an experimental input utility for the APL programming language. GRAPHPAK utilized a commercially available sonic pen and custom glass tablet. The pen emits a spark at the pen's point, soundwaves generated by the spark are detected by two condenser microphones, and the time delay is then converted to x and y coordinates [4].

In 1984, Minsky presented a prototype of a gesture recognition system that recognized some single-finger gestures which were then used to manipulated displayed objects. These gestures were recognized by a "gesture parser" and were used for interactive computing systems designed for young children [5].

Studies by Benko, Wilson, Baudisch with their Stretch, X-Menu, and Slider techniques have shown that direct data manipulation without intermediary devices has a particular allure to novice users, especially for public installations [6].

Many toolkits have been released to support gestural computing. Microsoft's Tablet PC SDK [7] uses three separate areas: ink collection, ink data management, and ink recognition. A similar library for Java is SATIN, developed by Hong and Landay. SATIN uses machine learning (Rubine's Recognizer) to recognize key gestures [8]. SATIN also handles support for dealing with ambiguities in touch screen/pen input (such as cusps).

2.2. Gesture-based Gaming

Gesture recognition gaming is an increasingly popular way to develop for today's gaming industry, as can be seen with the recent success of Nintendo's two latest systems, the Nintendo DS and Wii. Both of these systems allow for the users to interact with the game through novel forms of gesture recognition: with the DS, wielding a stylus allows the user to draw directly on the screen, and with the Wii, a remote control recognizes the three dimensional motion of the player's arm(s). Games that utilize such forms of interaction include the popular WarioWare games (Wii and DS) and Lost Magic (DS).

Fig. 1. User interaction in Lost Magic

In the game Lost Magic (Fig. 1), the user plays a magician character who casts spells to fight off enemies. Rather than assigning each of these spells to a unique button combination, the DS gives the player the ability to draw symbols to cast a spell. In addition, navigating the character and his warriors is also done via the stylus and not the conventional directional pad. To move a character, the user only needs to tap the character he would like to move and click an area on the screen into which he wishes to move the character.

Fig. 2. WarioWare

WarioWare, for both the DS and Wii systems, challenges users to interact with the system in new and unique ways, such as holding the Wii remote in an unconventional manner or using the stylus as a unique controller. Fig. 2 shows a user playing WarioWare on the Wii accompanied with the corresponding game

screen. Note that the user must hold the controller in a distinct manner in order to play the game correctly.

2.3. Multi-point Touch Systems

As a public installation, Planet Diggum encourages cooperative gesturing. This collaborative approach fosters increased participation and uses community reach space to extend the workable screen area [9].

Multi-point Touch Systems have been evolving since the early 1990's, but most current implementations are based on frustrated total internal reflection (FTIR) [10].

Using multiple fingers and hands to interact was demonstrated by Krueger's VIDEOPLACE [11]. This system was able to pass gestures so that remote users were able to see them as if they were in the same room.

Wellner's digital desk demonstrates the possibility of integrating physical objects with the computer display [12]. It allows for some multi-point techniques but is focused mainly on single user applications.

The DiamondTouch table supports multiple points and multiple users (up to four) [13]. It can differentiate between each user's actions, but at a cost, because the inputs can become ambiguous. Morris et. al's CollabDraw software was built on this technology [9].

SmartSkin also supports multiple points of contact and allows for many users; however, it is unable to differentiate their inputs [14]. This system is able to get a much more accurate image of the surface and the contact points.

Advances in resolution and latency have been made with TouchLight [3] -- which uses a pair of cameras in combination with a projection display -- and Han's FTIR Multitouch Interaction research [10].

3. Project Goals

Planet Diggum is an investigative gaming application being developed by Drexel Digital Media (DIGM) Master's students as part of their gaming curriculum and utilizes a multi-point display being built by Computer Science (CS) and Electrical and Computer Engineering (ECE) undergraduate students. This project is coordinated by Drexel's RePlay Lab, which is a multi-departmental lab researching play and gaming technology.

This project leverages open source software libraries, low-cost hardware, and standards-based 3D protocols to create a modular, extensible framework for rapid prototyping and experimentation.

This project evolved from the task of investigating novel mechanisms of multi-user gameplay, and was driven by several assumptions and observations:

- Gross body gestures are now being adequately explored by modern consoles (Wii and Sony Eye) as alternative to controller.
- Physical proximity of users can enhance gameplay experience (Space Wars)
- Most interactions with games involve physical disconnect between user and image
- Casual games have largest growth potential
- Gesture computing (base) technology has reached sufficient maturity for rapid prototyping.

Gross body gestures, such as those employed by the Wii, were abandoned in favor of a more intimate gaming experience to enable proximity-based collaborative play. It should be noted that physical proximity of game participants can enhance the experience [9] where participants are in actual competition for the controls/interface rather than solely in competition in the virtual space. An example of this is the original Space Wars arcade game where a numeric pad in the center of the two contestants dynamically changed gameplay, yet was simultaneously physically accessible to each [15].

Planet Diggum is similar in ways to the award winning "Black & White" and other "god" games being released in the game industry. "Black & White" allows the player to play as a god competing against other deities in the region to become their one true god. The user can do this by giving the inhabitants helpful blessings or cursing them with devastation. Each of these actions causes a subsequent reaction in the game. By being helpful, the inhabitants will praise the players help, but by causing destruction, the inhabitants will accept the god through fear of the consequences resulting from disrespect. Planet Diggum follows similar suit in the ability to manipulate the world and its inhabitants, the diggums. The user(s) are able to directly interact with the characters, and similarly to "Black & White", these actions will cause the inhabitants to either praise or fear the player(s).

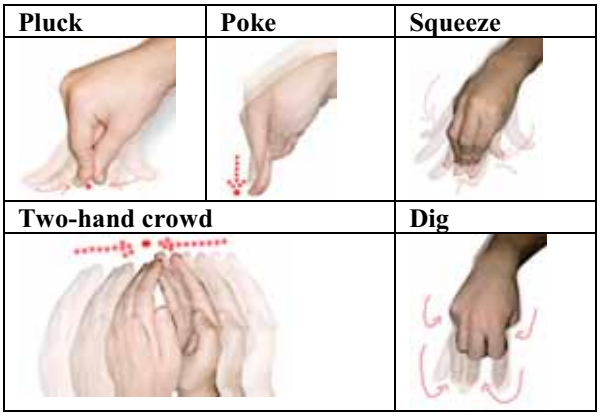


Fig. 3. Gesture examples

Planet Diggum differs from god games such as "Black & White" by allowing multiple players to play simultaneously to allow for a conflict between good and evil in the "god" character itself, for the players could either decide to work co-operatively with each other or oppose each other. This style of gameplay is made simple due to the use of a multi-touch screen kiosk for the project. The multi-touch screen kiosk allows for all actions in the game to be easily performed through simple hand gestures.

Planet Diggum takes the gesture recognition style of playing a step further by removing input devices such as controllers. This permits a smaller learning curve for the game as the hand motions provide a much more natural way of controlling the game rather than memorizing button combinations.

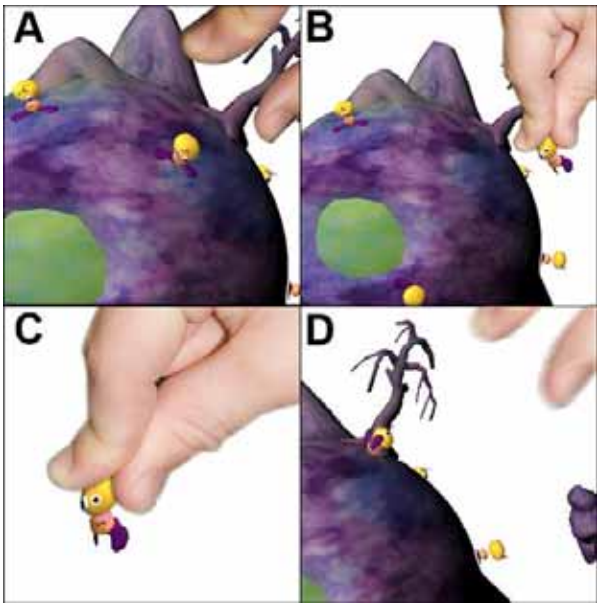


Figure 4: An example of user interaction with creatures in the environment.

This is crucial for the game, as it is being designed to be played in an environment where the amount of time each user would play the game would be very short (i.e. a few minutes), in comparison to a typical console game, which can average over 20 hours. For this reason the game must be very easy to approach, easy to learn the rules of the world and its controls, and entertaining all at the same time.

4. Design and Implementation Framework

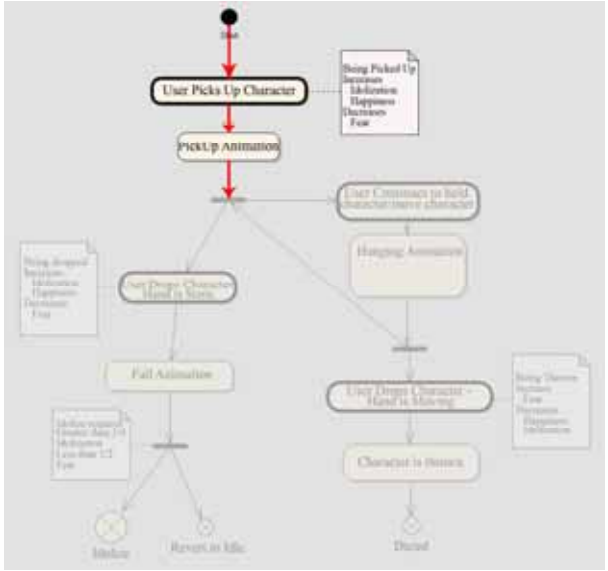


Fig. 5. Motion tree for Fig. 4B

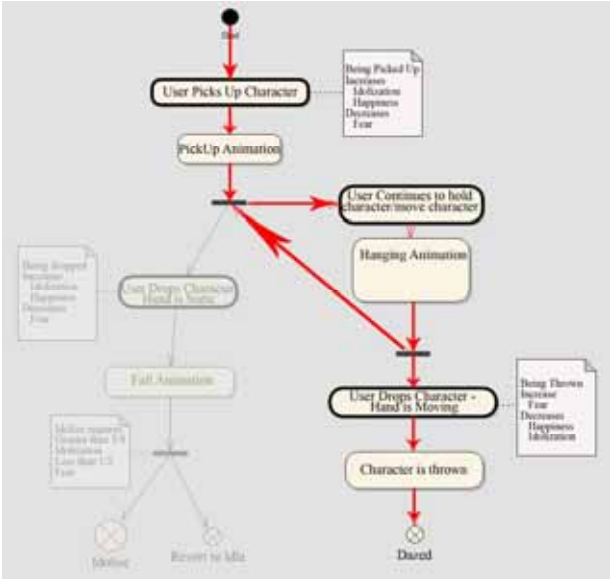


Fig. 6. Motion tree for Fig. 4D

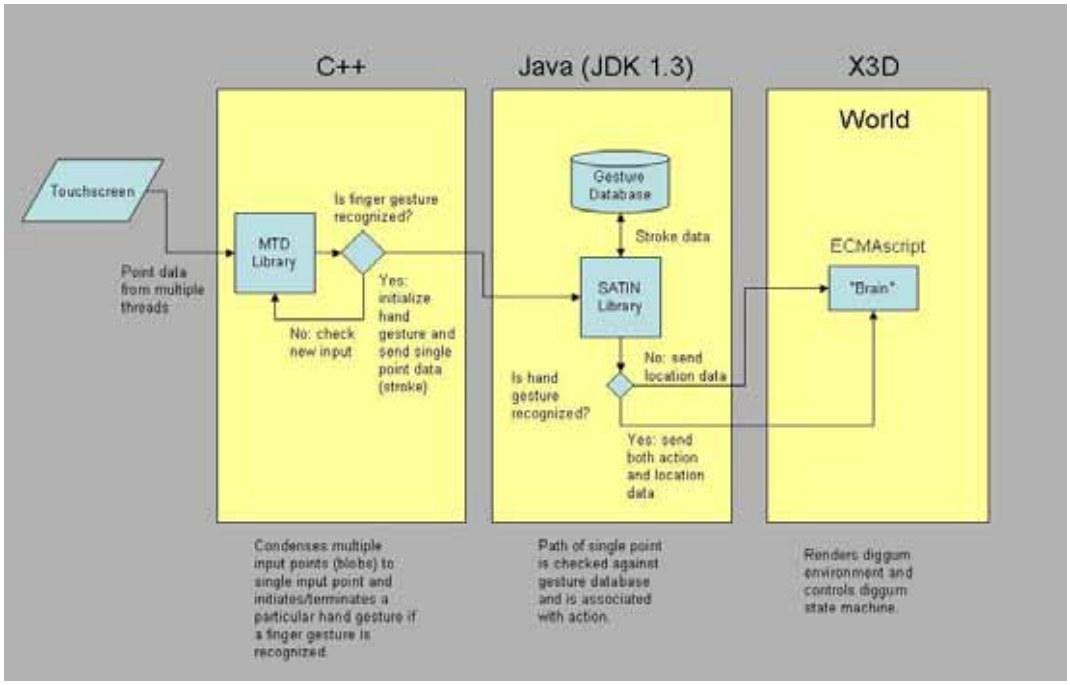


Figure 7: Architecture for system

4.1. Gesture-based Events

One of the problems with multi-point touch systems is the recognition and coordination of multiple touch points into a single recognizable gesture [16]. Our approach is to segment the problem by defining a glossary of finger and hand gestures which define time/space events. Finger gestures are used to define initiating and terminating events of a hand gesture; hand gestures map to corresponding actions. Fig. 3 lists some our recognized finger gestures.

Analogous to this is the mouse-click paradigm where a particular finger event corresponds to a button-down event, and a hand event can correspond to either a pull-down menu-selection and/or motion event. Figures 4, 5, and 6 are examples. Figures 5 and 6 show the state trees associated with a particular diggum during interaction in figure 4B and 4D. Fig. 4A shows the diggum in its natural habitat. 4B shows the "pluck" finger gesture (4C: close up). 4D shows a hand gesture, throwing the diggum into a tree.

4.2. System Components

Input from the multitouch screen will be handled via a C++ library (MTD) developed by the infinitouch team at Drexel University for their FTIR screen. MTD was based on OpenCV [17]. Finger gestures are determined by a fit of neighboring touches and the number of touch points. Multiple touch points can be reduced to a single event (i.e. left-mouse down) and a single location (i.e. centroid of touches). By reducing multiple finger touches down to a singe location, hand stroke recognition is greatly simplified. This information (movement of a single point and an event) is then passed to a Java application for hand gesture recognition. Our current design involves using the SATIN library for Java for gesture recognition (a drawn stroke linked to a command) [8]. When a gesture is performed, SATIN looks at a library of gestures, and returns an n-best list of possible matches, the highest ranked match being chosen. To create this gesture library, a Rubine Recognizer allows developers to "train" the system to recognize gestures through repetition [18]. If the movement of the single point is recognized as a hand gesture, this action is sent to our 3D environment. Otherwise, it is simply regarded as "dragging".

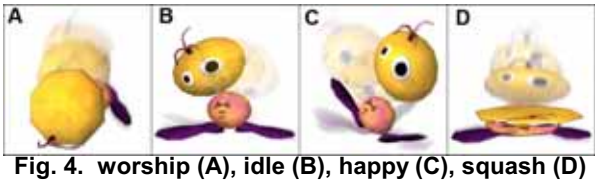


Fig. 4. worship (A), idle (B), happy (C), squash (D)

The diggums' environment is built in X3D [19], and their state machine is controlled by embedded ECMAscript. Fig. 7 outlines communication between these separate components.

4.3 Diggum Indirect Manipulation

Each diggum is governed by a series of variables that control their movement through an associated action tree. The variables are affected by both the actions of the users and the environment of the creature. These variables then in turn determine the actions of that particular diggum, which will then affect the variables of its neighbors. For example, when a user picks up a diggum it will become more fearful of the player, and begin to worship; likewise, nearby diggums viewing this interaction will also become fearful. Depending on the creature's status going into the ordeal and how it is returned to the world, either thrown or gently placed, it may either cower in fear or bow its head, idolizing the user. In this manner, whether or not there is user input, the characters' actions will vary from one another based on their past experiences.

5. Summary and Future Work

Through this gaming environment and novel gesture recognition technique (separation of finger and hand gestures), we are able to perform extremely complex actions through our interface. Future work includes recording and analysis of users interacting with the system to adjust the gesture vocabulary. In addition, issues such as movement latency with regard to tracking will be investigated using predictive techniques such as dead-reckoning.

6. References

[1] S. P. Kerr, D. Robert, B. Ording, J. P. Ive, P. J. Kennedy, A. M. Fadell and J. L. Robbin, "Proximity Detector in Handheld Device," U.S. Patent 20060161871, 2006.

[2] S. Oviatt, "Ten myths of multimodal interaction," *Commun. ACM.,* vol. 42, no. 11, pp. 74–81, Nov. 1999.

[3] I. E. Sutherland, "Sketch Pad," *Annu. ACM IEEE Design Automation Conf. Proc.SHARE Design Automation Workshop*, 1964.

[4] A. L. Jones and R. H. Katyl, "A Graphical Input Capability for APL," *Proc. 6th International Conf. APL,* Jan., 1974.

[5] M. Minsky, "Manipulating Simulated Objects with Real-world Gestures Using a Force and Position Sensitive Screen," *Computer Graphics,* July 1984, vol.18, no. 3, pp. 195-203.

[6] H. Benko, A. D. Wilson and P. Baudisch, "Precise Selection Techniques for Multi-Touch Screens" in *Proc. SIGCHI Conf. Human Factors in Computing Systems*, Montreal, Canada, 2006, pp. 1263-1272.

[7] Microsoft Corp. (2007). Tablet PC Development Guide. http://msdn2.microsoft.com/en-us/library/ms704849.aspx

[8] J. I. Hong and J. A. Landay, "SATIN: A Toolkit for Informal Ink-Based Applications," in *Proc. 13thAnnu. ACM Symp. User Interface Software and Technology,* San Diego, CA, 2000, pp. 63-72.

[9] M. R. Morris, A. Huang, a. Paepcke and T. Winograd, "Cooperative Gestures: Multi-User Gestural Interactions for Co-located Groupware" in *Proc. SIGCHI Conf. Human Factors in Computing Systems,* Montreal, Canada, 2006, pp 1201-1210.

[10] J. Y. Han, "Low-Cost Multi-Touch Sensing through Frustrated Total Internal Reflection," in *Proc. 18th Annu. ACM Symp. User Interface Software and Technology*, Seattle, WA, 2005, pp 115-118.

[11] M. Krueger, "VIDEOPLACE and the Interface of the Future" in *The Art of Human Computer Interface Design*, B. Laurel, Ed., Menlo Park, CA: Addison Wesley, 1991.

[12] P. Wellner, "Interacting With paper on the Digital Desk" *Commun. ACM*, vol. 36, no.7, pp. 87-96, July. 1993.

[13] P. Dietz, and D. Leigh, "DiamondTouch: a Multi-user Touch Technology" in *ACM UIST Symp. User Interface Software and Tech.*, Orlando, FL, 2001, pp. 219-226.

[14] J. Rekimoto, "SmartSkin: an Infrastructure for Freehand Manipulation on Interactive Surfaces" in *Proc. SIGCHI Conf. Human Factors in Computing Systems: Changing our World, Changing Ourselves*, Minneapolis, MN 2002, pp. 113-120.

[15] M. Markowitz. (2001, Dec. 13). Spacewar. Available: www3.sympatico.ca/maury/games/space/spacewar.html

[16] X. Wang, X. Zhang and G. Dai, "Tracking of Deformable Human Hand in Real Time as Continuous Input for Gesture-based Interaction," *Proc. 12th International Conf. Intelligent User Interfaces*, Honolulu, HI, 2007, pp 235-42.

[17] OpenCV. Online:opencvlibrary.sourceforge.net

[18] D. Rubine, "Specifying Gestures by Example," *Proc. 18th Annu. Conf. Computer Graphics and Interactive Techniques,* 1991, vol 25, no. 4, pp. 329-337.

[19] Web3D Consortium. (2004) Extensible 3D (X3D). http://www.web3d.org/x3d/specifications/ISO-IEC-19775-X3DAbstractSpecification/

Improving the Robustness of Diagrammatic Sketch Recognizers

Gennaro Costagliola, Vincenzo Deufemia, Michele Risi
Dipartimento di Matematica e Informatica, Università di Salerno
Via Ponte don Melillo, 84084 Fisciano(SA), Italy
{gcostagliola,deufemia,mrisi}@unisa.it

Abstract

Sketching is an activity that produces informal documents containing hand-drawn shapes highly variable and ambiguous. In this paper we present a diagrammatic sketch recognizer that is able to cope with the recognition of inaccurate hand-drawn symbols by exploiting error recovery techniques as developed for programming language compilers. The error recovery algorithms are able to interact with recognizers automatically generated from grammar specifications in order to obtain the information on missing or misrecognized strokes.

1. Introduction

Sketches represent an effective medium for facilitating conceptual design activities by enabling designers to focus on critical issues rather than on intricate details such as precise size, shape, location and color. In the last decade, many efforts have been put into developing software capable of understanding sketches with objects that can be represented using structural descriptions [2][4][6][11]. These recognizers enable users to create sketches using pen-based devices and to transform the edited sketches into input for more powerful design systems. However, existing sketch recognition techniques are error-prone or severely limit the user's drawing style.

In this paper we present a sketch recognition technique able to identify hand-drawn messy symbols of diagrammatic languages. In particular, it extends the approach introduced in [7] with two important features: recognizing hand-drawn symbols with missing strokes and automatically correcting errors of stroke misrecognition. The approach is based on Sketch Grammars [8] for modeling diagrammatic sketch notations and for the automatic generation of efficient recognizers whose parsing technique is based on LR parsing techniques [1]. The recognition of inaccurate hand-drawn symbols is faced by using error recovery techniques as developed for programming language compilers [15]. In particular, these techniques are able to recover the information on missing strokes from the symbol recognizers automatically generated from grammar specifications.

The sketch recognition system performs an on-line interpretation of the user strokes using the eager modality [5]. This means that the recognizers interpret the strokes immediately after they have been drawn, and provide a user with feedback of the recognized symbols as soon as possible. This approach is more robust and efficient, and consequently more usable, than the lazy one, where the recognition occurs only when explicitly requested by the user and it involves all strokes previously drawn. During the recognition process, the user validates or rejects the symbol interpretations progressively. Moreover, the user can obtain information on partially recognized symbols.

The paper is organized as follows. In Section 2 we describe the proposed approach for the recognition of hand drawn diagrammatic symbols. The related work is discussed in Section 3. Finally, the conclusion and further research are discussed in Section 4.

2. The proposed approach

As shown in Fig. 1, the proposed sketch recognition system interacts with the sketch interface to obtain the edited strokes, to provide the results of its interpretation process, and to receive user's feedback on the recognized symbols.

The *sketch recognition module* works in the eager mode and is composed by three sub-modules. The domain independent recognizer interprets the strokes as primitive shapes, such as lines, arcs, ellipses, etc. The symbol recognizers cluster the primitive shapes in order to identify possible domain symbols. In particular, when a symbol recognizer is able to parse a new stroke, it gives as output the new status of the symbol, which can be partially or completely recognized. The strokes not parsed by a symbol

recognizer are temporarily stored in its *unmatched strokes* repository. This repository contains both graphical and classification information of each unparsed stroke.

The interpretations produced by the symbol recognizers are analyzed by the language recognizer, which applies its recognition context for selecting the interpretations to be forwarded to the sketch interface.

The *sketch interface* visualizes as feedback the information obtained by the language recognizer. In particular, the interface requires users to accept or reject the recognized symbols. When a symbol is accepted, the language recognizer updates its parsing status and discards the interpretations of symbol recognizers that are in conflict with the accepted symbol.

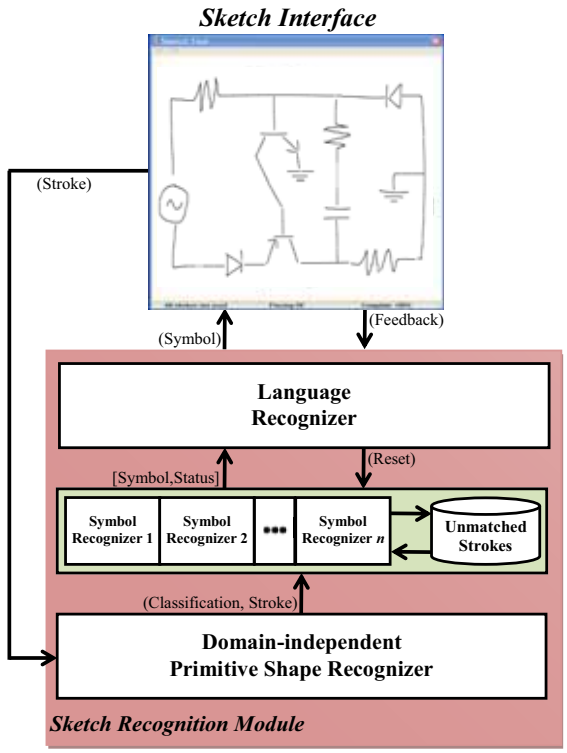


Figure 1. The recognition process

In the following, we first describe a grammar-based implementation of a symbol and language recognizer and then show how to integrate error recovery techniques to improve the robustness of the recognition.

2.1. Describing and recognizing sketched symbols

The formalism used to specify the sketched symbols of a domain language is the *Sketch Grammar* [7]. A Sketch Grammar represents an extension of string grammars, where also geometric and topologic relations are allowed. The grammar productions alternate symbols and relations, and cluster the input strokes into shapes of a domain language. As an example, the following symbol grammar productions specify a transistor symbol of electric circuit diagrams, represented in Fig. 2:

(1) npnTransistor → ELLIPSE
<contain(t_1), vertical(t_2), length(1,t_3), joint$_{1_1}$(t_4), joint$_{1_2}$(t_4)> ^{r1} LINE$_1$
<contain^1(t_1), rotate(90,t_5), joint$_{1_1}^1$(t_4), near$_2$(t_6)> ^{r2} LINE$_2$
<contain^2(t_1), rotate^1(45,t_5), joint$_{1_1}^2$(t_4), near$_2^1$(t_6)> ^{r3} wireUp

(2) npnTransistor → ELLIPSE
<contain(t_1), vertical(t_2), length(1,t_3), joint$_{1_1}$(t_4), joint$_{1_2}$(t_4)> ^{r1} LINE$_1$
<contain^1(t_1), rotate(90,t_5), joint$_{1_1}^1$(t_4), near$_2$(t_6)> ^{r2} LINE$_2$
<contain^2(t_1), rotate^1(-45,t_5), joint$_{1_1}^2$(t_4), near$_2^1$(t_6)> ^{r4} wireDown

(3) wireUp → LINE$_3$
<contain^3(t_1), rotate^2(-45,t_5), joint$_{1_1}^3$(t_3), near$_2^2$(t_6)> ^{r5} LINE$_4$
<contain^4(t_1), length(0.33,t_3), rotate(-45,t_5) > ^{r6} Arrow

(4) wireDown → LINE$_4$
<contain^3(t_1), rotate^2(45,t_5), joint$_{1_1}^3$(t_3), near$_2^2$(t_6)> ^{r7} LINE$_3$
<contain^4(t_1), joint$_{2_1}^1$(t_3), length(0.33,t_3), rotate^1(-45,t_5) > ^{r8} Arrow

(5) wireDown → LINE$_4$
<contain^4(t_1), joint$_{2_1}$(t_3), length(0.33,t_5), rotate(-45,t_5) > ^{r6} Arrow
<contain^5(t_1), rotate^3(45,t_5), joint$_{1_1}^4$(t_3), near$_2^3$(t_6)> ^{r9} LINE$_3$

(6) Arrow → LINE$_5$
< joint$_{1_1}$(t_1), length(1, t_2), rotate(90,t_5)> ^{r10} LINE$_6$

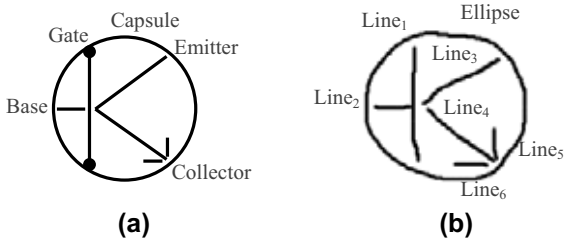


Figure 2. The npn Transistor symbol (a) and a hand-drawn sketched version (b)

The transistor symbol is composed by an ellipse and several lines as shown in Fig. 2(a-b). Each primitive shape has associated a set of attributes, which are used to relate a shape to the others, and their values depend on the "position" and "dimension" of the shape in the sketch. The relations specified in the productions are such that the strokes in the sketch in Fig. 2(b) satisfy them. The values $t_1,\ldots,t_6$ specify the error margin in the satisfaction of the relations.

As an example, the first relation in production 1 relates through two *joint* relations attribute 1 of ELLIPSE, which represents its borderline, to attribute 1 and 2 of a LINE, which represent its end points, highlighted by bullets in Fig. 2(a). Moreover, relations

vertical and *length* indicate that the first LINE has to be vertical with a length equal to the diameter of the ELLIPSE that completely contains it.

Given a sketch grammar for a symbol of a diagrammatic language it is possible to automatically generate the corresponding symbol recognizer. This recognizer parses a symbol scanning the strokes in the order defined by the productions. As an example, the previous productions describe an npn Transistor symbol with the sequence: capsule, gate, base, collector, and emitter. In order to perform a robust symbol recognition process, the recognizers should be able to parse symbols using more than one sequence. Since when drawing symbols the users employ only a subset all possible sequences [12], the sketch grammar for a symbol should only include the stroke temporal patterns mainly used. As an example, the npn Transistor is characterized by three main sequences.

The parser built from the grammar is based on an extension of LR-parsing. The parsing algorithm analyzes the input exploiting the information contained in a parsing table. The parsing table for the previous grammar of the npn Transistor symbol is shown in Table 1.

Table 1. Parsing table for npn Transistor grammar.

State		Action			Goto				Next
		LINE	ELLIPSE	$	npnTransistor	wireUp	wireDown	Arrow	
0			:sh1		:17				(start,npnTransistor)
1		:sh2							(r1, LINE)
2		:sh3							(r2, LINE)
3	1	:sh4				:6	:6		(r3, LINE)
	2	:sh5				:7	:7		(r4, LINE)
4		:sh8							(r5, LINE)
5	1	$\overline{r6}$, $\overline{r8}$:sh9						:12	(r7, LINE)
	2	:sh10						:12	(r6, LINE)
	3	:sh11						:12	(r8, LINE)
6		Production 1							-
7		Production 2							-
8	1	:sh10						:13	(r6, LINE)
	2	:sh11						:13	(r8, LINE)
9	1	:sh10						:14	(r6, LINE)
	2	:sh11						:14	(r8, LINE)
10		:sh15							(r10, LINE)
11		:sh15							(r10, LINE)
12		:sh16							(r9, LINE)
13		Production 3							-
14		Production 4							-
15		Production 6							-
16		Production 5							-
17		accept							-

The parsing table is composed by a set of rows and is divided into three main sections: *Action*, *Goto*, and *Next*. Each row is composed of a set of one or more sub-rows each corresponding to a parser state. The *Action* and *Goto* sections are similar to the ones used in the LR parsing tables for string languages [1], while the *Next* section is used by the parser to select the next stroke to be processed. In particular, an entry *Next*[*k*] for a state *k* contains the couple (*relations*, *x*), which drives the parser in selecting a symbol *x* satisfying *relations*. The special entry (*start*, *S*) is used to retrieve the first symbol to be parsed. The empty entries in the *Action* section halt the recognizer that enters in a waiting state.

Sketch grammars are used also to specify the *language grammars*, which define the sentences of the language as composition of the shapes defined by symbol grammars through spatial relations [7]. As an example, the following productions represent some of the language grammar productions for the circuit diagrams.

(1) Circuit → SubCircuit
(2) SubCircuit → SubCircuit <***any***> SubCircuit
(3) SubCircuit → SubCircuit <$\boldsymbol{joint_{1\_1}(t_1)}$> Wire <$\boldsymbol{joint_{2\_1}(t_1)}$> Component
(4) SubCircuit → SubCircuit <$\boldsymbol{joint_{1\_1}(t_1)}$> Wire
(5) SubCircuit → Component
(6) Component → npnTransistor
→ pnpTransistor
→ Lamp
→ Battery
→ Capacitor
→ Resistor
→ Ground
→ Switch

Production 2 states that a circuit can be composed of unconnected *SubCircuit*s. Production 3 recursively defines a *SubCircuit* as the composition of three language symbols: *subCircuit*, *Wire*, and *Component*. Production 4 specifies that a *SubCircuit* can be connected through a *Wire* to itself.

Figure 3 shows a sketch matching the previous productions. In particular, the recognized circuit uses the npn Transistor in place of the switch to control the flow of electrons from battery through the lamp.

Similarly to symbol recognizers, the language recognizer is automatically generated from the language grammar.

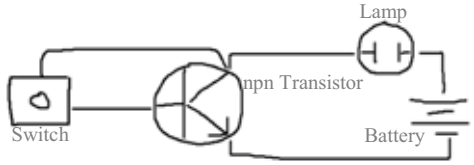


Figure 3. A switch circuit

2.2. Sketch Symbol Recognition with Error Recovery

The parsing approach previously described correctly works if the users completely drawn all the strokes composing a symbol. However, inaccuracy is intrinsic in hand-drawn sketches, thus the user could miss to drawn some symbol strokes, or could draw strokes that result to be difficult to identify, yielding the recognition approach ineffective. In order to cope with these difficulties in the following we introduce error recovery techniques in the previous parsing algorithm.

The goal of the error recovery algorithm is to overcome the problem of stroke linearization performed by the parsing algorithm, which prevents the recognition of incomplete symbols. To this aim, when a stroke of a symbol is missing (or misrecognized) the parsing algorithm actives an error recovery procedure that allows the parser to proceed in the recognition of the symbol. Nevertheless, the error recovery process should not be applied if it will not lead to the recognition of a symbol, i.e., many strokes of the symbol are still missing. In order to face with this requirement we associate to the terminals of the symbol grammars an *importance value* indicating the importance of the stroke in the described symbol. As an example, for the terminals of the previous npn Transistor symbol we can associate the following importance values, specified within parenthesis.

(1) npnTransistor → ELLIPSE(20) <…> ^{r1} $LINE_1$(15)
<…> ^{r2} $LINE_2$(15) <…> ^{r3} wireUp
...
(3) wireUp → $LINE_3$(10) <…> ^{r5} $LINE_4$(5) <…> ^{r6} Arrow
(4) wireDown → $LINE_4$(5) <…> ^{r7} $LINE_3$(10) <…> ^{r8} Arrow
…
(6) Arrow → $LINE_5$(25) <…> ^{r10} $LINE_6$(10)

The importance values indicate that the left head segment of the *Collector* wire has a weight (25) greater than the other segments. This allows the parser to discriminate *npn Transistor*s against *pnp Transistor*s, since the latter represents the Collector symbol in a inverse direction with respect to the first, as shown in Fig 4. Thus, when the strokes that allow to distinguish a symbol from the others are drawn they will have associated a high importance value.

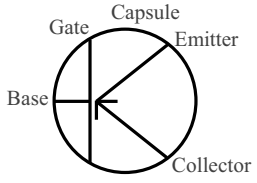


Figure 4. The pnp Transistor symbol

The importance values are stored in the *Next* section of the generated parsing table, associated to each next symbol to be processed.

Thus, the importance values associated to the complex symbols allow the generated recognizers to associate to the partially recognized symbols a value that can be used both to disambiguate the recognition of quite similar symbols and to pose a limit in the number of missing strokes admissible in a symbol. In particular, let t be a value such that if the sum of importance values associated to the recognized strokes of a symbol S is greater than t then S is (partially or completely) recognized. The error recovery process should terminate when the importance values associated to the missing strokes exceed the value 100-t. Indeed, in this case the recognized symbol will never exceed threshold t even if all the remaining symbol strokes have been drawn.

In the following we provide the algorithms implementing the error recovery technique.

```
Recovery(){
    PT = parsingtable[parser];
    state = stack[parser].currentState;
    while (state != null) {
        (r, s, imp_value) = PT.Next[state];//for multiple instances
                //the triple with lowest importance value is selected
        if (imp_value > 100-threshold[parser])
            exit;
        threshold[parser]= threshold[parser]+ imp_value;
        newStroke = Fit(r, s);
        state = ContinueParsing(newStroke);
    }
}

Fit(r, s) {
    rep = repository[parser];
    foreach(x in rep) {
        if (r is a valid relation for x) {//using complete classification
            delete x from rep;
            return x;
        }
    }
    simulatedStroke = use the constraint resolver to calculate a
            shape of s type compatible with the stack and relation r;
    return simulatedStroke;
}

ContinueParsing(s) {
    action[parser].shift(s);
    rep = repository[parser];
    input[parser]=rep;
    action[parser].continue; //reactivate the parser on rep
    input[parser]=PrimitiveShapeRecognizer; //restore the input
    if(state[parser] != accept and rep.empty == true)
            return null;
    if (state[parser] == accept)
            return null;
    return stack[parser].currentState;
}
```

The *Recovery* function is automatically invoked by a symbol recognizer when a syntax error occurs. In particular, if the parser can still recognize a symbol (i.e., the value 1-*t* has not been exceeded by the importance values of the recognized strokes) then the *Fit* and *ContinueParsing* functions are invoked, respectively, to look for a stroke able to reactivate the parser, and to continue the parsing of the symbol from the stroke in the symbol stroke sequence that follows the missing stroke.

Fit function first analyzes the repository of unmatched stroke to find a stroke, with low accuracy, able to solve the syntax error. If it is found, it is used as the next input for the parser. Otherwise, the function computes the information to simulate the missing stroke. The stroke given in output by *Fit* function is then used by *ContinueParsing* function for updating the parsing state and then reactivating the parser on the repository to recover from unparsed strokes.

After the recovery process, the symbol recognizer checks the acceptance state of the parser to verify if the symbol is completely or partially recognized.

2.3. An example

Let us suppose that the user has drawn the sketches shown in Fig. 5(a-e), and considering a threshold value of 70 for the recognition of the npn Transistor symbol.

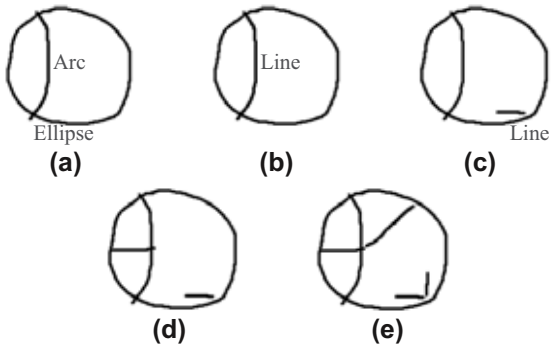


Figure 5. Error recovery on npn Transistor symbol

After the recognition of the Ellipse stroke in Fig. 5(a), the npn Transistor parser reaches state 1, but fails because the next symbol in the input is an Arc against an expected Line. The parser stores the unparsed stroke in the repository associated to the npn Transistor recognizer and then invokes the error recovery procedure. The *Recovery* algorithm invoked by the parser uses the triple (*r1*, *LINE, 15*) to check if the current importance value 15 exceeds 30 (i.e., 100-the threshold), and then finds in the repository a valid stroke compatible with relation *r1* and primitive shape *LINE*. Thus, thanks to the recovery algorithm the stroke is correctly interpreted as Line solving a misrecognition error, as shown in Fig 5(b). The parser reaches the state 2 and waits for the next stroke.

Fig 5(c) shows the sketch after the drawing of a horizontal line. The new stroke is not compatible with the triple (*r2*, *LINE, 15*) associated to state 2. Thus, the error recovery algorithm simulates the Base wire of the transistor, but cannot continue since will never reach the maximum admissible importance value (i.e., 30). In this case, a backtracking process restores the parser state to the one preceding the error recovery invocation. After the drawing of the Base wire stroke (see Fig 5(d)), the parser moves from state 2 to state 3, and simulates the Collector Wire with the triple (*r4*, *LINE, 5*), considering all the strokes provided as input and including the horizontal line previously sketched. However, the threshold value of 70 for the recognition of the npn Transistor symbol is not reached.

When the user draws the symbol in Fig. 5(e), the parser simulates the Collector Wire and reaches the acceptance state, and the symbol with missing strokes is completely recognized.

3. Related Work

A large body of work has been proposed for the recognition of freehand drawings using structural, syntactic, and temporal methods.

In [10] Kara and Stahovich present a multi-level parsing scheme that uses contextual knowledge to both improve accuracy and reduce recognition times. However, the recognition process is guided from "marker symbols", which are symbols easy to recognize, that they assume to exist always in the sketch. Moreover, the approach assumes that the hand-drawn diagram consists of shapes linked by arrows.

In [9], Gennari *et al.* present a circuit diagram recognition system that runs isolated symbol recognizers to generate an interpretation. In particular, the symbols are located by considering the areas with high density of pen strokes and the temporal information associated to the segmented input strokes. Then, the candidate symbols are classified using a statistical model constructed on training examples. On the other hand the segmentation algorithm used in this work does not handle interspersed drawing, so the user is required to finish an object before starting a new one.

A strategy quite similar to that proposed in this paper has been developed by Alvarado and Davis [3]. They describe a blackboard based architecture with a top-down recognition component based on dynamically constructed Bayesian networks that allows recovery from bottom-up recognition errors. In

particular, the approach allows to model low-level errors explicitly and use top-down and bottom-up information together to fix errors that can be avoided using context. However, the high computational cost of the whole method makes the system unsuitable for real-time recognition of realistic sketches.

To keep the search tractable, the spatial recognition method for text and graphics proposed in [14] makes some assumptions about the objects in the domain, such as, the objects have no more than eight strokes. However, these assumptions limit the applicability of the method to domains where objects vary in size and shape or where assumptions on the object size and scale might not hold.

Another approach to reduce the computational cost of sketch recognition is to exploit the preferred stroke orders. By the observation that when asked to draw a symbol more than once, people tended to draw it in the same order and that certain regularities exist between individuals, Sezgin and Davis construct a Hidden Markov Model based on these orders for recognizing each symbol [12]. The HMM-based approach exploits the regularities to perform very efficient segmentation and recognition. However, the recognition algorithm requires each object to be completed before the next one is drawn. The approach has been improved to exploit both stroke-level and object-level temporal orderings, to support continuous observable features, and to handle interspersed drawing [13].

4. Conclusions and Future Work

The paper introduces a recognition technique for diagrammatic sketches, which is able to recognize hand-drawn symbols with missing strokes and to correct stroke misrecognition errors automatically. The recognition system is composed of hierarchically arranged recognizers automatically generated from grammar specifications, which integrates error recovery techniques to improve the robustness of the recognition.

In the future we intend to evaluate the effectiveness of the propose approach by integrating it in our previously development sketch-based software application [7]. In particular, we intend to conduct several user studies to evaluate the performances of the recognition algorithm and the improvements in the recognition accuracy achieved with the proposed parsing approach.

Moreover, in order to reduce the number of active error recovery processes, and consequently the number of user feedback requests, we intend to exploit the feedbacks provided by the users to adapt the behaviour of the recognizers. In particular, we plan to modify the importance values associated to the terminals of the symbol grammar based on the acceptance/reject decisions of incomplete symbols.

5. References

[1] Aho, A.V., R. Sethi, and J.D. Ullman, *Compilers Principles, Techniques, and Tools*, Addison-Wesley Series in Computer Science, 1987.

[2] C. Alvarado and R. Davis, "A Framework for Multi-Domain Sketch Recognition", in *Proceedings of AAAI Spring Symposium on Sketch Understanding*, 2002, pp. 1-8.

[3] C. Alvarado, and R. Davis, "SketchREAD: A Multi-Domain Sketch Recognition Engine", in *Proc. of UIST'04*, 2004, ACM Press, pp. 23-32.

[4] D. Blostein and L. Haken, "Using Diagram Generation Software to Improve Diagram Recognition: A Case Study of Music Notation", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(11), 1999, pp. 1121-1136.

[5] D. Blostein, A. Lank, A. Rose, and R. Zanibbi, "User Interfaces for On-Line Diagram Recognition", in *Proc. of the 4th IAPR International Workshop on Graphics Recognition Algorithms and Applications* (*GREC'2001*), LNCS, Vol. 2390, Springer-Verlag, Berlin, 2001, pp. 92-103.

[6] A. Caetano, N. Goulart, M. Fonseca, and J. Jorge. "JavaSketchIT: Issues in Sketching the Look of User Interfaces", *AAAI Spring Symposium on Sketch Understanding*, 2002, pp. 9-14.

[7] G. Costagliola, V. Deufemia, M. Risi, "A Multi-layer Parsing Strategy for On-line Recognition of Hand-drawn Diagrams", in *Proceedings of IEEE Symposium VL/HCC'06*, Brighton, United Kingdom, 2006, pp. 103-110.

[8] G. Costagliola, V. Deufemia, and M. Risi, "Sketch Grammars: A Formalism for Describing and Recognizing Diagrammatic Sketch Languages", in *Proc. of International Conference on Document Analysis and Recognition* (*ICDAR'05*), Seoul, Korea, 2005, pp. 1226-1230.

[9] L. Gennari, L. B. Kara, T. F. Stahovich, K. Shimada, "Combining Geometry and Domain Knowledge to Interpret Hand-drawn Diagrams," *Computers & Graphics*, 29(4), 2005, pp. 547-562.

[10] L.B. Kara, and T.F. Stahovich, "Hierarchical Parsing and Recognition of Hand-sketched Diagrams", in *Proc. of UIST'04*, ACM Press, 2004, pp. 13-22.

[11] J. Lin, M. Newman, J. Hong, and J. Landay, "DENIM: Finding a Tighter Fit Between Tools and Practice for Web Site Design", in *CHI Letters: Human Factors in Computing Systems*, CHI 2000, 2(1), pp. 510-517.

[12] T.M. Sezgin and R. Davis, "HMM-Based Efficient Sketch Recognition", in *Proc. 10th Int'l Conf. Intelligent User Interfaces* (*IUI'05*), ACM Press, 2005, pp. 281-283.

[13] T. M. Sezgin and R. Davis, "Sketch Interpretation Using Multiscale Models of Temporal Patterns", *IEEE Computer Graphics and Applications*, 27(1), 2007, pp. 28-37.

[14] M. Shilman and P. Viola, "Spatial Recognition and Grouping of Text and Graphics", in *Proc. Eurographics Workshop Sketch-Based Interfaces and Modeling*, Eurographics Assoc., 2004, pp. 91-95.

[15] G. Snelting, "How to Build LR Parsers which Accept Incomplete Input", *SIGPLAN Notices*, 25(4), 51-58, 1990.

Interval-valued computing as a visual reasoning system

Benedek Nagy
University of Debrecen
Faculty of Informatics and Information Sciences
nbenedek@inf.unideb.hu

Sándor Vályi
University of Debrecen
Faculty of Health College
Department of Health Informatics
valyis@de-efk.hu

Abstract

We show the visual reasoning power of a recently developed unconventional computing model – the so-called interval-valued computing. It bears not only a high evidence of power in visual representation of Boolean algebraic calculations to show validity of propositional logical laws but also a natural way of faithful visual following of the process of the solution of a typical PSPACE-complete problem, namely, determining validity of quantified propositional formulæ.

1. Introduction

Diagrammatic and visual reasoning is a basic bridge between formal reasoning and human understanding. A basic method for visualizing Boolean algebraic calculations is the method of Venn diagramms. It is applicable for formulæ built from two or three propositional variables. There are good ideas to generalize Venn diagrams to a higher number of variables (see [3],[5],[6],[7],[10],[13]). We will employ the apparatus of interval-valued computing which is a recently developed computing paradigm ([9], [11], [12]) that works with finite sets of intervals (opposite to the interval-arithmetic which works with intervals taking care about uncertainty).

Moreover, we show that the interval-valued computation process is particularly suitable to visualize the process of evaluation of quantified propositional formulæ. This process is an unconventional but visually well interpretable decision process for a PSPACE-complete problem (whether the given quantified propositional formula is true). This problem is a basic example of a class of problems where the question is to determine whether there is a winning strategy for the first player in the given game and most of these problems are reducible to $QSAT$.

In Section 2, the notion of interval-values is introduced, also the interval-valued computations are described. In Section 3, we show a visualized proof of a propositional logical law with four variables. In Section 4, we demonstrate how visualize an interval-valued computation for the validation of a quantified propositional formula.

2. The interval-valued computation system

2.1. The idea

In [9], Nagy proposed a new discrete time / continuous space computational model, the so-called interval-valued computing. It involves another type of idealization than Turing machines – the density of the memory can be raised unlimitedly instead of its length. This new paradigm keeps some of the features of traditional Neumann-Turing type computations.

It works on specific subsets of the interval $[0, 1)$, more specifically, on finite unions of []-type subintervals. In a nutshell, interval-valued computations start with $\left[0, \frac{1}{2}\right)$ and continue with a finite sequence of operator applications. It works sequentially in a deterministic manner.

The allowed operations are motivated by the operations of the traditional computers: Boolean operations and shift operations. There is only an extra operator, the product. The role of the introduced product is to connect interval-values on different 'resolution levels'. Essentially, it shrinks interval-values.

So, in interval-valued computing systems, an important restriction is eliminated, i.e. there is no limit on the number of bits of a cell in the system; we have to suppose only that we always have a finite number of bits. Of course, in the case of a given computation an upper bound (the bit height of the computation sequence) always exists, and it gives the maximum number of bits the system needs for that computation process. Hence our model still fits into the framework of the Church-Turing paradigm, but it faces different limitations than the classical Turing model.

Although the computation in this model is sequential, the inner parallelism is extended. One can consider the system without restriction on the size of the information coded in an information unit (interval-value). It allows to increase the size of the alphabet unlimitedly in a computation. In this article we employ this inner parallelism to extend the visual expressiveness of calculations with interval-values. Long manipulations on the separate bits can be shown as a unit, acting uniformly to the whole stored data.

2.2. Interval-values

We note in advance that we do not distinguish interval-values (specific characteristic functions from [0,1) into $\{0,1\}$) from their subset representations (subsets of [0,1)) and we use always the more convenient notation.

The set $\mathbf{V}$ of *interval-values* coincides with the set of finite unions of $[)$-type subintervals of $[0,1)$.

The set $\mathbf{V}_0$ of *specific interval-values* coincides with $\left\{ \bigcup_{i=1}^{k} \left[\frac{l_i}{2^m}, \frac{1+l_i}{2^m}\right) | m \in \mathbf{N}, k \leq 2^m, \text{and } 0 \leq l_1 < ... < l_k < 2^m \right\}$.

We note that the set of finite unions includes the empty set ($k = 0$), that is, $\emptyset$ is also an allowed interval-value.
Our notion of interval-value is related to the notion of generalized interval ([4]). Our paper outlines not only a rule-based reasoning system, but an unconventional computation system whose computations represent and follow also visually the steps of the solution of the given problem more faithful than other known unconventional computing paradigms as membrane or DNA-computing. In [2] diagrams are used to solve problems on specific graphs.

2.3. Operators on interval-values

Similarly to traditional computers working on bytes, of course, we allow bitwise Boolean operations. If we consider interval-values as subsets of [0,1), then the operations negation, disjunction, conjunction coincide with the set-theoretical operations of complement ($\overline{A}$), union ($A \cup B$) and intersection ($A \cap B$), respectively. $\mathbf{V}$ forms an infinite Boolean set algebra with these operations. $\mathbf{V}_0$ is an infinite subalgebra of this algebra.

Before we add some other operators, we introduce a function assisting the formulation of the following definition. Intuitively, it provides the length of the left-most component (included maximal subinterval) of an interval-value A. We define the function $Flength : \mathbf{V} \to \mathbf{R}$ as follows. If there exist $a, b \in [0,1]$ satisfying $[a,b) \subseteq A$, $[0,a) \cap A = \emptyset$ and $[a,b') \not\subseteq A$ for all $b' \in (b,1]$, then $Flength(A) = b-a$, otherwise $Flength(A) = 0$.

The binary operators $Lshift$ and $Rshift$ on $\mathbf{V}$ are defined in the following way using the characteristic function notation of interval-values. If $x \in [0,1]$ and $A, B \in \mathbf{V}$, then

$Lshift(A,B)(x) =$
$\begin{cases} A(x + Flength(B)), \text{ if } 0 \leq x + Flength(B) < 1, \\ 0 \text{ in other cases.} \end{cases}$

$Rshift(A,B)(x) =$
$\begin{cases} A(frac(x - Flength(B))), \text{ if } x < 1, \\ 0 \text{ if } x = 1. \end{cases}$

Here the function *frac* gives the fractional part of a real number, i.e., $frac(x) = x - \lfloor x \rfloor$, where $\lfloor x \rfloor$ is the greatest integer which is not greater than x. Now we explain the so-called *fractalian product* on intervals.
Let A and B be interval-values and $x \in [0,1)$. Then the fractalian product $A * B$ includes x if and only if $B(x) = 1$ and $A\left(\frac{x-\underline{B}}{\overline{B}-\underline{B}}\right) = 1$, where $\underline{B}$ denotes the lower end-point of the B-component including x and $\overline{B}$ denotes the upper end-point of this component, that is, $[\underline{B}, \overline{B})$ is the maximal subinterval of B containing x.

The idea and the role of this operation is similar to that of unlimited shrinking of 2-dimensional images in [14]. It will be used to connect interval-values of various resolution. The fractalian product of two interval-values is the result of shrinking the first operand to each component of the second one.

2.4. Syntax and semantics of computation sequences

In this subsection, we formalize the interval-valued computations of [9]. This formalisation is of Boolean network style, since equality or similar tests do not seem to be easily implementable for interval-values, just like in the case of optical computing (no tests for equalities on images)[14]. As usual, the length of a sequence S is denoted by $|S|$ and its i-th element by S_i. If $j \leq |S|$ then the subsequence containing the j first elements of S is denoted by $S_{\to j}$.

An *interval-valued computation sequence* is a nonempty finite sequence S satisfying $S_1 = FIRSTHALF$ and further, for any $i \in \{2, \ldots, |S|\}$, S_i is (op, l, m) for some $op \in \{AND, OR, LSHIFT, RSHIFT, PRODUCT\}$ or S_i is (NOT, l) where $\{l, m\} \subseteq \{1, \ldots, i-1\}$. The *bit height* of a computation is the number of the applied $PRODUCT$ operators in it.

The semantics of interval-valued computation sequences can be defined by induction on the length of the sequences. The *interval-value* of such a sequence S is denoted by $\|S\|$. Let $\|(FIRSTHALF)\|$ be the interval-value $\left[0, \frac{1}{2}\right)$ and the value of longer sequences be composed by the corresponding operations on the interval-values.

2.5. Decidability

In this subsection, we give the definitions concerning interval-valued computability and complexity.

Let Σ be a finite alphabet and let $L \subseteq \Sigma^*$ be a language. We say that L is *decidable by an interval-valued computation* if there is an algorithm A that for each input word $w \in \Sigma^*$ constructs an appropriate computation sequence $A(w)$ such that $w \in L$ if and only if $\|A(w)\|$ is nonempty. Furthermore, we consider $\overline{L}$ also decidable in this case.

This last remark makes it possible to test emptiness and, by applying set-theoretical operators, also to test whether $\|A(w)\| = [0,1)$. In [9], SAT was solved by a linear interval-valued computation in the following meaning.

We say that a language $L \subseteq \Sigma^*$ is *decidable by a linear interval-valued computation* if and only if there is a positive constant c and a logarithmic space algorithm A with the following properties. For each input word $w \in \Sigma^*$, A constructs an appropriate interval-valued computation sequence $A(w)$ such that $|A(w)|$ is not greater than $c \cdot (|w|)$

and $w \in L$ if and only if $\|A(w)\|$ is nonempty. Again, deciding $\overline{L}$ instead of L itself is allowed.

3. Visual expressiveness of interval-valued computations: Propositional logic

In [9] it was proved that the problem whether a propositional formula is tautology or not, is decidable by a linear interval-valued computation. We will show that the sequence of interval-values produced by this computation represents visually the information needed to follow the Boolean algebraic calculation proving or disproving that the input formula (ϕ) is a tautology ($\phi \in TAUT$). This representation is no less natural than the representation by Venn diagrams.

The solution was to give an algorithm for constructing a computation sequence $K_1, \ldots, K_{3n+m+1}$ for any input formula ϕ that contains exactly the variables $x_1, \ldots, x_n$ and the number of its subformulae is m. The algorithm provides the above computation sequence in such a way that its interval-value will be [0,1) if and only if $\phi \in TAUT$.

Let K_1 be $FIRSTHALF$. For all positive integers $k \leq n$, we define

$K_{3k-1} = (PRODUCT, 1, 3k-2)$,
$K_{3k} = (RSHIFT, 3k-1, 3k-2)$ and
$K_{3k+1} = (OR, 3k, 3k-1)$.

The following statement can be established:
For all positive integer k, if $k \leq n$ then

$$\|K_{\to 3k-2}\| = \bigcup_{l=0}^{2^{k-1}-1} \left[\frac{2l}{2^k}, \frac{2l+1}{2^k}\right).$$

The n independent truth values of $x_1, \ldots, x_n$ will be represented by the interval-values $\|K_{\to 1}\|, \|K_{\to 4}\|, \ldots, \|K_{\to 3n-2}\|$. See Figure 1, lines 1–4 for an example with $n = 4$, X_1 is $\|K_{\to 1}\|$, x_2 is $\|K_{\to 4}\|$, x_3 is $\|K_{\to 7}\|$ and x_4 is $\|K_{\to 10}\|$. We also can use connectives $\to, \equiv$, as they can be defined by $\neg, \vee$ and $\wedge$ by the usual way.

By this correspondence, we can build a mapping i from $\{0,1\}^n$ to $\mathbf{V}$ in the following way. If $(t_1, \ldots, t_n) \in \{0,1\}^n$ then $i(w)$ is $\left[\frac{v}{2^n}, \frac{v+1}{2^n}\right)$ where $v = \sum_{k=1}^{n} t_k 2^k$.

Let $\Phi_1, \ldots, \Phi_m$ be an enumeration of all the subformulae of Φ satisfying that any formula is preceded by its subformulae (consequently, $\Phi_m = \phi$). The algorithm gives the next part of the computation sequence $(K_{3n-2+1}, \ldots, K_{3n-2+m})$ in the following way. For each $i \in \{1, \ldots, m\}$,

$K_{3n-2+i} =$
$\begin{cases} (AND, 3n-2+j, 3n-2+k) \text{ if } \phi_i = \phi_j \wedge \phi_k, \\ (OR, 3n-2+j, 3n-2+k) \text{ if } \phi_i = \phi_j \vee \phi_k, \\ (NOT, 3n-2+j) \text{ if } \phi_i = \neg\phi_j, \\ (AND, 3j-2, 3j-2) \text{ if } \phi_i = x_j. \end{cases}$

By induction on j the following statement can be verified: *For each $j \in \{1, \ldots, m\}$, $\|K_{\to 3n-2+j}\| =$*

$\{r \in [0,1) :$
$\phi_j\,[[r \in \|K_{\to 1}\|, r \in \|K_{\to 4}\|, \ldots, r \in \|K_{\to 3n-2}\|]] = 1\}$.

Here $\psi[[t_1, \ldots, t_n]]$ denotes the truth value of ψ by truth valuation $(x_1 \to t_1, \ldots, x_n \to t_n)$. By this fact, once we have i, we represent visually all the subformulæ of the input formula by a homomorphism h from the Boolean algebra of the propositional formulæ into $\mathbf{V}$ as follows.

$h(\psi) = \bigcup\{i(t_1, \ldots, t_n) : \psi[[t_1, \ldots, t_n]] = 1\}$.

In Fig. 1 line 5 shows the resulted interval-value of the propositional formula C_0. This representation is very natural and easily realizable even for 5-6 and more variables. In principle, this representation works with an arbitrary number of propositional variables, only the resolution of the visual representation give a limit. Having the interval-values of the propositional variables the solution of SAT or tautology problems is based on truth-tables: the interval-values represent bit-sequences. The following part of the representation is more interesting.

4. Visual solution of a PSPACE-complete problem

We show the visual reasoning power of the interval-valued computations by a PSPACE-complete problem, namely, the problem whether a quantified propositional formula is true. It is also decidable by a linear interval-valued computation ([11]). Also for this case, we show that the sequence of interval-values produced by this computation represents visually the full information needed to understand the solution of the given case of the problem.

We continue the computation sequence in the previous section with $K_{3n-2+m+1}, \ldots, K_{3n-2+m+8n}$ in such a way that for each integer $j < n$, the following holds:

$\|K_{\to 3n-2+m+8(j+1)}\| =$
$((Lshift(\|K_{\to 3n-2+m+8j}\|, \|K_{\to 3(n-j)-2}\|) \cap$
$\|K_{\to 3(n-j)-2}\|)$
$\cup$
$\|K_{\to 3n-2+m+8j}\|)$
$\cup$
$((Rshift(\|K_{\to 3n-2+m+8j}\|, \|K_{\to 3(n-j)-2}\|) \cap$
$\overline{\|K_{\to 3(n-j)-2}\|})$
$\cup$
$\|K_{\to 3n-2+m+8j}\|)$,
if $n-j$ is even,
and
$(Lshift(\|K_{\to 3n-2+m+8j}\|, \|K_{\to 3(n-j)-2}\|) \cap$
$\|K_{\to 3(n-j)-2}\| \cap$
$\|K_{\to 3n-2+m+8j}\|)$
$\cup$
$(Rshift(\|K_{\to 3n-2+m+8j}\|, \|K_{\to 3(n-j)-2}\|) \cap$
$\overline{\|K_{\to 3(n-j)-2}\|} \cap \|K_{\to 3n-2+m+8j}\|)$
in the other case.

In this definition, we did not specify all the intermediate expressions between $K_{3n-2+m+8j}$ and $K_{3n-2+m+8(j+1)}$, they are the subexpressions of $K_{3n-2+m+8(j+1)}$ needed

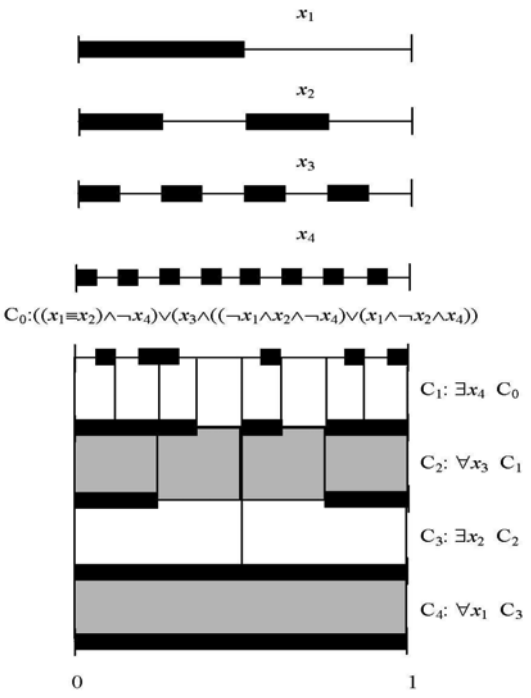


Figure 1. Example for testing validity of a quantified propositional formula

to express $K_{3n-2+m+8(j+1)}$ from $K_{3n-2+m+8j}$ and $K_{3(n-j)-2}$.

We assume without any further mention, that variables $t_1, t_2, \ldots, t_n$ range over the truth values. We recall that the quantifier sequence $Q_1, Q_2, Q_3, \ldots$ is defined as $\forall, \exists, \forall, \ldots$, respectively. With these notations, the following holds:

For each $j \in \{0, \ldots, n\}$ *and for all* $r \in [0, 1)$*:*

$r \in \|K_{\to 3n-2+m+8j}\|$ *if and only if*

$Q_{n-j+1} t_{n-j+1} \ldots Q_n t_n$

$\Phi\big[\big[r \in \|K_{\to 3\cdot 1-2}\|, \ldots, r \in \|K_{\to 3(n-j)-2}\|,$

$t_{n-j+1}, \ldots, t_n\big]\big] = 1.$

The last statement is the base of our visual representation. The sequence of interval-values $\|K_{\to 3n-2+m+8j}\|$, $(j \in \{0, \ldots, n\})$, shows visually the steps of computation corresponding to the application of propositional quantifiers.

A $\forall$-step means checking the interval AND its corresponding neighbour, while an $\exists$-step amounts to checking the interval OR its neighbour.

In this way the linear computation sequence of interval-values includes visually all the information needed to follow a proof for a quantified propositional formula.

In Figure 1 one can follow a detailed interval-valued computation deciding whether a given quantified propositional formula is true.

5. Conclusions

While interval temporal logic ([1]) deals with problems about interval-values representing non-contiguous events, our system efficiently computes classical decision problems with the help of computations on such interval-values. By the visual power of the system one can easily understand the reasoning itself. Our system does not use the interval-relations of [8], but some of these relations can be expressed by the logic of interval-values. To make explicit comparisons of these systems is a topic of future work.

6. Acknowledgements

The work has been supported by the grants OTKA T049409, by the Hungarian Ministry of Education and by the KPI and NKTH.

References

[1] J. Allen. Maintaing information about temporal intervals. *Communications of the ACM*, 26:832–843, 1983.

[2] B. Nagy and G. Allwein. Diagrams and Non-monotonicity in Puzzles. *Proceedings of Diagrams'2004, Cambridge, England, LNCS/LNAI* 2980:82–96, Diagrammatic Representation and Inference, eds.:A. Blackwell, K. Marriott, A. Shimojima, 2004.

[3] D. E. Anderson and F. L. Cleaver. Venn-type diagrams for arguments of n terms. *J. Symbolic Logic*, 30:113–118, 1965.

[4] P. Balbiani, J.-F. Condotta, L. F. del Cerro and A. Osmani. Reasoning about generalized intervals. *in: F. Giunchiglia (ed), Artificial Intelligence: Methodology, Systems and Applications, LNCS-LNAI* 1480:50–61, 1998.

[5] K. B. Chilakamarri, P. Hamburger and R. E. Pippert. Venn diagrams and planar graphs. *Geometriae Dedicata*, 62:73–91, 1996.

[6] A. W. F. Edwars. *Cogwheels of the Mind: The Story of Venn Diagrams*. John Hopkins University Press, 2004.

[7] D. W. Henderson. Venn diagrams for more than four classes. *American Mathematical Monthly*, 70:424–426, 1963.

[8] Z. Kulpa. A diagrammatic approach to investigate interval relations. *J.Visual Lang.and Computing*, 17:466-502, 2006.

[9] B. Nagy. An interval-valued computing device. *Proceedings of CiE2005, Computability in Europe: New Computational Paradigms, Amsterdam eds: S. B. Cooper, B. Loewe, J. Tucker, 166-177*.

[10] B. Nagy. Reasoning by intervals. *4th International Conference on the Theory and Applications of Diagrams, Stanford(CA), USA, LNCS/LNAI* 4045:145–147, 2006.

[11] B. Nagy and S. Vályi. Solving a pspace-complete problem by a linear interval-valued computation. *Proceedings of CiE2006, Computability in Europe: Logical Approaches to Computational Barriers, Swansea, Report no. CSR-7-2006, eds. A. Beckmann, U. Berger, B. Loewe, 216–225*.

[12] B. Nagy and S. Vályi. Interval-valued computations and their connections with PSPACE. *Theoretical Computer Science*, accepted.

[13] F. Ruskey and M. Weston. A survey of Venn diagrams. *Electronic Journal of Combinatorics*, 12, 2005.

[14] D. Woods and T. Naughton. An optical model of computation. *Theoretical Computer Science*, 334:227–258, 2005.

Visual Environment for the Taylor integration in 3D Stereo

Alexander Gofen

The Smith-Kettlewell Eye Research Institute, 2318 Fillmore St.,
San Francisco, CA 94115, USA

A cross-platform application called the Taylor Center implements advanced user interface and 2D or 3D stereo dynamic visualization of the solution. It integrates ODE's using the Modern Taylor method, having unlimited order of approximation, finite step, and high accuracy, allowing to explore or export the solution in different formats. Designed in Delphi-6, the package implements a powerful Graphical User Interface, and is portable to windowed versions of Linux too.

Introduction

What we expect from an application with advanced interactive visualization is:

(a) User interface whose style, controls, and handlers fit best for the specific model and operational tasks;
(b) Realistic visualization of the modeled process employing all appropriate faculties of the human perception, achievable with advanced hardware and multimedia.

With that in mind we consider a general purpose program called the Taylor Center, which performs integration of initial value problems for systems of Ordinary Differential Equations (ODE's) [7,8]. Unlike earlier packages [1,2], this sophisticated Taylor Solver is designed for PCs as an interactive application with extensive Graphical User Interface. Moreover, it was developed in a programming environment which exemplified the Visual programming and made its good name: the Borland's Delphi (a dialect of Pascal having such advance features as dynamic arrays, objects, and sophisticated graphics). Actually the software may be considered as a cross platform Windows/Linux project because Borland offers also a Linux implementation of Delphi (called Kilex).

Although not any mathematical object may be readily visualized, the solutions of ODE's are, as their geometrical interpretation is trajectories in Euclidian space $\boldsymbol{E}^n$ (planar curves for $n=2$). In a computerized world we can expect not just still images of trajectories, but also real time animation of the motion along them, and we can even expect visualization for $n=3$ too.

This All-In-One application is called the Taylor Center (according to the metaphor of a "Music Center") because it offers an interactive environment, where researchers can input (import) ODE's in symbolic format, vary the parameters, explore the Taylor expansions and their convergence radius, integrate the equations with high accuracy (up to all available float point digits), store or export the results, graph the solutions and "play" them dynamically as bullet trajectories in real time animation. The feature making this application unique is that it can display non-planar 3D trajectories as anaglyphic stereo viewed through Red/Blue glasses [6,7]. In particular, it implements a 3D cursor (controlled by a conventional mouse). This cursor with "tactile" audio feedback allows users to "touch" points of non-planar curves hanging in thin air, while viewing the current 3D coordinates. Here the multimedia is employed in its full capacity (especially because of the stereo viewing – not yet ubiquitous in the computer world).

Particularities of the 3D visualization

The perspective drawing of 3D world (isometry) is known for a couple of centuries. Human imagination easily interprets perspectively correct planar images of common objects, especially those with strong perspective cues, but that is not always the case, and then human capability of stereo perception and tools for stereo vision come to play.

Stereo vision is a powerful ability of our brain to recreate an image of 3D scene[1] "fusing" the two planar images received by the eyes [3, 6]. Mental image of 3D space belongs completely to the realm of imaginary sensations, because there is no place in our body where any kind of 3D replica of reality is materialized. The only physical replica we do

[1] Mathematically speaking, recreation of the 3D scene from a pair of 2D images is not a well defined problem, and it may have multiple solutions.

have is the two planar images on the retina of our eyes.

The real excitement of the 3D vision happens when we look at objects expected to be planar (a screen, a hologram), but the brain suddenly discovers a 3D scene in it, popping out from the planar surface or shifting back into the depth.

This is the 3D Stereo Vision, or the real 3D – not to be confused with the misnomer "3D Graphics" used for video cards and video libraries. This misleading term emerged when the game and video technology switched from simplistic projections (known as front, side and top views in drawing) to a more advanced oblique projection of 3D scenes. Photos, isometric drawings, artistic paintings (if their perspective is correct) all are examples of such projection. When we look at them, we perceive them exactly as planar 2D projections of 3D reality[2].

We have all become good at "reading" these projections and understanding the 3D reality they represent. Such projections were the closest we could come to 3D imagery until invention of the stereoscope in 19th century in Great Britain.

The two eyepieces of a stereoscope deliver the two planar images of a stereo pair directly to the respective eyes. Then a miracle occurs: viewers perceive the 3-dimesionality of the scene as real – so real that they wish to touch the objects hanging in air. This never happens when we look at 2D projection of the 3D world.

After the stereoscope, many other techniques were developed for displaying 3D stereo images [4]. Except for the holograph, they all use the same basic principle as the stereoscope: to deliver each of the two images of a stereo-pair to the proper (and only to the proper) eye. On the contrary, the holographic equipment completely reproduces the 3D light front, i.e. the 3D vector field of the same electromagnetic waves that would be reflected by the 3D scene if it were really there. Thus a hologram creates a sculptural "ghost" in the real world as a physical phenomenon – a 3D scene we view by a pair of eyes, while a stereo pair creates this scene in our mind – an ultimate addressee anyway. Another fundamental difference is in the quantity of information: a stereo pair is just two planar images, while a hologram encodes (almost) infinity of them.

Does the stereo vision give some advantages vs. a planar perspective projection of a 3D scene? The isometrics (or photo-imaging) suffice in many, but not in all situations. In displaying such objects as scenes with rectangular shapes, bodies with edges, skeletal structures, scenes with good perspective cues or with reflection hints, isometry is good enough. For them, the stereo viewing only adds certain excitement in perception.

Yet the isometrics works poorly for smooth surfaces without edges (smooth 2D manifolds). For example, to perceive a picture of a sphere or a torus, we additionally draw grids on their surfaces, or special shadows and reflections. The 2D projections become especially inefficient while visualizing non-planar curves (1D manifolds) – and that is what solutions of ODEs are. For example, trajectories of space probes launched to planets of Solar system are usually non-planar. To draw them in conventional isometry we need to add several auxiliary planes and referencing coordinate lines. The real stereo viewing eliminates necessity in all auxiliary hints (see the 3D demo samples). Thus, displaying non-planar solutions of ODEs in 3D stereo dramatically improves the perception[3].

This particular software employs the cheapest technical realization of stereo vision, known as anaglyph stereo, which requires nothing more than a conventional PC monitor plus cardboard glasses – the cheapest gear possible [6]. The two images of an anaglyphic stereo pair must be monochromatic – each in a different color (one of the basic Red/Green/Blue set). The Red/Blue pair is better because their spectrums are farther away from each other and better separable. When you look through the Red/Blue glasses at the Red/Blue images of a stereo pair that overlap on a screen, each eye receives only the proper image of the stereo pair.

How does the brain fuse stereo pairs presented in different colors? After all it evolved and is trained to match *corresponding* elements of stereo pairs, and the *corresponding* elements must be of the same color indeed.

Fortunately, the brain follows not a "rigid", but rather a "flexible" algorithm. Certain neurons fire if they find out recognizable elements in the image disregarding their colors, while the others

[2] Binocular viewing of a perspectively correct planar image creates perception of a planar drawing only: isometric or not, yet in a plane. That is because this plane is the *solution* corresponding to the 3D reality encoded in this stereo pair. Surprisingly enough, if the same isometry is viewed by one eye at the perspectively correct distance, with a little effort the brain does generate the stereo perceptions of this very scene reconstructing it out of the perspective cues – a phenomenon called monocular stereopsis.

[3] There are many other cases when stereo viewing becomes essential, being implemented in various equipment from micro- to telescopes [4]. For example it is used to visually detect counterfeit bills, or changes in aerial photos of landscapes.

fire if these elements happen to comprise a "meaningful" stereo pair. The signals of those "stereo-specialized" neurons dominate. As a result we easily fuse the 3D scene in spite of the fact, that each eye sees it in a different color. This works well both on a PC monitor for a single viewer, and with a screen projector for a large audience.

The modern technology nowadays offers a variety of other technical solutions with completely uncompromised stereo viewing.

Particularities of the integration method: more than tabulated functions

The method of integration implemented in this software is the modern Taylor method – to be distinguished from its classical counterpart in that the process of *n*-order differentiation was optimized and made feasible due to the concepts of Automatic Differentiation (AD) and (generalized) elementary functions [5].

Among other methods of integration, the Taylor method is situated somewhere between purely numerical methods and symbolic integration. However, unlike purely numeric methods, it delivers not only tabulated values of unknown functions in point of interests, but also their derivatives up to any specified order (default 30) and convergence radius at every point. And unlike symbolic integration, the Taylor method applies not only to very limited list of ODEs solvable in quadratures, but to any ODEs in regular points of the phase space.

From the programmatic point of view, any Taylor solver differs from conventional integrators in that the input – an Initial Value Problem for ODEs – should be provided not simply as a subroutine computing the right hand parts, but rather as arithmetic expressions themselves of those right hand parts in order to enable automatic differentiation (understood as optimized classical formulas for *n*-th order derivatives applied to sequences of basic expressions).

Correspondingly, the output and interaction of a Taylor solver with other applications has its specifics too. The result is not just tabulated values of the solution, but rather its expansion into the Taylor series (analytical elements), or a sequence of such elements (although the tabulated solution may be exported also).

The tasks and environment at a glance

This section walks you through the basic features of the Taylor Center, pre-loaded with fascinating classical problems. You will be able to "play" them in real time animation, to experiment with a 3D cursor, and to study effects of specific parameters. To install the program [8], unzip it into an empty folder of your choice and run TCenter.exe. To try first what makes this program unique, begin with the 3D stereo and have Red/Blue glasses ready.

Go to *Demo/Three Bodies/Disturbed/3D* in the menu. The compilation script runs and ends with a message "Compilation successful", displaying a knotty Red and Blue curves. Now put on your anaglyphic glasses (over those you usually use, if any) and maximize the Graph window.

What you hopefully perceive looks like a "fishing line" hanging in thin air between the monitor and your face. These are trajectories of three bodies moving under gravitational pull. More specifically, this is the so called disturbed Lagrange case. (See Lagrange case proper under *Three Bodies/Symmetrical*). This "fishing line" is a result of small disturbances applied perpendicularly to the initial plane (the plane of your screen).

However the program is capable of producing something more than "still life". Click the Play button, initiating real time 3D stereo motion of the bullets representing the three bodies, and watch how they accelerate, decelerate, and couple.

When they come to rest, you may try exploring the elements of the trajectories with a "tactile" 3D cursor. Move it into the scene, where it will transform into a small cross. The mouse always moves the stereo cursor in a plane parallel to the screen. To control its depth, use the mouse wheel. (If the mouse wheel is not available, *move* the mouse keeping depressed either Ctrl key to bring the cursor closer to your eyes, or Shift key – to move it away from you). Current 3D coordinates of the cursor always appear at the top window panel.

Now, controlling the 3D cursor, try to touch one of the trajectories in space with it. Switch the speakers On, and you will hear a clicking sound when the touch occurs: this is the so called "tactile" audio feedback, helping to explore points of interest in the curves.

You can rotate the curves in the space with the Turn controls. Given specific sizes of the parallelepiped, you may notice that the front side (controlled by MaxZ value) keeps the curves inside the parallelepiped "flattening" them. (Therefore increase MaxZ).

Already familiarized with the 3D stereo features of the package, you may try several other problems. Click Main Panel in the menu to re-visualize the main form, and go to *Demo/Four Bodies*. The two pairs of bodies with equal masses are all initially placed in a horizontal plane, parallel

to your desk (perpendicular to the screen). The horizontal components of the velocities provide near circular motion for each coupled pair, while the small vertical components push the two pairs into a large circular motion around the center of the masses (see the initial values in the Main window). At the beginning the trajectories spin into a braid looking as though they outline a torus (like the tiny braided rings of Saturn discovered by the Voyager probe). However this braid actually does not outline a torus: you can notice that both coupled pairs preserve their initially horizontal plane.

Another fascinating example of 3D motion is under *Demo/Möbius*. You can watch 4 bullets lined up in a straight line whose motion outlines a Möbius surface winded 1.5. To get the simplest one (winded 0.5), change value of *n=0.5* (in Constants), *Compile*, click button *Previous* (in Graph setting page), click *Clear* in Graph window, and finally click the *More* button.

You can explore several more 3D stereo examples opening their scripts. Click the Main Panel and go to *File/Open* script menu item. Here are files producing 3D stereo images:

PendulumApple.scr, PendulumFlower.scr (spherical pendulum);

KnotChain3D.scr, TrefoilKnot3D.scr;

MobiusLarge.scr;

Beside 3D stereo samples, there are also instructive examples in 2D, such as the recently discovered eight-shaped solution of the three body problem called "Choreography" (in *Three Bodies/Choreography*). Under *File/Open* script there are also two more classical examples in celestial mechanics: the Euler case with the bodies of equal masses (3EqBodEuler.scr) and the case when one mass is near zero (3NonEqBodEuler.scr). There are also scripts for single and double pendulums, and the Four body Lagrange case (4BodiesPlane.scr).

Conclusions

Summarizing all the features, with the current version of the Taylor Center users can:

- Specify and study the Initial Value Problems for virtually any system of ODE's in the standard format of explicit 1st order ODE's with numeric and symbolic constants and parameters;

- Perform numerical integration of Initial Value Problems with the high accuracy (up to all digits of 64-bit mantissa), while the step of integration remains finite and does not approach zero (because the order of approximation is high: by default 30, or higher).

- Obtain the solution as a set of analytical elements – Taylor expansions covering the required domain, exportable to other systems either as is, or in tabular format;

- Study Taylor expansions and the (heuristic) radius of convergence for the solution at all points of interest (with the only limitation that the terms in the series do not exceed the maximum value of about 10^{4932} implied by the 10-byte implementation of the real type *extended*);

- Perform integration either "blindly", or graphically visualized; either a given number of steps, or until an independent variable reaches a terminal value, or until a *dependent* variable reaches a terminal value;

- Switch integration between several versions of ODE's defining the same trajectory with respect to different independent variables. For example, it is possible to switch the integration by t to that by x or y in order to reach the terminal value (or zeros) of the dependent variable;

- Integrate piecewise-analytical ODE's;

- Specify different methods of controlling the accuracy and the step size;

- Specify accuracy for individual components either as an absolute or relative error tolerance, or both;

- Graph color curves (trajectories) for any pair of variables of the solution up to 7 on one screen, either as plane projections, or as 3D stereo images to be viewed through anaglyphic (Red/Blue) glasses. The 3D cursor with audio feedback (controlled by a conventional mouse) enables "tactile" exploration of the curves literally "hanging in thin air";

- "Play" dynamically the near-real time motion of bullets along the computed trajectories either as 2D or 3D stereo animation;

- Graph a field of directions with a grid of short lines or curves of variable length;

- Explore several meaningful examples supplied with the package such as the problem of Three or Four Bodies, "Choreography". Symbolic constants and expressions parameterize the equations and initial values making it possible

to try different initial configurations of special interest.

The current kernel version may be further developed in different directions (Fig 1):

- To implement other types of arithmetic (complex numbers, intervals);

- To generate compileble code in Assembler or in a high level language;

- To simultaneously integrate an *array* of Initial Value Problems;

- To compute derivatives by parameters.

- To implement special loops and array variables (to automate expressions of high complexity in the right hand sides such as sums of many terms).

References

[1] Chang, Y. F. ATOMFT User Manual. 1986. Claremont McKenna College, Claremont, CA, 91711

[2] Chang Y. F., Corliss G.: ATOMFT: Solving ODEs and DAE Using Taylor Series. Computers Math. Applic. Vol. 28, No. 10-12, 1994. pp. 209-233.

[3] Howard, I.P., Rogers B.J. Seeing in Depth, Vol. 2, Depth perception. Published by I. Porteous, Toronto, University of Toronto Press, 2002, 626 p.

[4] Ibid. Chapter 30, Stereoscopic Techniques and Applications, pp. 541-564

[5] Gofen, A. M.: ODEs and Redefining the Concept of Elementary Functions. In: P.M.A. Sloot et al. ICCS 2002, LNCS 2329. Springer-Verlag, Berlin 2002.

[6] Gofen, A. Stereo Vision on Your Home PC. Delphi Informant, Vol. 10, No. 1, 2004, pp. 8-15.

[7] Gofen, A. Interactive Environment for the Taylor integration (in 3D Stereo), in: "Proceedings of the 2005 International Conference on Scientific Computing" (CSC'05, June 2005, USA); Editors: H. R. Arabnia and G. A. Gravvanis, pp. 67-73.

[8] Gofen, A. M.: The Taylor Center Demo for PCs and the User Manual. www.ski.org/gofen/TaylorMethod.htm

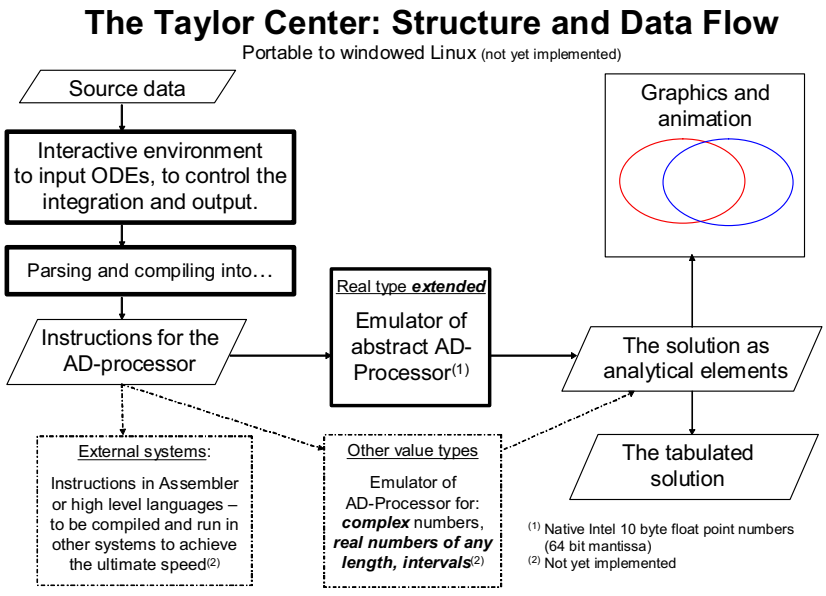


Fig 1. The outline of the Taylor center kernel. The source ODE's, symbolic constants and initial values are entered in editor boxes or from a file. The interactive environment controls how to integrate, what to graph, and what to export. The dotted boxes represent features possible in future developments.

Chorem Editing - Visual Summary of Spatial Database Content

Vincenzo Del Fatto[1, 2], Robert Laurini[2], Karla Lopez[2],
Monica Sebillo[1], Genoveffa Tortora[1], Maurizio Tucci[1], Giuliana Vitiello[1]
[1] *DMI, Università di Salerno, Italy*
[2] *LIRIS, INSA-Lyon, France*
{msebillo, tortora, mtucci, vitiello}@unisa.it
{Vincenzo.Del-fatto, Robert.Laurini, Karla.Lopez}@insa-lyon.fr

Abstract

The research described in this paper is part of an international project launched among several research institutions in order to define cartographic solutions able to better summarize spatial database content and represent geographic knowledge.
The solution proposed to achieve this aim is based on the chorem concept and its capability to associate a visual notation with a schematized representation of territory.
In this paper we focus on the embedded visual environment for chorem editing meant to enable expert users, such as cartographers, to add details for map customization.

Keywords: visual metaphors, spatial databases, digital cartography.

1. Introduction

Visual tools represent a notable support for several human activities, where the complexity of reality of interest can require considerable efforts by users with no familiarity with specific procedures.

Recent literature has demonstrated that an approach based on advanced (visual) applications embedding visual languages and environments results more effective in achieving the expected results, because it takes into account users' expertise and allows to customize their interaction in agreement with usability criteria. In particular, the usage of visual metaphors represents a solution when dealing with geographic data and applications. Here the double nature of data, which assembles descriptive and spatial properties, implies the inner complexity of many applications, such as geographic information systems, spatial decision support systems and spatial data mining. Then, it should be more productive to exploit proper visual representations to locate problems and new patterns, in order to bridge the gap between the expected applications and domain expert users.

To this aim a research program was launched among several research institutions in order to test whether cartographic solutions based on *chorems* can be more satisfying in representing visual summaries of datasets. Proposed in [1, 2], chorems are defined as a schematized representation of territory, where some salient aspects of the global vision are emphasized. Figure 1 gives an example of the France chorem map where the some aspects are stressed, namely the geometric shape is simplified, only some big cities are mentioned and two flows are depicted, which represent the trend of people migration.

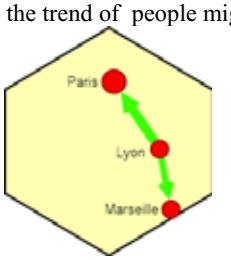


Figure 1. An example of chorem map

Then, this capability of chorems to visually describe the thematic content of a spatial dataset is the starting point to construct maps for spatial decision making.

The research we are carrying out arises from this assert, i.e. the necessity of providing expert users with a (semi)automatic system to build chorems, starting from datasets, even heterogeneous in data format and source. In particular, in this paper we describe the visual environment which will be embedded in the system in order to allow expert users to edit both chorems and chorem maps. It adds

a vector graphic editor the capability to jointly handle a descriptive component, expressed as a set of attributes associated with graphic features. The chorem editor imports an XML-like format and exports an SVG format, which can be then used by a traditional editor/browser.

The paper is organized as follows. Section 2 describes the main components of the explorative system, focusing on the *ChorML* language. In Section 3 the visual environment embedding the Chorem Editor is described, showing some functionality. Conclusions are given in Section 4.

2. An explorative system for chorem extraction and management

The concept of chorem, introduced by Brunet in [1, 2], is used mostly by geographers to manually depict basic features of a territory. The knowledge they represent is essentially coming from their familiarity with the territory, its history, the climatic constraints and the main sociological and economic problems. Features such as cities, population and specific phenomenon trends, represent the most common concepts that can be sketched by chorems.

Since then, the increasing popularity of chorems as metaphor to represent elements of interest, has involved new research fields, where the usage of chorems may represent an effective tool both to get an immediate idea of data in terms of visual summary and to derive new information starting from elementary visual patterns.

Recent literature has shown that when chorems represent either a visual summary of spatial database contents, or a global vision of a spatial database [4, 5], or visual geographic knowledge, it is possible to envision comprehensive software tools able to extract and associate chorems with a layout [5].

To this end, we have designed the explorative system whose architecture is shown in Figure 2. It consists of two main subsystems, namely the *Chorem Extraction Subsystem*, whose aim is to discover geographic patterns starting from spatial datasets, and the *Chorem Visualization Subsystem*, which assigns a visual representation to knowledge previously discovered.

The *Chorem Extraction Subsystem* provides for both traditional and spatial data mining techniques, each of them devoted to accomplish a specific task. In particular, the spatial pattern discovery task is meant to extract spatial knowledge by applying clustering and aggregation procedures together with *SELECT*s queries to the database. Such a knowledge may be then translated into a set of chorems, which are expressed by an GML-like representation, named *ChorML0*.

Once both inter-chorem and topological relationships are computed, a *ChroML1* version is obtained where proper tags are added in order to handle information about how chorems are spatially bound.

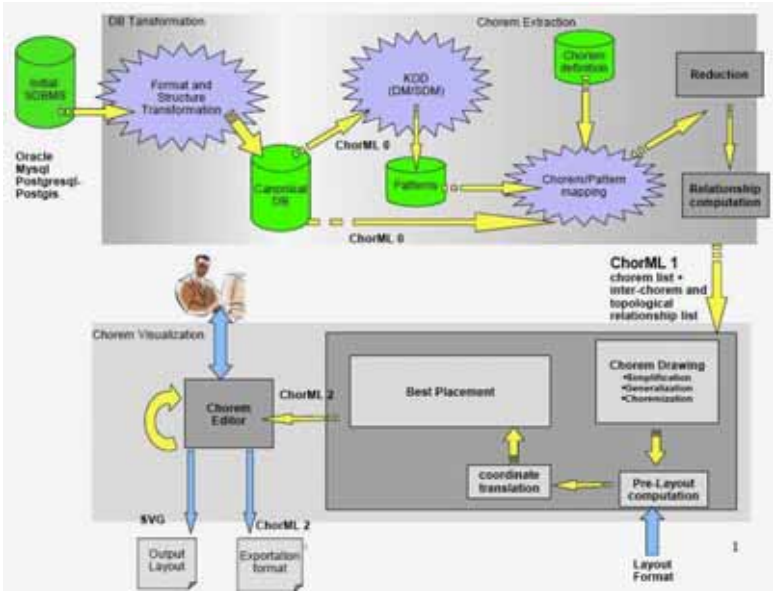


Figure 2. The architecture of the system prototype

Once the list of chorems and the set of constraints among them are obtained from the *Chorem Extraction* Subsystem, they are sent to the *Chorem Visualization* Subsystem in order to derive a visual representation of chorems and chorem maps, both in terms of layout and semantic content. As shown in Figure 2, five different tasks are performed by this subsystem, namely chorem drawing, coordinate translation, best-placement of chosen chorems, pre-layout computation and chorem editing. As for the chorem drawing, it is performed through three, not necessary interconnected, steps, named *simplification*, *choremization* and *generalization*, where some procedures and spatial operators are invoked in order to derive sketched representation of the involved elements.

Once the drawing of the chorem is obtained, users are asked to specify details about the output map, such as the number of colours and the final layout format (for instance A4). The latter affects the number of chorems that can be introduced onto a map, since it is necessary to guarantee the readability requirement.

The goal of next step consists of aggregating chorems onto the output map. This is accomplished by a multi-agent system whose aim is to spatially arrange chorems onto the chosen visualization format and determine their best placement [3], preserving structural and topological constraints among them. In order to guarantee the best placement requirement, independent sets of interrelated chorems may be aggregated onto different maps, thus providing users with more intuitive and readable chorem maps.

Some difficulties can occur regarding chorem placement and layout, as well as further refinements affecting semantic and graphic properties may be required by users. To this aim, users are provided with a tool for chorem editing as described in Section 3, which allows them to refine the expected output map.

3. The visual environment for chorem editing

The final goal of the explorative system we are developing, consists of providing expert users with visual representations of the spatial knowledge they own about spatio-temporal phenomena.

This phase may require the human intervention in order to adapt the chorem map to better express the geographic knowledge associated with it. It may be necessary because a basic dictionary is usually provided which contains a set of chorem types covering general concepts in a domain of interest. However, on specific demands more details should be added in order to enrich the semantics of a chorem, by modifying the graphic representation and/or the descriptive component, as well as the underlying layout.

The Chorem editor represents the tool suitable to the refinement process, which is the last step of this phase. In particular, the Chorem Editor may perform several tasks, from the import of a list of chorems, to their display and modification, from the export of *ChorML2* code to the generation of SVG-based representations.

First, it accepts a list of chorems positioned onto a chorem map and a set of constraints both expressed in terms of *ChorML1*. The editor transforms the tags related to the spatial position of elements into a graphical representation based on SVG [6]. Such a transformation is illustrated in Figure 3, where the involved tags are highlighted.

```
<?xml version="1.0" encoding="utf-8"?>
<chorml>
 <ChoreMaticMap>
  <MapId>map1</MapId>
  <MapName>La France</MapName>
  <GeneralInformation>
   <Date/>
   <Author/>
   <Format>
    <Height>100</Height>
    <Width>200</Width>
   </Format>
  </GeneralInformation>
  <ChoremList>
   <Chorem>
    <ChoremElementList>
     <ChoremElement>
      <ElementId>elem1</ElementId>
      <ElementName>Paris</ElementName>
      <ElementSize>4</ElementSize>
      <svg>
                <ellipse style="fill: #ff0000" cx="10.7406"
          cy="4.15625" rx="0.55625" ry="0.55625"/>
       <ellipse style="fill: none; fill-opacity:0;
          stroke-width: 0.1; stroke: #ff0000" cx="10.7406"
          cy="4.15625" rx="0.55625" ry="0.55625"/>
       <ellipse style="fill: none; fill-opacity:0;
          stroke-width: 0.01; stroke: #000000"
          cx="10.7406" cy="4.15625" rx="0.55625"
          ry="0.55625"/>
       <text style="fill: #000000; text-anchor:start;
          font-size: 0.7; font-family: sans; font-style: normal;
          font-weight: normal" x="8.5" y="4.375">Paris</text>
      </svg>
     </ChoremElement>
```

Fig. 3. An example of *ChorML2* code

At this point, users may interact with the environment in order to derive a visual representation that better matches with their mental model. They can modify both visual representation and semantic structure of chorems, without loss of consistency between them. Moreover, users may also solve problems regarding chorem placement and layout by changing chorem position, colours and shape.

Once a suitable visual representation for a chorem map is obtained, it can be saved in *ChorML2* for further refinements and manipulations. Finally, an SVG file can be exported, containing the necessary

information for the chorem map visualization through a traditional graphic editor/browser.

The Chorem Editor has been built as an extension of the *Magelan Graphics Editor*, an open source 2D vector graphics editor, based on Java programming language. It consists of two working areas, namely a property window and a visualization window, and a toolbar containing both a set of buttons and a tabbed list by which functionality may be invoked. In particular, the property window allows users to interact with and modify chorem properties, also affecting the visual representation. Analogously, the visualization window, which is meant to display the chorem map under construction, allows users to manipulate its graphic components, also affecting properties displayed into the property window.

Figure 4 shows the visual interface of the Chorem Editor, where a chorem map is under construction, containing two points with different sizes representing cities, a flow between them, representing a migration, and a region which spatially contains all these chorems.

In the Chorem Editor, chorems are distributed on different layers, thus allowing users to handle each graphic object independently of the others. Moreover, every change performed on the graphic properties affects also the corresponding attributes and vice versa.

4. Conclusions

The usage of chorems as visual summary of spatial databases supports expert users to obtain an immediate synthesis of semantic data content. Such a synthesis may be the starting point for further processing tasks aimed to derive spatial analysis data, as well as to support expert users in decision making.

This paper describes a system prototype embedding a visual environment for chorem editing, whose functionality is defined according to some exhibited guidelines, such as the consistency between spatial and descriptive components and the provision of a domain-based vocabulary of elementary chorems.

Further modules are being implemented, namely the Legend Editor which associates a chorem map with a set of symbols useful to interpret it, and the *ChorML* editor for users able to directly manage map content through the textual language.

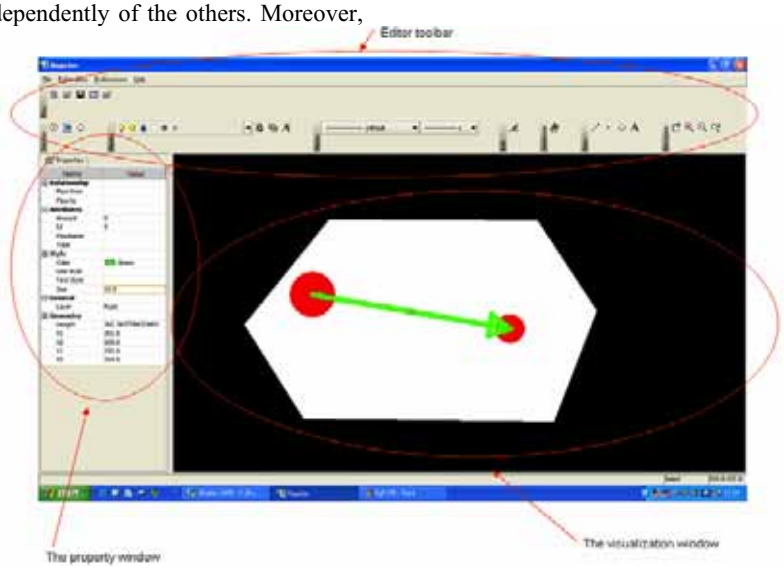


Figure 4. The Chorem Editor interface

5. References

[1] Brunet R. (1986) "*La carte-modèle et les chorèmes*", Mappemonde 86/4 pp. 4-6.

[2] Brunet R. (1993), "Les fondements scientifiques de la chorématique", in *La demarche chorématique*, Centre d'Études Géographiques de l'Université de Picardie J. Verne.

[3] Jones C.B. (1989) "*Cartographic Name Placement with Prolog*". IEEE Computer Graphics and Applications. Volume 9, Issue 5 (September 1989), pp. 36 – 47.

[4] Laurini R., Milleret-Raffort F., Lopez K. (2006) "*A Primer of Geographic Databases Based on Chorems*". In proceedings of the SebGIS Conference, Montpellier, Published by Springer Verlag LNCS 4278, pp. 1693-1702.

Reviewers' Index

A
Giovanni Adorni
Nizar Akil
Saud Aldarmi
Frederic Andres
Alfio Andronico

B
Aziz M. Barbar
Elisa Bertino

C
Jinwei Cao
Alfonso F. Cardenas
Heng-Shuen Chen
Ing-Ray Chen
Shu-Ching Chen
Luigi Colazzo
Kendra Cooper
Maria Francesca Costabile
Philip Cox

D
Stephan Diehl
Chabane Djeraba
Jing Dong
David H. C. Du
Jean-Luc Dugelay

E
Larbi Esmahi

F
Filomena Ferruci
Farshad Fotouhi
Rita Francese

G
Moncef Gabbouj
Forouzan Golshani

H
Masahito Hirakawa
Kwei-Liang Ho
Maolin Huang
Kao-Shing Hwang

I
Pedro Isaias

J
Joemon Jose

K
Mohan Kankanhalli
Kusai Kayali
Souhail Khawatmi
Ammar Kheirbek
Jun Kong
Zenon Kulpa

L
Robert Laurini
Michael Lew
Fuhua Lin
Chien-Tsai Liu
Jonathan Liu
Benjamin Lok

M.
Kim Marriott
Brahim Medjahed

Gabriele Meiselwitz
Atif Memon
Rym Mili
Piero Mussio

N
Marc Najork

O
Vincent Oria

P
Sethuraman Panchanathan
Luca Paolino
Minho Park
Joseph J. Pfeiffer
Giuseppe Polese

R
Syed M. Rahman

S
Giuseppe Scanniello
Ishwar Sethi
Timothy K. Shih
Hyunju Shim
Peter Stanchev
Nenad Stankovic
Georgi Stojanov
David Stotts
Yu-Huei Su
Hari Sundaram
Mahbubur R Syed

T
Gnag Tan
Genny Tortora
Noam Tractinsky
Goran Trajkovski

U
Joseph E. Urban

V
Athina Vakali
Athanasios Vasilakos
Son T. Vuong

W
Ellen Walker
Huahui Wu

X
Kui Xie

Y
Ling-Ling Yang
Changhe Yuan

Z
Kang Zhang
Taieb Znati

Authors' Index

DMS 2008 Call For Papers

The Fourteenth International Conference on Distributed Multimedia Systems

Hyatt Harborside at Logan Int'l Airport, Boston, USA
September 4 - September 6, 2008
Organized by
Knowledge Systems Institute Graduate School

SCOPE

The DMS conference is an international conference series, which covers a wide spectrum of technique discussions, demonstrations, and student program/paper contests in the fields of distributed multimedia computing. Started in 1994, the series of conference has been held at Hawaii, Hong Kong, Vancouver, Taipei, Aizu-Wakamatsu, Japan, San Francisco, Miami and Canada. This time, the conference will be held in the Hyatt Harborside at Boston's Logan Int'l Airport, Boston, USA.

TOPICS

DMS 2008 focuses on techniques, systems, applications, and theories in the fields of distributed multimedia computing. The conference organizers seek contributions of high quality papers, panels or tutorials, addressing various aspects of distributed multimedia systems and applications, for presentation at the conference and publication in the proceedings.

Topics of interest include, but are not limited to:

audio and video compression
MPEG, Quicktime, Windows API standards
image, video, audio content analysis and indexing/retrieval
image, video and audio watermark
3D audio and video
computer graphics and animation
modeling and analysis of distributed multimedia systems
OS support for distributed multimedia systems
distributed multimedia databases and computing
distributed object management
multi-paradigmatic information retrieval
multimedia human-computer interaction
multimedia communications and network architectures
mobile networks and mobile computing
multimedia stream synchronization and QoS control/scheduling
multimedia software engineering
multimedia processors and ASIC
multimedia technologies for people with disability
intelligent multimedia computing
intelligent multimodal interaction
multi agents, mobile agents and their applications
communication encryption
security of distributed computer systems
sensor networks for multimedia computing
Web servers and services
XML applications
Java, VRML and multimedia languages
visual and multidimensional languages for multimedia applications
multimedia digital libraries and mail systems
multimedia applications for CSCW
multimedia authoring tools and intelligent tutoring
cultural heritage multimedia applications
tele-conferencing, tele-medicine and tele-lecture
virtual reality, distributed virtual environment, and their applications
virtual school, virtual university, and virtual learning community
distance learning methodology, tools and systems
e-commerce, e-education and e-entertainment

The use of prototypes and demonstration video for presentations is encouraged.

WORKSHOPS AND SPECIAL SESSIONS

This year, DMS 2008 will be held in conjunction with workshops, conferences and special sections. Papers submitted to workshops or special sessions are invited by the program committee chairs of the workshops/sections. This joint organization aims to collect research results from different perspectives. The following workshops and conferences are being planned:

International Workshop on Visual Languages and Computing
International Workshop on Mobile Systems, E-Commerce and Agent Technology
International Workshop on Distance Education Technology

Please contact the conference program co-chairs if you are interested in organizing a workshop or a special session. A one-page proposal with the following items is required:

Title of the workshop/special session
Name of the program committee chair(s)
A list of program committee members
E-mail address of the corresponding program committee chair
A brief description of the theme

Each special session should include at least 5 invited (or accepted) papers. Each workshop should include at least 2 special sessions (10 papers). Paper review process of the workshop/special session will be arranged by the individual program committee chair. It is the responsibility of the individual chairs to ensure the quality of papers in the workshop/special sessions. The schedule of workshops/special sessions will be the same as the main conference (See Important Dates below). Each individual program committee chair can set up the Web site of the workshop/special session. However papers will be printed in the same volume as the main conference.

SPECIAL ISSUE OF JOURNALS

The DMS conference is closely coordinated with the International Journal of Software Engineering and Knowledge Engineering, the Journal of Visual Languages and Computing and the International Journal of Distance Education Technologies. Usually the best ranked papers from the conference, after rigorous reviews, extensive revisions and further enhancements, will appear in one or more special issues of the above journals. Papers suggested for the special issue(s) will be reviewed by external reviewers following the standard procedure of review stipulated by the respective journal.

CONFERENCE SITE (HOTEL INFORMATION)

The DMS 2008 Conference will be held in the Hyatt Harborside at Boston's Logan Int'l Airport, Boston, USA. The hotel has made available for these limited dates (9/3/2008 - 9/7/2008) to DMS 2008 attendees a discount rate of $149 US dollars for single/double, not including sales tax.

INFORMATION FOR AUTHORS

Papers must be written in English. An electronic version (Postscript, PDF, or MS Word format) of the full paper should be submitted using the following URL: http://conf.ksi.edu/dms08/submit/SubmitPaper.php. Please use Internet Explorer as the browser. Manuscript must include a 200-word abstract and no more than 6 pages of IEEE double column text (include figures and references).

INFORMATION FOR REVIEWERS

Papers submitted to DMS'08 will be reviewed electronically. The users (webmaster, program chair, reviewers...) can login using the following URL:
http://conf.ksi.edu/dms08/review/pass.php.
If you have any questions or run into problems, please send e-mail to:
dms08@ksi.edu.

DMS 2008 Conference Secretariat
Knowledge Systems Institute Graduate School
3420 Main Street
Skokie, IL 60076 USA
Tel: 847-679-3135
Fax: 847-679-3166
E-mail: dms@ksi.edu

If you cannot submit electronically, please send four copies of the complete manuscript to the above postal address.

IMPORTANT DATES

March 1, 2008 *Paper submission due*
May 1, 2008 *Notification of acceptance*
June 1, 2008 *Final manuscript due*
June 1, 2008 *Early conference registration due*