DMS 2013

Brighton - Seafront, United Kingdom
August 8-10

Proceedings of the Nineteenth International Conference on Distributed Multimedia Systems

PROCEEDINGS

DMS 2013

The 19th International Conference on Distributed Multimedia Systems

Sponsored by

Knowledge Systems Institute Graduate School, USA

Technical Program

August 8-10, 2013

Holiday Inn, Brighton - Seafront, United Kingdom

Organized by

Knowledge Systems Institute Graduate School

ISBN 1-891706-34-9 (Paper)
ISSN 2326-3261 (Print)
ISSN 2326-3318 (Online)

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

Printed by Knowledge Systems Institute Graduate School

DMS 2013 Foreword

Welcome to DMS 2013 – the 19th International Conference on Distributed Multimedia Systems!

Multimedia is now ubiquitous in computing and continues to evolve, advance and expand in many exciting new areas. The main theme of the 19th International Conference on Distributed Multimedia Systems (DMS 2013) is multimedia inspired computing. With the recent evolutions of media, social media and semantic computing, a large range of new research areas and applications have arisen. New challenges include: data modeling and analytics, data rendering and presentation, user interaction, massively parallel programming, big data, semantic and intelligence, simulation and reasoning, and beyond. The new challenges and research strands are particularly important for an increasing number of end-users, domain experts, IT professionals, practitioners. Together, we are going to have a large impact on productive activities, making positive changes in the economic trend and combating recession.

The conference is organized in sessions to focus on a variety of specialized themes, including: Slow Intelligence Systems, Multimedia Content Modeling, Software Development and Cloud Services, Multimedia & Gaming, Reasoning and Visualization of Computational Processes, Collaboration & Assessment, Visual Art and Design & Human-Machine Interface Design, Learning Objects, Visual Languages & Visual Programming. Two workshops which extended the main conference offerings, and complemented the conference program are: the International Workshop on Distance Education Technologies (DET 2013), and the International Workshop on Visual Languages and Computing (VLC 2013).

The paper selection was based upon a rigorous review process, with an acceptance rate of 43%. The conference program also includes short papers, posters and demo to discuss ongoing research activities and applications.

DMS 2013 is pleased to welcome Dr Roberto Laurini as our keynote speaker. Roberto is a Professor Emeritus at the University of Lyon, France, and a KSI Fellow. He is internationally renowned in the domains of geographic information systems, geographic knowledge management and multimedia information systems.

The DMS conference continues to be an international research hub. This year we are expecting authors and guests from 16 countries: Australia, Austria, Brazil, Canada, China, Czech Republic, Germany, Ireland, Italy, Japan, Mexico, Saudi Arabia, South Africa, Taiwan, UK, and USA.

As program co-chairs, we would like to express our appreciation and gratitude to the dedicated international program committee members, conference support staff, particularly the Steering Committee Chair Dr S.K. Chang and his team from the Knowledge Systems Institute, all authors who submitted papers, presenters and participants. Thank you all for contributing to the success of the conference.

We are delighted to welcome you to the 19th International Conference on Distributed Multimedia Systems (DMS 2013). We hope that you will find this year's conference exciting and rewarding with synergetic and enthusiastic exchange of latest research ideas, and we hope that you will also find some time to enjoy the beautiful sights, sounds, and tastes of the south coast of Great Britain.

Paolo Nesi and Kia Ng
DMS 2013 Program Co-Chairs

The 19th International Conference on Distributed Multimedia Systems (DMS 2013)

August 8-10, 2013
Holiday Inn, Brighton - Seafront, United Kingdom

Conference Organization

Steering Committee Chair

Shi-Kuo Chang, *University of Pittsburgh, USA*

Conference Co-Chairs

Gem Stapleton, *University of Brighton, UK*
Giorgio Valle, *Univ. of Milan, Italy*

Program Co-Chairs

Paolo Nesi, *University of Florence, Italy*
Kia Ng, *University of Leeds, UK*

DMS'13 Demo Co-Chairs

Francesco. Colace, *University of Salerno, Italy; Demo co-chair*
Pierpaolo Di Bitonto, *Univ. of Bari, Italy; Demo co-chair*

Program Committee

Junia Coutinho Anacleto, *Federal University of Sao Carlos, Brazil*
Arvind K. Bansal, *Kent State University, USA*
Andrew Blake, *University of Brighton, UK*
Augusto Celentano, *Universita Ca Foscari di Venezia, Italy*
Ing-Ray Chen, *Virginia Tech (VPI&SU), USA*
Shu-Ching Chen, *Florida International University, USA*

Publicity Chair

Conference Secretariat

International Workshop on Distance Education Technologies (DET 2013)

August 8-10, 2013
Holiday Inn, Brighton - Seafront, United Kingdom

Conference Organization

Steering Committee Chair

Shi-Kuo Chang, *University of Pittsburgh, USA*

Workshop Co-Chairs

Qun Jin, *Univ. of Waseda, Japan*
Pierpaolo Di Bitonto, *Univ. of Bari, Italy*

Program Co-Chairs

Angela Guercio, *Kent State University, USA*
Tim Arndt, *Cleveland State University, USA*

Program Committee

Tim Arndt, *Cleveland State University, USA*
William A. Beasley, *Cleveland State University, USA*
Maiga Chang, *Athabasca University, Canada*
Yuan-Sun Chu, *National Chung Cheng University, Taiwan*
Pierpaolo Di Bitonto, *University of Bari, Italy*
Rita Francese, *University of Salerno, Italy*
Angela Guercio, *Kent State University, USA*
Robert Heller, *Athabasca University, Canada*
Pedro Isaias, *Open University, Portugal*
Qun Jin, *Univ. of Waseda, Japan*
Lian Li, *HeFei University of Technology, China*
Hong Lin, *University of Houston-Downtown, USA*

Paolo Maresca, *University Federico II, Napoli, Italy*
Andrea Molinari, *University of Trento, Trento, Italy*
Ignazio Passero, *University of Salerno, Italy*
Michele Risi, *University of Salerno, Italy*
Teresa Roselli, *University of Bari, Italy*
Veronica Rossano, *University of Bari, Italy*
Giuseppe Scanniello, *University of Salerno, Italy*
Mahbub Syed, *Minnesota State University, USA*
Kazuo Yana, *Hosei University, Japan*
Chengjiu Yin, *Kyushu University, Japan*

International Workshop on Visual Languages and Computing (VLC 2013)

August 8-10, 2013
Holiday Inn, Brighton - Seafront, United Kingdom

Conference Organization

Steering Committee Chair

Shi-Kuo Chang, *University of Pittsburgh, USA*

Workshop Co-Chairs

Metin Sezgin, *Koc University, Turkey*
Levent Burak Kara, *Carnegie Mellon University, USA*

Program Co-Chairs

A. J. Delaney, *University of Brighton, UK*
Jennifer Leopold, *Missouri University of Science & Technology, USA*

Program Committee

Danilo Avola, *University of Rome, Italy*
Paolo Bottoni, *Universita Sapienza, Italy*
Jeff Browne, *UC Santa Barbara, USA*
Alfonso F. Cardenas, *University of California, USA*
Gennaro Costagliola, *University of Salerno, Italy*
Philip Cox, *Dalhousie University, Canada*
Sergiu Dascalu, *University of Nevada, USA*
Aidan Delaney, *University of Brighton, UK*
Vincenzo Deufemia, *University of Salerno, Italy*
Filomena Ferrucci, *University of Salerno, Italy*
Manuel J. Fonseca, *INESC-ID, Portugal*
Vittorio Fuccella, *University of Salerno, Italy*
Joaquim A Jorge, *Instituto Suoerior Tecnico, Portugal*

Table of Contents

Foreword iii

Conference Organization iv

Keynote

A Conceptual Framework for Geographic Knowledge Management
Professor Roberto Laurini xv

Multimedia Modeling and Content

A dynamic spine model for interactive hypermedia synchronization
Augusto Celentano 3

Multimodal Feature Matching for Event Synchronization
Matt Benatan, Kia Ng 9

Spatial Temporal Reasoning Using QSR, Physics, and Image Processing
Nathan Eloe, Jennifer L. Leopold, Chaman L. Sabharwal 14

Institutional Services and Tools for Content, Metadata and IPR Management
Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, Michela Paolucci 20

Emotion-Based Mashup for Social Media Contents
Mohamed Abdur Rahman 26

Improving the Search Experience in a Social Network with Cross Media Contents
Daniele Cenni, Paolo Nesi 32

Towards a Semantic Multimedia Content Retrieval Framework (S)
Mohamed A. Ahmed 38

Slow Intelligence Systems

A PetCare System designed by Slow Intelligence Principles
Shi-Kuo Chang, Bin Gao, Liqun Kuang, Yingze Wang **43**

Slow Intelligent Segmentation of Chinese Sentences using Conceptual Interval (S)
Tiansi Dong, Peiling Cui **51**

Software Development and Cloud Services

Supporting CMMI assessment using distributed, non-invasive measurement and process mining
Saulius Astromskis, Andrea Janes, Alberto Sillitti, Giancarlo Succi **55**

Analyzing Version Control Open Source Software Survivability
Bee Bee Chua **61**

Service Design for Cloud Services of Dental Clinics
Kang-Lin Peng, Chia-Yung Lin, Yu-Chee Tseng, Jui-Yuan Tsai, Ji Cheng **66**

Ontologies in Global Software Development (S)
Rodrigo G. C. Rocha, Ryan R. Azevedo, Sergio Mendonça, Alex N. Borges, Catarina Costa, Silvio Meira **71**

Manitou: An Open Framework for Multimodal Interaction (S)
Roman Hak, Tomas Zeman **75**

Playing on large displays to foster children's interest in archaeology (S)
Carmelo Ardito, Giuseppe Desolda, Rosa Lanzilotti **79**

DET Workshop

Multimedia & Gaming

A web-based phonetics tutor using generative CALL (S)
Greg Lessard, Michael Levison, Jonathan Reid, Boyang Fu **85**

Visualizing Geographic Learning Objects Through a Mobile Learning Application
Vincenzo Del Fatto, Gabriella Dodero, Ashraf ul-Islam Chowdhury, Luca Geraci **90**

Design and Development of ePortfolio Using Gamification to Increase Engagement (S)
Yuki TERAWAKI, Taketo SHIMOHIGOSHI, Makoto WATANABE, Toshiyuki KAMADA, Yasushi KODAMA, Kazuo YANA **95**

Educational MMORPG for Computer Science: DeBugger, a Virtual Lab on PC and Smart Phones
Ilmi Yoon, Arno Puder, Gary Ng, Manori Thakur, Hunvil Rodrigues, Jae H. Paik, Eun-Young Kang **100**

Scenes extraction from videos of telementored surgeries
Paolo Buono, Giuseppe Desolda, Rosa Lanzilotti **106**

Collaboration & Assessment

An Investigation of the Attitudes of Instructors and Students to On-line Assessment in Mathematical Subjects
Elizabeth Acosta-Gonzaga, Niels R. Walet **112**

Methodology and Software Support for Evaluating Teams' level of collaboration (S)
Luigi Romano **118**

Modeling Multiple Common Learning Goals in an ETC*plus* Educational Project
Angela Guercio, Paolo Maresca, Lidia Stanganelli **122**

Closing the Distance in ODL - e-Portfolio Assessment (S)
Petra le Roux **128**

Learning Objects

msMLO:A Novel Approach for Selecting and Fusing Learning Objects
Aldo Ramirez-Arellano, Elizabeth Acosta-Gonzaga **131**

Virtual learning communities: yet another LMS?
Luigi Colazzo, Andrea Molinari **137**

VLC Workshop

Reasoning and Visualization of Computational Processes

A Fast Intersection Detection Algorithm For Qualitative Spatial Reasoning (S)
Chaman L. Sabharwal, Jennifer L. Leopold **145**

Visual Thinking Design Patterns
Colin Ware, William Wright, Nicholas J. Pioch **150**

Multi-level Visualization of Concurrent and Distributed Computation in Erlang
Robert Baker, Peter Rodgers, Simon Thompson, Huiqing Li **156**

Visual Art and Design & Human-Machine Interface Design

Pencil-Like Sketch Rendering of 3D Scenes Using Trajectory Planning and Dynamic Tracking
Günay Orbay, Levent Burak Kara **162**

Exploring Local Optima in Schematic Layout
Daniel Chivers, Peter Rodgers **168**

From use cases to task trees through resources (S)
Paolo Bottoni **176**

Visual Languages & Visual Programming

Formalizing Concept Diagrams
Gem Stapleton, John Howse, Peter Chapman, Aidan Delaney, Jim Burton, Ian Oliver **182**

A Locally Nameless Visual λ-calculus
Peter Chapman **188**

GPGPU Implementation of a Generative Modelling Language (S)
Martin Griffin, Aidan Delaney, Karina Rodriguez Echavarria **194**

Comparing Purely Visual with Hybrid Visual/Textual Manipulation of Complex Formula on Smartphones (S)
Annemarie Harzl, Vesna Krnjic, Franz Schreiner, Wolfgang Slany **198**

VLC Demo/Poster

KeyScretch on Android Tablets and Smartphones (P)
Gennaro Costagliola, Mattia De Rosa, Vittorio Fuccella **A-3**

Reviewer's Index **A-5**

Author's Index **A-7**

Poster/Demo Presenter's Index **A-9**

Note: (S) indicates a short paper.
(P) indiecates a poster or demo, which is not a refereed paper.

Keynote
A Conceptual Framework for Geographic Knowledge Management

Professor Roberto Laurini
KSI Fellow

Abstract

In many applications, the management of geographic knowledge is very important especially for urban and environmental planning. However there are several practical problems hindering the efficiency, some of them being technical and other being more conceptual. The goal of this talk is to present a tentative conceptual framework for managing practical geographic knowledge taking account of accuracy, rotundity of earth, the mobility of objects, multiple-representation, multi-scale, existence of sliver polygons, differences in classifying real features (ontologies), the many-to-many relationship of placenames (gazetteers) and the necessity of interoperability.

Therefore, geographic objects must be distinguished into several classes of objects with different properties, namely geodetic objects, administrative objects, anthropic (manmade) objects and natural objects.

Regarding spatial relations, in addition to conventional topological and projective relations, other relations including tessellations and ribbon topology relations are presented in order to help model geographic objects by integrating more practical semantics.

Any conceptual framework is based on principles which are overall guidelines and rules; moreover, principles allow at making predictions and drawing implications and are finally the basic building blocks of theoretical models (theories). But before identifying the principles, one needs some preliminary considerations named prolegomena. In our case, principles will be essentially rules for transforming geographic knowledge whereas prolegomena will be assertions regarding more the foundations of geographic science.

Based on those considerations, twelve principles are given, preceded by twelve prolegomena. For instance, some principles deal with the transformation of spatial relationships based on visual acuity, with the influence of neighboring information and cross-boundary interoperability.

New categories of geographic knowledge types are presented, spatial facts, cluster of areas, flows of persons, goods, etc., topological constraints and co-location rules.

To represent knowledge chunks, three styles are presented, based respectively on descriptive logics, XML and visual languages. To conclude this talk, an example of visual language to manage geographic knowledge is proposed.

About the Speaker

Dr. Robert Laurini (sometimes called Roberto) is presently Professor Emeritus at the Computing Department of INSA-Lyon, University of Lyon, France. He is an expert in the domains of geographic information systems, geographic knowledge management and multimedia information systems. He has overall worked on spatial indexing, conceptual modelling of geodata, quality control of geodata, field-oriented databases, interoperability of geographic database and chorems as visual tools for representing geographic knowledge and visual summaries of geodatabases. He got two doctorates from the Claude Bernard University of Lyon, worked as researcher in the University of Cambridge, UK, and visiting professor at the University of Maryland, College Park, USA. From 1995 to 2008, he had also a teaching position at the IUAV University of Venice, Italy. He is fluent in French, English, Italian and Spanish. He has tutored more than 40 PhD on those domains. He is/was associated editor of the following journals, « Journal of Visual Languages and Computing », « Computers, Environment and Urban Studies », « GeoInformatica », « International Journal of Geographical Information Sciences », etc. He received two awards from ACM one in 1998 and the second in 2010 for his works. He wrote several books, among other « Fundamentals of Spatial Information Systems » with Derek Thompson which was a best seller. During 10 years, he was the president of the steering committee of the ACM GIS conference. From 1980, he his vice-president of UDMS (Urban Data Management Society) whose goal is to promote urban information systems. In addition, he was member of scientic committees of more than 100 conferences. He has been involved in PhD committees in 17 countries throughout the world. Recently, he has created an NGO « Academics Without Borders » which is a network of volunteer academics consultants the main goal of which is to assist universities in developing countries.

Proceedings
International Conference on Distributed Multimedia Systems (DMS 2013)

A dynamic spine model for interactive hypermedia synchronization

Augusto Celentano
Università Ca' Foscari, Venezia, Italia
auce@dsi.unive.it

Abstract

This paper presents a model for interactive hypermedia synchronization based on an extension of the concept of spine *introduced by the* IEEE 1599 *standard for music description. The model draws from the early hypermedia models some basic concepts framing them in current world wide web scenario, and integrates the user interaction into dynamic media behavior in a coherent and seamless way.*

1 Introduction

The conception, design and development of hypermedia systems has produced the most innovative results around the '90s of the last Century, when basic concepts have been analyzed and formalized coming to a systematic study of a rapidly growing field. The theoretical models such as the Dexter Hypertext Reference Model [7] and the Amsterdam Hypermedia Model [9], to cite only the first two reference models for hypertext and hypermedia, set the milestones in the development of hypermedia systems by defining and consolidating the concepts of link, anchor, navigation and synchronization. Early systems like, e.g., Intermedia [14] and Notecards [8], payed great attention to modeling the user activity during information exploration, generating navigation assistants like maps, hierarchies, indexes, histories, etc..

With the advent of the world wide web these concepts and they implementation have been standardized but also simplified due to the need of implementing a unique model, accommodating the needs of a large and heterogeneous population of users. Concepts like link typing, bidirectional linking, global and local views, history, maps, were implemented in early hypermedia systems to help users to navigate in a semantically rich environment without being overloaded by the information. In the world wide web many of these concepts have been removed from the document and navigation implementation, and associated to content design, thus leading to two contrasting results: the possibility of specifying in a more precise and accurate way the context of a link by *describing* it explicitly in the document but, conversely, the *need* to specify in the document text not only its content but also its structural context.

A similar evolution can be observed concerning media synchronization, that has its focal model in AHM, the Amsterdam Hypermedia Model [9] from which the W3C standard SMIL language [13] has been designed. AHM defines temporal links that allow a set of documents to be temporally coordinated, i.e., related in time instead of space. Mutual time dependencies are associated to links, that become complex structural elements carrying the change of context between their source and target ends. AHM, however, still considers the user a *navigator in space*, while in the time dimension his/her actions are limited to little more than media play and stop. A better role was given to the user by the Firefly system[4], but a rich user interaction has not been considered an essential part of hypermedia functionality until recent times.

Indeed, the problem of putting the user in the hypermedia synchronization loop has been approached long since with formal models and authoring systems implementations. Dynamic timed Petri nets are introduced in [12] to describe the synchronization of media streams after user actions like skipping parts of a presentation, reversing it, pausing, resuming, scaling speed, etc.. Media scheduling in terms of duration and ordering is analyzed in [10] by considering similar user actions, including hyperlink jumps, forward and backward jumps, speedig-up and slowing-down, etc.. In these works the structure of a multimedia document is not affected by the user, who can only act on its temporal properties. A more complex scenario is presented in [5], where the user is able to select the way a multimedia presentation is played by personalizing its structure, selecting and navigating parts of the content on a logic rather than temporal base, narrowing the gap between the author's and the reader's roles.

In this paper we present a model for interactive hypermedia synchronization; the model is based on two key ideas: a generalization of the concept of *link* as a carrier across media items of the events triggered by dynamic media and user interaction, and an extension of the concept of *spine*, a time segment introduced by the IEEE 1599 standard for music description, here used to dynamically host, in a time ordered way, the events as they happen during a hypermedia presentation.

The paper is organized as follows. In Section 2 hypermedia synchronization and user interaction are discussed. Section 3 presents the synchronization model. Section 4 briefly introduces the main implementation issues. A discussion and the concluding remarks are in Section 5.

2 Synchronization of dynamic media with user interaction

While in HTML the primary action associated by design to an anchor is the navigation towards another document or another section, anchors are indeed general purpose interaction elements. The Web 2.0 paradigm induced a shift towards programming applications as dynamic documents: while the HTML/CSS part takes care of the presentation interface, javascript and other similar languages implement the application logic by changing the internal state of the application and the user view according to the user commands. Traversing a link to another document or section is only one of the possible actions associated to an anchor, often not the most frequent.

Nevertheless, from an external point of view a wide range of information based applications still rely on navigation, that occurs in several ways: in space, by changing the locality of the view; in structure, by changing the appearance of the view; in time, by playing, pausing and stopping continuous media like audio and video.

In many applications coherent rules for moving within an information space exist, as demonstrated by the large number of application generators and information based prototyping environments. They provide ready-made models to access, manipulate and visualize information coming from structured or semistructured data repositories, and also interface templates suited to such models. What is still missing is a model able to unify the relationships among different dynamic media and between them and a user, and able to help the designer not only to design such relationships without relying on low-level programming, but also to validate them and to check for the coherence of the whole hypermedia.

As discussed above, one of the drawbacks of old hypertext/hypermedia models is that user interaction and navigation was mainly considered in spatial terms. Time-related actions like pause, speed-up, reverse, etc., are provided without considering the need for changes in media sequencing and linking. In other words, the focus is on controlling the individual multimedia components rather than the overall non linear document structure. Yet, there are several cases in which the user needs to act in more sophisticated ways then simply changing the speed of a presentation or playing it in reverse.

For example, in educational multimedia the user browses multimedia documents either driven by the application logic or freely, exploring information at his/her own pace. The structure of educational applications is far from being linear, and fits well the hypertext paradigm: surveys, deepening of concepts, related readings, experiments, references are common and intermixed. In such an environment the user must be able to control not only the dynamics of separate media (by playing, pausing and resuming audio and video clips) but also, and mostly, by activating and synchronizing them in a coherent way.

For example, if a student brings into the interface and starts reading a relevant text while watching an educational video, the video should be automatically paused. While reading the text, the student could find useful to browse other material, that could be of multimedia nature, requiring proper playing and pausing of continuous media without conflicts and interference among them. All such dependencies should be expressed at hypermedia design time rather than left to the care and attention of the user.

3 A model for synchronization specification in interactive hypermedia

We introduce a model for describing synchronization in interactive hypermedia based on a set of related concepts: media, event, anchor, link, user interaction and spine. Most of these concepts are drawn from existing (and sometimes old) hypermedia models. From each model we have taken the concepts most suitable to be integrated into a comprehensive and homogeneous framework facing all the aspects related to media synchronization and user interaction. A special attention has been given to the possibility of implementing the model using current web-based technology.

The model focuses on a discrete, event-based representation of synchronization [3, 6]. With respect to timeline based models such as SMIL, the main advantage is the ability to address synchronous and asynchronous events in a uniform way, and user interaction is by definition asynchronous with respect to dynamic media evolution. The model does not address fine-grain synchronization, for which good technical solutions based onstandards formats and protocols exist. It covers the design of the relationships between media items on a discrete scale, with a focus on the handling of the events raised by user interaction during the hypermedia fruition.

3.1 Media and events

A hypermedia document is composed of media items: text fragments, audio and videoclips, images, etc.. For simplicity, and without loss of generality, we can assume that each media item is stored and identified separately, even if this is not strictly required and pertains to the implementation more than to the model itself.

Media items can be static or dynamic. *Dynamic* media (audioclips, videoclips and animations) evolve by changing their state according to a time base which is proper of the media item. They have a duration defined by the content and play modes. *Static* media (text and images) do not change their state unless an external event triggers some action (e.g., making an image disappear or change colors). Static media have no time base, their duration is in principle infinite [6].

A media item can be active or inactive. If it is *active* it exists in the hypermedia presentation interface, can evolve in time (for dynamic media) and can receive and trigger events. *Inactive* media do not exist in the hypermedia presentation interface (or exist as bare placeholders, e.g., "greyed out" fields and widgets), do not evolve and cannot trigger events until activated. A dynamic media, while active, can be run, stopped, paused and resumed. A static media can only be active or not. A media item can be *interactive* or *non interactive*. Interactive media can receive user input, while non interactive media cannot.

An *event* is something that happens as a consequence of time flow or user action. Events can trigger changes in the state of media items by executing *actions* on them, e.g., stopping a video or displaying a text. Media can be related in the sense that one media item (*master*) can cause other media (*slave*) to change their behavior as a consequence of events triggered by it. In principle in a complex hypermedia presentation more than one media item can act as a master, but for simplicity we shall constrain the discourse to one master at a time. The presence of multiple masters can cause incoherences in the control of media if events are triggered which produce conflicting actions, and must be solved by proper design methodology and by priority rules.

3.2 Links and anchors

The concepts of anchors and links are mutually related as in the hypertext models. An anchor is a point in a media item to which an event is associated (*source anchor*) affecting some other media item, or the effect of an event occurred elsewhere is associated (*target anchor*); a link is a connection between two anchors, such that the event at the source media item anchor causes an action at the target media item anchor. Links are directional and describe how events are propagated among media to give the hypermedia document its global behavior.

Anchors can be temporal or spatial. A *temporal anchor* is defined in a dynamic media item as a time coordinate relative to the media item start time; when the media execution arrives at that time the anchor triggers an event which is transported across the link to the target anchor point. Actions are associated to the link, with parameters that describe the action details. For static media, which have no temporal behavior other than their bare existence, temporal anchors can only be associated to media activation and deactivation.

Spatial anchors can be defined in static and dynamic media, and identify spatial extents occupied by the media item in the user interface (or identified by a placeholder associated to the whole media item as in, e.g, audio clips) on which a user action can be associated (e.g., a mouse click, a touch), triggering actions towards the linked media.

Any media item has an implicit target anchor associated to the whole item, like in conventional Web navigation. Anchors can be associated to several prototype media types, each representing a class of media with similar properties:

video a prototype for a dynamic media item with a spatial extension that can hold temporal and spatial anchors;

audio a prototype for a dynamic media item without a spatial extension that can hold only temporal anchors;

text a prototype for a static media item with a spatial extension that can hold spatial anchors associated to fragments of its content;

image a prototype for a static media item with a spatial extension that can hold only one spatial anchor associated to the whole content;

external a prototype for a media instance external to the hypermedia presentation, such as, e.g., a webpage, which is under the user control in an unspecified way.

Being prototypes, they can represent different actual media types; for example, an image map is equivalent to a text prototype because it holds several spatial anchors, and a text without anchors in it is equivalent to an image.

The *user* is also associated to a specific media type, a dynamic media with temporal anchors which are not predefined in time. The user and his/her interaction is described in Section 3.3.

A link has a series of properties/attributes which describe the action to be taken on the media item at the target anchor:

- *activate*/*deactivate* the media item at link target anchor, deactivate the media item at link source anchor;
- *set*/*reset* interaction on link source and/or target anchors; if reset, the media goes to a non interactive state and cannot receive further user interaction until reverted to interactive state. A common use of this property is to transfer the user control from the source to the target media item by resetting interaction on source and setting it on target;
- *pause*/*resume*/*play*/*stop* the media item at link source and/or target (meaningful only for dynamic media). A common use of this property is to pause or stop the media item at link source and resume or play the media item at link target.

Links could set also a time delay from the trigger to the action, but this additional parameter does not change substantially the model, and will not be discussed here.

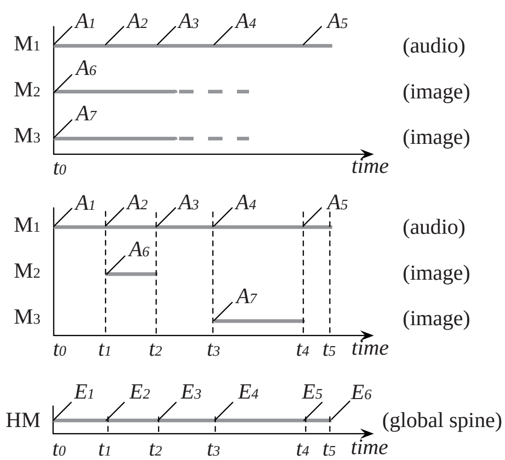


Figure 1. Local and global spines

3.3 User interaction

The user is part of the model as any other (dynamic) media item, except that it has no pre-defined time behavior. Temporal anchors associated to the user depend on the time he/she interacts. The evolution of the hypermedia presentation is defined both by the autonomous dynamics of media and by the user operations, which can integrate, alter or contrast the events generated by the media items dynamics. From a practical point of view, the user interacts through a series of input devices such as buttons, menus, touch, gestures, which depend on the type, style and rules of the interface. We do not enter into such details, and consider the user interaction from an abstract perspective.

The user triggers temporal source anchors, which are the starting point of links to media items. Hence, the user can activate/deactivate media items, make them interactive or not, play/pause/resume/stop dynamic media, etc., as any other dynamic media, with the difference that the time at which events occur are not pre-defined but unpredictable and depend only on the user activity.

As in media relationships, bad design can produce incoherent scenarios that must be solved by proper methodology and priority rules. As a general remark we can assume that the user has priority over conflicting behavior naturally produced by the media. There are cases, like, e.g., educational applications, in which the opposite behavior is desirable, to force the user to access the hypermedia content in an ordered and constrained way.

3.4 Spine

The temporal synchronization of the media composing the hypermedia presentation is achieved through a component called *spine*. It is derived from the concept of spine of the IEEE 1599 standard for music description [1, 2]. Basically, a spine is a (segment of a) temporal axis where the events

Table 1. Links between anchors of Figure 1

Link	Source	Target	Action
L_1	A_2	A_6	activate target media M_2
L_2	A_3	A_6	deactivate target media M_2
L_3	A_4	A_7	activate target media M_3
L_4	A_5	A_7	deactivate target media M_3

defined by dynamic media temporal anchors and user interaction are placed.

Spines are of two types: dynamic spines and static spines. *Dynamic spines* are segments of a time axis associated to dynamic media items, whose time span is the media item duration. They contain at selected positions the temporal anchors defined in the media items. *Static spines* are associated to static media, and are defined only for completeness. Since static media have no defined duration they are represented by a *ray*, an open-ended time interval whose real duration is set by actions following events triggered by other media or by the user interaction. Each spine contains also a set of events corresponding to spatial anchors, which are not associated to time, but are placed in the spine at current time as the anchors are activated. For consistency, the target anchor associated to the whole media item is placed at the item starting time t_0.

These spines are *local spines*, i.e., associated to individual media items. A *global spine* is a dynamic structure built as the hypermedia unfolds in time. It is a time axis with selected positions representing events resulting from the union of the local and static spines active at any moment, and evolves by adding new events as the media are activated/deactivated and played/paused/resumed/stopped. User interaction events are also placed on the global spine as they occur. In such a way the global spine is a recording of past events (a history) in the part preceding the current time, and the description of following behavior (less user interaction) in the part following the current time. User interaction is defined only at the current time upon occurrence of an explicit user operation; after execution it belongs to the hypermedia history.

Figure 1 illustrates a simple non interactive example. The hypermedia is composed of three media items, an audio M_1 and two images M_2 and M_3. The audio activates and deactivates the images during playback, with an interval between them. Table 1 defines the links and their properties. The top diagram in Figure 1 shows the media items with their local spines on which the anchors are placed. The middle diagram shows the temporal behavior of the static media M_2 and M_3 as defined by the events associated to the temporal anchors of the audio M_1, the only dynamic media item. The bottom diagram shows the global spine at the end of the execution. E_1–E_6 are the events raised: E_1 corresponds to the hypermedia start, i.e, the activation of the

Table 2. Links introduced by the user interaction

Link	Source	Target	Action
L_5	U_1	A_1	pause target media M_1
L_6	U_2	A_1	resume target media M_1
L_7	U_3	A_9	activate target media M_4
L_8	U_3	A_1	pause target media M_1 reset interactivity on M_1
L_9	U_4	A_9	deactivate target media M_4
L_{10}	U_4	A_1	resume target media M_1 set interactivity on M_1

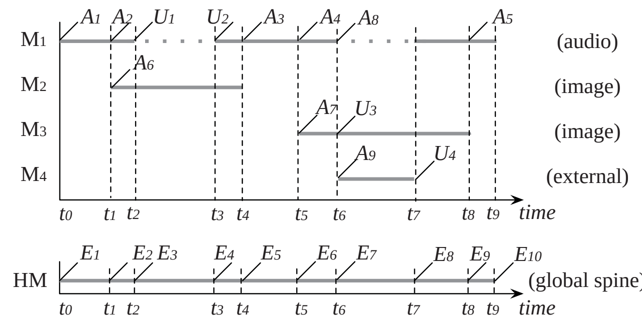


Figure 2. Local and global spines with user interaction

audio M_1, while E_6 is the natural end of M_1, causing the end of the whole hypermedia. E_2–E_5 are, respectively, the events triggering the links L_1–L_4: at time t_1 the first image is displayed, until time t_2. At time t_3 the second image is displayed, until time t_4. At time t_5 the audio stops and the presentation ends.

In case of user interaction, the behavior may change. Suppose that the user pauses and resumes the audio while watching at image M_2, and that image M_3 is associated to a spatial anchor leading to a website in a separate window; the user can freely browse the website by automatically pausing the audio while the browser window is open. The paused audio remains in a non interactive state while the user consults the web page (and possibly other linked pages). When the browser windows is closed, the audio resumes playing and returns in an interactive state. The webpage is represented by an *external* media item M_4, with a target anchor associated to it as in any other static media item.

Table 2 lists the additional links between user operations (source anchors) and target media. Figure 2 shows an instance of the local and global spines after the interaction has been completed (pauses in dynamic media are represented by dotted lines). For clarity, anchors corresponding to user interaction are drawn upon the media to which they are linked, at the time where they occur; they are named U_1–U_4. A more faithful graphical representation would represent the user anchors on a separate timeline for consistency with the concept that the user behaves as a dynamic media item, but the graphic would be cluttered.

As shown in the figure, pausing a media item has the effect of *shifting* the remainder of its local spine to a future time. The shift amplitude depends on the pause time. A more concrete view, which is adopted in the implementation, is that the local spine is *suspended* during the pause, i.e., the segment not yet executed is *removed* from the global spine and *copied back* at the time in which media execution is resumed.

4 A prototype implementation

A thorough discussion about the model implementation is out of the size of this paper, hence only a survey is given here. One of the goal of the model is the ability to implement it with current web-based technology, using the basic mechanisms used in HTML pages rendering and navigation. Media items, links, anchors, interactivity, are usual components of multimedia applications, hence the synchronization and interaction related issues should be implemented by relying on the standard protocols and caching mechanisms of the current world wide web, without sophisticated, ad-hoc media control systems.

A prototype of an authoring system has been built [11]. The authoring system does not cover the layout design, which is out of the model's scope, but implements a number of javascript functions which build up the set of actions for media operations defined in the model. Function calls are associated to anchors, which change the properties of target media and allow a user to interact with a hypermedia presentation in the way described here.

The prototype defines a standard interface, hence is not able to cope with a general multiplicity of media, but covers all the prototypical media types defined in Section 3.1. The prototype has been implemented using part of the project *Popcorn.js* by Mozilla (http://popcornjs.org), a javascript library for media control in HTML5 documents, which has been properly modified and extended to meet the needs of the synchronization model mainly for what concerns user interaction.

The whole hypermedia (except external media) is an HTML document whose content is structured in sections corresponding to the component media items. The concepts of anchor, media and link of the model have been translated to HTML5 constructs: the anchor is implemented by the anchor tag `<a>`. A media item is a section of the document, structured as a `<div>` with a unique referencing *id*. Format and visual properties can be assigned and modified as usual through javascript functions. Links associated to source anchors correspond to the execution of a javascript function which sends the event to the target media item.

The behavior of the media item are changed through an extension of the Popcorn.js library, which converts media item to interactive javascript objects able to send and receive events.

5 Discussion

The first question naturally raising in discussing this model is: *why another synchronization model for hypermedia documents?* The question is legitimated by the huge research work existing on the subject and by the current standards and programming practices for incorporating active media and interactive commands into HTML-based documents.

The answer comes from the lack of a unifying model which includes also the user as an active component driving the temporal behavior of a multimedia document through a reasonable and coherent set of interactive actions.

It is quite evident from the analysis of the research work on hypermedia in the '90s that the goal was primarily to define standards and rules for *structuring* complex documents, i.e., putting together the different parts so that a user could navigate in them coherently. While dynamic media were included in documents by the Amsterdam Hypermedia Model and other models, they were considered as self-contained information chunks whose relation with other media were not under the user control.

The document model(s) introduced by the WWW environment and refined until today's versions (the so-called Document Object Model, DOM) face problems related to the manipulation of the *structure* of a document and its content, still without approaching interaction issues. Such issues are partly implied by the rigid navigation structure, and more recently they are delegated to algorithms programmed in javascript and other languages to manipulate the document components and their rendering.

Hence, this model fills a gap between the conceptual representation of content and structure and the conceptual representation of time relationships induced both by dynamic media and the user.

A second reason justifying this model is the close relationship between the model's components and the actions implemented in a conventional HTML environment, that allow a programmer to use templates for enacting the anchors and links behavior. Such close relationship is the base for developing authoring systems for the end user not only with an eye to the document structuring, which is a goal amply covered by CMSs, but also with a reference to interaction design.

Finally, a model (even if not completely formal) helps a designer to check the consistency of its design before coming to a prototype, and favors an organized maintenance activity. Consistency is, however, a problem not yet fully investigated and is the main future work, together with an automated translation of the link definitions from the tabular style of Tables 1 and 2 into function calls with proper parameters.

Acknowledgments

Alessandro Mantovani has implemented a prototype authoring system using HTML5-based features, demonstrating the model feasibility.

References

[1] D.L. Baggi, G.M. Haus, The New Standard IEEE 1599, Introduction and Examples, *Journal of Multimedia* 4, 1, pp. 3–8, 2009.

[2] A. Baratè, G. Haus, L.A. Ludovico, *IEEE 1599: a New Standard for Music Education*, in *ELPUB2009, Proc. of the 13th Int. Conf. on Electronic Publishing*, pp. 29–45, 2009.

[3] E. Bertino and E. Ferrari, Temporal Synchronization Models for Multimedia Data, *IEEE Trans. Knowledge and Data Eng.* 10, 4, pp. 612–613, 1998.

[4] M.C. Buchanan, P.T. Zellweger, Specifying Temporal Behavior in Hypermedia Documents, in *Proc ACM Conference on Hypertext*, pp. 262–271, 1992.

[5] D. Bulterman, User-Centered Control within Multimedia Presentations, *Multimedia Systems Journal* 12, 4-5, pp. 423–438, 2007.

[6] O. Gaggi, A. Celentano, Modelling Synchronized Hypermedia Presentations, *Multimedia Tools and Applications*, 27, pp. 53–78, 2005.

[7] F. G. Halasz, M. Schwartz, The Dexter Hypertext Reference Model. *Comm. ACM* 37, 2, pp. 30–39, 1994.

[8] F. G. Halasz, Reflections on NoteCards: seven issues for the next generation of hypermedia systems. *ACM J. Comput. Doc.*, 25, 3, pp. 71–87, 2001.

[9] L. Hardman, D.C.A. Bulterman, G. van Rossum, The Amsterdam Hypermedia Model: Adding Time, Structure and Context to Hypertext, *Comm. ACM* 37, 2, pp 50–62, 1994.

[10] W. Liao, V.O.K. Li, Synchronization of Distributed Multimedia Systems with User Interactions, *Multimedia Systems Journal* 6, 3, pp. 196–206, 1998.

[11] A. Mantovani, *Back to Hypermedia: A System for Dynamic Media Structuring Based on Web Standards*, Master Thesis, MSc Program in Computer Science, Università Ca' Foscari Venezia, Italy, 2012.

[12] B. Prabhakaran, S.V. Raghavan, Synchronization Models for Multimedia Presentation with User Participation, *Multimedia Systems Journal* 2, 2, pp 53–62, 1994.

[13] *Synchronized Multimedia Integration Language (SMIL 3.0)* W3C Recommendation, 2008: http://www.w3.org/TR/smil

[14] N.Yankelovich, B.J. Haan, N.K. Meyrowitz, S.M. Drucker, Intermedia: The concept and the construction of a seamless information environment, *Computer*, 21, 1, 81–96, 1988.

Multimodal Feature Matching for Event Synchronization

Matt Benatan and Kia Ng
ICSRiM – University of Leeds
School of Computing,
Leeds LS2 9JT, United Kingdom
info@icsrim.org.uk, www.icsrim.org.uk

***Abstract*— With the on-going increase in processing power and range of multimedia, multimodal systems are becoming more prevalent in a broad variety of applications. These include a wide range of entertainment applications such as games and mu-sic, as well as new developments within usability and human computer interaction. This project proposes a multimodal synchronization approach through the application of feature detection and association. This paper discusses features of interest, analysis and similarity measures, together with techniques such as Dynamic Time Warping and Active Appearance Modelling for synchronization purposes. In order to demonstrate the framework, this paper presents a case study using a multimodal music application which combines voice data with facial gesture detection for a virtual vocal coaching application.**

Keywords-multimodal; synchronization; edutainment; music; audio; visual

I. INTRODUCTION

With the increase in computing power over recent years there has been a growing interest in developing multimodal control systems. This has led to numerous developments within interactive multimedia in which multiple data sources are harnessed to create more interactive and immersive user experiences across a broad range of areas including gaming, sports science, accessibility and smart interfaces. While many multimodal systems have been developed, there have been fewer investigations into frameworks for use within this area.

This paper introduces a generic framework for feature matching and multimodal synchronization. This is presented in the context of a multimodal system utilizing audio and visual information for the control of an interactive music system. The system uses user input, in the form of vocal audio and head movements, in order to control the system's musical output.

The framework discussed here has been developed to match features from multiple simultaneous data sources for synchronization and use within multimodal audio/visual systems. To do so, signal segmentation, feature detection and signal warping techniques have been used.

II. BACKGROUND

Interactive multimedia systems typically use a range of multimodal data in order to provide a better understanding of user interactions. In this way, more intuitive control systems can be developed. With the advent of more complex commercially available gesture recognition systems [1, 2], research into multimodal gestural interaction has increased dramatically. This, coupled with continued increase in processor capabilities, has resulted in a broad range of new interactive multimedia systems, including everything from music interfaces to novel approaches in human computer interaction and smart interfaces.

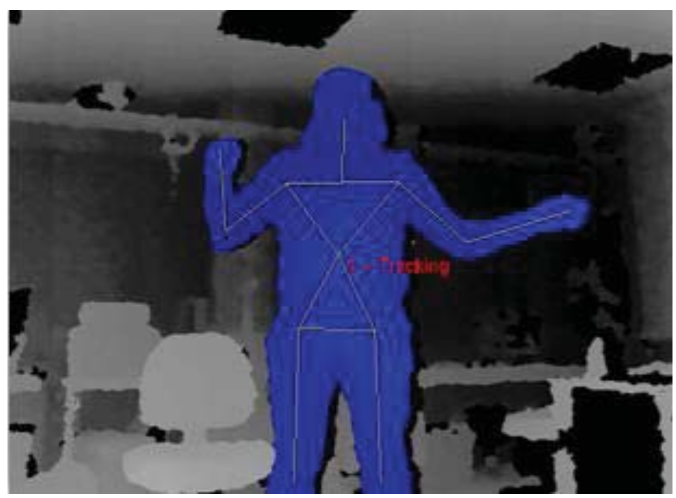


Figure 1. Kinect depth map used within Music via Motion (MvM).

A number of recent developments in interactive music systems have utilized new technologies for use within gesture-based systems, such as Music via Motion (MvM) [3] and Crossole [4]. The MvM system uses several input sources including video camera, sensors and Microsoft's Kinect within a range of interactive music applications. Through following the user's gestures in 3-dimensional space (see Figure 1), the user is able to control virtual instruments, or even a virtual orchestra using the Conducting interface, in real-time. In a similar way, Crossole [4] makes use of the Kinect system to create a novel form of musical interaction: a musical crossword that allows users to create unique compositions by manipulating blocks in a virtual space (Figure 2).

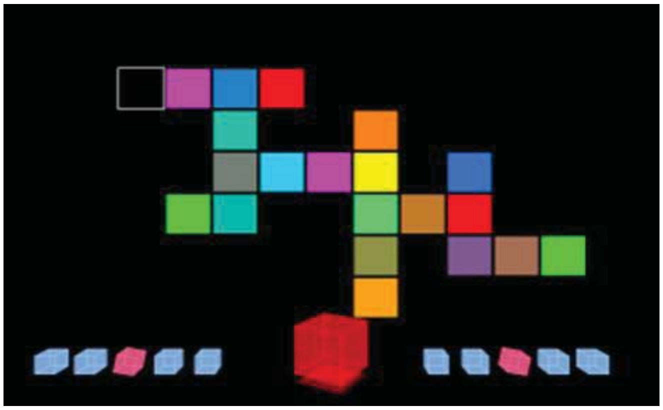


Figure 2. Crossole gestural interface.

Recent developments at SIEMPRE have also gone on to refine these technologies through the Gesture Agents framework [5]. Through coordinating concurrent multi-user interaction, this builds on current interactive multimodal system technologies through providing a framework to enable multiple users' to share the same interactive space. In doing so, this allows several users to interactive with one musical interface for real-time computer-music collaboration.

Multimodal systems have also been widely explored within a range of interactive systems including developments in human-robot interactions, such as Shimon [6]. This musical robot incorporates a multi sensor system with a variety of improvisation algorithms in order to emulate human-like musical interactions. Through exhibiting human gestures, such as head bobbing, Shimon is able to provide musical collaborators with gestural cues, much like a human would. Shimon's multi-sensor system is also able to interpret music – allowing it to respond to rhythms and melodies. This combination of multi-sensory input and gestural interaction results in highly intuitive human-robot musical collaboration.

Another area of application is within music pedagogy – one such example is a poly-phonic pitch tracking system used for automatic guitar notation [7]. The system combines visual input with audio signal analysis in order to map out the notes being played for automatic transcription.

Multimodal systems have also been applied to explore human interaction [8]. Investigations carried out by Varni *et al.* utilized a multi-sensor system in order to perform real-time analysis of human interactions to detect the synchronization of affective behavior and the emergence of functional roles within a group of musicians; thus providing an innovative method of gathering data for psychological research.

As well as full-body gesture recognition, interest has also been building in the area of head/face tracking and facial expression detection. This has many applications, from controlling effects used within musical performance [9] to developments with-in usability systems, such as vision-based facial expression interfaces [10]. These systems provide hands-free approaches to computer interaction, utilizing face-tracking algorithms to develop solutions to accessibility issues within HCI.

Similar approaches have been exploited within gaming, whereby face-tracking has been used to control a variety of computer games [11]. Through the development of a real time, low resource head-tracking system, Da Silva *et al.* have explored the use of head-tracking within game interaction. The system utilizes a number of techniques, including gaze tracking and head pose estimation in order to parameterize head and facial gestures for use as an input device. This has been successfully applied within a variety of games, including Tetris, racing games and adventure games.

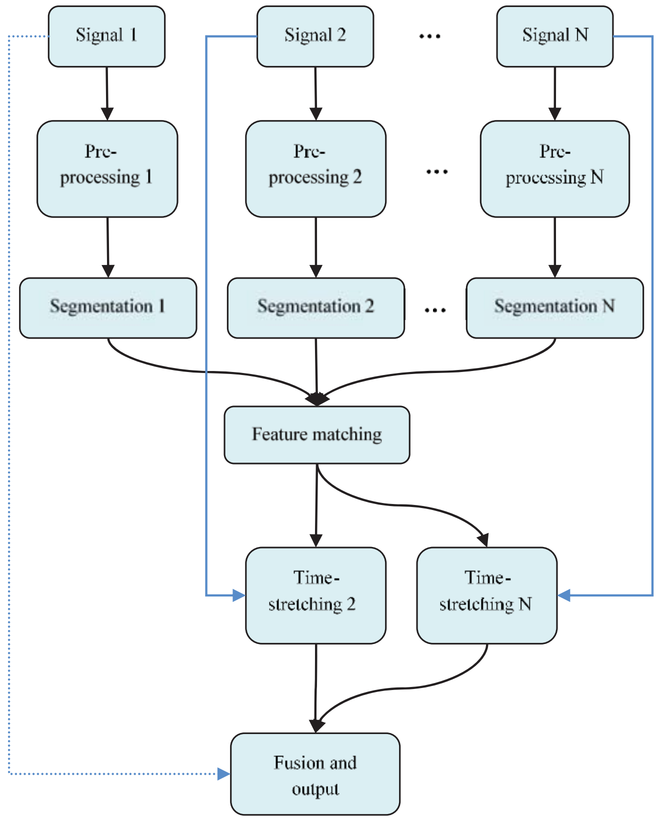


Figure 3. An overview of the framework.

III. DESIGN

Figure 3 illustrates the architecture of the framework. Here, each signal is synchronized to a master reference signal, in this case, signal 1. Before the main stages of the synchronization process can take place, the signals are pre-processed. The pre-processing stage includes (optional) smoothing and normalization as necessary. A segmentation stage is then applied. At this stage a collection of segmentation algorithms can be selected to breaks the signal up into appropriate/meaningful regions that will then be used for the next stages including correlation and synchronization.

A. Segmentation

A range of algorithms can be applied for segmentation, depending on the context and type of signal that is being segmented. For audio signals and gesture data, salient point detection can be applied; using cluster-based, wavelet-based or feature-based algorithms to identify key signal components [12]. For video data, algorithms such as mean-shift clustering or Haar-like feature detection [13] can be used to identify regions of interest. The segmented data can then be processed to identify correlating features between the data streams.

B. Feature Matching

Feature matching is applied in order to detect key features within the segmented data. As with segmentation, a range of algorithms are available depending on the context. For facial expression within a given region of interest (identified during segmentation), Active Appearance Modelling (AAM) [14] can be used to detect gestural information. In a similar way, spectral analysis [15] and statistical modeling, such as Hidden Markov Models [16] can be integrated to identify features within an audio signal. Dynamic Time Warping technique has also been integrated to find similarities between data sets.

C. Synchronization

Once the signal features have been identified, they are synchronized through cross-correlation and comparison against matching features. In the case of facial detection and audio, these variables may be mouth movement and speech – using audio recognition and mouth modelling as the synchronization mechanism. The signals are then synchronized via time-stretching, whereby the signal(s) are altered in the time domain in order to match the master reference signal. This results in a synchronized multimodal system. Following the development of the framework, it was applied within a test application so that it could be validated in a practical context.

IV. CASE STUDY

The synchronization framework was applied within a pedagogical music application designed as a vocal training tool. The application monitors both the user's vocal (via a microphone) and their facial expressions (using a video camera). With the association between facial gestures and vocal theory [17], the application aims to improve the singer's execution through providing audio cues. This is achieved through both audio signal processing and computer vision analysis which compares the user's facial expression and musical pitch to an ideal model. For example, higher notes are associated with wider mouth shapes (e.g. an exaggerated smile) and low notes are associated with long 'o'-like expressions [17]. In this way, the application helps to "coach" the user to achieve their potential range while also aiding their level of control, to improve their understanding of how facial expression can impact their vocal output.

In order to associate the vowel sounds with their corresponding facial gestures, the system first conducts frequency analysis on a training data set (see Figure 4). The same is then done for a data set of facial gestures (see Figure 5), after which these are combined to build a model of vowel sounds and facial expressions using a classification tree.

Once the model has been established, it can be used to evaluate real-time audio and visual data. This achieved through applying the synchronization framework to the audio and video data streams, after which it can be evaluated against the training data and used to provide feedback to the user.

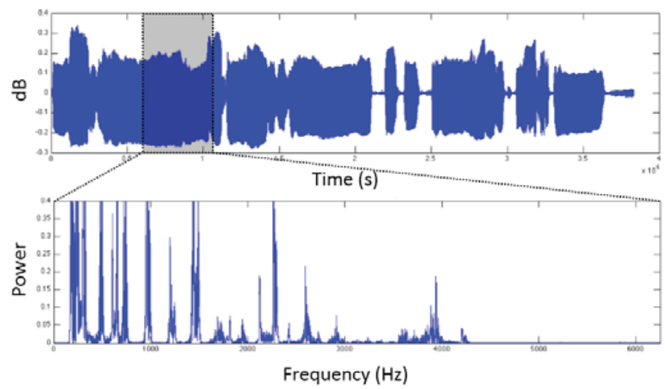


Figure 4. Top: audio signal from sung vocal. Bottom: frequency content of audio signal sample.

A. Segmentation and Feature Matching: Video Processing

The audio data stream is first segmented using cluster-based feature detection. This breaks up the incoming signal into discrete audio components for analysis using Fast Fourier Transforms (FFTs) in order to provide frequency domain information (Figure 4). This data is then analyzed for the prominent frequency bands to create a distinct feature vector for evaluation against the training data.

B. Segmentation and Feature Matching: Video Processing

At this stage, the visual information is processed by a Haar classifier in order to detect the region of the face within the image. This provides a bounding box to limit the search area for the active appearance model. The active appearance model is then applied to detect specific facial features. This results in pixel locations within the image which can be used to detect movement. To detect mouth movement, the upper-lip, bottom lip and both corners of the mouth (left and right) are used.

Through comparing these in relation to each other mouth movement can be detected, e.g. the distance between the upper and lower lip indicates mouth height, while the distance between the corners indicates mouth width. As the mouth moves, the pixel locations change, and the movement can be detected through as the distances vary. This data is output in the form of pixel unit measurements (Figure 5) for comparison against the training data.

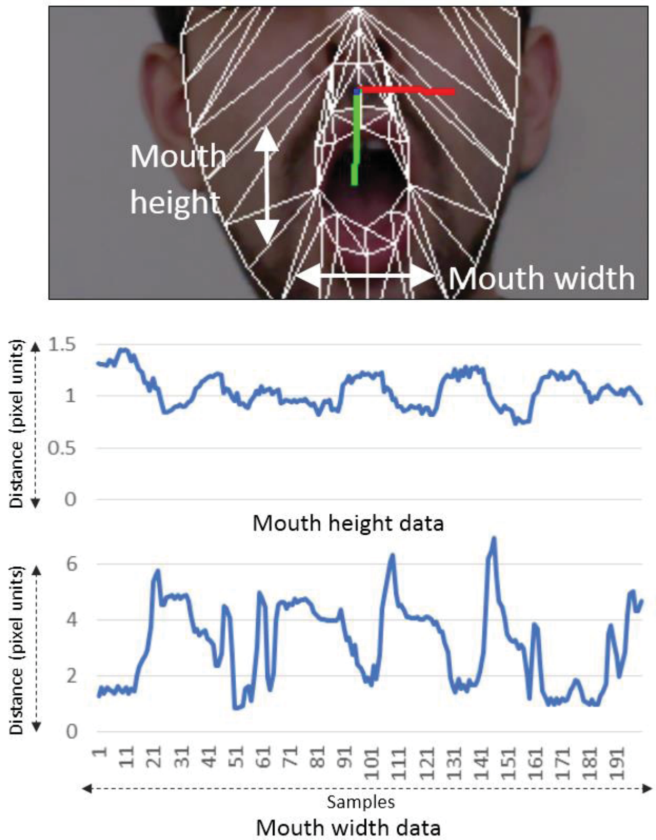


Figure 5. Top: active appearance model (AAM) used for facial gesture detection. Middle: AAM mouth height data. Bottom: AAM mouth width data.

C. *Synchronization*

It is important that the two data streams are synchronized in order to ensure that the user feedback is accurate. To achieve this, the two data streams are evaluated using the classification model built from the training data. This model contains information associating facial expressions with vocal formants, and can thus be applied to correctly match the audio signal sample with the accompanying segment from the video feed using a classification tree [19].

D. *User Feedback*

Once the data is synchronized the system evaluates the user's vocal performance in order to provide appropriate constructive feedback. If the incoming musical note does not align with its target pitch according to the tuning system chosen, the system evaluates whether it is sharp or flat (in relation to the nearest musical note). If the note is flat an audio cue is activated encouraging the user to accentuate their facial expression horizontally (smile more). If the note is sharp, the same process is applied, though the resulting cue will signal that they emphasize their expression vertically (towards a more "o"-like shape).

V. Discussion and Conclusions

This paper has discussed a multimedia synchronization framework that has been designed to synchronize multiple data sources to facilitate multimodal systems.

A range of different approaches for the core stages of the framework are being integrated to include more techniques and algorithms for segmentation, similarity measures and feature detection. Through applying the framework to a range of different case studies, we aim to develop a comprehensive synchronization solution capable of handling a range of multimodal synchronization tasks through simple configuration. The configuration process will follow a modular workflow through which the most appropriate components can be selected for segmentation, feature matching and synchronization. A case study using the framework has been presented in the form of a pedagogical application for singing. While validation of the case study is ongoing, the framework has already proven to be successful within a virtual conducting application [18], demonstrating that it is capable of streamlining the development process for multimodal systems. Several other applications are currently being explored in different areas including gaming, sports science, and accessibility.

References

[1] Z. Zhang, "Microsoft Kinect Sensor and Its Effects", IEEE Multimedia, Vol. 19, Issue 2, pp 4-10 (2012).

[2] Leap Motion (2013), www.leapmotion.com (02/05/2013)

[3] J. Leonard and K. Ng, "Music via Motion: "A Distributed Framework for Interactive Multimedia Performance", Proceedings of the International Conference on Distributed Multimedia Systems, Florence, Italy, 2011.

[4] S. Şentürk, S. W. Lee, A. Sastry, A. Daruwalla and G. Weinberg, "Crossole: A Gestural Interface for Composition, Improvisation and Performance using Kinect", Proceedings of the International Conference on New Interfaces for Musical Expression, Ann Arbor, Michigan, USA, 2012.

[5] C. F. Julià, S. Jordà and N. Earnshaw, "GestureAgents: An Agent-Based Framework for Concurrent Multi-Task Multi-User Interaction", Proceedings of the International Conference on Tangible, Embedded and Embodied Interaction, Barcelona, Spain, 2013.

[6] G. Hoffman, and G. Weinberg, "Interactive Improvisation with a Robotic Marimba Player", in Solis and Ng (eds.) Musical Robots and Interactive Multimodal Systems, Springer Tracts in Advanced Robotics,

ISBN-10: 3642222900, Chapter 4, pp. 233-251, Springer, Heidelberg, Germany (2011).

[7] G. Quested, R. Boyle and K. Ng, "Polyphonic Note Tracking Using Multimodal Retrieval of Musical Events", in Proceedings of the International Computer Music Conference, Belfast, Northern Ireland, 2008.

[8] G. Varni, G. Volpe and A. Camurri, "A System for Real-Time Multimodal Analysis of Nonverbal Affective Social Interaction in User-Centric Media", in IEEE Transactions on Multimedia, Vol. 12, pp. 576-590 (2010).

[9] D. Merrill, "Head-Tracking for Gestural and Continuous Control of Parameterized Autio Effects", in Proceedings of New Interfaces for Musical Expression, Montreal, Canada, 2003.

[10] G. C. De Silva, M. J. Lyons, S. Kawato and N. Tetsutani, "Human Factors Evaluation of a Vision-Based Facial Gesture Interface", in Proceedings of Computer Vision and Pattern Recognition Workshop, Madison, WI, USA, 2003.

[11] M. P. Da Silva, V. Courboulay, A. Prigent and P. Estraillier, "Fast, Low Resource, Head Detection and Tracking for Interactive Applications", in PsychNology Journal, Vol. 7, pp. 243 – 264 (2009).

[12] N. Sebe, Q. Tian, E. Loupias, M.S. Lew, and T.S. Huang, Evaluation of Salient Point Techniques, in Image and Vision Computing Vol. 21, pp 367-377 (2002)

[13] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", in Proceedings of Computer Vision and Pattern Recognition, Kauai, HI, USA, 2001.

[14] J. G. Bueno, "Facial Gesture Recognition Using Active Appearance Models Based on Neural Evolution", in Proceedings of the International Conference on Human-Robot Interaction, Boston, MA, USA, 2012.

[15] J. W. Pitton, K. Wang and B-H. Juang, "Time-Frequency Analysis and Auditory Modeling for Automatic Recognition of Speech", in Proceedings of the IEEE, Vol. 84, pp. 1199 – 1215 (1996).

[16] H. Jiang, "Large Margin Hidden Markov Models for Speech Recognition", in IEEE Transactions on Audio, Speech and Language Processing, Vol. 14, pp. 1585 – 1595 (2006).

[17] P. Rubin, E. Vatikiotis, "Measuring and Modeling Speech Production", in S.L. Hopp, M.J. Owren and C.S. Evans (eds.) Animal Acoustic Communication, ISBN-10: 3642762220, Chapter 8, pp. 251-290, Springer, Heidelberg, Germany (1998).

[18] M. Benatan and K. Ng, "Feature Matching of Simultaneous Signals for Multimodal Synchronization", in Proceedings of International Conference on Inormation Technologies for Performing Arts, Media Access and Entertainment, Porto, Portugal, 2013.

[19] J. Shotton (2011) "Real-Time Human Pose Recognition in Parts from Single Depth Images", IEEE Computer Vision and Pattern Recognition 2011, Colorado, USA, 21-25 June 2011. IEEE, Colorado, USA.

Spatial Temporal Reasoning Using QSR, Physics, and Image Processing

Nathan Eloe
nwe5g8@mst.edu

Jennifer L. Leopold
leopoldj@mst.edu

Chaman L. Sabharwal
chaman@mst.edu

Computer Science Department
Missouri University of Science and Technology
Rolla, MO 65409, USA

Abstract

Qualitative spatial reasoning (QSR) is a powerful tool in automated computer reasoning, a necessary step forward in fields like computer vision and media analysis. Stereo graphical media has rapidly become a prevalent part of technological culture, and the amount of these kinds of data that exists is staggering. Humans interpret depth information using prior knowledge that a computer lacks. This prior knowledge stems from remembered observance of the basic laws of physics. While the computer lacks the intuitive understanding of these principal physical properties, it is capable of determining more precise information through calculation. Herein the authors explore the information that can be gained from an amalgamation of QSR methods and physics, and present some preliminary results from an implementation based on this powerful combination.

1 Introduction

Human perception of three dimensions is a complex field. A person determining the shape of an object must rely heavily on visual cues in the form of shading, edges, and depth information [1]. Where this information is insufficient to unambiguously identify an object, the observer must make a judgment that may not be consistent with that of all other observers [1]. Research has shown that an individual's desires can influence the way things are perceived ("Wishful Seeing") [2].

This raises some important questions in the field of computer vision, including:

- As a binary entity, how can the computer deal with ambiguous visual information?
- With no base of previous experience, what information should the computer use to reduce ambiguity?
- How can the information gain of the system be maximized while the computational cost is minimized? In other words, what calculations should be done to obtain the most information with the least work?

With stereoscopic media becoming pervasive in the form of 3D movies and consumer electronics (e.g. televisions and portable gaming devices), there is a growing, urgent need for computational analysis of these data. Such technology could have impact in physical security (e.g. analysis of images from multiple sources such as security cameras), robotic vision, and defense (e.g. identification of potential dangers and suspicious behavior from stereoscopic information).

Image processing techniques can provide insight into a system recorded stereoscopically, but only about what can be seen by the cameras. Humans use experience and prior knowledge to make assumptions about parts of the scene that are hidden from view. One example of this is a speaker behind a podium; an audience member would know that a human standing behind a podium most likely has legs, and that the podium does not continue back into infinity because the objects in the scene (human, podium), are known quantities that have been encountered before. The computer does not have this same experience; image processing techniques alone would only allow the computer to know about what is directly visible.

In this paper, the authors explore the use of Qualitative Spatial Reasoning (QSR) methods and basic physical properties in addition to visual information from the scene to reduce the amount of incomplete visual information. Spatial information gained from QSR and physics is retroactively applied to the scene to further reduce ambiguity; present knowledge about a system is used to revise past assumptions, which improves the precision of current and future data.

2 Background and Related Work

2.1 Image Processing and Disparity

Image processing is an important field in computer and robotic vision. A large portion of research in this area has been devoted to finding computationally efficient algorithms; images are inherently two dimensional, which implies that most naive algorithms are at best $O(m \times n)$ in their computational complexity for an $m \times n$ image. The persistence of high resolution images (full high definition already being common and 5k resolution beginning to emerge) means that these algorithms will be computationally expensive, especially with many image formats being 4-channel (giving an $m \times n \times 4$ data structure size to hold RGBA or HSVA (Hue Saturation Value Alpha) information, two popular information formats).

Disparity [3, 4, 5] and the parallax effect are two concepts exploited in image processing to mimic human perception of depth. It uses the fact that objects (and distances) appear smaller the farther away from the observer they are. Thus, by determining the distance between occurrences of an object in each of a pair of stereo images, the relative distance from the cameras to the object can be determined. Disparity has also been used to estimate the motion of objects [5]. It is an invaluable tool in determining spatial information from multiple observations of the same scene.

2.2 Human Perception in Three Dimensions

Human 3D perception is fascinating: by all reckoning, such a feat should be mathematically impossible with the abstract data the brain receives from the eyes [1]. Regardless, humans are capable of making relatively consistent judgments about shapes and motion in three dimensions using only data from two “cameras” (the eyes) and a base of experience. Learned behavior such as object permanence [6] show that prior knowledge is required to make judgments about three dimensional space. Mimicking human perception with a computer is an important facet of computer and robotic vision.

2.3 Qualitative Spatial Reasoning (QSR)

Qualitative Spatial Reasoning (QSR) has varying applications in Geographic Information Systems (GIS), visual programming language semantics, and digital image analysis [7, 8]. Systems for spatial reasoning over a set of objects have evolved in both expressive power and complexity. The design of each system focuses on certain criteria, including efficiency of computation, ease of human comprehension, and expressive power.

The spatial reasoning system chosen for this investigation is VRCC-3D+ [9], an expansion and implementation of the RCC-3D [10] system designed by Albath et al. As opposed to other RCC systems (many of which have no implementation), the relations in VRCC-3D+ express both connectivity (in 3D) and obscuration. Obscuration will change from viewpoint to viewpoint, but connectivity is a global property that can be used to discern new information at every perspective in the system.

For this work, the authors focus on the obscuration element of the VRCC-3D+ relation. The connectivity portion of the relation will become important as the system is expanded to handle an arbitrary number of cameras and vantage points. VRCC-3D+ identifies four basic kinds of obscuration: no obscuration (*nObs*), partial obscuration (*pObs*), complete obscuration (*cObs*), and equal obscuration (*eObs*). The system further breaks each base obscuration into three different classes: regular obscuration (object A obscures object B), converse obscuration (object A is obscured by object B), and equal/mutual obscuration (object A and object B obscure each other). At this point in the investigation, this further classification is unimportant; it only matters if obscuration is present between two objects, not which object is being obscured.

2.4 Inertia and Conservation of Mass and Energy

There are a multitude of physical properties that can be used to discern information about spatial relationships. Every property used to derive spatial information introduces a new computational cost and has an upper limit to the amount of information it can deduce. The ideal property would be one that would give insight into the system without requiring any new calculation. When this is impossible, the goal should be to maximize the ratio of information gain to computational cost. One of the goals of this research is to discover a combination of physical properties that maximizes this ratio.

As a starting point, two physical properties will be examined: inertia and conservation of mass and energy. Inertia is best described by its colloquial definition: an object at rest tends to stay at rest, an object in motion tends to stay in motion. More formally, inertia is the resistance a physical object has against a change in its state of motion or rest. This can provide useful insight into the physical relationship of two objects. Given two objects, if one passes behind another, it can be used to determine whether or not the objects collided at any point. In terms of spatial connectivity, this collision will correspond to a change from a disconnected (DC) state to an externally connected (EC) state. This in turn gives useful information because it defines a known point on the (possibly hidden) boundary of one or both objects.

Conservation of mass and energy will also be used in conjunction with inertia to gain additional information. If an object becomes obscured by another object, its trajectory can be estimated. If the actual trajectory is different than the calculated trajectory, then something must have changed the state of motion or rest of one or both objects. Using the difference in expected and actual position at a given time to revise earlier calculations results in a corrected physical model that yields additional information about the entire system.

2.5 Current Work in Qualitative Spatial and Temporal Reasoning

Qualitative spatial and temporal reasoning has been an active field in recent years. Takahashi [11] explored using a new expansion to RCC-8 in which he uses two specific vantage points at right angles to a scene. Connectivity and obscuration were determined from each location to give a more precise determination about the objects in the scene. Takahashi's work differs from this work in that he uses a front and bird's-eye ("side" and "upper") view to obtain information. In contrast, this work focuses on emulating human sight using stereo images, which will be expanded to include information from additional visual sources.

Renz [12] proposed efficient algorithms for determining tractable subsets of RCC-8 and the Interval Algebra by phrasing the problem as a consistency satisfaction problem CSPSAT(S) and refining the set when necessary. Directly determining the relations between objects in space and time is not a direct consequence of these tractable subsets, but any reduction in the size of the subset of possible relations can increase the efficiency of determining actual relations between objects [13]. These tractable subsets can be used to aid in the disambiguation of information from multiple sources and will be exploited in this research.

Renz and Ligozat [14] performed a theoretical analysis of spatial temporal reasoning systems and showed that if a system exists such that weak composition does not result in actual composition, path consistency no longer applies. In these cases, algebraic closures of the system must be used to determine composition. They examine the effects of weak composition on spatial temporal reasoning systems and provide a methodology to analyze spatial and temporal calculi. While purely theoretical, this work benefits qualitative spatial and temporal reasoning. Path consistency and composition are two important attributes of a QSR system that have been exploited to aid in automated reasoning; analysis of this work to show these facets of spatio-temporal reasoning are not violated will be important to the continued usefulness of the system.

Ye and Hua [15] explored using depth cameras to determine three dimensional spatial relations. They did not apply their work to a series of images over time, and use specialized depth finding cameras to determine depth (the Xbox Kinect). As the research presented in this paper is expanded to include additional visual sources, Ye and Hua's work may be investigated as another kind of information source.

In 2007, Santos [16] investigated a framework in which the depth and motion of an object may be reasoned with while accounting for the observer's viewpoint. He presented a formal logic based approach to reasoning about depth and motion that he used in a robotic vision application called the Depth Profile Calculus (DPCC). DPCC uses depth maps obtained through disparity calculations to determine information about three dimensional space, but ignores many other visual cues available (such as color, lighting, and other physical properties). In this work, the authors use similar methods, but incorporate additional information to get a more correct view of the world.

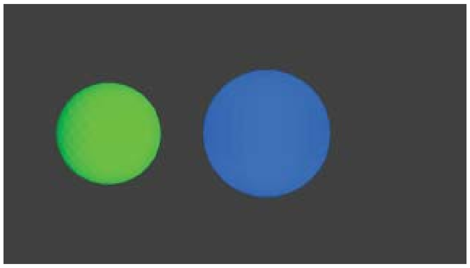
3 Computational Spatial and Temporal Reasoning

As an initial exploration, the authors constrained the system of interest as follows:

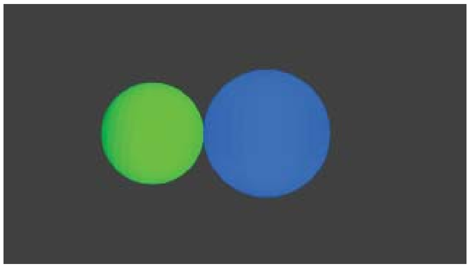
- The system is modeled as a single rolling green sphere that passes behind a stationary blue sphere but does not collide (Fig. 1).
- The system is simulated using Blender 2.64 [17] with two separate camera positions to guarantee that frames from the cameras would be showing different perspectives of the same point in time.
- The cameras were aligned such that the direction of views were parallel and the top row of the left camera's image corresponded to the top row of the right camera's image. This differs from human vision slightly, as the computer does not need to "focus" on an object by pointing both cameras at it; its visual information is more complete than a human's over the entire image.
- The floor of the system was transparent.
- The moving sphere's trajectory was perpendicular to the view direction of the cameras.

The preceeding constraints were placed on the system to allow simplifications that are considered to be unimportant in the context of this work:

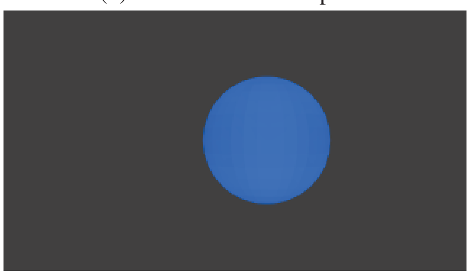
- Masking the image using HSV (Hue Saturation Value) values was used for image segmentation into objects.
- Disparity was calculated for each object by finding the center of the matching bounding box of objects and determining the difference in the x direction.



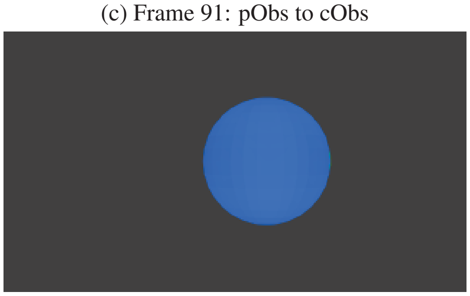
(a) Frame 33: nObs



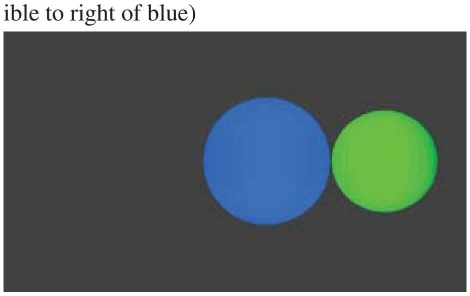
(b) Frame 50: nObs to pObs



(c) Frame 91: pObs to cObs



(d) Frame 101: cObs to pObs (Green barely visible to right of blue)



(e) Frame 145: pObs to nObs

Figure 1: Images from analyzed video: as seen from the left camera. The green sphere is further from the cameras than the blue sphere, and as such appears smaller.

Analysis of stereoscopic videos is a three step process: *frame analysis*. *obscuration analysis*, and *object analysis*.

3.1 Frame Analysis

In the context of this work, a *frame pair* is a pair of stereo images from a left and right oriented camera that portray the same moment in time from different perspectives. For every frame pair in the videos the following actions are taken:

- The images are converted from the default representation to HSV
- Range filtering is used to determine the locations of both objects in the images.
- The disparity and bounding rectangle are calculated for each object.
- The bounding rectangle and disparity for each object are stored, along with the frame number.

3.2 Obscuration Analysis

For this paper, obscuration and object analysis occur with respect to the left camera. The results could be refined by using information from both cameras.

The following pseudocode is used to determine the obscuration from the left camera at every step. The list of bounding rectangles and disparities from the frame analysis is stored in steps. The green sphere corresponds to object A in the pseudocode, and the blue sphere is object B, and bbox refers to an object's bounding box

```
obss = [] #the list of obscurations
for s in steps:
  if object A has a bbox in s:
    xa = A's bbox x location in s
    wa = A's bbox width in s
    if object B has a bbox:
      xb = B's bbox location in s
      wb = B's bbox width in s
      if the bboxes overlap:
        lastO = 'pObs'
      else:
        lastO = 'nObs'

    else:
      lastO = 'cObs'
  else:
    lastO = 'cObs'
  obss.append(lastO)
```

The eObs obscuration type is combined with cObs; not enough information exists in this experiment to distinguish between cObs and eObs.

Note that there is no distinction as to which object obscures the other, just that some obscuration occurs. It was visually verified that this code correctly identified changes in obscuration with respect to the left camera's video feed. Fig. 1 shows the frames identified as changes in obscuration occurred from the left camera's perspective.

3.3 Object Analysis

In the object analysis step, the positions and depths of each object are determined. Position is determined using the right most edge of the bounding box. If no obscuration is detected from either perspective, the depth and position of the object are directly recorded. Otherwise a polynomial is fit to the previous values recorded and used to estimate the current location. Every direction of movement (x,y,z) is handled independently. Due to the simplicity of the nature of this system, a linear fit was used; as the system is generalized, the order of the polynomial can be increased to handle differing kinds of acceleration and forces.

4 Experimental Results

Fig. 2 shows the positions of objects from a birds eye view of the system. Every marker on the graph shows an observed or estimated object location of a particular object. This information can be remarkably helpful in learning about the structure of the system. For example, it may be possible to determine from using only the stereo images that the blue sphere does not extend into infinity due to the perspective nature of the projections. However, depending on the intrinsic properties of the camera, there could be a large area in space that may or may not contain the blue sphere. Using the information gained from projecting the path of the green sphere behind the blue sphere, it can be concluded that the green sphere did not collide with the blue sphere, so an upper bound is placed on how far back the blue object can extend.

This figure illustrates that this line of inquiry shows promise: a relatively accurate extrapolation of the green ball's location is feasible with a relatively small number of data points. This estimation could be improved further by including the observed position of the green ball in later frames, then using that information to retroactively correct the estimations of the balls location. This will allow information inferred from that estimation to be refined even further.

5 Conclusions and Future Work

Using physical properties in conjunction with QSR and image processing methods is a promising direction in the field of computational vision and spatio–temporal reasoning. This could have applications in physical security (automated CCTV analysis), media analysis, and many other multimedia fields.

In this paper, the authors have initiated an exploration into using these three areas to accomplish automated spatio–temporal reasoning. The results of this initial research are encouraging but leave room for improvement and refinement. This work will be continued to allow analysis of systems with fewer constraints, add additional physical properties that are considered, and eventually be applied to video feed of live events from cameras, not just rendered physical simulations. Different combinations of physical properties and image processing techniques will be investigated to find a high information gain to computational cost ratio.

References

[1] J. T. Todd, "The visual perception of 3D shape," *TRENDS in Cognitive Sciences*, vol. 8, no. 3, pp. 115–121, Mar. 2004.

[2] D. Dunning and E. Balcetis, "Wishful Seeing: How Preferences Shape Visual Perception," *Current Directions in Psychological Science*, vol. 22, no. 1, pp. 33–37, Feb. 2013.

[3] S. T. Barnard, "Disparity analysis of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 4, pp. 333–340, Jul. 1980.

[4] K. Mühlmann, D. Maier, J. Hesser, and R. Männer, "Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation," *International Journal of Computer Vision*, vol. 47, no. 1–3, pp. 79–88, Apr. 2002.

[5] D. Demirdjian and T. Darrell, "Motion Estimation from Disaprity Images," in *Proceedings of the Eighth IEEE International Conference on Computer Vision ICCV*, vol. 1, 2001, pp. 213–218.

[6] R. Baillargeon, "Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants," *Cognition*, vol. 23, no. 1, pp. 21–41, 1986.

[7] A. G. Cohn, B. Bennett, J. Gooday, and M. M. Gotts, "Qualitative Spatial Representation and Reasoning with the Region Connection Calculus," *Geoinformatica*, vol. 1, no. 3, pp. 275–316, Oct. 1997.

[8] D. A. Randell and M. Witkowski, "Using Occlusion Calculi to Interpret Digital Images," in *Proceedings*

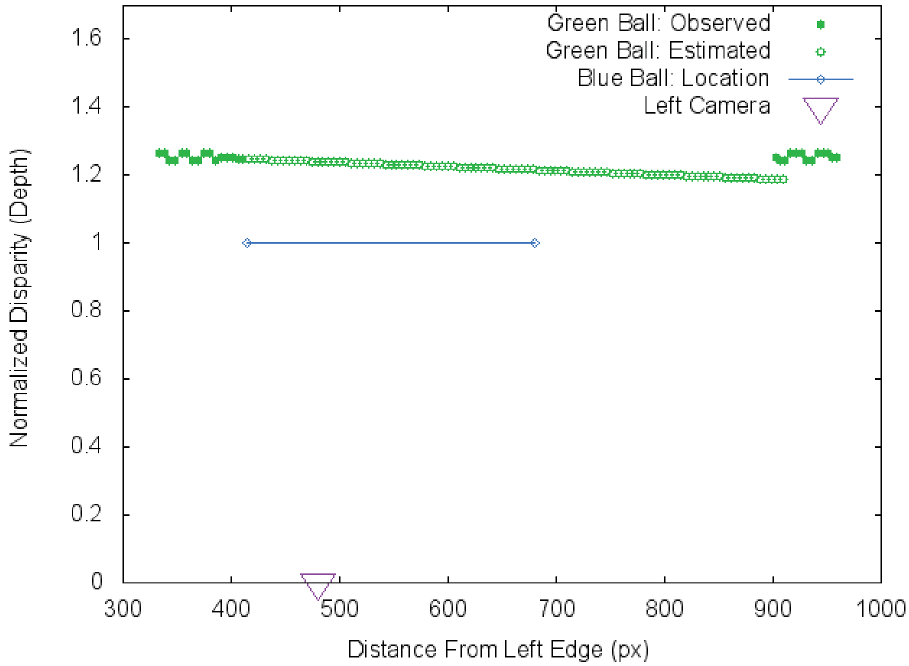


Figure 2: Observed and extrapolated positions. Each data point is a frame. Motion is from left to right.

of the 17th European Conference on Artificial Intelligence, Aug. 2006, pp. 432–436.

[9] C. L. Sabharwal, J. L. Leopold, and N. Eloe, "A More Expressive 3D Region Connection Calculus," in *Proceedings of the 2011 International Workshop on Visual Languages and Computing (in conjunction with the 17th International Conference an Distributed Multimedia Sysems (DMS 11))*, Aug. 2011, pp. 307–311.

[10] J. Albath, J. L. Leopold, C. L. Sabharwal, and A. M. Maglia, "RCC-3D: Qualitative Spatial Reasoning in 3D," in *Proceedings of the 23rd International Conference on Computer Applications in Industry and Engineering*, Nov. 2010, pp. 74–79.

[11] K. Takahashi, "Reasoning about Relative Relationships in 3D Space for OObject Extracted from Dynamic Image Data," in *Proceedings of the 2012 ECAI Workshop on Spatio–Temporal Dynamics*, Aug. 2012, pp. 45–51.

[12] J. Renz, "Qualitative Spatial and Temporal Reasoning: Efficient Algorithms for Everyone," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, Jan. 2007, pp. 526–531.

[13] N. Eloe, J. L. Leopold, C. L. Sabharwal, and D. McGeehan, "Efficient Determination of Spatial Relations Using Composition Tables and Decision Trees," in *2013 IEEE Workshop on Computational Intelligence for Visual Intelligence*, 2013, pp. 1–7.

[14] J. Renz and G. Ligozat, "Weak Composition for Qualitative Spatial and Temporal Reasoning," in *Principles and Practice of Constraint Programming – CP 2005*, ser. Lecture Notes in Computer Science, P. Beek, Ed., 2005, vol. 3709, pp. 534–548.

[15] J. Ye and K. A. Hua, "Exploiting Depth Camera for 3D Spatial Relationship Interpretation," in *Proceedings of the Multimedia Systems Conference*, Feb. 2013.

[16] P. E. Santos, "Reasoning about Depth and Motion from an Observer's Viewpoint," *Spatial Cognition and Computation: An Interdisciplinary Journal*, vol. 7, no. 2, pp. 133–178, Dec. 2007.

[17] Blender Foundation, "blender.org - Home," http://blender.org, Apr. 2013.

Institutional Services and Tools for Content, Metadata and IPR Management

Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, Michela Paolucci
DISIT, Distributed Systems and Internet Technology Lab,
Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Firenze,
Firenze, Italy
{pbellini,ivanb,paolucci}@dsi.unifi.it, paolo.nesi@unifi.it

Abstract **— Multimedia services need to be supported by content, metadata and workflow management systems to efficiently manage huge amount of content items and metadata production as in the case of cultural institutions. Online digital libraries and cultural heritage institutions and the other portals of publishers need an integrated multimedia back office in order to aggregate content and provide them to national and international aggregators. Different technologies need to be integrated to improve existing content management and workflow systems in order to efficiently organize and manage large amount of data and processes to cope with them. The aim of this paper is to formalise and discuss about requirements, design and validation of an institutional aggregator for metadata and content, coping with IPR models for providing content towards Europeana, the European international aggregator. ECLAP, the European Collected Library of Artistic Performance, has been realised respecting all these features and taking into account the problems connected to the cultural heritage cross media content on Performing Arts domain. In order to establish the quality of the institutional services and tools for content described, a set of measures have been made and reported.**

Keywords: institutional archive; content aggregator; grid computing; workflow; metadata enrichment; metadata validation; semantic computing; IPR management.

I. INTRODUCTION

Multimedia services need to be supported by content, metadata and workflow management systems to efficiently manage huge amount of content items and metadata production. With the introduction of web 2.0/3.0, and thus of data mining and semantic computing, including social media and mobile technologies most of the digital libraries and museum services became rapidly obsolete and were constrained to rapidly change. In most cases, the cultural institutions see their content ingested, promoted, distributed and exploited by final users via online commercial partners (e.g., Google, Amazon, YouTube), that may take benefits to commercial resell and/or via advertising. Thus the online digital libraries and cultural heritage institutions such as ACM, PubMed and IEEE and the other portals of publishers need an integrated multimedia back office. In Europe and in US, most of the cultural heritage institutions aggregate content and provide them to national and international aggregators such as Europeana in Europe [1], Library of Congress in US [2]. Europeana has more than 24 million of contents coming from about 2200 different content Providers, and about 100 content aggregators. Specific workflow and metadata processing, enrichment (name resolving, date analysis, linking with open data, creation of relationships, etc.) are needed to cope with the content aggregation and publication. To this end, specific solutions and architectures are needed that present some aspects or can be integrated with traditional workflow management systems [3], [4]. In [15] a visual tool for defining authorization workflow models for e-commerce application has been proposed.

In this paper, the requirements, the design and the results of the ECLAP aggregator of Europeana are reported. ECLAP stands for European Collected Library of Artistic Performance, [10]. It has been started as an European Commission project CIP-ICT-PSP.2009.2.2, Grant Agreement N°250481. ECLAP has been designed and developed to ingest content coming from 35 different cultural international institutions in 13 languages thus creating a considerable online archive for all the performing arts in Europe [12]. Until now, it has ingested and processed about 180.000 objects (images, document, video, audio, 3D, Braille, e-books, etc.) with 940.000 items (pages, images, video, audio).

The paper is organized as follows. In section II, general requirements of the back office and tools for cultural heritage content aggregator are reported. Section III provides ECLAP overview describing the architecture and ECLAP institutional services and tools, while a report on tools usage is shown in Section IV. Conclusions are drawn in section V.

II. GENERAL REQUIREMENTS

The main requirements of the back office tool for cultural heritage content aggregator are reported as follows. The back office has to be capable to:

*1) **Ingest** a large range metadata formats (XML based or Dublin Core, METS, MPEG-21, etc.) coming from different channels (http, ftp, oai-pmh, etc.). To ingest implies to get the metadata and content, link them together, collect them on a suitable storage, and ingest IPR information. The ingestion may be performed from web pages uploads, and/or from batch processing and/or crawling.*

*2) **Perform human content enrichment,** such as metadata translations, validation; addition of comments; social media promotion; voting/rating; promoting; publication to other portals; editing and performing corrections; quality assessment, etc. Multiple activities for users imply to have the possibility of granting different authorizations.*

*3) **Perform automated activities,** such as: estimation of technical parameters (duration, size, etc.), extraction of descriptors, indexing, automated translations, searching for VIP names, geonames resolutions, linking with LOD, metadata assessment (completeness and consistency [6], [14]), IPR (Intellectual Property Rights) verification. Among these activities, it has to be included the content adaptation and repurpose according to a large set of the distribution channels. For example, ingesting video in any format and producing the multiple formats for distribution on different devices.*

*4) **Harmonising the activities of human and automated processing** among the above mentioned. For example, identifying when the human actions are needed, taking trace of the performed manual activities, blocking the humans when the automated elaboration has locked the resource.*

*5) **Scale up of the back office architecture** to cope with large number of transactions on metadata information and activities in the back office. In most cases, the back office has to be capable to process large data sets per day, and thus the execution of massive processing on distributed resources as GRID is needed [7].*

*6) **Support and model one or more workflows** according to each specific content life cycle. Workflow management also means to have different user roles for different activities. For example, not all ingested content is ready to be published at the same time. Thus, some users may be entitled to move forward the status of their content, while other may need approval.*

*7) **Cope with the IPR** modelling, assignment and verification. It also implies that the IPR model may regulate uses/accesses to the digital content, and the exploitation of rights about the content manipulation and reuse according to the owner rights [11].*

Moreover, a large number of detailed requirements have been identified as reported in [8]. Among the additional requirements, we noticed the need of (i) logging and keeping trace of metadata versioning: to keep trace of the work done and changes performed, (ii) formalizing and managing a number of different roles / capabilities to be assigned to the ECLAP users (i.e., enricher, publisher and validator), (iii) providing user accessible tools for: multilingual metadata editing, IPR managements, massive content editing of some specific object status associated with metadata (workflow status, IPR, public/non-public, tags, groups, etc.).

III. ECLAP OVERVIEW

The ECLAP architecture for content and metadata management (see Figure 1) consists of three main areas: Metadata Ingestion Server, ACXP back office services [7] and ECLAP Portal. The Metadata Ingestion Server, which collects metadata provided by digital archives and libraries (realised by using MINT metadata mapping tool, [8]). There the metadata coming in different schemas are mapped according to the ECLAP semantic model and are made available through the OAI-PMH protocol. ACXP back office services provide automated procedures for content and metadata processing (harvesting, ingestion, analysis, production, adaptation, validation, publishing, etc.). The ECLAP portal is the Drupal based front end, which provides front-office tools to work on content and metadata, IPR models definition, content management and Europeana publishing. ECLAP provides a social media style front end service (BPN, Best Practice Network) with more than 2200 users; directly linked via service oriented interface to the more complex back office (based on AXCP) capable to really cope with the complexity of managing a complex workflow and content processing and ingestion. The most important activities addressed by ECLAP are performed in the automated back office.

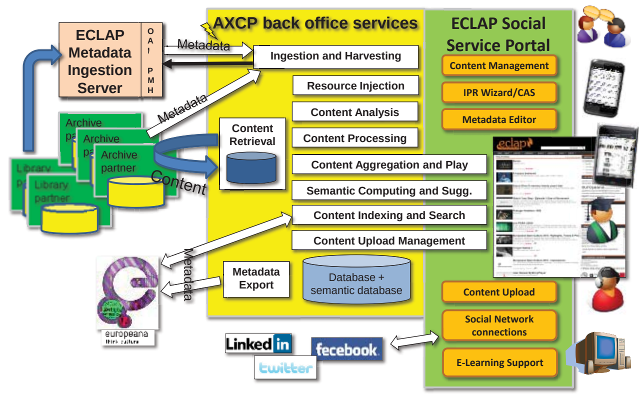


Figure 1 - ECLAP architecture

In order to better understand the content and metadata management, it is useful to describe the ECLAP Overall Scenario in terms of workflow, rules, procedures, etc., which each Content Provider follows to publish content on ECLAP and then provide it to Europeana (see Figure 2). All content managed in the ECLAP are associated with a specific workflow. In event of Europeana based ECLAP workflow, content has to be: (i) uploaded/ingested; (ii) enriched through metadata (some metadata must be sent to Europeana and others are necessary to describe and manage the content in the ECLAP); (iii) associated with an IPR Model (through the IPR Wizard, as described in next sections) [9].

The content uploaded/ingested is initially available on the ECLAP BPN with maximum restrictions, while metadata are immediately available for indexing and searching for all kind of ECLAP users. Only content presenting a (i) sufficient set of metadata (e.g., Europeana mandatory metadata) and (ii) IPR information license defined (one from the set admitted by "europeana:rights" in [13]), can be published on Europeana, [1].

Front office and back office tools of ECLAP allow covering the whole content life-cycle. The AXCP grid solution adopted provides a scalable back office capable to cope with a huge amount of content and metadata processing capabilities and features [7].

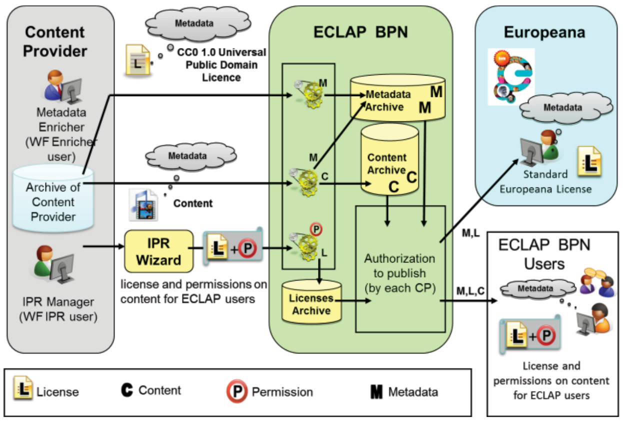


Figure 2 - ECLAP Flow Overall Scenario

A. Front office tools

The ECLAP front office allow users covering the whole content life-cycle: content upload, enrichment, validation, IPR modelling and editing, content and metadata assessment and management, publication, etc.

WEB based content upload allows users uploading content on the portal through the Upload web page. The page shows a form where users input: DCMI metadata consisting of a set of Dublin Core fields; Taxonomy as multiple classification terms selection; Groups assignment (ECLAP Groups); Resource data by selecting one or multiple files from user's HD device or a valid URL (via http or ftp); Workflow type associated with the life-cycle of content and IPR model, if accessible.

Metadata Editor is the tool for editing, enriching and validating metadata. According to the user role, the editor works in the "enrichment mode" or in "validation mode", respectively for those enabled users. Metadata editor allows editing any kind of metadata organized in specific panels to work on. All changes made on metadata are tracked to maintain the history of changes and who made it and when.

IPR wizard tool allows creating IPR Models that formalize the owner rights related to publishing content online in the ECLAP context. The IPR Logic Model implemented takes decisions for the IPR Managers according to the relationships: among user roles and among permissions [9]. The IPR manager has just to select one or more permissions for a user role that he/she wants to associate with an IPR Model (and therefore to a set of contents) and the wizard automatically selects also the permissions implied by those chosen (e.g. download imply play, see Figure 3 for the list of permissions). When some restrictions are applied Creative Commons Licenses [10] cannot be associated with the IPR Model, so the user can choose the license from one of the restricted licenses allowed by Europeana ("Unknown copyright status" or "Right Reserved – Restricted access"). In ECLAP, many different set of content permissions (rights) can be imposed by the content owners, which are the ECLAP Content Providers. For example: content and metadata upload methods; metadata standards and formats; IPR on content (licenses, permissions, etc.); collection topics; etc. Permissions managed on the ECLAP Portal can be referred to the following aspects:

- access to the content (e.g., the content can be accessible via progressive download and/or download)
- user device (e.g., the content can be played via a PC and/or a mobile device, iPad, etc.)
- content resolution (e.g., the content can be accessible only in a reduced Low Resolution and/or in High resolution).

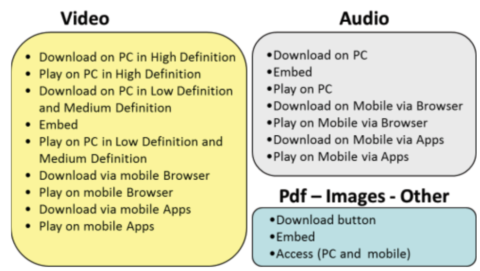


Figure 3 - ECLAP Permissions

Content Management tool allows users to manage contents and publish them to Europeana and to perform massive editing on large set of content elements.

B. Back office services

The ECLAP back-office services consists of a set of grid processes that run automated workflow processes both on a single and on multiple contents.

Automated ingestion – It ingests metadata and content coming from ECLAP partners and Digital Archives and from the external metadata mapping tool MINT. The process allows ingesting both massively and singularly metadata and digital resources. When resources are big file, they are provided by using physical device. In this case, ECLAP just starts with metadata ingestion and when the digital resources are available, the joining is performed.

Content production and adaptation - This process works with the digital resource and metadata uploaded via web or ingested. Metadata and digital resource are retrieved from the CMS or the storage area or downloaded from the provided URL. Incoming metadata (Dublin Core, Taxonomy, Groups, Performing Arts metadata, workflow type, user) are enriched with technical metadata built by analysing the digital resource: (i) content format (document, audio, video, image, crossmedia), (ii) content type (file format), (iii) structural information (size, duration, number of pages). The produced enriched metadata and digital resource are aggregated and published in the publication database. Metadata are indexed to make the content ready for access on the portal. The production process works also when the digital resource has to be replaced with a new one (Update). To make the incoming digital resource accessible by different devices Content Adaptation processes are exploited: (i) Content adaptation to different resolutions produces content accessible by different devices (iPhone, iPad, Android, Windows Phone, etc. and on the ECLAP portal, any browser.); (ii) Video adaptation produces the Low, Medium and High Definition versions of a video; (iii) Metadata Translation translates Dublin Core metadata and missing metadata in different languages by using tool or web service for text translation.

Content management - During the life-cycle of content, massive actions on content could be needed: changes in the workflow status, changes in the metadata, addition of details in the metadata sets, etc. Specific actions are also needed to maintain and manage the content and work both on single content and multiple such as: delete content, update metadata, and publish content uploaded by common users.

ECLAP back-office services and front-office tools work both on content and metadata. However, such processes have to work in concurrency: back-office services could access and process content in parallel to the user activities on the front-end. Activities of translation, enrichment, validation, IPR definition and assessment cannot be performed by more than one process at time on the same content. On the other hand, sequential processing is too expensive and time consuming to sustain the content workflow and ingestion. In ECLAP, several thousands of new content per day have to be processed. To this end, a workflow state diagram has been modelled, formalized and implemented. Therefore, to manage the concurrency and to guarantee a safety access to the content a mechanism of lock-unlock access has been defined. The general workflow state diagram is coded as described in Figure 4.

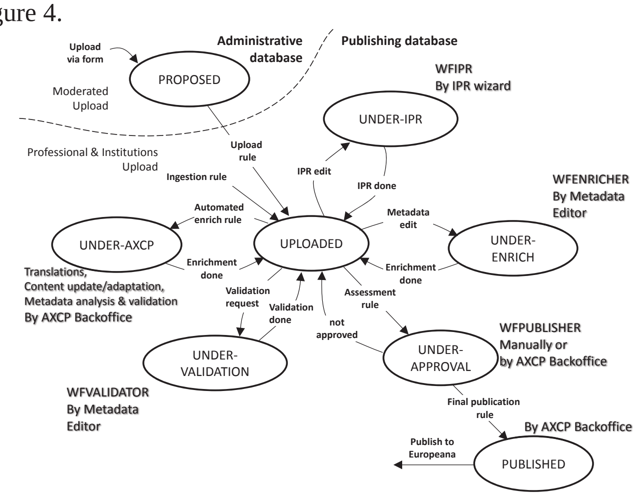


Figure 4 - ECLAP Workflow diagram

C. *User Roles and Workflow Model*

The front-office tools have to allow working on metadata in different ways. In order to implement am high quality content enrichment process, each specific activity has to be granted to specific people according to their skill, language and the identification of the institutional Content Provider (CPID). Moreover, the solution has to keep trace of each single metadata change to have the evidence of the work performed and eventually recover wrong situations. To this end, the following user roles have been defined with their parameters:

- **WFIPR (CPID):** authorizes the definition and validation of IPR models, and IPR assignment to the content of the CPID; by using the **IPR Wizard** and during the Upload for the IPR Model Assignment.
- **WFENRICHER (CPID, {languages}):** authorizes the metadata enrichment and changes in the specified languages (add, edit metadata).
- **WFVALIDATOR (CPID, {languages}):** authorizes the validation of the metadata for the identified language. The metadata fields can be singularly validated until the object may pass the whole approval phase.
- **WFPUBLISHER (CPID):** to take the final decision for publishing on ECLAP and on Europeana. The publishing of single or groups of content can be performed by using the **Content Management Tool and AXCP,** together with much other functionalities, plus eventual new actions to be programmed on the same tools.

Back-office services are not associated with a specific user role since they are performed as root user by rules on AXCP computing grid background automated processes on content and metadata.

IV. Report on tools usage

In this section, the results about the ECLAP back office activities performed on the content, metadata and IPR until April 2013 is reported. This service allows users and automated workflow processes to interoperate securely and to increase the quality and accessibility of content and metadata, without any creation of conflicts each other. It is currently in use by 31 institutions. The number of state transitions and their distribution in the time period put in evidence the whole activity of the portal on content and metadata and allow analysing singularly both the back-office and the user activities. Some results are reported in the temporal domain considering the "month" as a time period unit.

A. *Workflow Users*

Actually, there are 29 workflow qualified users. Each user may have single or multiple user roles (grant authorization). The workflow user roles are distributed as: 24 enrichers (WFENRICHER), 6 validators (WFVALIDATOR), 23 IPR users (WFIPR) and 9 publishers (WFPUBLISHER).

B. *Workflow Transitions*

At current date, 706,052 workflow transitions have been handled for 117,861 content items with an average of 6 transitions per content and a maximum of 104 transitions for a single content. These transitions were performed in 653 days with an average of 1,014 transitions per day and a maximum of 13,162 transitions in one day, with a maximum of 14 different virtual nodes in the AXCP grid, on DISIT Cloud.

TABLE I. Distribution of workflow transitions

From	To	Number of Transitions
'Uploaded'	'Under-AXCP'	179912
'Under-AXCP'	'Uploaded'	179912
'Proposed (creation)'	'Uploaded'	117861

'Uploaded'	'Under-Approval'	113549
'Under-Approval'	'Published'	111362
'Uploaded'	'Under-IPR'	929
'Under-IPR'	'Uploaded'	929
'Uploaded'	'Under-Enrichment'	611
'Under-Enrichment'	'Uploaded'	611
'Under-Approval'	'Uploaded'	212
'Uploaded'	'Under-Validation'	38
'Under-Validation'	'Uploaded'	38
'Published'	'Uploaded'	3

C. Back-Office services

The back-office services consist of a set of grid processes that run periodically automated workflow processes both on a single and on multiple contents.

1) Content and Metadata Ingestion

The number of content ingested and processed by the back-office has been 106,525 corresponding to the UPLOADED workflow state of content.

2) Metadata Analysis

Metadata analysis for assessment or automated translation performs a transition to the UNDER-AXCP in order to lock the content and avoid that a user could be access to it for manual editing or validation. In total, 179,912 of these transitions were performed.

3) Metadata Validation

Every time content passed the metadata analysis the back-office performs a transition to the UNDER-APPROVAL. In total, 113,549 of these transitions were performed.

4) Content Publication

Every time the back-office performs the publication of content in the UNDER-APPROVAL workflow state it performs a new transition to the final state: PUBLISHED In total, 107,598 of these transitions were performed.

D. Front-Office tools

In this section the analysis of the activity performed by users via front-office tools is reported.

1) Web Page Upload

11336 content has been manually uploaded by users via the Web Page Upload. Once uploaded the process is passed to the back-office.

2) Metadata Editor

In order to evaluate the usage of Metadata Editor for the enrichment and validation activities, both the number of workflow transitions from UPLOADED to UNDER-ENRICH and from UPLOADED to UNDER-VALIDATION have been considered. The former transition gives a measure of manual enrichment activity, while the latter of the manual validation activity. The transitions related to enrichment were 611, and 38 for validation. Figure 5 reports the distribution in time of the enrichment activity.

3) IPR Wizard, IPR definition model

ECLAP *IPR Wizard* is largely used by more than 35 partners in Europe. To evaluate the usage of IPR Wizard, the number of workflow transitions from UPLOADED to UNDER-IPR over time have been tracked. The transitions were 929, and their distribution is reported in the Figure 6. Comparing Figures 5 and 6, it is evident that the IPR tool has been much more used than the metadata editing tool. This is due to the fact that, in most cases the content metadata where ingested by stable institutional databases and archives, while the IPR model was missing on those archives; and Europeana constrained them to formalized the IPR aspect before the content submission.

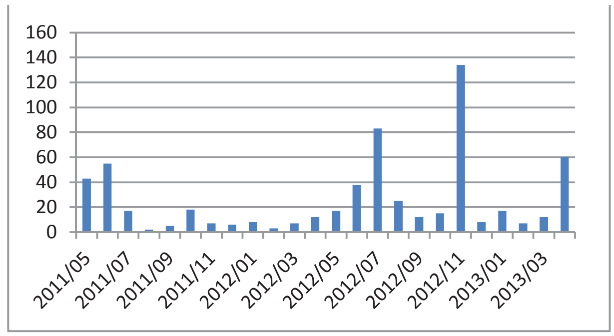


Figure 5 - Enrichment Activities

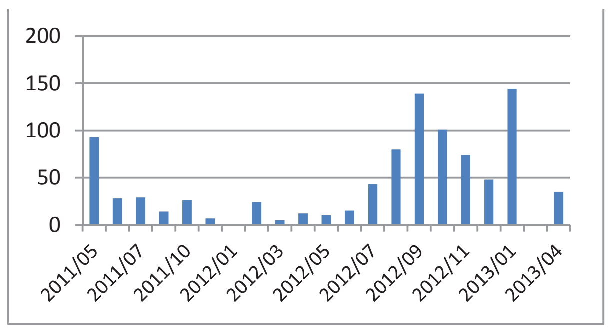


Figure 6 - IPR Wizard Activities

For the IPR aspects, 67 different IPR models/licenses have been used. 40 of them are restrictive not public models, while 27 are public models allowing the full content access. Most content providers used 1, 2 or 3 different IPR models/licenses for their content, while a few partners used 4, 8 or 12 models. Figure 7 reports he number of files used per IPR model. It can be seen that the most diffuse two models cover more than the one half of the whole content collection. On the other hand, the semantic flexibility of the IPR model proposed allowed to cope with many different needs of the content owners that impose the IPR according to legal rights they can really provide.

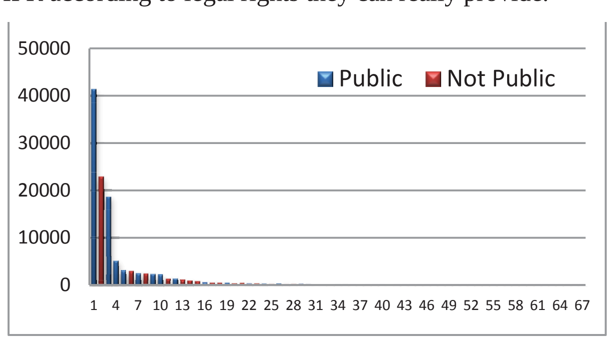


Figure 7 - Statistics on IPR Models.

The 68% of content is associated with a public IPR model. Regarding the 30 restrictive IPR models defined by a Content Provider, in 11 cases they restricted the access to only educational group users and in 6 cases restricted the access to group users only (educational and not educational). Moreover, 23 models have been used to not allow the download of the digital resource for some kind of resource type (regardless of the user type).

TABLE II. IPR MODELS ALLOWING PERMISSIONS BY USER TYPE

Permission	User type		
	public	*group*	*educ./research*
only play/access	11	13	19
download & play	3	8	11
no permission	19	12	4

Table II reports, for the three user types (public, group and educational), how many IPR models allow only play/access of the digital resource or allow the download and play of it or no permission are provided. It can be seen that in most cases the models are used to restrict access from the public users (19 over 30) and to limit the download of the resource.

4) Content Management Tool

To evaluate the usage of Content Management tool for publication activity we measured the number of workflow transitions from UNDER-APPROVAL to PUBLISHED made by partners. The transitions were 3764 and distributed by month as reported in Figure 8.

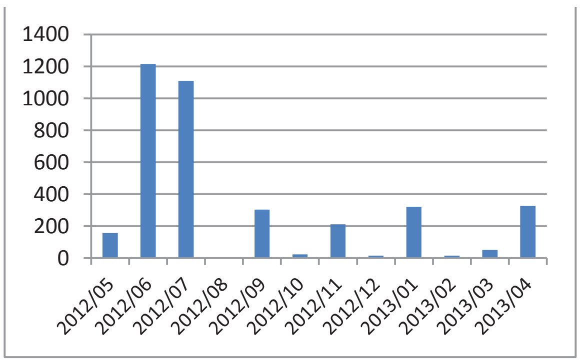


Figure 8 - Publication Activities.

V. CONCLUSION

The paper discussed about requirements, design and validation of ECLAP which is an institutional aggregator for metadata and content coping with IPR models for providing metadata towards Europeana, the European international aggregator. The proposed solution takes into account issues connected to the cultural heritage cross media content on Performing Arts domain and integrates front office tools and an automated back-office based on a Grid. The solution allows to cope with high quality and provides large scale multimedia services to manage huge amount of content and metadata, coming in turn from several national and local institutions: museum, archives, content providers. Finally, the usage analysis puts in evidence the whole activities of ECLAP on content, metadata and IPR until April 2013. It underlines that the huge activity on content and metadata aggregation, analysis and validation to match the Europeana requirements has been mainly automated and performed by the back-office, thus allowing to keep content processing cheap and sustainable. Regarding the front office side, the most used tools by content providers have been associated with IPR, namely IPR Wizard and the Content Management since they allow users to finalise the rights and to provide a connection of the content versus Europeana. Most of the metadata provided were already in a good shape and less than the 1% of content has been corrected from that point of view. On the other hand, the IPR details requested by Europeana constrained the content provider to associate to the100% of the content a new IPR model. This huge effort has been kept under control by exploiting the IPR Model, and applying only 67 models to the whole set of more than 120.000 different content coming from more than 35 different collections and institutions.

ACKNOWLEDGMENT

The authors want to thank all the partners involved in ECLAP, and the European Commission for funding ECLAP in the Theme CIP-ICT-PSP.2009.2.2, Grant Agreement No. 250481.

REFERENCES

[1] Europeana, http://www.europeana.eu

[2] The Library of Congress, http://www.loc.gov

[3] Yu, J., & Buyya, R.. A taxonomy of workflow management systems for grid computing. Journal of Grid Computing, 3(3-4), 171-200, 2005

[4] W.M.P. van der Aalst and K.M. van Hee. Workflow Management: Models, Methods, and Systems. MIT Press, Cambridge, MA, USA, 2002.

[5] European Library of Artistic Performance, ECLAP, http://www.eclap.eu/

[6] Bellini E., Nesi P. "Metadata Quality assessment tool for Open Access Cultural Heritage institutional repositories", Proc. of the ECLAP 2013 conference, 2nd Int. Conf. on Information Technologies for Performing Arts, Media Access and Entertainment, Springer Verlag LNCS, 2013.

[7] Bellini P., Bruno I., Cenni D., Nesi P., "*Micro grids for scalable media computing and intelligence on distributed scenarious*", IEEE Multimedia, 18 Aug. 2011, IEEE Computer Society Digital Library, ISSN: 1070-986X.

[8] "ECLAP DE3.1 infrastructure: ingestion and processing content and metadata", 2011, ECLAP Project, http://www.eclap.eu/urn:axmedis:00000:obj:a345a84f-6fdf-4f84-a412-88094ce363e2

[9] Bellini P., Nesi P., Paolucci M. (2013). IPR Management Models for Cultural Heritage on ECLAP Best Practice Network, IEEE International Conference on Communications, 9-13 June, Budapest, Hungary.

[10] Creative Commons, http://creativecommons.org

[11] Wang X., "MPEG - 21 rights expression language: Enabling interoperable digital rights management," IEEE Multimedia, 11(4):84–87, 2004

[12] ECLAP Partners. List and information on ECLAP Partners available at: http://www.eclap.eu/drupal/?q=node/3578

[13] 'Guidelines for the europeana:rights metadata element', v4.0 – 20, 9 February 2012, http://pro.europeana.eu/documents/900548/0d423921-23e0-45fa-82a4-2ac72b3e6f38

[14] Park Jung-Ran, Metadata quality in digital repositories: A survey of the current state of the art, "Cataloging & classification quarterly", vol. 47, nos. 3-4 (April 2009), p. 213-228

[15] S.K. Chang et al, "Visual Authorization Modeling with Applications to Electronic Commerce", IEEE Multimedia, (2003), Vol. 10, No. 1, pp. 44-54.

Emotion-Based Mashup for Social Media Contents

Mohamed Abdur Rahman
Advanced Media Laboratory, Computer Science Department
College of Computer and Information Systems
Umm Al-Qura University
Makkah Al Mukarramah, Kingdom of Saudi Arabia
marahman@uqu.edu.sa

Abstract— **In this paper, we propose an open source web-based emotion retrieval service for textual contents, called MediaTagger, which can extract emotional value from online content coming from diversified Internet-based services. MediaTagger mashes up state of the art emotion services and allocates the right emotion retrieval service with several emotion visualization metaphors based on the content of each service. MediaTagger also incorporates a flexible and adaptable emotion authoring service based on Naïve Bayes machine learning theorem. The authoring service within the system also helps creating new domains of emotion extraction services. Finally, we share some quantitative and qualitative results that show the viability of our system as well as the experience and satisfaction of regular end-users.**

Keywords-emotion analysis; emotion authoring; emotion mashup; multimedia

I. Introduction

Nowadays, we are surrounded with a lot of user generated content, thanks to the widespread acceptance of social networks. Examples include blogs, RSS news feeds, social networks such as Twitter, Email messages, image/video sharing services, etc. Contents from each source have emotional value to the information producer [1]. People leave their emotional footprint mostly in diversified social network services in the form of reviews, comments and answers through varieties of media. For example, people from all over the world upload videos of numerous domains such as weather, technology, news, sports, different products etc. People make comments about those entities while expressing their emotions. However, text containing user emotion about weather for example does not have same emotional value as the user comments about the review of a Microsoft Kinect for XBOX sensor. This leads to the fact that for every domain of knowledge, user emotion extraction requires separate emotion extraction knowledge.

In [3], the author analyzes the emotion classifications and states from cognition perspective. The work provides a 3D circumplex model that describes the relationships between different emotion possibilities. The authors in [4] present a platform called SenticNet for mining online opinions and discovering human emotions using common sense reasoning, polarity concepts and their developed characterization model. In [5], the authors discuss the topic of tagging in general especially handling image and photo tags in Flickr/ZoneTag online services. With the ever growing use of online Web 2.0 tools and especially customer reviews, the authors of [6, 7, 8] analyzed the effect of online customer reviews and emotions on new customer purchases and on the branding image, perception and marketing strategies of companies and vendors.

In [9], the authors' work tries to classify the sentiment of Twitter posts and trends using machine learning techniques. However, the algorithm needs to refine noisy knowledge and provide option of feedback and control mechanism to update and insert new knowledge. Meanwhile, the authors in [10] analyze keywords within microblog feeds such as Twitter, Plurk and Jaiku to learn about the sentiments of those keywords while visualizing the results using audiovisual interface through music tones (using dynamic arousal and valence values) that represent the sentiment of each microblog post. They make use of several factors such as response, context and friendship in deciding the sentiment labels. Some research is also done on evaluating the mood sentiments of video contents such as the work in [11] where the authors try to utilize low-level video features such as color and sound that are mapped to their corresponding Valence-Arousal values to determine the emotion within standalone video contents. However, the main issues with that work are that it targets standalone video files and the accuracy rate they provided is relatively low (about 60%).

Existing emotion extraction services provide support of a subset of domains as it needs specialized algorithms relevant to each domain. This limits the horizontal scalability. People have to search for those services that provide that type of emotion extraction. Mashing up existing services into one platform would give service consumers a great relaxation in emotion service consumption. In addition to the mash-up of existing services, an authoring process of providing facility to create new domain of emotion extraction service would usher in new era of emotion computing for multimedia contents. We aim to realize those goals in this proposed paper. We also leverage existing open source APIs to mash-up emotion services for both text and image contents. To provide authoring facility, we use Naïve Bayes [12] theorem, which provides both horizontal as well as vertical elasticity of emotion extraction capability. Naïve Bayes theorem is proved to provide excellent and fast results to classify new event given its associated features and using available classified training sets. Thus using Naïve Bayes theorem, we can add new dimension of emotion extraction verticals, by incorporating a dynamic knowledge base that gets adapted continuously through a feedback process that makes use of user expertise, locale, and preferences. As a proof of concept, we particularly process Twitter messages, YouTube video reviews, Email messages, weather status feedback, movie reviews, and facial expression from images.

The remainder of this paper is organized as following. In Section II we show the design of our system. Section III illustrates the implementation details of our developed system. In Section IV we provide our collected test results followed by concluding remarks along with our future objectives in Section V.

II. DESCRIPTION OF THE SYSTEM

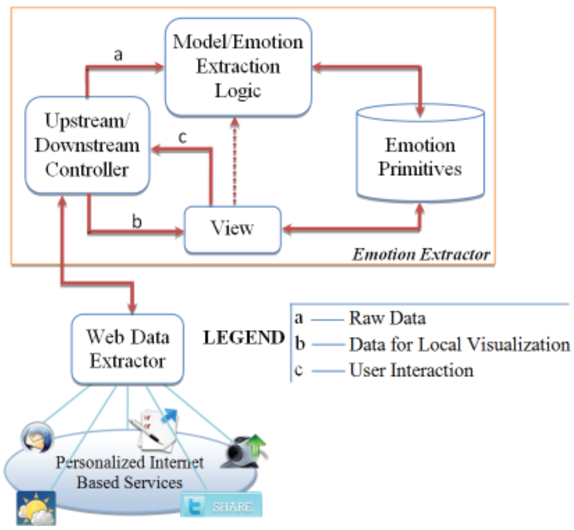


Figure 1. High level components of MediaTagger.

As shown in Figure 1, the MediaTagger system has two main modules namely: Web Data Extraction Module and Emotion Extraction Module. The components and functions of each module are described as following:

A. Web Data Extractor

Web Data Extractor module is responsible of collecting raw contents from their respective sources. We leverage our open source multimedia content extraction framework called SenseFace [2] to extract live content from heterogeneous Internet-based sources. The framework embeds a suite of protocols and algorithms that can communicate with complex and proprietary sources of existing heterogeneous Internet-based services and retrieve online multimedia content. The framework provides state of the art multimedia content extraction services from diversified mail servers, web sites, blogs, video sharing, and image sharing sites. To manage load balancing and scalability, the framework uses proxy servers where each proxy server actually listens to each type of content retrieval service request and depending on the number of concurrent service requests, more number of proxy servers may be employed within the system. A proxy server stores the list of content retrieval services available within the framework in XML format.

B. Emotion Extractor

This module is responsible for pre-processing and analyzing content, extracting emotion, indexing the emotion primitives, presenting the results to the user, and in the case of Naïve Bayes theorem, adapting the emotion value from the user feedback to train the system. The module utilizes the MVC (Model/View/Controller) design pattern to extract and present emotions. In its upstream data collection path, the Controller receives raw web content from the Web Data Extractor. The Controller also issues request to extract new media content in downstream path. Upon receiving raw media content, the Controller delegates the content to the model component i.e. Emotion Extraction Logic. This component mashes up all the emotion extraction services available within the framework and delegates the content to the most optimal service, depending on the media and user requirements. It also leverages the metadata of each API in the form of types of media support, response type per unit content, size of each payload per request, types of request and response i.e. JSON, XML, REST, number of requests per API call, type of domain knowledge supported, types of functionalities supported, types of emotional values supported (+ve, -ve, neutral), ranges of emotion value, and sentiment attributes such as affection friendliness, sadness, amusement, contentment and anger, to name a few. The unit could be horizontally enriched with N number of services.

The Emotion Extraction Logic APIs in MediaTagger are of two types: external APIs and those using Naïve Bayes learning theorem. In the former case, the Emotion Extraction Logic simply uses the I/O API methods without any training. In the latter case, it uses Naïve Bayes theorem. Thus, it uses three different working phases. These phases are:

1) Training phase: The system makes use of a Naïve Bayes theorem, which is a supervised learning method, to classify the emotion of the retrieved content assuming conditionally independent classification features. Classifications would include positive, negative or neutral sentiments. We use this theorem to evaluate the posterior probability of sentiment membership to classify new input sample according to its associated features (i.e. content text keywords). We do the same for all possible classifications. Thus, it would be easy to classify the new event to be the classification with highest posterior probability. The Naïve Bayes formula is summarized in equation 1 where C is the number of possible classifications.

$$p(y|x) = \frac{p(x,y)}{p(x)} = \frac{p(x|y)p(y)}{\sum_{y'=1}^{C} p(x|y')p(y')} \quad \text{--------(1)}$$

The nature of each media is different from one source to the other. For example, text nature within Email messages is different to a great extent from text content of Twitter posts and YouTube review, etc. So, the system was initially installed with large corpus of knowledge base that is used by the Naïve Bayes classifier. This attached information has predefined corpus of words and phrases that are tagged with certain classification. We have then trained and tested the system using targeted topics such as weather comments especially about the cities of the Kingdom of Saudi Arabia. The training can be done for single user comments or it can be done in a bulk volume such as all the user comments of a certain YouTube video that shows weather information. Any decision that was misclassified or undefined gets corrected by the user (i.e. trained into the system).

2) Execution Phase: When the system receives new content to be classified, it tries to analyze the data before classification. Using the Naïve Bayes theorem, we try to get the probability of a certain classification given the input data. Naïve Bayes theorem is a fast and reliable technique to classify certain data with great certainty based on the given data and against the training data set even with the possibility of noises in the trained data.

3) Feedback phase: The use of the Naïve Bayes theorem provides highly accurate results to classify the emotion tag for the input feed. However for different reasons, we might receive incorrect or undefined decision. Thus, we added the capability for the user to train the system at run-time to refine the algorithm's knowledge base and improve its overall efficiency.

C. Emotion Primitives

Emotion Primitives repository stores the emotion primitives, which are the outputs of Emotion Extraction Logic. Each API stores its result to a separate repository. Some APIs use the stored emotion primitives as a training dataset and use their stored emotion data as an input to the emotion extraction logic, which in this system is the Naïve Bayes theorem. This data set gets enriched throughout the lifecycle of the emotion extraction service. The richer this database is, the more accurate the logic would behave.

D. View

The logic of the *Emotion Extraction Logic* unit is transparent to the user through the *View*. Through the view interface, a user can give his/her feedback by either accepting or refuting the outcome of the emotion value. At the end of user interaction through the View or user interface, the user feedback is stored in the *Emotion Primitives repository*. In order to aid in visualizing the emotions, we quantize the emotional states for each category by adopting the Hourglass of Emotions, which is an affective categorization model proposed in [4]. A sample of categorization according to the above model is shown in figure 2.

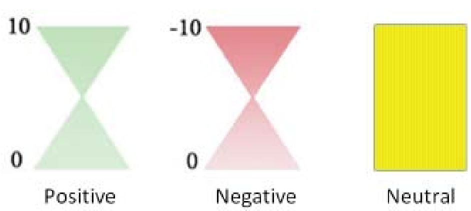


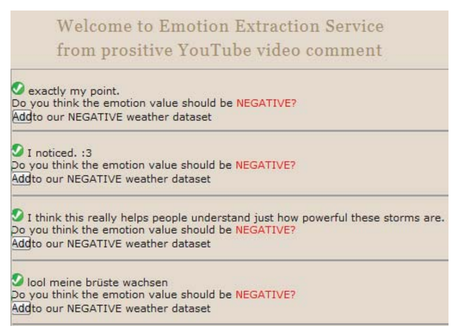
Figure 2. Color coded box according to the emotional values.

The maximum and minimum range of the hourglass depends on the chosen API. Some APIs operate in the range [+10, -10], some in [+5, -5], some in [+1, -1] …etc. The model component adjusts the threshold value accordingly. To extend to new domains, the authoring interface includes a function to capture details about the type of information e.g. news, sports, weather, etc. and sources of the information e.g. YouTube, Twitter, RSS, etc. We have already captured quantitative parameters such as delay, miss ratio of the emotion as well as user experience through many usability tests. Details of the results are presented in Section IV.

III. IMPLEMENTATION

The Web Data Extraction Module is built using open source Apache Web Server v. 2.4.2, and PHP v. 5.4.7. The system includes a user management component for users to register and start utilizing the services over the Web. Each user, within his profile, could customize the needed parameters for the raw social network content retrieval function. He might elect to incorporate some or all of the defined sources such as his Gmail Email credentials, his Twitter info, the keywords that he targets on YouTube and the city or country of interest in certain Weather networks channels.

(a)



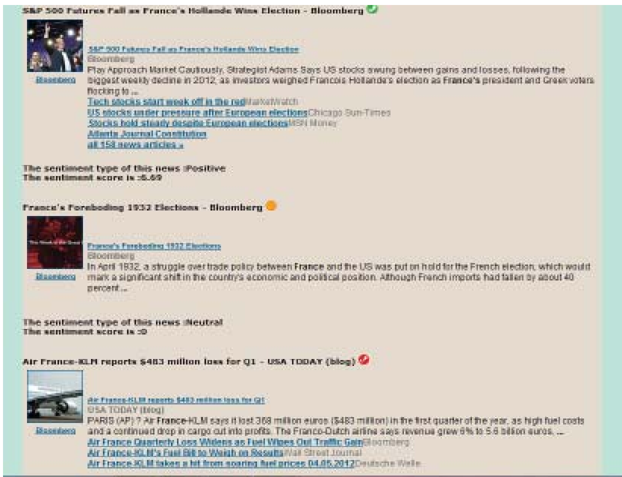
(b)

Figure 3. (a) MediaTagger main window (b) part of MediaTagger authoring interface

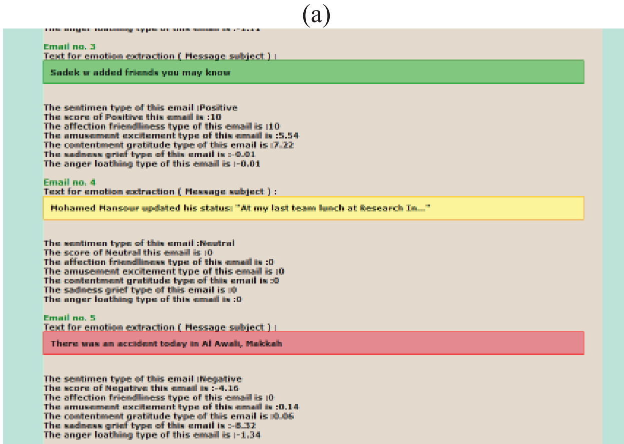
Meanwhile, the Emotion Extraction Module is built using PHP server-side scripting language. The existing APIs that MediaTagger mashes up are Viralheat [1], SentiStrength [2], mashape [3], alchemyApi [4], Open Dover [5], Tweet sentiments [6],

[1] http://www.viralheat.com/
[2] http://sentistrength.wlv.ac.uk/
[3] http://www.mashape.com
[4] http://www.alchemyapi.com/

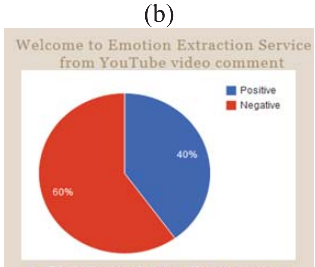
Face.com[7], Imgur[8], and lymbix[9]. Each of these APIs has different strengths in different horizontal and vertical dimensions. The Mashup service utilizes them based on the source of the content. As a proof of concept and to add authoring capability, we've also implemented the Naïve Bayes theorem in PHP language to implement our own emotion extraction API. Users can add any new emotional dimension using our authoring interfaces.



(a)



(b)



(c)

Figure 4. Some of the MediaTagger emotional metaphors: (a) tags shown as tick mark besides News heading (b) tags shown as rectangle box for Google email and Twitter message (c) pie chart to tag YouTube comments

[5] http://developer.opendover.nl/page/Get_started_now
[6] http://twittersentiment.appspot.com
[7] http://developers.face.com/
[8] http://imgur.com/
[9] http://lymbix.com

Figure 3 shows the main web page and the authoring interface that includes the different options of the system. As mentioned, we currently support certain content sources such as Twitter feeds, Gmail Email messages, Weather feeds, Image emotion analysis, YouTube video comments and News RSS feeds. Figure 3 also shows a snapshot where a user can optionally take part in authoring the emotional content by providing his/her feedback regarding those content that are absent from the emotion data set.

Figure 4 shows the emotion tags in different metaphors that leverages the Hourglass concept [4]. As mentioned and as an option, a user can insert his/her preferred emotion tags during the authoring phase as well.

IV. Test Setup and Results

We have tested the individual components of the system shown in Figure 1. The testing has been performed over a 9 months period between March 2012 and December 2012. Tests have been conducted at different network traffic conditions. The test results reveal several points: first, how efficient the Web Data Extractor is to collect the content from different social networks and internet based services. Second, the performance of the emotion extraction subsystem is primarily dependent on the content extraction subsystem. Hence, the clearer and less noisy the contents are, the better the emotion extraction subsystem will perform. Finally, we have measured the time delay in extracting the contents, evaluating emotional values using either API calls or the Naïve Bayes algorithm, adding appropriate emotional tags and rendering them dynamically to respective services as shown below:

Average Delay of Extracting Emotional Tag for a Service = $\sum$ (*Delay in retrieving the media content + The delay in extracting the emotional value from the content using the sentiment analysis API or Naïve Bayes algorithm + The delay in rendering emotional tag to the content and display results*) / *Total Number of Test Instances* --------(2)

TABLE I. Average Delay in Rendering Emotional Tags in Different Services

Service	Total Delay (sec)	Avg. no. of messages or images per test instance	Types of Emotion Extraction Algorithm
Twitter Message	16.55	20 messages	External API
News Feed	15.17	20 news slices	External API
Gmail	74.55	20 emails	External API
Movie Review	0.72	20 comments	Naïve Bayes Algorithm
Facial Emotion	5.29	1 image	Face.com & Imgur API
Weather Comments	0.017 3	20 comments	Naïve Bayes Algorithm
YouTube Video Comments	15.71	20 comments	External API

Table I shows the average emotion extraction delay for about 15000 test instances. After we have analyzed the test result, we found that the Naïve Bayes algorithm always outperforms the external APIs. This can be observed in Table I where movie review dataset comprises of 150 records whereas the weather comment dataset had about 1000 records. As a result, weather comments produced correct results in less period of time. As shown in Table I, the delay time has different value from one service to another. There are many reasons for that delay. For example, the maximum value of delay is in the Gmail service because there is an extra delay in its authentication steps. There are some other external causes that are not within the control of the proposed system such as the network traffic, thus the weak networks can make the delay time worse. We are still conducting more testing to measure more emotional metrics. For example, we are working on evaluating quantitative metrics that are related to comparing our implemented Naïve Bayes algorithm with that of external APIs.

In order to test whether the emotional tags augmented with the contents reflect the accurate emotional value to the users, we have conducted several qualitative usability tests using a group of volunteers of different ages and professions through filling up a prepared questionnaire. The evaluation considered qualitative aspects such as system accessibility, interface layout, technical functionalities and user satisfaction. These measures were obtained by the combination of logging data as a result of user interaction with the system and the subsequent completion of online questionnaires. The evaluation was conducted between April 15, 2012 and September 9, 2012. The number of test subjects was 128. The majority of them belong to an age group older than 20 with about 10 per cent under 20 years. All test users used various Internet-based services in different scales. About 20 per cent of the test users were not computer savvy and they used the Internet sparingly only at their home desktops or laptops.

TABLE II. USER'S FEEDBACK ABOUT DIFFERENT EVALUATION METRICS OF MEDIATAGGER

SA – Strongly Agree, A – Agree, NS – Not Sure, D – Disagree, SD – Strongly Disagree

Variant	SA	A	NS	D	SD
Overall user feedback regarding adding emotional tag to the social networks	89	6	4	1	0
Users view about aggregation of diversified media tagger in one framework	78	19	2	1	0
Users' feedback regarding delay in rendering emotion tags	52	33	14	1	0
Users feedback about emotion tagging metaphors	71	22	7	0	0
Quantifying emotional tag and attaching with a message based on its emotional value	57	34	9	0	0
Users feedback about emotion authoring service	74	15	11	0	0

Table II shows each variant that was evaluated by the users and the corresponding results in percentage values. As shown in Table II, we see that the testing users were satisfied about certain aspects of our systems such as the metaphors and results of the system. They mostly liked the idea of combining different emotion tagging functions of different services within one Mashup as depicted in the first row. The least satisfactory feature of the system was the timing delay to receive the results which is mostly based on certain elements that were out of our hands such as networking and authentication delays. We are working on several solutions to improve the shortcomings pointed by the testing users. One approach would be to incorporate batch pre-processing and analysis tasks within the content and emotion extraction modules.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented our system called MediaTagger, which is a Web-based application that is developed to automatically handle online social media contents and add Emotion Tags for them. MediaTagger mashes up several emotion extraction services. The system also makes use of a reliable supervised learning technique which is the Naïve Bayes theorem to offer authoring options. Currently the system supports certain content sources such as YouTube reviews, Twitter posts, Weather feeds, Gmail Email messages, image mood and RSS News feeds.

We are planning to incorporate new media analysis tools that could process other media types such as generic image and video contents. On a different front, we will incorporate a timeline interface that shows the emotions and sentiments of certain topic or keywords along the time dimension. This could be very useful for example to marketing and branding companies that need to see the effect of certain announcements and news on the brand of certain products or services and possibly forecasting the best times to launch new campaigns and releases.

ACKNOWLEDGMENT

This project was supported by the NSTIP strategic technologies program (11-INF1703-10) in the Kingdom of Saudi Arabia. The author would like to thank Dr. Mohamed Ahmed, Dua, Omayma, Shatha, Jayda, and Ahdab of Umm Al-Qura University for helping in collecting data and testing diversified services.

REFERENCES

[1] Alam, K. M., Abdur Rahman, M., El Saddik, A. and Gueaieb, W., "Adding Emotional Tag to Augment Context-Awareness in Social Network Services," In Proceedings of the IEEE Instrumentation and Measurement Technology Conference I2MTC'2011, (Hangzhou, China), May 10-12 2011.

[2] Abdur Rahman, M., Hamdan, S., El Saddik, A. and Gueaieb, W., "Context-Aware Social Networks Mashup: A Personalized Web Perspective," In Proceedings of the IEEE Instrumentation and Measurement Technology Conference, (Austin, Texas), I2MTC'2010, 1228 – 1233, 2010.

[3] Plutchik, R., "The nature of emotions," American Scientist, 89(4), 344–350, 2001.

[4] Cambria, E., Speer, R., Havasi, C., and Hussain, A., "SenticNet: A Publicly Available Semantic Resource for Opinion Mining," AAAI Fall Symposium Series on Common Sense Knowledge (FSS10), 14-18, 2010.

[5] Ames, M., and Naaman, M., "Why we tag: motivations for annotation in mobile and online media," In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07) (New York, NY), 971-980, 2007.

[6] Garcia, D., and Schweitzer, F., “Emotions in Product Reviews-Empirics and Models,” In proceedings of IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom), (Boston, MA, USA), 483-488, Oct 9-11 2011.

[7] Baldoni, M., Baroglio, C., Patti, V., and Rena, P., “From Tags to Emotions: Ontology-driven Sentiment Analysis in the Social Semantic Web,” In Proceedings of the 5th International Workshop on New Challenges in Distributed Information Filtering and Retrieval (CEUR), (Palermo, Italy), Sept 2011.

[8] Domenico, C., “Emotions That Influence Purchase Decisions And Their Electronic Processing,” In Proceedings of international conference "Challenges of contemporary knowledge-based economy (ICMEA), 2(11), 996-1008, 2009.

[9] Prasad, S., “Micro-blogging Sentiment Analysis Using Bayesian Classification Methods,” Technical Report, Stanford University, 2010.

[10] Li, C., Wang, C., Tseng, C. and Lin, S., “MemeTube: a sentiment-based audiovisual system for analyzing and displaying microblog messages,” In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, (Portland, Oregon), 32-37, June 19-24 2011.

[11] Asawa, K., Agarwal, N. and Agarwal, R., “Therapeutic video tagging model,” In Proceedings of 2012 World Congress on Information and Communication Technologies (WICT), (Noida, India), 485-490, Oct. 30-Nov. 2 2012.

[12] Zhang, C., Xue, G., Yu, Y. and Zha, H., “Web-scale classification with naive bayes,” Proceedings of the 18th international conference on World Wide Web, (Madrid, Spain), 1083-1084, 2009.

Improving the Search Experience in a Social Network with Cross Media Contents

Daniele Cenni, Paolo Nesi

DISIT Lab, Department of Information Engineering
University of Florence, Florence, Italy
{cenni, nesi}@dsi.unifi.it, http://www.disit.dsi.unifi.it

Abstract

With the constant growth of Social Networks, researchers are facing new challenges and issues difficult to deal with. Multilingual and heterogeneous sets of metadata, taxonomical associations, document parsing and translation, are among the commonest ingredients, requiring solutions for an effective and reliable searching experience. Though relevant test collections and evaluations series are useful for ad hoc assessment of IR (Information Retrieval) systems, real IR services in production environments need further tools in order to improve the retrieval effectivenes, based on the digital archives at disposal. This paper describes an optimization strategy and infrastructure, developed to improve the searching experience in a digital archive of Cross Media Contents. The tests were implemented using stochastic approaches, including simulated annealing, genetic algorithms and other related techniques. The above strategy resulted in a considerable improvement of the effectiveness of the IR system, with comparable results among the different stochastic schemas. The research was conducted in the context of the EU funded project ECLAP (`http://www.eclap.eu`*).*

Keywords: *cross media content; indexing; searching; search engines; information retrieval; cultural heritage; stochastic optimization, test collections; social networks*

1. Introduction

In the context of Information Retrieval applied to Cross Media contents, a key factor is represented by dealing with sets of heterogeneous metadata, with the goal of finding things that satisfy an information need. In such a scenario, IR has to cope with unstructured or structured documents stored in huge digital archives, featuring multilingual metadata and various formats, while providing results with a satisfactory level of relevance. Typical IR approaches include the Boolean Retrieval Model and the Vector Space Model. Other models include the Binary Independence Model [14], the Probabilistic Relevance Model [12], and the Uncertain Inference Model [15]. In a Boolean Retrieval Model a query clause is a combination of boolean clauses, and document retrieval is performed by *grepping* the information with a linear scan through documents. In a Vector Space Model [11], each document is represented as a vector of index terms (e.g., words, keywords, phrases), used for computing the so called cosine similarity. In order to improve the performances of an IR system, it is a common task to deal with its *effectiveness*, meaning the overall quality of its produced results. This process includes computing standard IR measures, and plotting the so called *Precision-Recall curve*. Relevant test collections and evaluations series, for ad hoc assessment of IR systems, include TREC [1], GOV2 [2] and CLEF [3]. Nevertheless a major challenge consists in finding optimal estimates for every parameter involved in a real IR system.

For this purpose, the effectiveness of the IR infrastructure of the ECLAP Best Practice Network was evaluated by defining a set of 50 topics (i.e., a common choice for TREC runs, see for example [13]), in the field of Performing Arts. The relevant topics were collected using a list of popular queries, obtained from a query log analysis. The number of topics is chosen above a threshold, in order to get reliable results [4]. For each topic a query was defined, and then a set of relevance judgments, collected with a pooling strategy (i.e., items considered to be relevant are retrieved by choosing a limited subset of the whole set). This method is reliable with a pool depth of 100, although limiting the pool depth to 20 [8] or 10 may change precision results, while not affecting the relative performances of an IR system. A detailed study of IR performances is feasible, even with a relatively short list of relevance judgments [7].

Table 1: ECLAP Indexing Model

Media Types	DC (ML)	Technical	Performing Arts	Full Text	Tax, Group (ML)	Comments, Tags (ML)	Votes
# of Index Fields*	468	10	23	13	26	13	1
Cross Media: html, MPEG-21, animations, etc.	Y_n	Y	Y	Y	Y_n	Y_m	Y_n
Info text: blog, web pages, events, forum, comments	T	N	N	N	N	Y_m	N
Document: pdf, doc, ePub	Y_n	Y	Y	Y	Y_n	Y_m	Y
Audio, video, image	Y_n	Y	Y	N	Y_n	Y_m	Y_n
Aggregations: play lists, collections, courses, etc.	Y_n	Y	Y	Y/N	Y_n	Y_m	Y_n

* = (# of Fields per Metadata type) ∗ (# of Languages)
ML: Multilingual; DC: Dublin Core; Tax: Taxonomy

2. Metadata Model and Indexing Schema

In the context of the ECLAP Best Practice Network, a metadata schema was designed to build the IR infrastructure of a Social Web Portal, consisting of four set of metadata (see [5, 6] for further details): *Dublin Core*, *DCTerms* (a Dublin Core extension), *Technical* (used for administrative purposes) and *Performing Arts* related. When uploading contents to the above portal, each resource's metadata is extracted and indexed by an indexing service (e.g., content type, language, title), following the indexing model of Table 1. The indexing and searching platforms support a wide range of content types such as MPEG-21 (ISO/IEC TR 21000-1:2001), web pages, forums, comments, blog posts, images, rich text documents (e.g., doc, pdf, xls), collections, and play lists. The indexing service is deployed in a distributed parallel architecture, providing massive ingestion and indexing capabilities. The proposed solution includes monitoring and logging features, for further data analysis and validation (e.g., IR effectiveness and user behavior assessment).

Notation used in Table 1, Y_n: yes with n possible languages (i.e., n metadata sets); Y: only one metadata set; Y/N: metadata set not complete; T: only title of the metadata set, Y_m: m different comments can be provided, each of them in specific language. Comments may be annidated, producing a hierarchically organized discussion forum.

3. Weighted Query Model

The indexing model consists of 554 metadata fields, belonging to 8 categories. The most semantic relevant metadata (i.e., text, title, body, description, contributor, subject, taxonomy, and Performing Arts related metadata) are cloned into 8 multilanguage fields, for searching purposes. The search platform applies field boosting at query time (i.e., documents matching a query term have their score multiplied by a weight factor). A boolean weighted clause b is defined as

$$b := (title\colon q)^{w_1} \vee (body\colon q)^{w_2} \vee (description\colon q)^{w_3} \vee (subject\colon q)^{w_4} \vee (taxonomy\colon q)^{w_5} \vee (contributor\colon q)^{w_6} \vee (text\colon q)^{w_7} \quad (1)$$

where i) $(w_1, w_2, \ldots, w_7)$ are the boosting weights related to each query field; ii) *title* is the Dublin Core name of the resource; iii) *body* is the parsed content of a html resource; iv) *description* is a Dublin Core account of a resource content (e.g., abstract, table of contents, reference); v) *subject* is a Dublin Core content topic (e.g., keywords, key phrases, classification codes); vi) *taxonomy* is a content related hierarchy term; vii) *contributor* is a contribution to the content (e.g., persons, organizations, services); viii) *text* is a full text content, parsed from the resource (e.g. doc, pdf); ix) q is the search query.

4. Monte Carlo Analysis

The indexing model introduces a number of constraints difficult to deal with. Full text searches are performed through the most semantically relevant fields (i.e., *title*, *body*, *subject*, *description*, *text*, *taxonomy* and *contributor*). The optimization test must take into account the required computation time, with respect to the performance requirements. Every solution has to be tested by averaging over the recall levels, in order to compute the mean average precision of the IR system. At the same time, each simulation's run requires a certain time to complete, which is constrained by the latencies of a real IR system. Having seven metadata fields, setting the field's weight range interval $R = \{0, 100\}$, with an increase step $i = 0.01$, and hypothesizing a run simulation time $r = 120$ s, an exhaustive optimization search would require $\sim 3.8 \times 10^{22}$ years to explore all the energy states. The chosen approach was to analyze with a heuristic strategy the surface regions constituting a n^{th}-dimensional hyperplane.

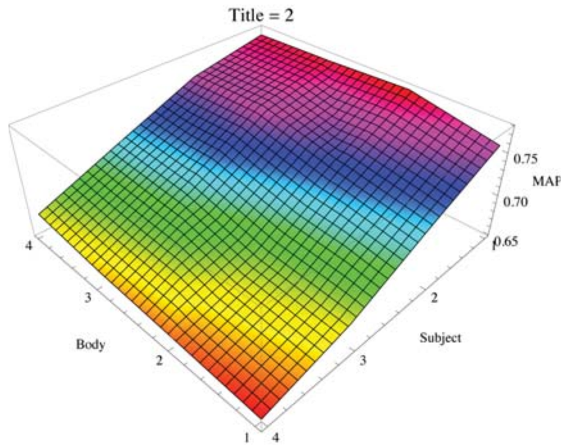


Figure 1: Field Title, weight 2.0

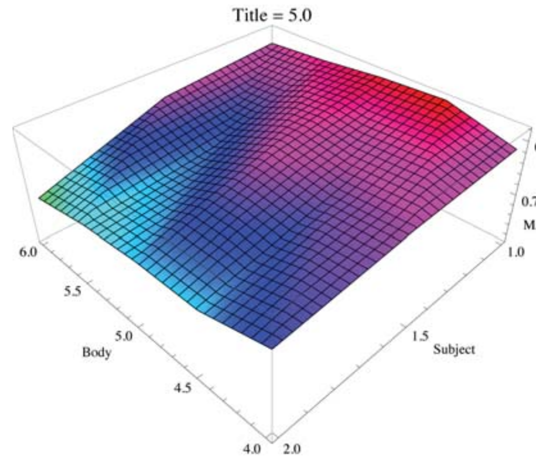


Figure 2: Field Title, weight 5.0

The range intervals for every field's weight in (1) may be analyzed with a Monte Carlo analysis, in order to discriminate trap states and local minima. A search analysis of local minima (or maxima) was performed, exploring the system at sampled values of some field's weights. A preliminary test was conducted using t_w, b_w, $s_w \in \{1, 4\}$, where t_w is the title's weight, b_w is the body's weight and s_w is the subject's weight. Figure 1 represents the weight surface when the title's weight equals 2.0, with respect to the other weights.

Another simulation test was performed using $t_w \in \{4,6\}$, while other field's weights (i.e., subject, description, text) were randomly chosen in the intervals $\{1,2\}$ and $\{4,6\}$. The test was scheduled to run with 10 000 random iterations, and every random weight was rounded with 5-digits precision. Figure 2 depicts the surface plot when the title's weight equals 5.0, with respect to the other weights. The observed patterns, in each surface plot, point out that there is a considerably wide interval with local maxima, for each tested configuration of weights, preventing to make any conjecture about the optimal ranges.

5. Simulated Annealing

The platform to analyze the optimal estimations for each index field's weight in (1), included a stochastic minimization test, that was designed and implemented. Considered the relatively high number of variables, the test implemented a simulated annealing strategy, testing various annealing schedules, initial state conditions and allowed transitions per temperature. Every simulation run took place by defining the state of the system as a vector of field weights $\vec{w_i} = (w_1, w_2, \ldots, w_7)$. A run of 50 queries was issued for each state condition, in order to retrieve the corresponding search results with the most relevant IR measures. For each run, the Mean Average Precision (*MAP*) was computed and (1−*MAP*) was set as the energy for the current state. *MAP* is defined as the arithmetic mean of average precision for the information needs, so that it can be thought as an approximation of the area under the Precision-Recall curve. Considering the *Metropolis Criterion*, a state transition probability p_t is defined by

$$p_t = \begin{cases} 1, & \text{if } E_{i+1} < E_i \\ r < e^{-\Delta E/T}, & \text{otherwise} \end{cases} \qquad (2)$$

where E_{i+1} and E_i are respectively the energy states of w_{i+1} and w_i, T is the *synthetic temperature*, $\Delta E = E_{i+1} - E_i$ is the *cost function*, r is a random number in the interval $\{0,1\}$. The *annealing schedule* was defined as $T(i+1) = \alpha T(i)$, with $\alpha = 0.8$. Two hundred random transitions were tested for the temperature of each iteration. A smoother annealing schedule is more likely to exhibit convergence, but generally requires a bigger simulation time. Stopping conditions were assumed by counting the number of successful transitions occurred during each iteration. Other popular choices include logarithmic schedules such as $T(i) = c/log(1+i)$ [10, 9].

Figure 5 reports the best simulation configuration, exhibiting convergence and system equilibrium. Some metadata fields were found to have a limited relevance weight, with respect to the relevance score (i.e., subject, taxonomy and contributor). Reducing the number of boolean clauses to be processed by the IR system is indeed an advantage that would result in a higher search speed. Scatter plots of field weights vs *MAP*, collected during the test run, showed a relevant dispersion across a huge range of high energy values (see Figures 8, 9). The observed pattern thus suggests a relevant sensitivity to initial conditions and random seeds. The minimization strategy resulted in an energy minimum at $w_1 = 68.4739$, $w_2 = 31.7873$, $w_3 = 0.2459$, $w_4 = 9.8633$,

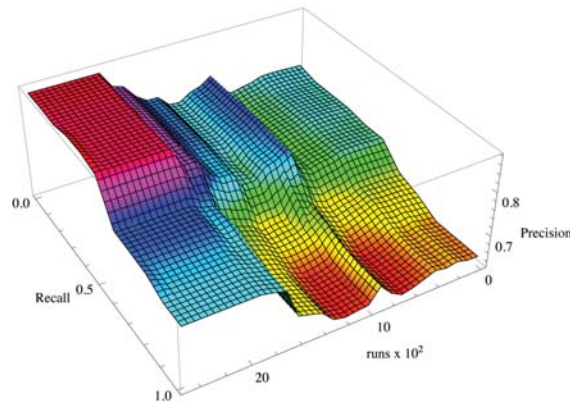


Figure 3: Precision-Recall vs test runs (Annealing)

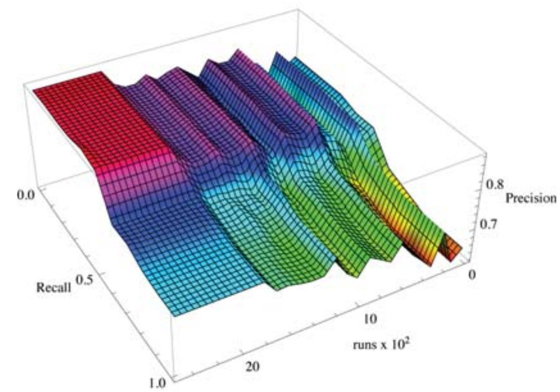


Figure 4: Precision-Recall vs test runs (GA)

$w_5 = 13.2306$, $w_6 = 2.1720$, $w_7 = 3.9720$, with *MAP* $= 0.8223$ (see the Precision-Recall curve in Figure 7, *MAP* scatter plot in Figure 6, and IR measures in Table 2). The behavior of the Precision-Recall curve, during some test runs, is depicted in Figure 3.

Before the optimization tests, the weight values used in the production server ($w_1 = 3.1$, $w_2 = 0.5$, $w_3 = 1.7$, $w_4 = 2.0$, $w_5 = 0.5$, $w_6 = 0.8$, $w_7 = 0.8$), produced *MAP* $= 0.7552$. The optimization strategy yielded an increase in *MAP* of $\sim 8.885\%$.

6. Genetic Algorithms

Another stochastic approach to IR optimization makes use of Genetic Algorithms. Each field weight, that constitutes the boolean query expression in (1), was defined as a gene of the sample chromosome. The test was built with a population of 100 chromosomes (the more the chromosomes, the larger the number of solutions, but with a longer computation time, due to the fact that the population will require more time to evolve for each round). The upper limit of maximum allowed evolutions was set to 10 000. The field's weight values of the fitness function f, were evaluated by computing their corresponding gene values for the current chromosome. For each vector of weights $\vec{w_i} = (w_1, w_2, \ldots, w_7)$ the indexing service was queried in order to find the corresponding Mean Average Precision.

The fitness function f was then normalized so to exaggerate the difference between the higher values, assuming $f_n = 2^{10f}$. Figure 5b shows the convergence of *MAP* across the test runs. Table 2 shows the most relevant IR measures collected for this simulation strategy. Also in this case a considerable dispersion across a huge range of energy values was noticed, for every index field (see Figures 12, 13). The MAP_{GA} value was

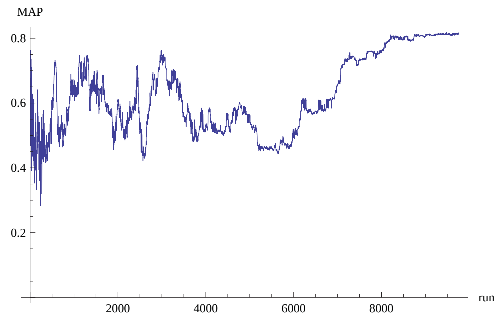
Table 2: IR measures for the optimal run

Measure	Simulated Annealing	Genetic Algorithm
# of queries	50	50
# of documents retrieved for topic	4312	4319
# of relevant documents for topic	85	85
# of relevant documents retrieved for topic	84	84
MAP	0.8223	0.8210
Geometric MAP	0.7216	0.7169
Precision after retrieving R docs	0.7658	0.7657
Main binary preference measure	0.9886	0.9884
Reciprocal Rank of the first relevant retrieved document	0.8728	0.8747

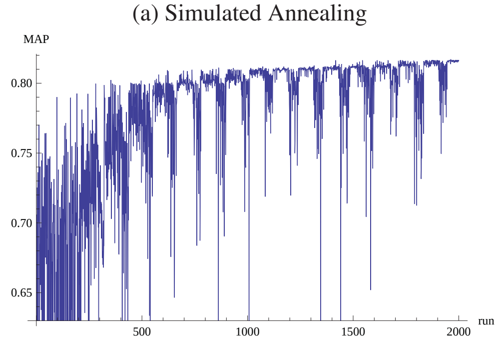
consistent to what obtained with the annealing strategy. The estimated *MAP* difference (Δ*MAP*) between the two approaches was $\sim 0.15\%$ ($\Delta MAP = MAP_{Annealing} - MAP_{GA} = 0.8223 - 0.8210 = 0.0013$). Figures 10, 11 shows respectively the *MAP* scatter plot, and the Precision-Recall curve, obtained during the best performing simulation run. The progress of the Precision-Recall curve, collected through some test runs, is depicted in Figure 4.

7. Conclusions and Future Work

This paper presented an optimization approach, in the field of Information Retrieval, applied to Cross Media contents. After a preliminary Monte Carlo analysis of the IR system, the search platform of the ECLAP digital archive was evaluated with different stochastic strategies (i.e., simulated annealing and genetic algorithms), in order to find the best tuning parameters for the IR model. For every test run, the most relevant IR measures were computed. The above analysis resulted in a marked improvement of the IR effectiveness of the system and the



(a) Simulated Annealing



(b) Genetic Algorithm

Figure 5: MAP vs test runs

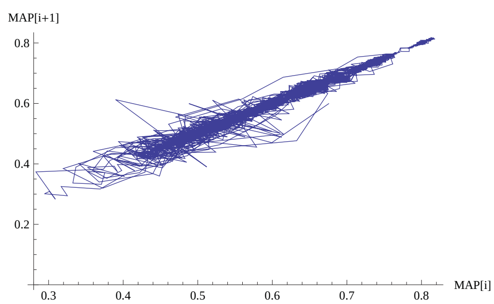


Figure 6: MAP scattering (Annealing)

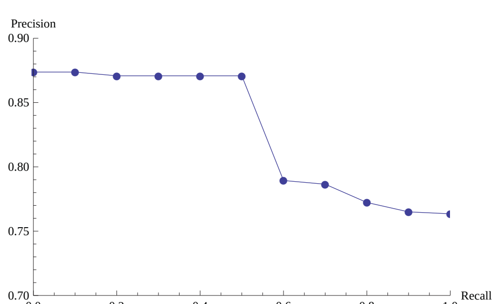


Figure 7: Precision-Recall (Annealing)

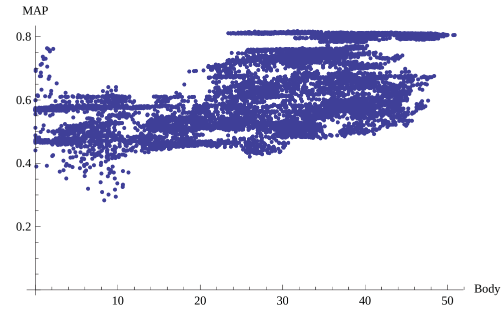


Figure 8: Body vs MAP (Annealing)

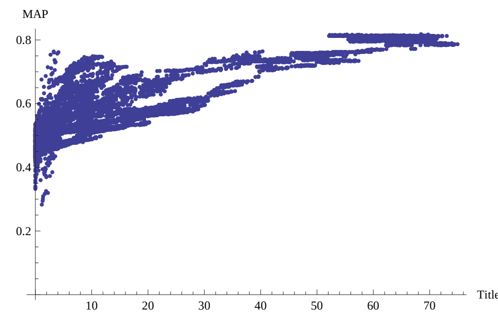


Figure 9: Title vs MAP (Annealing)

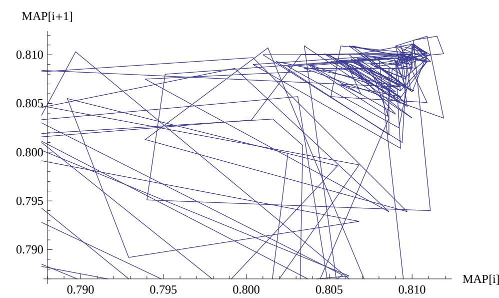


Figure 10: MAP scattering (GA)

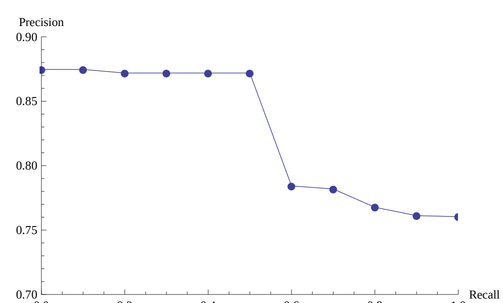


Figure 11: Precision-Recall (GA)

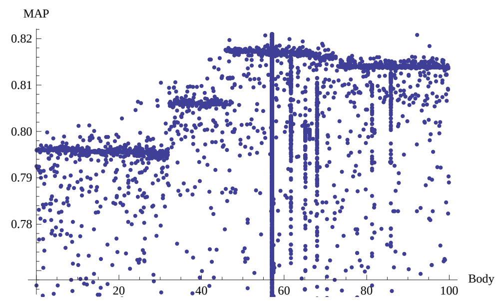


Figure 12: Body vs MAP (GA)

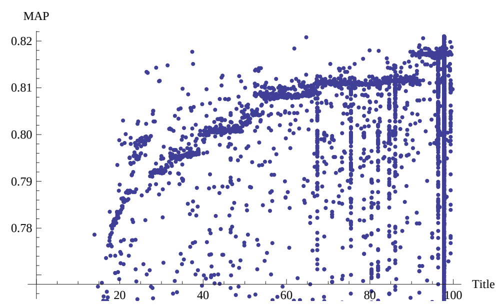


Figure 13: Title vs MAP (GA)

search experience. Further improvements include the application of metadata related techniques, for weighting and calibrating the contributes provided by different retrieval approaches, in the context of Content Based Image Retrieval (CBIR).

8. Acknowledgments

The authors want to thank all the partners involved in ECLAP, and the European Commission for funding the project. ECLAP has been funded in the Theme CIP-ICT-PSP.2009.2.2, Grant Agreement No. 250481.

References

[1] `http://trec.nist.gov`.

[2] `http://www-nlpir.nist.gov/projects/terabyte`.

[3] `http://www.clef-initiative.eu`.

[4] Timothy G. Armstrong, Alistair Moffat, William Webber, and Justin Zobel. Improvements that don't add up: ad-hoc retrieval results since 1998. In *Proceedings of the 18th ACM conference on Information and knowledge management*, CIKM '09, pages 601–610, New York, NY, USA, 2009. ACM.

[5] Pierfrancesco Bellini, Daniele Cenni, and Paolo Nesi. On the effectiveness and optimization of information retrieval for cross media content. In *KDIR, 4^{th} International Conference on Knowledge Discovery and Information Retrieval*, pages 344–347. SciTePress, 2012.

[6] Pierfrancesco Bellini and Paolo Nesi. A linked open data service for performing arts. In *ECLAP, 2^{nd} International Conference on Information Technologies for Performing Arts, Media Access and Entertainment*. Springer Verlag LNCS, 2013.

[7] Ben Carterette, James Allan, and Ramesh Sitaraman. Minimal test collections for retrieval evaluation. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '06, pages 268–275, New York, NY, USA, 2006. ACM.

[8] Nick Craswell, Dennis Fetterly, and Marc Najork. The power of peers. In *Proceedings of the 33rd European conference on Advances in information retrieval*, pages 497–502. Springer-Verlag, 2011.

[9] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-6(6):721 –741, nov. 1984.

[10] Bruce Hajek. Cooling schedules for optimal annealing. *Math. Oper. Res.*, 13(2):311–329, May 1988.

[11] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.

[12] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3):129–146, 1976.

[13] Stephen Robertson. On the contributions of topics to system evaluation. In *Advances in Information Retrieval*, volume 6611 of *Lecture Notes in Computer Science*, pages 129–140. Springer Berlin / Heidelberg, 2011.

[14] Thomas Roelleke and Jun Wang. Probabilistic logical modelling of the binary independence retrieval model. In *Proceedings of the 1st International Conference on Theory of Information Retrieval (ICTIR 07) - Studies in Theory of Information Retrieval*, 2007.

[15] C. J. Van Rijsbergen. A non-classical logic for information retrieval. *The Computer Journal*, 29(6):481–485, 1986.

Towards a Semantic Multimedia Content Retrieval Framework

Mohamed A. Ahmed

Advanced Media Laboratory, Umm Al-Qura University
Makkah, Saudi Arabia
mamahmed@uqu.edu.sa

ABSTRACT

In this paper, we address the challenges of automatically extracting multimedia objects from diversified online repositories in order to mine semantic media information. We propose algorithms to mine heterogeneous multimedia objects from diversified domains and extract multimedia metadata and semantic information to dynamically recommend context-aware multimedia objects. We have developed this framework as part of an ongoing e-Learning project where we need to search and index multimedia objects from diversified online sources that are suitable for children with Hemiplegia disability. We share the implementation details in building a real life proof of concept open source software environment and usability results.

1. INTRODUCTION

Recent advancements in online digital media sharing have made the Internet a place of giant online multimedia repository. People on a daily basis upload or share diversified types of multimedia objects through their personal web pages, blogs, corporate web sites, and online repositories. Searching multimedia such as plain text, audio, video, animation, documents and other media types over these vast and diversified sources poses a great challenge. One reason is that there are numerous available multimedia data types; most of them are proprietary. With the advent of Web 2.0, a lot of people are regularly adding new media contents on their websites or online storage on a daily or weekly basis. Once a robust web service is assumed that can search media over the web, the next challenge is to mine each media and extract metadata to gain semantic as well as technical attributes of the media. For example, if we are interested about all the videos regarding the MonaLiza over most popular video sharing networks such as YouTube, Vimeo, Veoh, and DailyMotion, to name a few, one needs to search the video metadata available from the metadata of the video, download the video itself and then analyze the video frames to find key attributes of the MonaLiza.

Metadata has been used to store information about various multimedia objects such as books in library, inventory in businesses, media in entertainment industry, digital objects in classrooms, etc. Metadata can be classified into different types, depending on the type of digital object. Examples of such types are semantic metadata, emotional metadata, bibliographic metadata, technical metadata, classification metadata, evaluative metadata, educational metadata, semantic metadata, usage metadata, sequencing and relational metadata, interaction metadata, and rights metadata, to name a few. Each owner of file type provides different types of metadata, which does not follow any one particular standard. For example, although there are several metadata standards such as IEEE LOM[1] and Dublin Core[2], people neglect inserting metadata with digital objects due to lack to awareness. As a result, extracting metadata from digital media remains a challenging task.

In this paper, we propose the following contributions: first of all, we combine a set of algorithms to extract metadata from diversified types of digital media from large web domains. Secondly, we have developed algorithms to automatically combine different types of metadata of particular file type as described above by employing more than one web services. Thirdly, we have developed algorithms to mine semantic metadata from online media. As an example, we use emotional metadata in the form of positive and negative classifications to add new dimensions of semantics that would enrich many applications.

The rest of the paper is organized as following: in Section 2 we have briefly presented some of the closely related works. In Section 3, we will present the framework architecture and the description of salient components of the framework. In Section 4, we have discussed the implementation details followed by subjective test results in Section 5. Section 6 concludes the paper with a future vision.

[1] http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
[2] http://dublincore.org/

2. RELATED WORK

Authors in [1] describe a useful metadata extraction system which performs a powerful combination of speech and speaker recognition tasks. In order to make indexing and retrieval more flexible and efficient, the extracted metadata is stored using the MPEG-7 multimedia content description interface. The system has been successfully tested on the recordings of the plenary sessions of the Catalan Parliament. An e-Learning Multimedia Objects system is developed by authors in [2] where they have developed a digital library architecture for enabling the reuse of audiovisual documents in an e-Learning context. Authors in [3] have developed Biblio, which is an adaptive system that automatically extracts metadata from semi-structured and structured scanned documents. Instead of using hand-coded templates or other methods manually customized for each given document format, it uses example-based machine learning to adapt to customer-defined document and metadata types. Automatic Metadata Extraction for Archival Description and Access has been suggested by William Underwood in [4]. The author has developed techniques for automatically extracting metadata from electronic records that is necessary for automatically describing items, file units and records series and for supporting access to these records. Automated Metadata Indexing and Analysis (AMIA) project [5] aims to provide an effective digital asset management (DAM) tool for large digital asset databases. In addition, multimedia metadata has been used for semantic analysis. Authors in [6] have used metadata to extract semantic information about the objects around a person and then search those entities in a semantic web environment. In the context of emotional semantic metadata, authors in [7] have employed three types of metadata to add emotional label extracted from facial expressions in video frames. An image ranking application has been proposed in [8], which is based on the analysis of textual description of the image, technical metadata available from the image metadata analysis and the semantic information from the visual features. In [10], the authors make use of user comments and reviews in tagging audiovisual contents especially for Youtube videos while using time-series techniques.

3. FRAMEWORK DESIGN

Our automatic metadata extraction system is equipped with the following high level components shown in Figure 1.

Media Source: This is the external source of the digital media, which is outside the proposed framework. This can be a website where digital media is embedded, it can be a complete domain or it can be a certain media repository or URL (such as YouTube or Flickr) of media content.

Pre-processor: The Pre-processor performs several salient tasks. First of all, it receives the source of media from the user and indexes it in local full text search engine. If the source is a single web page, it stores the URL of the page and also performs full text indexing. In case the URL is a domain address, it starts crawling the domain, provided that the crawling option is enabled, and indexes the domain web site URLs as well as the included text. In the case of domain website as source, this process is done offline. Acually, the framework supports two modes: instantaneous and offline. In the case of instantaneous mode, the framework only supports a single web page while in the case of bulk volume of web sites, the framework supports offline preprocessing. The framework also supports non-indexed and indexed mode. In the former case, the framework simply starts the metadata preprocessing steps, without going through indexing phases. While in the latter option, the framework first indexes the URL, mines all the digital objects such as audio, video, images, and texts within the preprocessor database.

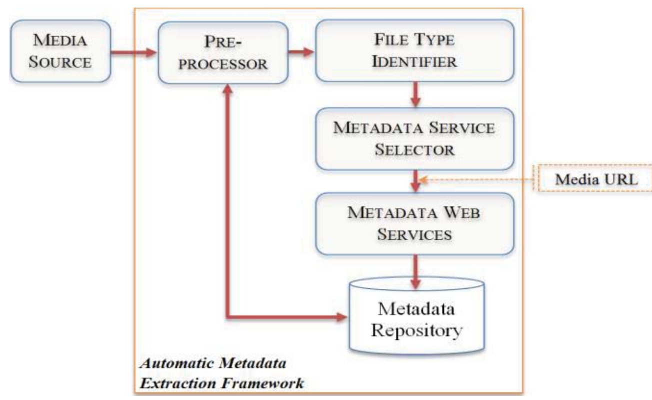


Fig. 1. Automatic metadata extraction system

File Type Identifier: At the end of preprocessing stage, a user can view the media elements or files available or embedded in a web page or repository and choose complete or a subset of those media elements (e.g. see Figure 3.a) for metadata extraction. *File Type Identifier* runs different algorithms to automatically identify the type of media. This is necessary to precisely identify the file type so that appropriate metadata extraction service can be called. For example, when a .pdf file is identified, the metadata extraction algorithm for .pdf file will be instantiated. For example, the algorithm to identify a media object from its URL is different from identifying a YouTube video or a Flickr image. One pseudo code example for finding and extracting YouTube video content is shown in figure 2.

Metadata Service Selector: This component receives an input URL of a particular media element and helps in calling the appropriate metadata extraction service depending on the type of input file. It also allocates the necessary resources for performing the extraction and analysis later on. The actual metadata extraction algorithms might run in one server or it can be a distributed service provided by a third party. Sometimes, if more than one service is available for a particular file type, a wrapper is used to keep track of a number of services for that particular file type. For example, there are more than twenty different types of image metadata formats available, for which numerous third party metadata extraction tools are available. This component will intelligently call the appropriate wrapper for each such option.

```
function extractYouTubeVideo  ($url) {  →  passing the link of the media
      $html = file_get_contents($url);  →  reading entire file into a string

preg_match_all('/(http\:\/\/www\.youtube\.com\/v\/\w{11})|(http\:\/\/www\.youtube\.com\
/watch\?v=\w{11})/', $html, $matches);
      $yVidArray = array_values(array_unique($matches[0]));
      return $yVidArray;  →  returning the link if there was a match

   }
```

Fig.2. Example: finding and extracting Youtube content.

Metadata Web Services: Each *Metadata Web Service* is responsible for the actual extraction of metadata for a particular file type. For example, for .pdf file one service, for .docx file there is different service, for .avi file there is a third service, and so on. As discussed above, if more than one service is available for a particular file type, a wrapper will register all of them. Because each tool might vary in metadata extraction capability, sometimes it might be needed to employ more than one tool to get a rich set of metadata by integrating and filtering the different results.

Each web service employs a local repository where the automated metadata extraction tool will download the digital media temporarily or permanently prior to the actual metadata extraction, depending on the digital media rights. The idea is that the media which is subject to the extraction process needs to be downloaded from a remote URL to the local *Metadata Web Services* component where the automated metadata extraction process will take place. However, if a particular file metadata is already available in the *Metadata Repository*, it does not need to extract the metadata again. Each metadata service provider is designed to output the extracted metadata in varieties of formats such as JSON and XML. A sample output of such process is shown in Figure 3.b. Another important feature of the *Metadata Web Services* is that a particular wrapper can call a set of available services for a particular file type to Mashup the results and make a rich set of metadata types. In addition to textual metadata, it can employ computer vision algorithms or other semantic media mining algorithms to extract semantic metadata as well. For example, Figure 3.c shows the metadata about an image available from the publisher or user comments available from a social network, while Figure 3.d shows the output of a face recognition algorithm that provides us semantic information by analyzing one's face. Similarly, we have developed several services to parse content from diversified types of files such as MS Office/Open Office/Pdf documents. We have configured *Metadata Web Services* to provide the following types of metadata: emotional metadata, bibliographic metadata, technical metadata, classification metadata, evaluative metadata, educational metadata, semantic metadata, usage metadata, sequencing and relational metadata, interaction metadata, and rights metadata [9].

Metadata Repository: After the extraction of metadata, two scenarios might happen. First, a link of the extracted metadata in XML/JSON format is sent back to requesting application, which is used for visualization purposes or for indexing. Secondly, the metadata is stored in metadata repository and a copy of metadata schema is attached with the actual digital media. This is then made available to the applications requesting different types of metadata to decide which media is suitable for what context; for example, in our current context of automatically searching images from Flickr[3] and Imgur[4], videos from YouTube[5], and audio/video from Archive[6] that is suitable for children having hemiplegia disability.

4. IMPLEMENTATION

The framework has been implemented using several open source packages. As a web server, we use XAMPP server that would include Apache 2.4.2, MySQL 5.5.27, and PHP 5.4.7. In order to index the websites per domain, we use open source called sphider[7]. As video repository and metadata repository, we use Fedora[8]. Video mining algorithms have been implemented by designing a custom web crawler using regular expressions and XPATH. The framework mashes up several open source metadata extraction tools in addition to our own algorithms developed in PHP language. One of the most powerful open source tools that we take benefit from is ExifTool[9]. Special algorithms using Python have been developed to download videos and other huge-size files from the Internet. Figure 3.a shows the interface where media mined from a web page is shown. The interface allows a user to choose all or a subset of media elements for metadata extraction. Once a user chooses the media he/she wants to extract, the framework outputs the extraction results as shown in Figure 3.b.

---

[3] http://www.flickr.com/
[4] http://imgur.com/
[5] http://youtube.com/
[6] http://archive.org/
[7] http://sphider.eu/
[8] http://fedora-commons.org/
[9] http://www.sno.phy.queensu.ca/~phil/exiftool/

Fig. 3. Snapshots of different user interfaces of the system. (a) interface showing a summary of media elements available per web page, (b) summary after the metadata extraction process, (c) interface showing metadata about a particular media element, (d) semantic information mining from a YouTube video available from service provider, (e) result of semantic metadata extraction service from an image, (f) services to extract emotional metadata, and (g) emotional metadata extracted from analysis of a user comments about a YouTube video.

Meanwhile as shown in Figure 3.b, the resources pane shows the media and the Metadata pane provides a hyper link that presents the metadata; an example of which is shown in Figure 3.c. It also shows a response summary if the metadata extraction process fails. Figure 3.d and 3.e show a sample scenario where we employ semantic metadata extraction process from video and image respectively. As an extension for supporting newer type of metadata, we have investigated emotional metadata which reflects user's comments about the quality or certain aspects of a media in terms of positive and negative emotion classification. These automatic classifications are based on Naïve Bayes machine learning algorithms. Figure 3.f shows the developed semantic emotional metadata extraction services targeting different domains such as text and image media. A sample emotional metadata result is shown Figure 3.g. We are also currently investigating the above metadata extraction framework to dynamically extract media to be incorporated in a Second Life-based e-Learning framework suitable for children having hemiplegia disability.

5. TEST RESULTS

We have done some qualitative tests to analyze the viability of our system. We have assembled a test group sample containing 150 persons of different ages and technical background. We have designed a survey questionnaire that includes six questions. Prior to filing out the questionnaire, we have deployed the semantic multimedia retrieval framework in a test web server. We then asked the test subjects to test the metadata framework using our online website and then answer the questionnaire according to their experience with the system. Table 1 shows a summary of the subjects' responses for the given six questions given in the "Variant" column. As depicted in the results, we see that the subjects are mostly satisfied with the framework. They have approved the search and metadata functions and capapbilities. The main problem they faced was the difficulties in understating the concept of different types of metadata and their classification. It seems the website was not intuitive and easy enough for them especially to the non-technically savvy persons. We are planning to resolve that in the next version of the website design. We plan to add a wizard-based component that would make the process more streamlined and easy to follow.

TABLE 1. USER'S FEEDBACK ABOUT DIFFERENT EVALUATION METRICS

Variant	Agree	Not Sure	Disagree
Do you like the interface?	60%	13%	27%
Were the guidelines helpful?	89%	7%	4%
Do you face any difficulties in using the website?	7%	33%	60%
Did you find the media you were looking for?	87%	0%	13%
Did you find the metadata for media you want?	87%	0%	13%
Are you satisfied with our web services?	60%	27%	13%

6. CONCLUSION AND FUTURE WORK

Extracting metadata from diversified digital media is a challenging task and still not very widely utilized. Meanwhile, automated extraction can not only help improve efficiency in time and resource management within the digital media preservation systems, but it also reduces the problems associated with the metadata data entry**.** In this paper, we described our system that retrieves, indexes and searches diversified multimedia contents available within a whole web domain, a certain web page or other types of repositories. Tests show strong user support. In future, we plan to extend the system to be incorporated within our context-aware 3D game environment geared for disabled children.

ACKNOWLEDGEMENT

This project was supported by the NSTIP strategic technologies program (11-INF1703-10) in the Kingdom of Saudi Arabia.

7. REFERENCES

[1] H. Delgado, J. Serrano, and J. Carrabina, "Automatic metadata extraction from spoken content using speech and speaker recognition techniques," *FALA2010*, pp. 201 - 204, Vigo, Spain, Nov. 2010.

[2] P. Bolettieri, F. Falchi, C. Gennaro, and F. Rabitti, "Automatic metadata extraction and indexing for reusing e-learning multimedia objects," *Workshop on multimedia information retrieval on the many faces of multimedia semantics (MS '07), ACM*, New York, NY, USA, pp. 21 - 28, 2007.

[3] C. Staelin, M. Elad, D. Greig, O. Shmueli and M. Vans, "Biblio: automatic meta-data extraction," *Intl. Journal on Doc. Analysis and Recog.*, Springer-Verlag Berlin, Heidelberg, pp. 113 – 126, 2007.

[4] W. Underwood, "Automatic Metadata Extraction for Archival Description and Access," *Society of American Archivists Research Forum*, pp.1 - 10, 2008.

[5] P. Y. Chiang et al., "Technologies and the development of the Automated Metadata Indexing and Analysis (AMIA) system," *Journal of Visual Comm. and Image Repres.*, vol. 21, no. 3, pp. 200 - 209, 2010.

[6] W. V. Woensel, S. Casteleyn, E. Paret, and O. De Troyer, "Mobile Querying of Online Semantic Web Data for Context-Aware Applications," *IEEE Internet Computing, IEEE*, vol. 15, no. 6, pp.32 - 39, Nov. 2011.

[7] M. Tkalcic, A. Odic, A. Kosir, and J. Tasic, "Affective Labeling in a Content-Based Recommender System for Images," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 391 - 400, Feb. 2013.

[8] F. Schroff, A. Criminisi, A. Zisserman, "Harvesting Image Databases from the Web," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 754 - 766, April 2011.

[9] S. Downes, "Resource Profiles," *Journal of Interactive Media in Education*, vol. 2004, pp. 1 - 32, May 2004.

[10] C. Eickhoff, W. Li, and A. de Vries, "Exploiting User Comments for Audio-visual Content Indexing and Retrieval," *34th European Conf. on Information Retrieval*, Moscow, Russia, pp. 38 - 49, March 2013.

A PetCare System designed by Slow Intelligence Principles

Shi-Kuo Chang[1], Bin Gao[1], Liqun Kuang[2], Yingze Wang[1]
[1]Department of Computer Science, University of Pittsburgh, Pittsburgh, PA, USA
[2]Department of Computer Science & Technology, North University of China, Taiyuan, Shanxi, China
{chang, big5, lkuang, yingzewang}@cs.pitt.edu

Abstract

In this paper we describe the design of a PetCare system following slow intelligence principles, so that its performance will improve gradually over time. The PetCare system uses a PC with webcam to monitor pet(s) at home. When the pet is motionless or super-active for a prolonged period of time, the owner will be alerted. The architecture of the PetCare system allows easy expansion to add more functionalities. Experimental results confirm this slow intelligence approach leads to performance improvements. Similar slow intelligence principles can be applied to the design of a Senior Patient Care system that share many similar characteristics of the PetCare system.

Keywords

Component-based software engineering, slow intelligence principles, super component, pet care system

1 Introduction

In the past few years, the lead author of this paper has developed the basic slow intelligence principles. A slow intelligence system (SIS) [2] is a general-purpose system characterized by being able to improve its performance over time through a process involving *enumeration, propagation, adaptation, elimination and concentration.* A SIS continuously learns, searches for new solutions and propagates and shares its experience with other peers. A SIS is a system with multiple decision cycles such that actions of slow decision cycle(s) may override actions of quick decision cycle(s), resulting in poorer performance in the short run but better performance in the long-run.

A SIS is characterized by employing *super components*, in the sense that multiple components can be activated either sequentially or in parallel to search for solutions. In [3] we explained in detail a visual specification approach using dual visual representations, and described the user interface design to produce and manage the dual visual representations for the generic SIS system. In this paper we further apply the slow intelligence principles to the design of a PetCare system, which uses a PC with webcam to monitor pet(s) at home. When the pet is motionless or super-active for a prolonged period of time, the owner will be alerted. The architecture of the PetCare system allows easy expansion to add more functionalities. The design of the system follows slow intelligence principles, so that its performance will improve gradually over time. The experimental results confirm the slow intelligence approach leads to performance improvements.

The paper is organized as follows: In Section 2 we present the architecture and the operational scenarios for the component-based PetCare system. The PetCare system also provides a Web-based user interface described in Section 3. In Section 4 we apply the slow intelligence approach to design the PetCare System and present the experimental results. Some discussions are provided in Section 5.

2 PetCare System Architecture

2.1. System Architecture

Figure 2.1 illustrates the PetCare system architecture comprising the key components of the system. The system administrator, the end user and the SIS creator (the software engineer who creates the slow intelligence components) can access the PetCare system via the Internet/Intranet.. The Certificate management server issues a digital certificate for each VPN (Virtual Private Network) equipment and software with USB key, which is stored either in a device's flash or as a USB internal key. The PetCare Imaging and Database Server stores and manages the images as well as other types of multimedia information. Devices such as the Webcam and the Kinect interface can be connected to the PetCare Imaging and Database Server.

The PetCare SIS Server controls slow intelligence components such as Enumerator, Eliminator, Concentrator and Time Controller. The PetCare Management Server handles other PetCare administrative and management functions. A virtual private network extends a private network of the PetCare system across public networks such as the Internet. It enables a host computer to send and receive data across shared or public networks as if they were an integral part of the private network of the PetCare system, with all the functionality, security and management policies of the PetCare system.

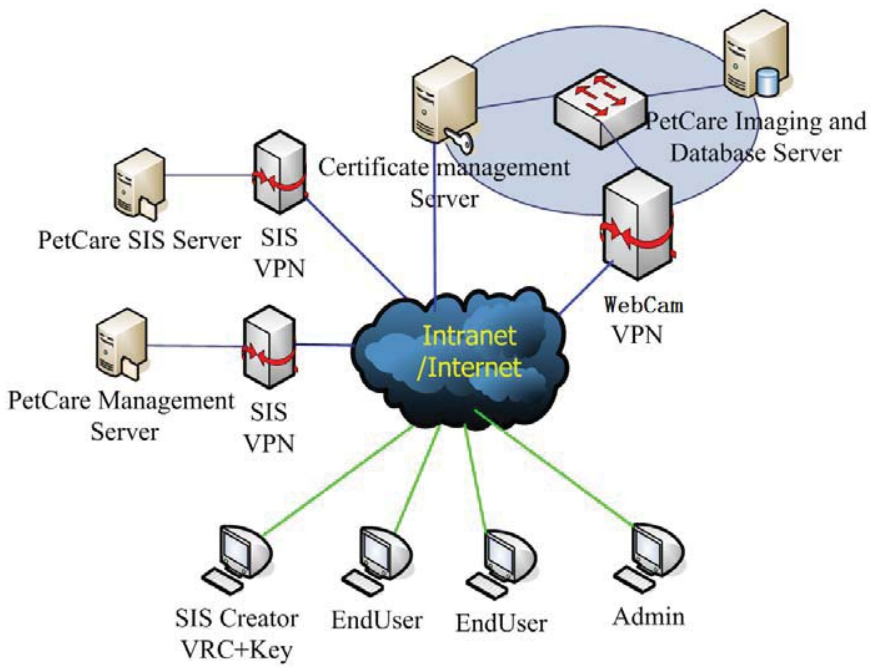


Figure 2.1. System architecture of PetCare System.

2.2. End User's Scenario of PetCare System

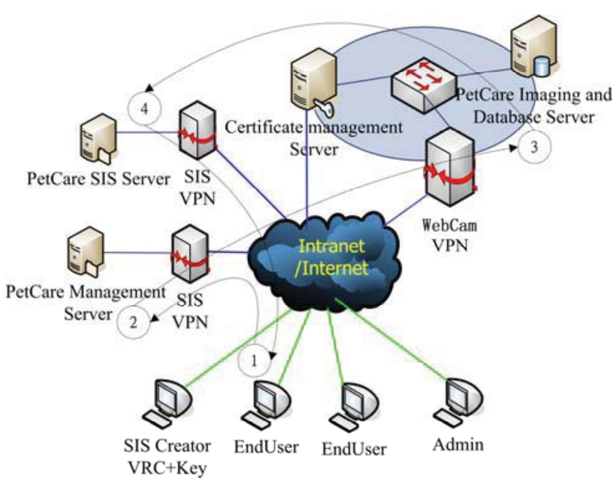


Figure 2.2. End user's scenario of PetCare System.

Step 1: EndUser(1)->PetCare Management Server(2)

Description: Verify user log in credentials, retrieve user configuration parameters, and display the welcome page. At the welcome page, the user can change its password and configuration parameters. The user can also request the images.

Step 2: PetCare Management Server(2)->PetCare Imaging and Database Server(3)

Description: The PetCare Management Server determines the user permissions, then sends the request to the Imaging and Database Server for images.

Step 3: PetCare Imaging and Database Server(3) ->PetCare SIS Server(4)

Description: The imaging and Database Server forwards the image sequence to the SIS Server for further processing.

Step 4: PetCare SIS Server(4)->EndUser(1)

Description: The SIS Server delivers the images to the end user and also notifies the end user related events after processed the data.

2.3. Administrator's Scenario of PetCare System

Step 1: Admin(1)->PetCare Management Server(2)

Description: Verify admin log in credentials, and display the admin page. At the admin page, admin can perform user management functions.

Step 2: PetCare Management Server(2)->Admin(1)

Description: The Management Server performs the operation that admin requested, and notifies the admin whether the operation is successful or not.

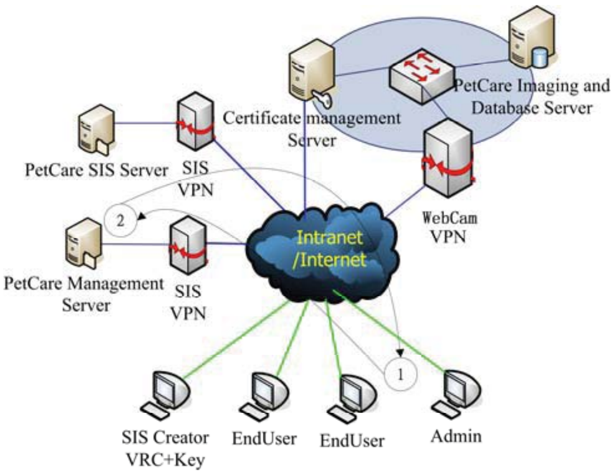


Figure 2.3. Administrator's scenario of PetCare System.

2.4. Component-based PetCare System

Following the above architecture, the experimental component-based PetCare system is illustrated by the UML deployment diagram shown in Figure 2.4. The SIS server provides algorithms to detect different states of the pet such as PetAbsence and PetNoMovement, an InputProcessor to process input from the WebCam, and an Uploader to upload data to the Internet. The WebServer provides PetCare Management services to the end user and the administrator (to be described in Section 3). A universal interface, PrjRemote), is provided for debugging purpose, which can be replaced by the WebServer in normal operations.

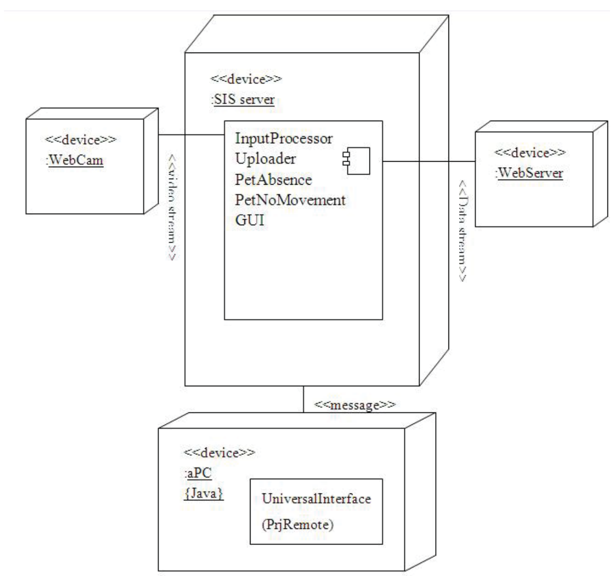


Figure 2.4. Component-based PetCare System.

3 PetCare System Web Interface

The PetCare System Web Interface is the client-side gateway of the PetCare System. It provides the access points to the three modes of the PetCare System, namely the *Preparation Mode*, the *Training Mode* and the *Working Mode.*

(Note: Before accessing the web interface, the PetCare System must be up and running on the server, and the browser must be enabled with javascripts.)

The PetCare System Web Interface provides the client-side interface to interact with the server-side PetCare System. The three modes of the PetCare System can be accessed individually through the web interface. The three modes of the PetCare System can also be modified to form a closed cycle of the PetCare System, the Automatic Retraining Mode. In the Automatic Retraining Mode, the modified and saved image sequence from the Working Mode will be passed to the Training Mode to retrain the SIS system. Again, the retrained SIS system will provide better state predictions in the Working Mode. Thus, the whole system forms an online, fully automatic cycle.

3.1. Authentication

Figure 3.1: The Login Page.

The default login page is shown in Figure 3.1. The user can log into the system through two different types of accounts, the admin account and the user account. The admin account provides all the features of the system including adding new users and changing parameters of the PetCare System, while logged in using the user account will lead the user directly to the working mode as shown in Figure 3.2.

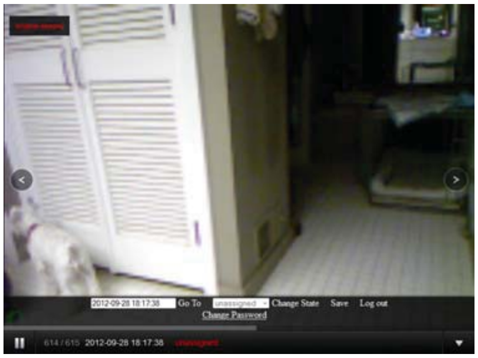


Figure 3.2: The Working Mode after Logged In using the User Account.

3.2. Modes of the PetCare System

There are three modes of the PetCare System, the Preparation Mode, the Training Mode and the Working Mode. When executed individually, they present different facets of the system. Together, they form a cycle of the execution path of the PetCare System. After logged in using the admin account, the user will be presented with the Mode Selection page as shown in Figure 3.3. Through the Mode Selection page, the user can enter different modes of the system.

Figure 3.3. The Three Modes of the PetCare System.

Preparation Mode

The Preparation Mode can be entered by selecting a file and then clicking the Preparation Mode button on the Mode Selection page.

Figure 3.4. The Preparation Mode.

The Preparation Mode is shown in Figure 3.4. Different portions of the web interface provide different functions.

1. Backward button, using this button to move backward to the previous image.
2. Forward button, using this button to move forward to the next image.
3. Jump panel, clicking in the textbox to display the DateTime picker as shown in Figure 3.5, then click the *Go To* label to jump to the image closest to the time selected.
4. State Selector, clicking in the drop-down list to select a state, then click the *Change State* label to change the state of the current image. Note that the status bar will change accordingly.
5. Save button, using this button to save the current states into the event file.
6. Log out button, using this button to log out of the system.
7. Fast Forward button, using this button to quickly change states of the images. When the user clicks the Fast Forward button, the Preparation Mode will automatically jump to the next image and change the state of the next image to the state indicated in the State Selector. The same action will continue applying to the displayed images until the user clicks the Fast Forward button again.
8. Status Bar, it displays the index of the current image, the total number of images, the time stamp of the current image and the state of the current image.
9. The Collapse/Expand button, using this button to collapse or expand the function panel.

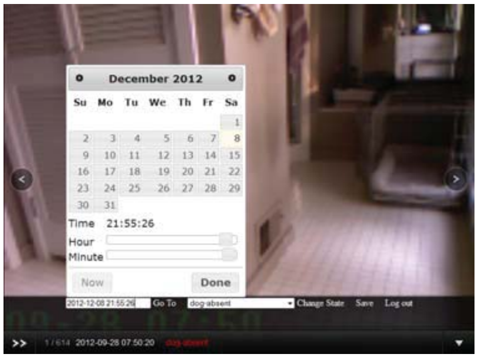


Figure 3.5. The DateTime Picker.

Training Mode

The Training Mode can be entered by selecting a file and then clicking the Training Mode button on the Mode Selection page. After clicking the Training Mode button on the Mode Selection page, the user will be led to the Training Mode page as shown in Figure 3.6, which indicates that the user-selected training file has been saved. In the next step, the PetCare System on the server will be able to train the SIS system automatically by using the user-selected training file.

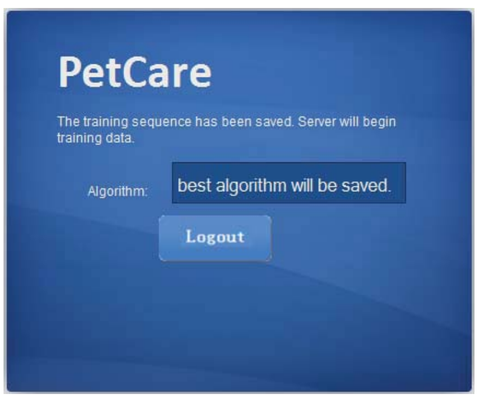


Figure 3.6. The Training Mode.

Working Mode

The Working Mode can be entered by clicking the Working Mode button on the Mode Selection page. The Working Mode will use current date data file by default. (Note: the file drop-down list indicates the offline working mode, which now is a server-side function only.)

The Working Mode interface is very similar to the Preparation Mode, except for some minor differences.

1. Alert panel, displays the alert associated with the current image.
2. State Selector, disabled when the working mode is loading images continually. It will be activated when the working mode is paused.
3. Pause/Play button, using this button to pause or resume the working mode. When paused, the user will be allowed to use the state selector.

3.3. Admin Functions

There are four additional user functions associated with the admin account, and one additional admin function associated with the user account.

When logged in using the admin account, the admin will be presented with options to change the password, change the parameters of the PetCare System, change alert-related parameters, and to add users.

3.4. Web Interface Installation

All the source code of the PetCare System Web Interface can be downloaded from the URL at: https://petcaresystemwebinterface.googlecode.com/svn/trunk/. In order to set up the web interface on a window-based host computer, a php server such as WampServer is required, which can be downloaded from

Figure 4.1. Five typical dog states in two scenes

www.wampserver.com. Once the WampServer is installed, copying all the source code to the hosting directory will complete the setup, and the web interface will be accessible from the computers connected to the host machine. The default hosting directory of the WampServer is C:\wamp\www. The SIS system must be installed in the same directory of the PetCare Web Interface in order for them to interact.

4 Slow Intelligence for PetCare

In this section, we describe how to apply slow intelligence system principles to design the Petcare system. Our task is image recognition by identifying the dog-states such as ‘dog-sleeping', 'dog-standing', 'dog-eating', 'dog-playing', 'dog-absent' and so on, in the image sequences. There are various methods for this kind of pattern recognition task. Our objective is to utilize these algorithms and evolutionarily select the best one.

4.1. Pattern Recognition Models

Hidden Markov Models (HMM)

Hidden Markov model is well known for its application in temporal pattern recognition such as gesture recognition. A hidden Markov model (HMM) is a triple (Π, A, B) where

$\Pi = (\pi_i)$ is the vector of the initial state probabilities

$A = (a_{ij})$ is the state transition matrix: $\Pr(x_{i_t} \mid x_{j_{t-1}})$

$B = (b_{ij})$ is the emission matrix: $\Pr(y_i \mid x_j)$

In our problem setting, we have several sequences of images. The observable states (y) are features extracted from each image, which will be explained in detail later. The hidden states (x) correspond to five possible states including 'dog-sleeping', 'dog-standing', 'dog-eating', 'dog-playing', 'dog-absent'. The parameter learning task in HMM is to find the best set of state transition and output probabilities (Π, A, B), given an output sequence or a set of such sequences. The task is usually to derive the maximum likelihood estimate of the parameters of the HMM given the set of output sequences. We adopt the k-Means clustering algorithm implemented by jahmm-0.6.1 [4] to learn HMM model. This algorithm first packs the observation sequences in *clusters* using the general purpose clustering class KMeansCalculator and iteratively returns a better approximation of a matching HMM until a fix point is reached (*i.e.* the clusters doesn’t change anymore). After obtaining the particular HMM model, we can determine from an observation sequence of images the most likely sequence of underlying hidden pet states that may have generated it. The Viterbi algorithm [6] provides a computationally efficient way of analyzing observations of HMMs to recapture the most likely underlying state sequence. We also adopt the jahmm-0.6.1 implementation for Viterbi algorithm.

Improved Hidden Markov Models (IHMM)

When we have the training sequences of images with labeled hidden states for each sample, the learning step by k-means clustering algorithm uses only observation states of image sequences. The improved Hidden Markov Model utilized both observable states and corresponding hidden states to learn HMM model in the training stage. We estimate (Π, A, B) using maximum likelihood estimation. We use the implementation of Part-of-Speech tagging techniques for HMM [5]. We first count the number of occurrences of each hidden state, then count the number of occurrences of two continuous hidden states pair, e.g. (dog-playing, dog-playing), (dog-playing, dog-sleeping), (dog-standing, dog-playing) and etc. Thus there are 25 two continuous hidden states pair. Finally we analyze the counts for each observation state-hidden state pair. We use the Viterbi algorithm [6] to decode an observation sequence in the testing stage.

Principal Component Analysis (PCA)

Principal component analysis is the mathematical procedure for dimension reduction that is widely applied in pattern recognition. In our problem setting, each image can be represented by several features that will be explained later. If the feature size is very large, we can firstly apply PCA to reduce the dimension and then use support vector machine to learn the classifier classifying each image to its corresponding state ('dog-sleeping', 'dog-standing', 'dog-eating', 'dog-playing', 'dog-absent'). We use LIBSVM [1] to implement SVM for our multiclass classification task.

All of these three models can be applied for our Petcare application. The SIS system can automatically select the best algorithm for different training samples and utilize this algorithm for testing experiment.

4.2.Dataset Description

Our offline dataset contains images taken in four separate days. For each day, the camera is set at a fixed position and captures the image every minute from early morning around 8:00am to late afternoon around 6:00pm. Each image is of the size of 160*120. The images in these four days include two scenes, one is taken in the kitchen (2012-9-28, 2012-10-5) while the other is taken in the living room (2012-10-18, 2012-10-19). By pre-analyzing the data, we identify five typical dog states in the pictures shown in Figure 4.1, 'dog-sleeping', 'dog-standing', 'dog-eating', 'dog-playing', 'dog-absent'. We use Petcare interface Preparation mode to manually annotate each picture of these four days, which are further used as training and testing data. Any picture with an undetermined state is given the "unassigned" label.

4.3. Experimental Setup and Results

As described in the previous section, we should extract feature from each image. In HMM and IHMM, the observable states are the image features. We equally separate each image to 16 regions and assign specific number to each sub-region. The observable state of image is the sub-region (number) most frequently occupied by the dog. For example, in Figure 4.2, most part of the dog is in region 16, thus the observable state for this picture is 16. For PCA, the image is split equally into 16 regions, and the (R,G,B) histogram of the color image is used as features. In particular, we extract the histogram counts for (R,G,B) with 256 bins for each sub-region and then we concatenate 16 regions with 256*3 features from each sub-region to construct our feature vector. In our experiment, we reduce the dimensions to 50.

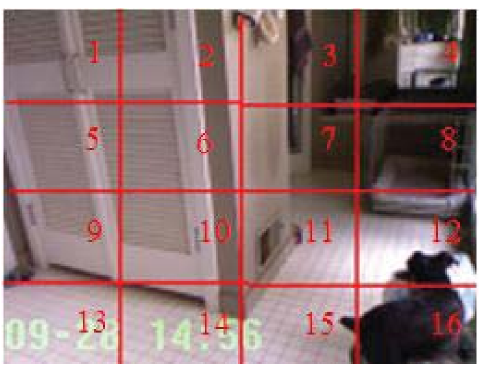


Figure 4.2. Observable state of an example image

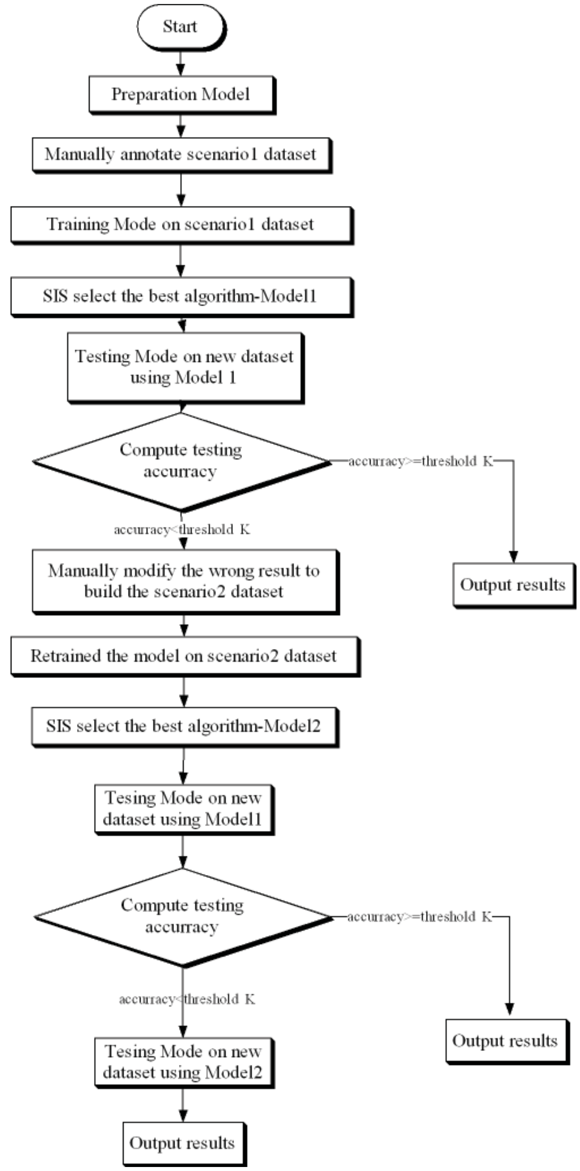


Figure 4.3. The SIS-based experimental scenario.

In the experiment, we use each of these four days data (2012-9-28, 2012-10-5, 2012-10-18, 2012-10-19) as the training set, and test the learned model on the other three days of data.We apply three algorithms HMM, IHMM, PCA in each training and testing phases and compute the accuracy. In the training stage, the data is split into two sets: the training set (50%) to learn model and the validation set (50%) to compute the training accuracy. The experimental results are show in Table 4.1.

From the Table 4.1, we can observe that:

1) The model learned from one scene has significantly better testing performance on the same scenic data than on the different scenarios.

2) IHMM performs better than PCA on kitchen dataset while PCA is better on living room dataset. Generally, the best model with most training accuracy has the best performance for testing data also. The only exception is marked in red in Table 4.1, where IHMM is the best algorithm in the training stage on 2012-10-5 dataset but performs worse than PCA on 2012-9-28 testing dataset.

Based on these observations, we can utilize our Petcare SIS system to automatically select the best algorithm for each training set and apply it to the testing data. Moreover, when the scenarios of training data (environment) change, SIS can self-learn and evolutionarily choose the most suitable algorithm according to corresponding dataset. In our Petcare SIS system, we can first manually annotate one day data in preparation mode, and automatically select best algorithms according to different scenic data. We assume that there are two kinds of scenic data such as kitchen dataset and living room dataset. The SIS-based experimental scenario is illustrated in Figure 4.3.

5 Discussion

The experimental results confirm the slow intelligence approach leads to performance improvements. Similar slow intelligence principles, and lessons learned from the PetCare system, can be applied to the design of a Senior Patient Care system that share many similar characteristics of the PetCare system.

Acknowledgements

The Web Interface source code is based on and modified from Sam Dunn’s Supersized Fullscreen Slideshow jQuery Plugin(www.buildinternet.com/project/supersized).

References

[1] C.-C. Chang and C.-J. Lin, “LIBSVM: A Library for support vector machines”, ACM Transactions on Intelligent Systems and Technology, 2:27:1-27:27, 2011.

[2] Shi-Kuo Chang, "A General Framework for Slow Intelligence Systems", International Journal of Software Engineering and Knowledge Engineering, Volume 20, Number 1, February 2010, 1-16.

[3] Shi-Kuo Chang, Yingze Wang and Yao Sun, “Visual Specification of Component-based Slow Intelligence Systems”, Proceedings of 2011 International Conference on Software Engineering and Knowledge Engineering, Miami, USA, July 7-9, 2011, 1-8.

[4] http://code.google.com/p/jahmm/

[5] Tapas Kanungo, "UMDHMM: Hidden Markov Model Toolkit,", Cambridge University Press, 1999.

[6] Viterbi AJ (April 1967). “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm”. *IEEE Transactions on Information Theory* **13** (2): 260–269.

Training Data	Training accuracy			Test Data accuracy											
				2012/9/28 Kitchen1			2012/10/5 Kitchen2			2012/10/18 Living1			2012/10/19 Living2		
	IHMM	HMM	PCA	IHMM	HMM	PCA	IHMM	HMM	PCA	IHMM	HMM	PCA	IHMM	HMM	PCA
2012/9/28 Kitchen1	**0.647**	0.0203	0.59661				**0.859**	0.02	0.857	**0.199**	0.033	0.0304	**0.051**	0.008	0.002
2012/10/5 Kitchen2	**0.79**	0.036	0.75676	**0.637**	0.039	0.66610169				**0.879**	0.033	0.0304	**0.136**	0.008	0.002
2012/10/18 Living1	0.8131	0.0093	**0.9346**	0.0661	0.039	**0.191525**	0.045	0.02	**0.074**				0.112	0.008	**0.919**
2012/10/19 Living2	0.8196	0.0157	**0.8824**	0.1119	0.039	**0.191525**	0.035	0.02	**0.074**	0.818	0.037	**0.879**			

Table 4.1. The experimental results.

Slow Intelligent Segmentation of Chinese Sentences using Conceptual Interval

Tiansi Dong
In der Wehrhecke 42
53125 Bonn, Germany
Email: tdong@acm.org

Peiling Cui
Department of Sinology
Universität Bonn, Germany
Email: pcui@uni-bonn.de

***Abstract*—The segmentation problem of Chinese sentences is revisited. The term *conceptual interval* is proposed as an implementation of a generated understanding of the *cue* concept in psycho-linguistics. A *conceptual interval* has at least one of the following three conceptual interpretations: ontologies, markers, and frames. A frame delivers a concrete meaning, if all of its roles are filled with anticipated ontologies. Markers help the fill of ontologies into frames. A Chinese sentence is *meaningful*, if it has pairwise disjointed *conceptual interval*s, such that all roles of a frame are filled by ontologies. The slow intelligent process starts with enumerating all *conceptual interval*s of a sentence, then filling ontologies into frames, with or without the help of markers. Pruning is carried out based on the *word order* information, markers, and the property of pairwise disjointedness among *conceptual interval*s. An evaluation formula is presented to quantitatively rank preferences of different segmentation.**

I. Introduction: Segmentation problem in Chinese

The basic element of Chinese sentences is character. One or more characters form a word, which carries concepts. A Chinese sentence is simply a sequence of characters with no space between them. If English sentences were also so structured, *wewouldwritethisway* – no spaces among *we would write this way*. To understand a Chinese sentence, we need to partition the sentence into several groups, each group forms a word. This is by no means trivial for understanding Chinese sentences. The sentence “北京大学生在北京大学唱歌” has 12 characters. Possible words can be: “北”, which may mean *north*, *be defeated*, “北京”, the current capital city of China, “北京大学”, *Peking University*, “大学生”, *university student*, “生”, which may mean *give birth to*, *raw*, etc. This follows different segmentation of the sentence. The preferable segmentation is “北京|大学生|在|北京大学|唱|歌”, which means *university students in Bejing sing songs at Peking University*; a non-preferable segmentation would be “北京大学|生|在|北京|大学|唱|歌”, which can be put into English as *Peking University gives birth to Beijing, University sings songs*. The reason to choose the first segmentation is that this segmentation delivers a *meaningful interpretation*. to group “北” and “ 京” to denote a city or to group “北,京,大” and “学” to denote *Peking University* depends on other groupings and the understanding of all these groups. That is, a local segmentation is determined by the meaning integration of all segmentation. Chinese sentence segmentation can be a slow intelligent decision cycle, Chang (2010), as it needs to enumerate possible segmentation, then to choose a preferable one, which has a *meaningful interpretation*.

What does the term *meaningful interpretation* mean? Let us back to the example. The second segmentation does not has a *meaningful interpretation*, for the first word “北京大学”(Peking University) is neither an animal nor a person, and can neither perform the action of “生”(give birth to) nor be raw. “生”(give birth to) here serves as a *meaning frame*, which has anticipations. *Meaningful interpretation* means that each anticipation is fulfilled. For example, in the first segmentation, “唱”(sing) is a *meaning frame*, which anticipates a person as its actor and a melody as its object. “大学生”(university student) fulfills the actor and “歌”(song) fulfills the object. A segment is an interval of sentence, which provides either anticipations, or possible fulfillment, or helping of certain fulfillment. Such a segment is a *conceptual interval*.

II. Conceptual Interval

A. Forms

The form of a Chinese sentence is both its characters and the spatial relations among these characters. Let $\mathbb{C}$ be the set of Chinese characters, S be a sentence with n characters, written as $S = [s_1 s_2 \dots s_n], s_i \in \mathbb{C}$. $S[i,j]$ represents the sub-sequence from ith character to jth. The word set, W_S, is the set of all such $S[i,j]$s that have meanings. For the convenience of reading, we write $S[i,j]$ as $s_i \dots s_j[i,j]$. For the above example, $S =$ [北京大学生在北京大学唱歌]; $W_S = \{$北[1,1], 北京[1,2], 北京大学[1,4], 大学[3,4], 大学生[3,5], 学[4,4], 学生[4,5], 生[5,5], 生在[5,6], 在[6,6], 北[7,7], 北京[7,8], 北京大学[7,10], 大学[9,10], 唱[11,11], 歌[12,12]$\}$.

B. Meanings

One word may have several different meanings. For example, “生” can mean `GIVE_BIRTH_TO`, `RAW`, `LIVING`, `UNFAMILIAR_WITH`, `VERY`, etc. We use `mathtt` style to represent meanings. The English word *raw* may mean *not cooked*, or *being in or nearly in the natural state*, or *lacking covering*, etc. What does the `RAW` in our semantic system differ to the word *raw* in English? Our semantic world is well-structured, formal, and bounded, in which each element is well defined and plays a unique role: the protagonists are frames; the costars are ontologies; some servants are markers.

1) Frame: A frame is an event structure, which schematizes a structure of an event happened in the world or in mind. In our example, possible frames are BE_DEFEATED, LEARN, GIVE_BIRTH_TO, BE_BORN_IN, and SING.

```
BE_DEFEATED
  Object : PERSON
  ∗Place : LOCATION
  ----------------
  Object ≪ BE_DEFEATED
```

```
LEARN
  Actor : PERSON
  Object : SKILL ⊔ KNOWLEDGE
  ∗Place : LOCATION
  ----------------
  Actor ≪ LEARN
```

```
GIVE_BIRTH_TO
  Actor : PERSON ⊔ ANIMAL
  Object : PERSON ⊔ ANIMAL
  ∗Place : LOCATION
  ----------------
  Actor ≪ GIVE_BIRTH_TO
```

```
BE_BORN_IN
  Object : PERSON ⊔ ANIMAL
  Place : LOCATION
  ----------------
  Object ≪ BE_BORN_IN
```

```
SING
  Actor : PERSON
  Object : MELODY
  ∗Place : LOCATION
  ----------------
  Actor ≪ SING
```

BE_DEFEATED schematizes the fact that a person is defeated. This frame has an object role. The object role shall appear before the frame in the sentence, written as '*Object* ≪ BE_DEFEATED'. When this role is filled, this frame delivers a meaningful interpretation. Roles starting with ∗ are optional. The knowledge-base of frame $\mathbb{F}$ is the collection of all frames.

2) Ontologies: A frame delivers a meaning, if its roles are filled. Role-filling needs ontologies. An ontologies can fill to a role, if it is either a sub-category or an instance of an anticipated category of this role. For example, STUDENT can fill the actor role of SING, for STUDENT is a sub-category of PERSON. A concrete person, say G.W.Bush, can also sing and fill the actor role of sing. Two basic relations among ontologies are "being an instance of"(isa and "being generalized by"(genls). The isa relation is non-transitive; while the genls relation is transitive. We list possible ontologies in the above example as follows: NORTH, PERSON, BEIJING, PKU, STUDENT, CITY, UNI_STUDENT, LOCATION, SKILL, SONG, MELODY. The isa relations are listed as follows: (BEIJING, LOCATION), (NORTH, LOCATION), (PKU, LOCATION), (SINGING, SKILL). The genls relations are listed as follows: (UNI, LOCATION), (SONG, MELODY), (UNI_STUDENT, STUDENT), (STUDENT, PERSON).

C. Cues in Chinese

A cue is a relationship between forms and meanings, following MacWhinney et al. (1984) and Li et al. (1993). According to Li et al. (1993), Chinese cues include the passive marker 被(by), the object marker 把(hold), the indefinite marker 一(one), animacy, and word order.

1) Marker: A marker can explicitly demarcate to which role of a frame a concept shall fill. For example, the passive marker 被(by) explicitly demarcates that the ontologies following this marker shall fill to the actor role. The object marker 把(hold) explicitly demarcates that the ontologies following this marker shall fill to the object role. In the above example, "在"(in) serves as a location marker, written as ZAI_LOC[1]: if there is a spatial entity following it (ZAI_LOC ≪ *Object*) and no frame between them (*frame* ∉ [ZAI_LOC, *Object*]), this ontologies shall fill the place role. We append LOC after ZAI, because "在" can also serve as a present tense marker. The function of ZAI_LOC can be represented in a schema as follows.

```
ZAI_LOC
  Object : SPA_ENTITY
  ----------------
  ZAI_LOC ≪ Object → Place
  frame ∉ [ZAI_LOC, Object]
```

2) Category: The animacy cue says that an actor role of a frame shall be filled with something with animacy. In the sentence "骨头这狗啃了"(The dog has gnawed on a bone), "the dog"(这狗) is something with animacy and can perform the action "gnaw". So it shall be the actor, although it appears behind "a bone"(骨头). We generalize the animacy cue into the category cue: each role of a frame shall be filled only by elements within a specific category. The actor role of the "gnaw" frame shall be filled only by something of animacy. Its object role shall be filled with something of nutrition.

3) Word order: In Chinese, it is common that an actor appears before a frame and that the object appears after the frame, i.e., Sun and Givón (1985). The word order cue is especially useful, when both of the actor and the object are in the same category. For the sentence 狗(DOG)在追(RUNN_AFTER)猫(CAT), category cue fails, as both DOG and CAT can fit to the actor role. The order among DOG, RUNN_AFTER, and CAT works: RUNN_AFTER separates the actor and the object. The actor is located before the frame, while the object is after it. In Chinese, object can be put at the beginning of a sentence. For example, 猫(CAT), 老鼠(MOUSE)在追(RUNN_AFTER). We may improve the word

[1] ZAI is the pronunciation of "在"

order cue as follows: the actor is directly before the frame, and the object may either after the frame or before the actor. This is unfortunately still not the whole story. For the sentence: 兔子(RABBIT), 猫(CAT)和狗(DOG)在追(RUNN_AFTER). Based on the improved word order cue, DOG shall be the actor as it is directly before the frame, RABBIT and CAT shall be the object, as they are before the actor. The fact is that CAT shall be an actor. To solve this case, we may either use the comma between RABBIT and CAT as a marker cue, or use category cue among RABBIT, CAT, and DOG by saying that CAT is ontologically more closer to DOG than to RABBIT, therefore it shall be grouped with DOG, instead of RABBIT. Word order cue is by no mean simple in Chinese, as well as in other languages, i.e., Uszkoreit (1987).

4) Cue dictionary: A sub-sequence of a sentence can be understood as a frame (tagged by **F**), or an ontologies (tagged by **O**), or a marker (tagged by **M**). We use a *cue dictionary*, $\mathbb{D}$, to link forms, ontologies $\mathbb{O}$, and frames $\mathbb{F}$.

Form	Tag	Meaning	shortened for
北	**C**	NORTH	NTH
北	**F**	BE_DEFEATED	B_D
北京	**C**	BEIJING	BJ
北京大学	**C**	PKU	PKU
大学	**C**	UNI	UNI
大学生	**C**	UNI_STUDENT	U_S
学	**F**	LEARN	LRN
学生	**C**	STUDENT	STU
生	**F**	GIVE_BIRTH_TO	G_B
生在	**F**	BE_BORN_IN	B_B
在	**M**	ZAI_LOC	Z_L
唱	**C**	SINGING	S_I
唱	**F**	SING	SI
歌	**C**	SONG	SO

D. *Conceptual intervals*

For a Chinese sentence, an interval can be any of its sub-sequence. If such of an interval has a meaning, e.g. frame, ontologies, marker, we name it *a conceptual interval*. A conceptual interval is represented by the associated meaning and the location of the interval within the sentence. For example, NTH$[1, 1]$ and PKU$[1, 4]$ are two conceptual intervals of the sentence S = [北京大学生在北京大学唱歌]. NTH$[1, 1]$ is an interval starting from the first character to the first character of the sentence. The associated concept is NTH. PKU$[1, 4]$ is an interval starting from the first character to the fourth character. Its associated concept is PKU. We write CIF for conceptual intervals of frames, CIO for conceptual intervals of ontologies, and CIM for conceptual interval of markers.

III. Slow intelligence process of meaningful segmentation of Chinese sentences

A. *Enumeration possible conceptual intervals*

Given the sentence S = [北京大学生在北京大学唱歌], we enumerate all possible conceptual intervals as follow: NTH$[1, 1]$, B_D$[1, 1]$, B_J$[1, 2]$, PKU$[1, 4]$, UNI$[3, 4]$, LRN$[4, 4]$, U_S$[3, 5]$, STU$[4, 5]$, G_B$[5, 5]$, B_B$[5, 6]$, Z_L$[6, 6]$, NTH$[7, 7]$, B_D$[7, 7]$, B_J$[7, 8]$, PKU$[7, 10]$, LRN$[10, 10]$, SI$[11, 11]$, S_I$[11, 11]$, SO$[12, 12]$.

B. *Matrix of conceptual intervals*

We construct a matrix of conceptual intervals as follows: Each column is a CIF or CIM; each row is a CIO. If a CIO can be filled into a role of a frame, the cell of this column and this row is assigned to the role name, otherwise 0. For example, as U_S can be generalized by PERSON, it can fill to the actor role of B_D, so the cell with row U_S$[3, 5]$ and column B_D$[1, 1]$ is set to Object. This assignment is purely based on category cues. For the above example, we have the matrix of conceptual intervals as follows.

	B_D[1,1]	LRN[4,4]	G_B[5,5]	B_B[5,6]
NTH[1,1]	Place	Place	Place	Place
B_J[1,2]	Place	Place	Place	Place
PKU[1,4]	Place	Place	Place	Place
UNI[3,4]	Place	Place	Place	Place
U_S[3,5]	Object	Actor	Actor, Object	Object
STU[4,5]	Object	Actor	Actor, Object	Object
NTH[7,7]	Place	Place	Place	Place
B_J[7,8]	Place	Place	Place	Place
PKU[7,10]	Place	Place	Place	Place
UNI[9,10]	Place	Place	Place	Place
S_I[11,11]	0	Object	0	0
SO[12,12]	0	Object	0	0

	Z_L[6,6]	B_D[7,7]	LRN[10,10]	SI[11,11]
NTH[1,1]	Place	Place	Place	Place
B_J[1,2]	Place	Place	Place	Place
PKU[1,4]	Place	Place	Place	Place
UNI[3,4]	Place	Place	Place	Place
U_S[3,5]	0	Object	Actor	Actor
STU[4,5]	0	Object	Actor	Actor
NTH[7,7]	Place	Place	Place	Place
B_J[7,8]	Place	Place	Place	Place
PKU[7,10]	Place	Place	Place	Place
UNI[9,10]	Place	Place	Place	Place
S_I[11,11]	0	0	Object	0
SO[12,12]	0	0	Object	Object

C. *Pruning by word order cue within and cross frames*

Word-order cues help to prune incorrect assignments. We remove B_D$[1, 1]$, LRN$[4, 4]$, G_B$[5, 5]$, and B_B, for the violation of the word order cue within the frame. The frames in the following matrix are filled, and deliver a meaning.

One word order of Z_L is that the frame shall not be located between Z_L and the spatial entity ($frame \notin [\text{Z\_L}, Object]$). In the above matrix, B_D$[7, 7]$ is located between Z_L$[6, 6]$ and four possible spatial entities. We eliminate column B_D$[7, 7]$.

	Z_L[6,6]	~~B_D~~[7,7]	LRN[10,10]	SI[11,11]
NTH[1,1]	~~Place~~	Place	Place	Place
B_J[1,2]	~~Place~~	Place	Place	Place
PKU[1,4]	~~Place~~	Place	Place	Place
UNI[3,4]	~~Place~~	Place	Place	Place
U_S[3,5]	0	Object	Actor	Actor
STU[4,5]	0	Object	Actor	Actor
NTH[7,7]	Place	Place	Place	Place
B_J[7,8]	Place	Place	Place	Place
PKU[7,10]	Place	Place	Place	Place
UNI[9,10]	Place	Place	Place	Place
S_I[11,11]	0	0	Object	0
SO[12,12]	0	0	Object	Object

D. Pruning isolated roles with markers

The function of markers is to explicitly state the role of CIOs. If a marker demarcates some CIOs for a role, we will remove other CIOs filled for the same role without the demarcation of this marker. The marker and the demarcated CIOs are integrated as one ontologies for the role. In our example, we will remove those CIOs, which are filled into the place role, but not demarcated by Z_L[6,6], and integrate Z_L[6,6] with its demarcated CIOs. For example, integrating Z_L[6,6] with NTH[7,7] results in NTH[6,7].

E. Pruning CIOs which intersect with CIF

If one conceptual interval is chosen as a correct segment, other conceptual intervals which intersect with this interval shall be eliminated. For each column, we eliminate CIOs which intersect with CIF in this column.

F. Maximizing lengths of filled and non-filled CIOs

If one interval is part of the other, and both can fill a role of frame, we choose the longer interval for the role. For example, STU[4,5] and U_S[3,5] can fit the actor role of LRN[10,10], and [4,5] is part of [3,5], we choose U_S[3,5] to fill the role. We prune again intersected intervals along all CIOs: As [3,5] is chosen, we remove [3,4] and [1,4]. If two intervals are not filled to any role, and one is part of the other, we also remove the shorter interval, as illustrated below.

	LRN[10,10]		SI[11,11]
~~NTH[1,1]~~	~~Place~~	~~NTH[1,1]~~	~~Place~~
B_J[1,2]	~~Place~~	B_J[1,2]	~~Place~~
~~PKU[1,4]~~	~~Place~~	~~PKU[1,4]~~	~~Place~~
~~UNI[3,4]~~	~~Place~~	~~UNI[3,4]~~	~~Place~~
U_S[3,5]	Actor	U_S[3,5]	Actor
~~STU[4,5]~~	~~Actor~~	~~STU[4,5]~~	~~Actor~~
~~NTH[6,7]~~	~~Place~~	~~NTH[6,7]~~	~~Place~~
B_J[6,8]	Place	B_J[6,8]	~~Place~~
PKU[6,10]	~~Place~~	PKU[6,10]	Place
UNI[6,10]	~~Place~~	UNI[6,10]	~~Place~~
S_I[11,11]	Object	S_I[11,11]	~~0~~
SO[12,12]	Object	SO[12,12]	Object

G. Evaluating segmentation quality of each CIF column

For each CIF column, we sort all intervals, and count numbers of *un-filled* and *unknown and un-filled* CIOs. The less the number of *un-filled* CIOs is, the higher the segmentation quality is; the less the number of *unknown and un-filled* CIOs, the higher the segmentation quality is. Let M be the total number of intervals, N be the number of *un-filled* CIOs, and P be the number of *unknown un-filled* CIOs. The segmentation quality κ is defined by the formula as follows.

$$\kappa = \frac{M-N}{M} * \frac{M-P}{M} \quad (1)$$

For column LRN[10,10], we have B_J[1,2] U_S[3,5] B_J[6,8] [...] L[10,10] S_I[11,11] SO[12,12]. $M = 7, N = 2, P = 1$, so, $\kappa = (7-2)/7 * (7-1)/7 \simeq 0.612$. The corresponding segmentation is "北京|大学生|在北京|大|学|唱|歌". 大(big) is an unknown character. If we ignore the meaning of 大(big), the meaning would be *university students of Beijing learn singing and songs in Beijing*. If we know the meaning of 大(big), the meaning will be *university students of Beijing learn singing and songs very hard in Beijing*.

For column SI[11,11], we have B_J[1,2] U_S[3,5] B_J[6,10] SI[11,11] SO[12,12]. $M = 5, N = 1, P = 0$, so, $\kappa = (5-1)/5 * (5-0)/5 = 0.8$. The corresponding segmentation is "北京|大学生|在北京大学|唱|歌". The meaning is *university students of Beijing sing songs at Peking University*. This is the more preferable understanding than the first one, for 大(big) is seldom used as *very hard* before 学(learn).

IV. Acknowledge

We proposed here the slow intelligent approach to the old difficult problem of Chinese segmentation, and still a long way to go. We are indebted to two anonymous reviewers for the critical comments and encouragement.

References

Chang, S. K. (2010). A general framework for slow intelligence systems. *International Journal of Software Engineering and Knowledge Engineering*, 20(1):1–18.

Li, P., Bates, E., and MacWhinney, B. (1993). Processing a Language without Inflections: A Reaction Time Study of Sentence Interpretation in Chinese. *Journal of Memory and Language*, 32:169–192.

MacWhinney, B., Bates, E., and Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23:127–150.

Sun, C. F. and Givón, T. (1985). On the So-Called SOV Word Order in Mandarin Chinese: A Quantified Text Study and its Implications. *Language*, 61(2):329–351.

Uszkoreit, H. (1987). *Word Order and Constituent Structure in German*. CSLI Lecture Notes, Center for the Study of Language and Information, Stanford University, California.

Supporting CMMI assessment using distributed, non-invasive measurement and process mining

Saulius Astromskis, Andrea Janes, Alberto Sillitti, and Giancarlo Succi
Centre for Applied Software Engineering
Free University of Bozen/Bolzano, Bolzano, Italy
{saulius.astromskis, andrea.janes, alberto.sillitti, giancarlo succi}@unibz.it

Abstract

The reputation of lightweight software development processes such as Agile and Lean is damaged by practitioners that claim benefits of such processes that are not true.

Teams that want to demonstrate their seriousness, could benefit from matching their processes to the CMMI model, a recognized model by the industry and the public administration. CMMI stands for Capability Maturity Model Integration and provides a reference model to improve and evaluate processes according to their maturity based on best practices.

On the other hand, particulary in a lightweight software development process, the costs of a CMMI appraisal are hard to justify since its advantages are not directly related to the creation of value for the customer.

This paper presents Jidoka4CMMI, a tool to – once a CMMI appraisal has been conducted – allow to document the assessment criteria in form of executable test cases. The test cases, and so the CMMI appraisal, can be repeated anytime, without additional costs.

The use of Jidoka4CMMI increases the benefits of conducting a CMMI appraisal. We hope that this encourages practitioners using lightweight software development processes to assess their processes using a CMMI model.

1. Introduction

Promoters of so called lightweight software development processes (e.g., Extreme Programming, Scrum, Lean software development) face the problem that luring consultants claim that if you use a lightweight development process, you do not need to document anything, can solve every problem, produce bug-free software, turn your customers into friends, maintain deadlines easily, and will have a 9 to 5 job [1].

Moreover, some companies call themselves Agile, Lean, etc., not because they pursue agility, but because it is fashionable to do so [2]. The picture of a young, dynamic, flexible, and lean team is what customers expect and, as a consequence, what companies convey within their marketing material.

In summary, the described development damages the reputation of lightweight software development processes. The conscientious part of the community needs to develop methods and tools to objectively assess their method to produce software and hence, to understand in which context it works and in which not. The availability of such methods and tools will help to distinguish the serious practitioner from the quacksalver.

Agile and Lean teams could benefit from having a reference model, accepted by the industry and the public administration, that defines what it means to develop "good" software. Teams that adhere to such model could claim that what they do corresponds to the best practices of software development.

The CMMI for Development is such a model that describes typical elements of effective processes [3]. Therefore, a proposed solution to increase the reputation of Agile and Lean teams is to embrace both: CMMI and Agile [3].

The comparison of the development process of a given organization with the recommendations of the CMMI helps to evaluate the maturity of the analyzed process, find gaps between the performed activities and the suggested ones, identify improvement possibilities, and demonstrate their maturity towards criticizers using a model recognized throughout the industry and public administration.

According to Hillel Glazer, one of the autors of the paper "CMMI or Agile, Why not embrace both?" [3], a CMMI appraisal for a small business with less than 100 developers that is completely ready to be appraised can cost from $36.000 to $60.000, the exact price depends on a variety of factors. If the team needs help to prepare everything to get appraised, clients typically spend about $125.000 over about a year's time or less [4].

Agile and Lean practitioners are attentive on the elimination of any unnecessary activities to improve their efficiency. Hence, there is a certain level of skepticism towards the CMMI since its value is indirect: it does not directly contribute to the creation of value for the client. Only if team is able to use the results of the appraisal to optimize the software development process, and finally the software it produces, it was worth it. For many, a CMMI appraisal represents an investment that, considering the costs, is too risky.

We want to encourage practitioners using lightweight software development processes (this are our **envisioned users**) to assess their processes from the CMMI perspective. Therefore we propose a tool that once configured, constantly evaluates the CMMI compliance of the ongoing processes, without generating additional costs.

The goal of our tool, called Jidoka4CMMI, is – similar to JUnit[1] or Hudson[2] – to provide a framework to define test cases that verify if a process conforms the CMMI recommendations.

As JUnit or Hudson, our tool is based on the Jidoka principle (see Section 2.1), i.e, once configured, it does not require *any* effort to assess a process.

2. Background

2.1. Jidoka

One of the pillars of Lean Management is "Jidoka", the introduction of quality inside the production process and product.

Jidoka is often translated with "autonomation" or "automation with a human mind" and is usually illustrated making the example of a machine that can detect a problem with the produced output and interrupt production automatically rather than continue to run and produce bad output [5].

Some authors translate Jidoka with "quality-at-the-source" [6] meaning that quality is inherent in the production system, which is in contrast to the traditional approach that checks quality after the process. In essence Jidoka is composed by two parts: a mechanism to detect problems, i.e., abnormalities or defects, and a mechanism to notify when a problem occurs [7].

[1]http://www.junit.org
[2]http://hudson-ci.org

Figure 1 illustrates the idea: produced parts are delivered into a box by an assembly line. On the way a switch – a mechanism to detect a problem – ensures that the parts are not to big. If the switch is enabled, the assembly line is stopped and a lamp – a mechanism to notify – informs the operator that something is wrong.

In this work we propose to use the concept of Jidoka to ensure that the processes in place correspond to the recommendations of the CMMI.

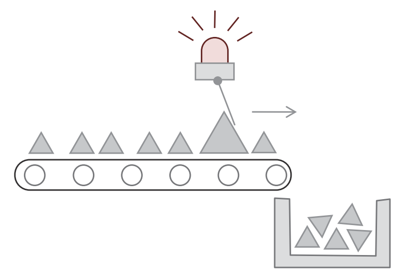


Figure 1. Jidoka

2.2. CMMI

The CMMI for Development is a model that describes typical elements of effective processes [3]. It is used to assess the maturity of a process. A CMMI model describes typical work products and typical practices that - if performed - indicate a certain maturity level.

CMMI models are not precise process descriptions. The actual processes used in an organization depend on many factors, including application domains and organization structure and size. The process areas of a CMMI model typically do not map one to one with the processes used in the organization [8].

To automate the assessment of processes against the recommendations of the CMMI, we adopt non-invasive data collection.

2.3. Non-invasive data collection

The term "non-invasiveness" originates from the medical field, in which it describes a treatment that does not require "the entry into the living body (as by incision or by insertion of an instrument) [9]".

We use this term to describe a measurement methodology that does not require *any* participation by the software developer. As depicted in figure 2, Non-invasive data collection relies on a set of sensors or probes that extract the data from other systems (active) or that get triggered by the observed events (passive) [10].

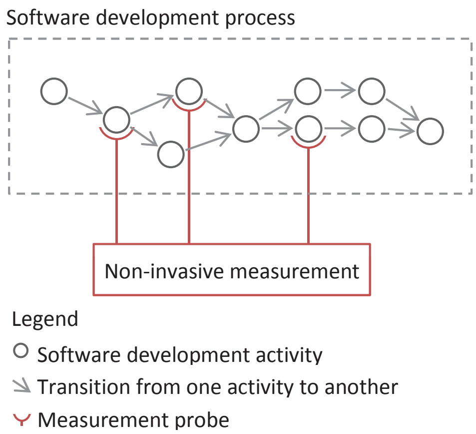


Figure 2. Measurement probes

The measurement probe logs events and report them to a message queue over the Internet using the REST [11] protocol.

We use a message queue to minimize the workload of the client computers. Developers do not want that their machines slow down because some measurement program is running in the background. By using a message queue (we use Apache ActiveMQ[3]), we can minimize the time the client machine is busy uploading data. The data are then read from the message queue, processed, and inserted into the data warehouse. Due to the large amount of data we collect, we use a NoSQL database, Apache Cassandra[4]. The data are then extracted from the data warehouse and processed to feed it to the process mining algorithm and its results visualized on a dashboard (see figure 3).

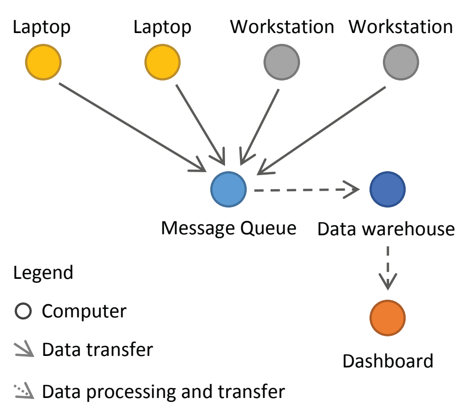


Figure 3. Data flow in the Jidoka4CMMI tool

[3]http://activemq.apache.org
[4]http://cassandra.apache.org

Figure 3 contains the same elements as figure 1: a mechanism to detect a problem and a mechanism to notify everybody that a problem exists: in figure 1, the switch under the lamp detects the problem and the lamp notifies everybody. In figure 3, the measurement probes report problems to the data warehouse and the dashboard informs everybody about the current status of the system.

3. The Jidoka4CMMI tool

The software engineering challenge we propose to address, is to allow practitioners to define test cases for their processes that are monitored by Jidoka4CMMI. Such a test case defines criteria to detect a problem and a mechanism to notify stakeholders when a problem occurs.

An analogy to our tool is to set up a mouse trap: we want to set it up so that it snaps when there is a problem.

The architecture of our system is shown in figure 4. The measurement probes continuously extract data from the development process about the employed resources, the produced output, and the carried out activities.

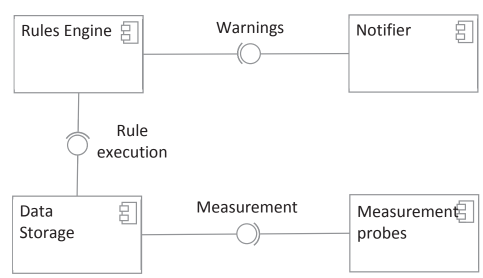


Figure 4. Component diagram of the Jidoka4CMMI tool

We implemented measurement probes that record all interactions developer have with popular integrated development environments such as Eclipse[5] or Microsoft Visual Studio[6], office automation suites such as Apache Open Office[7] or Microsoft Office[8], and other tools developers frequently use (e.g., web browsers or chat applications). Developers interested in adding additional data to the system can develop a plugin implementing a defined interface.

[5]http://www.eclipse.org
[6]http://www.microsoft.com/visualstudio
[7]http://www.openoffice.org
[8]http://office.microsoft.com

The collected data (a log of all detected events) are stored in the Data Storage. The Rules Engine retrieves the data from the data storage, analyses them, and generates warnings if a rule is violated. The Notifier executes the actions to perform if a rule is violated.

3.1. How to use Jidoka4CMMI

Typically, a CMMI appraisal team identifies practice implementation indicators, i.e., evidence that provides a basis for verification of the activity or practice implementation.

These practice implementation indicators (since version 1.3 called "objective evidence") indicate how well a practice was implemented [3]. The concept of "objective evidence" relies on the idea that the implementation of a practice results in "footprints" that can be observed and that prove the implementation of the practice.

A user that wants to write a test case for a process using Jidoka4CMMI has to:

1. specify a Goal Question Metric (GQM) model [12] that defines what to collect, how to collect it, and why to collect it;

2. specify the conditions that the collected "footprints" have to fulfill using JavaScript[9] or Java[10];

3. define how a detection of non-conformance has to be notified.

In step 1) we require the user not only to specify what data to collect, but also to document which question this data answers and how to interprete the answer of the question. The result of the test case is then linked back to the measurement goal to understand what the result means.

The CMMI model for Development specifies requirements for processes. Jidoka4CMMI contains process mining support to allow the user to evaluate if a process conforms to the defined process or, if we find out that the defined process is not followed, to discover the real process.

Process mining is a research discipline that sits between machine learning and data mining on the one hand, and process modeling and analysis on the other hand [13]. Process mining supports:

- Process discovery, in which an algorithm generates a process model based on the data coming from event logs.

- Process conformance, in which an algorithm locate discrepancies between the actual and the ideal process. Process conformance requires the definition of the ideal model in addition to the event logs.

- Process enhancement, in which the current process is analyzed to find opportunities to enrich the model, e.g., to make it reflect the reality better or to align it to the strategy of the organization.

The detection of non-conformance can be notified via e-mail or in a user configurable dashboard. The dashboard summarizes the results of the test cases.

4. Example

The initial validation of our approach occurred through the implementation of proof-of-concepts. In this paper we present one. Each proof-of-concept consists of five parts: the description of the context, the GQM model that describes why, how, and which data is collected, examples of the collected data, the description how the data is processed, and examples of possible results of the analysis.

The following proof-of-concept demonstrates how to write a test case for the specific practice 1.1, "Objectively Evaluate Processes" that is part of the "Process and Product Quality Assurance" process area [8].

In this proof-of-concept the process to evaluate is the requirements management process. Each requirement has to go through the states "Backlog", "Selected", "Development", "Done", "Deployment", and "Live".

The product owner adds ideas for requirements to the requirement backlog. Requirements that should be picked up by developers are set to the status "Selected". When a developer begins implementing a requirement he or she sets the status of the requirement to "Development" and after finishing to "Done". When a features gets deployed to the live system, its status is changed to "Deployment" and finally, after finishing the deployment to "Live".

Figure 5 shows the Goal Question Metric model that specifies how Jidoka4CMMI should verify if the requirements management process fulfills the specific practice 1.1 of the "Process and Product Quality Assurance" process area.

The model verifies if requirements follow the defined process through process mining. The log of status changes is compared with the defined process model using process conformance checking. Moreover, using process discovery, the number of loops that requirements performed is evaluated. If requirements are moved from "Deployment" to "Development" too often, it might indicate that the team has difficulties with

[9] http://www.ecmascript.org
[10] http://www.java.com

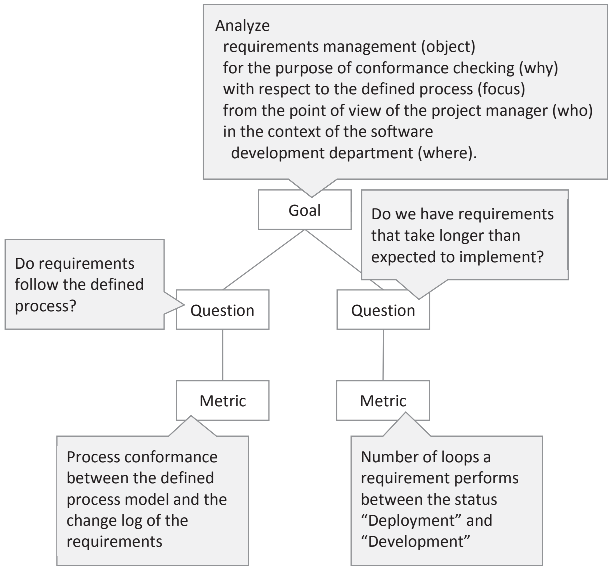


Figure 5. An excerpt of a Goal Question Metric model describing how a given specific practice is evaluated

some type of requirements, that the requirements are not precise enough, etc.

The data used to evaluate this GQM model comes from a probe that logs all status changes of the requirements. The probe reports data as in table 1.

Table 1. Example data

Timestamp	Requirement	Status change
01.01.2012	25	Backlog to Selected
02.01.2012	26	Development to Deployment
03.01.2012	25	Selected to Development

The data is passed to the process mining engine that evaluates the metrics defined in the GQM model.

To extract the visual process model from the event log, we use an algorithm called the Heuristics miner because this algorithm is suited for real world data [14]. The Heuristics miner algorithm is able to deal with noisy data [15], ignores rare events and activities, and therefore generates simple models representing only the most common behavior.

Examples of evaluation results are:

- The process conforms to the defined criteria
- Requirement 25 exceeded the limit for loops between “Deployment” and “Development”.
- Requirement 3 did not follow the defined process, it was directly added with the status “Development”.

5. State of the art

From a bird’s-eye perspective, a CMMI appraisal consists of two steps: collecting data and interpreting it according to the CMMI model [16]. Past research can be classified into these two groups, i.e, methods and tools that a) support the data collection or b) the data interpretation.

Data collection tools alleviate the burden of manual data collection through forms and templates, extract data generated by already existing systems, or infer properties about the executed processes using techniques based on data mining (e.g., [17, 18, 19, 20]).

Data interpretation tools help to infer the achieved maturity level interpreting the collected data (e.g., [21, 22, 23, 24]).

Our work differs from previous works in that it combines non-invasive data collection and process mining to support the maturity assessment of an Agile or Lean organization.

6. Conclusion

In this work, we present a novel scenario in the assessment of light weight software development processes. We show, how quality assurance can be built into the process, as proposed by lean thinking. We give an example of how to integrate the continuous monitoring of produced artifacts, ongoing development activities, and consumed resources with the production process together with the knowledge about unacceptable values of the monitored data.

References

[1] A. Janes and G. Succi, “The dark side of agile software development,” in *Proceedings of the ACM international symposium on New ideas, new paradigms, and reflections on programming and software*, ser. Onward! ’12. New York, NY, USA: ACM, 2012, pp. 215–228.

[2] M. Ceschi, A. Sillitti, G. Succi, and S. D. Panfilis, “Project management in planbased and agile companies,” *IEEE Software*, pp. 21–27, 2005.

[3] H. Glazer, J. Dalton, D. Anderson, M. Konrad, and S. Shrum, “Cmmi or agile: Why not embrace

both!" Software Engineering Institute, Carnegie Mellon University, Tech. Rep., 2008.

[4] H. Glazer. (2011, Feb) Roughly how much does an cmmi evaluation cost for a small business (less than 100 developers)? http://www.quora.com/Roughly-how-much-does-an-CMMI-evaluation-cost-for-a-small-busin/ess-less-than-100-developers.

[5] T. Ohno, *Toyota Production System: Beyond Large-Scale Production*. Productivity Press, 1988.

[6] C. Standard and D. Davis, *Running todays factory: a proven strategy for lean manufacturing*. Hanser Gardner Publications, 1999.

[7] Y. Monden, *Toyota Production System*, 2nd ed. Industrial Engineering Press, 1993.

[8] CMMI Product Team, "Cmmi® for development, version 1.3," Software Engineering Institute, Carnegie Mellon University, Tech. Rep., 2010.

[9] Merriam Webster, "Definition of invasive," http://www.merriam-webster.com/dictionary/invasive, 2010.

[10] A. Janes, M. Scotto, A. Sillitti, and G. Succi, "A perspective on non invasive software management," in *Instrumentation and Measurement Technology Conference, 2006. IMTC 2006. Proceedings of the IEEE*, april 2006, pp. 1104 –1106.

[11] R. T. Fielding and R. N. Taylor, "Principled design of the modern web architecture," *ACM Trans. Internet Technol.*, vol. 2, no. 2, pp. 115–150, May 2002.

[12] V. R. Basili, G. Caldiera, and H. D. Rombach, "The goal question metric approach," in *Encyclopedia of Software Engineering*. Wiley, 1994.

[13] W. Aalst, A. Adriansyah, A. K. Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. Brand, R. Brandtjen, J. Buijs *et al.*, "Process mining manifesto," in *Business Process Management Workshops*, 2012, pp. 169–194.

[14] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 1st ed. Springer Publishing Company, Incorporated, 2011.

[15] J. D. Weerdt, M. D. Backer, J. Vanthienen, and B. Baesens, "A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs," *Information Systems*, vol. 37, no. 7, pp. 654 – 676, 2012.

[16] SCAMPI Upgrade Team, "Standard CMMI appraisal method for process improvement (SCAMPI) a, version 1.3: Method definition document," http://www.sei.cmu.edu/library/abstracts/reports/11hb001.cfm, 2011.

[17] T. Sunetnanta, N.-O. Nobprapai, and O. Gotel, "Quantitative CMMI assessment for offshoring through the analysis of project management repositories," in *Software Engineering Approaches for Offshore and Outsourced Development*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, vol. 35, pp. 32–44.

[18] I. Garcia and C. Pacheco, "Toward automated support for software process improvement initiatives in small and medium size enterprises," in *Software Engineering Research, Management and Applications 2009*, J. Kacprzyk, R. Lee, and N. Ishii, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, vol. 253, pp. 51–58.

[19] N. Chen, S. C. H. Hoi, and X. Xiao, "Software process evaluation: A machine learning approach," in *Proceedings of the 2011 26th IEEE/ACM International Conference on Automated Software Engineering*. Washington, DC, USA: IEEE Computer Society, 2011, pp. 333–342.

[20] J. R. Curanas, "Process mining opportunities for cmmi assessments," http://www.tue.nl/en/publication/ep/p/d/ep-uid/276454/?no_cache=1, 2010.

[21] Accenture. Measuring and managing the cmmi journey using gqm. http://www.sei.cmu.edu/library/abstracts/presentations/Prasad-SEPG2006.cfm.

[22] L. Buglione, "Strengthening CMMI maturity levels with a quantitative approach to root-cause analysis," in *Proceedings of the 5th Software Measurement European Forum, Milan, Italy*, 2008.

[23] M. West, *Real process improvement using the CMMI*. Auerbach Publications, 2004.

[24] M. Khraiwesh, "Validation measures in CMMI," *World of Computer Science and Information Technology Journal*, vol. 1, no. 2, pp. 26–33, Mar. 2011.

Analyzing Version Control Open Source Software Survivability

Bee Bee Chua

University of Technology,Sydney

1 Broadway, P O Box 123

NSW 2000

Abstract— Survivability is a significant topic that is vital to the Open Source Software (OSS) community. User download is a metric suggested for measuring the user population size, use and download, to determine software popularity, successability and survivability. The size of the user download is significant as it can help determine a product's popularity and users' trust. This paper makes two important contributions to the literature: identification of low survival issues in relation to existing projects during survival growth and the significance of project type differences underpinning product age and survival impacts.

Keywords— OSS survivability, user download, survival issues, product age .

1. INTRODUCTION

Low development cost, short development time, quick bug detection and correction process, as well as code reusability are economical and operational reasons that justify why developers voluntarily develop and contribute to OSS. In general, developers are referred to as two groups: technical users and non-technical users [1, 2] .

Non-technical users are interested in using, learning, exploring and adapting new OSS functionality to support their work and research. In contrast, technical users want to learn new programming skills from others. Consequently, free virtual collaboration and cooperation from users motivates companies to shift their focus on software development from closed to open-based platforms [3, 4].

Increasing the market of users interested in downloading and using OSS can be challenging because in an open source market there are multiple OSS products of the same type to choose from and not all OSS is suitable for each user's environment, especially if the software does not satisfy the user's needs. Hence, the decision-making process to select appropriate OSS can be difficult.

For inexperienced users, product reviews and feedback can help them decide; for experienced users, decisions can be based on OSS popularity.

An increase in user download can not only increase users' trust relating to projects but can also help to boost business investors' confidence. From a service perspective, high user download means users are interested but also that support should be provided. It can also mean that users are satisfied with the software. Whereas, low and/or poor user download can mean the software has not received much attention from users and developers.

A high user download for a particular OSS means the product can perform and meet market demand, that is, it is able to attract user interest and meet the expected outcomes to satisfy those users. Low user download, on the other hand, is defined as a product that is unable to meet user demand, with a lack of performance in producing what is expected to satisfy users. In other words, high user download equates to a high chance of survival, success and popularity.

Unfortunately, a zero or low user download signals a risk to success, popularity and gaining user interest. To avoid no or low user download, an understanding of its effect on survivability is important. This understanding aids developers in developing better codes, acknowledging requirements for software, environmental characteristics, user demographics and logistics.

In this paper Section 2 discusses related work on user download effects, Section 3 presents our exploratory framework and Section 4 introduces the research methodology. In Section 5, results are presented and discussed and the last section updates our conclusions and future work.

2. RELATED WORK

User download is one of the metrics for measuring interest by users, developers and the wider community, and can predict OSS success and survival [5-10]. According to Stewart et al [11] and Grewal et al [9] the count of software and user download is therefore a good indicator for measuring user interest.

Wang [12] asserted that user download can measure user interest by analysing survivability effects. She confirmed the size of user download can be used to estimate the number of users but it is not a good indicator of individual group interest Unfortunately, Wang's work did not further explain user download impacts and effects on the relationship between target users, trust and experience.

There have been no studies confirming if low user download can negatively impact survival. The no or low user download metric is not sufficiently reliable to conclude OSS survivability effects, as survivability definitions vary especially in the context of social networks, community, contributors, artefacts, etc. [12-15].

According to Wang [12], every OSS has two periods of survival: initial survival and growth survival. Initial survival refers to initial hosting and the growth survival refers to continuing survival. Monitoring OSS longevity at this stage can be complicated as many factors can influence survivability. User download, developer download, network download, unrestricted licenses, programming language, project sponsorship, team coordination and communication have all been found to have an impact. [12, 13, 16]

Our prime focus in this research was to investigate whether there is a significant relationship between user download, project longevity and project type. Does a zero user download category only occur at project birth? And if not, which survival period can trigger the status of user download to zero? Further, under what circumstances would a zero user download happen during the different survival periods, initial or growth?

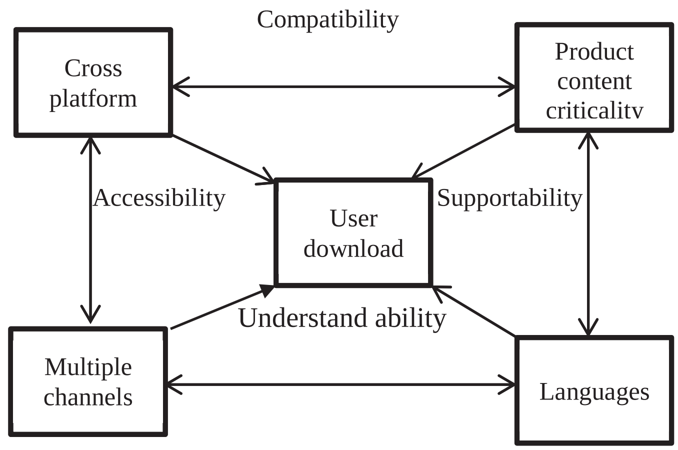
2.1 RATIONALE

Hosting is the first stage of OSS initial survival, also known as the 'project birth'. A new (birth) project takes time to attract and lock in users. Hence, it is common to find many birth projects without user download. The second stage is OSS growth, where user download is one qualifying variable that can be used as an input indicator for determining the performance of a particular OSS. Other reliable metrics that can be used for measuring OSS growth include license non-restriction, network size, developer size, programming language, etc. During growth survival, a zero user download is unusual but in this paper we present evidence showing zero user download can occur even after project birth.

3. FRAMEWORK

Our conceptual user survivability download (USD) framework is based on viewing the user download as one of the most influential critical success factors for successability and survivability in terms of usability characteristics of the software.

Figure 1: User Survivability Download Framework



The proposed framework shows how cross-platform, multiple channels, product content viability and languages can all impact on user download size. To define these terms: 'cross-platform' means different operating system environments at user locations; 'multiple channels' refers to users from different countries; 'product content viability' refers to the product's clear description and features; and 'languages' means both the programming language and native language. The four variables' inter-relationship shows functionality support, and the relationship between user download and the four variables means users base decisions on which OSS product to download on specific criteria. There is a higher chance of users downloading a product if: the content is well written; the content is viable to the user; the product is compatible and can support their platform, including having a flexible operating system that is easily accessible to users; and the product's native language can be easily understood by users.

4. RESEARCH METHODOLOGY

Our research process was divided into three stages: 1) pre-selection; 2) during selection; and 3) after selection. For pre-selection we first asked questions as to how many projects are enough for conducting survival analysis? Which type of projects will we investigate and why? And lastly, where can we gather project variables, and what variables must they be?

In terms of which projects we should include, we looked to previous work. Somaladas et al [15] used a large sample of projects to investigate survival analysis, particularly their evolution from website hosting to current. Their holistic survival analysis of a large sample did not include specific types of projects or describe individual projects. Grewal, Lilien and Mallapragada [9] also used a large sample of 108 projects and 490 developers to study network embeddedness. Stewart and Gosain [8] collected data from questionnaires completed by 67 project administrators.

Other examples of large sample sizes included reviews of OSS projects by Lerner and Tirole [17] and Subramaniam et al. [18]. The former conducted a cross-sectional analysis using data collected at a single time point and the latter conducted a longitudinal study using monthly data. However none of these studies conducted an analysis of user download as a measure of success at the product level.

Our approach differs from other researchers in that we focused on whether one project group from Sourceforge was a sufficient sample to conduct survival analysis and confirm if project type significantly influences product age.

We selected the version control category from Sourceforge and tested the input variable of user download. Version control was chosen because it is a code repository system used for maintaining OSS; therefore developers need tools to manage changes on the source code files. Open Source Version Control tools are important to developers for collaboration in their source code development.

In this paper, we measured each project size by the number of programming languages contributed by developers. In other words, a version control project that has more programming languages will have a larger developer base compared with a project with one programming language. Sixty-four projects had one or more than one programming language. They included versions of C programming language and non-C programming language, including Javascript, PHP, Python, Perl, Unix Shell, Delphi, Smalltalk, TCL, visual basic scripts and basic scripts. Among the sixty-four projects, fifty-four projects had one programming language and the remaining ten projects had multiple programming languages.

Developers use some version control tools to establish collaborations, for review and code management of open source and private development projects, as well as for code repositories. For example, Github is a popular choice of code repository system for developers. Concurrent Versions System (CVS), Git (GIT) and Subversion (SVN) are three types of revision control systems in Sourceforge.

For the second stage, OSS selection, the following procedures were followed:

1. Download 64 projects using Outwit Hub, a successful OSS (11 CVS, 8 GIT and 45 SVN). Figure 2 shows a breakdown of version product types by two categorisations; no user download and with user download.
2. Tabulate 12 variables related to user download, product age, product categorisation, product description, product feature, user interface, audience, user country, native language, programming language, operating system and unrestrictive license.
3. Apply the survival analysis, using Kaplan Miere on SPSS. [19]

Figure 2 Breakdown of Version Control Product Types

Project Type	No user download		With user download	
	(New (Birth) project released) in 2012	Projects released before 2012	(New (Birth) project released) in 2012	Projects released before 2012
GIT	0	1	2	5
CVS	0	1	0	10
SV (version control)	3	5	3	34

5. RESULTS

It is understandable and acceptable that birth projects have zero user download, as they are new. However it is surprising when zero download is found for an existing project, as was seen in this study, with three version control projects having zero user download after a period of 2 years, 4 years and 6 years respectively (see Graph 1).

Although these projects survived, it was difficult to analyse and predict survival growth and efficacy based on user download as an input metric, as some of them had no user download throughout the survival period (initial or ongoing). Some projects may be supported by one developer who contributed and developed codes only. Because of project size and complexity, we need to consider and use alternate metrics to measure growth for projects .

These projects had common survival issues. Users were local and not global, so the software was not written in English – this was a criterion outlined in our proposed framework as a potential negative influence on user download survivability. Additionally, the operating system interfaces of these three software were proprietary-based (QT) and the other one was written with TK; unfortunately these are not totally compatible with Windows, Linux or client-based systems. These findings confirmed our framework discussed earlier on the essential requirements for software design.

Our study results showed projects with long-term survival are correlated to global user software written in English and operating system that are proprietary. Short term survival projects correlated with factors such as local users, software not written in English and operating systems that are non-proprietary.

Graph 1 Project Age Survival Distribution

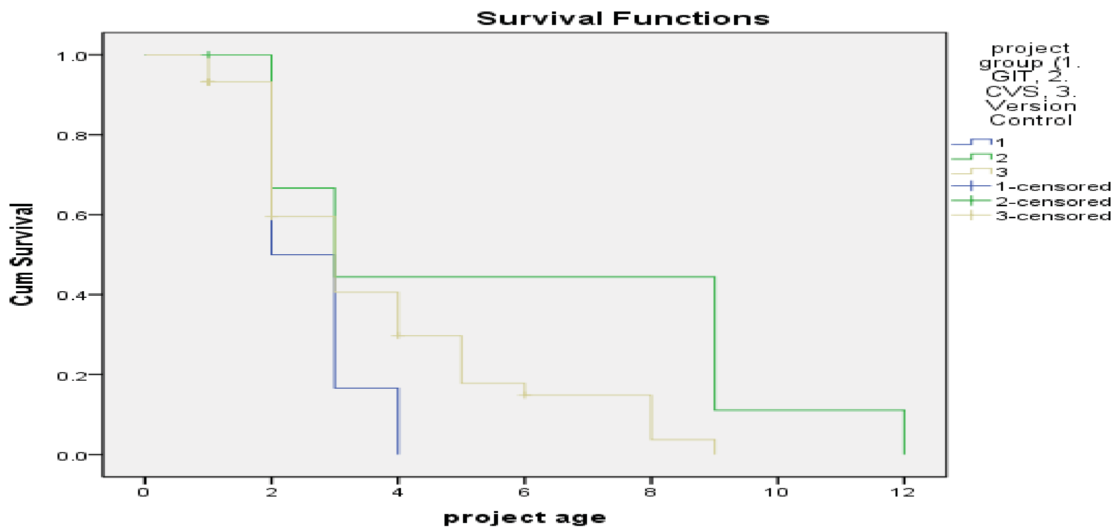


Figure 3 survival analysis distribution

project group (1.GIT, 2. CVS, 3.Version Control	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
1	2.667	.333	2.013	3.320	2.000	.	.	.
2	5.667	1.333	3.053	8.280	3.000	.745	1.539	4.461
3	3.744	.371	3.017	4.471	3.000	.421	2.175	3.825
Overall	3.981	.374	3.248	4.715	3.000	.316	2.380	3.620

a. Estimation is limited to the largest survival time if it is censored.

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	6.202	2	.045
Breslow (Generalized Wilcoxon)	2.007	2	.367

Test of equality of survival distributions for the different levels of project group (1.GIT, 2. CVS, 3. SV (version control))

Graph 1 shows results in answer to our research questions. A censored project refers to a project in which user download activity becomes inactive. In other words, the occurrence of zero user download. Figure 3 shows the results for the three types of projects. The results confirm there was a statistically significant difference between project age and project groups on survival distribution equality (P value < 0.045). In other words, if a software is still live after the hosting stage, it will survive. Using user download as the metric for measuring popularity and successability is acceptable; unfortunately, if it is used as an independent variable for determining survivability effects, we acknowledge that it can be useful only for checkpoints on finding causes.

Evaluating OSS survival growth based on user download is insufficient because there other variables can attribute to OSS survival rate, according to Wang [12]. Hence, to conduct a more robust good measurement of OSS survival growth, we suggest combining other variables with user download to strengthen evaluation impacts.

6. CONCLUSIONS AND FUTURE WORK

This study focuses on user download categorisation to explore projects' survivability effects. When reviewing project age and user download threshold, including a zero download category can assist in finding survival causes. In addition, our results show product age is a strong correlate factor on project type. These results lead to further work investigating OSS survivability as related to a product in a specific context. We could then debate whether generalising OSS survivability by relying on user download is an ideal option for predicting survivability and/or extinction. However we must not neglect other characteristics associated with project survival, including product design issues, market expectation and user interest group.

Our future work will also include validating these methods on nine other project categories in Sourceforge to determine low survival causes in relation to zero or low user download. Other user-oriented open source repository systems, like Fresh Meat and Ohloh, will also be considered for validation.

References

[1] A.H.Ghapanchi and A.Aurum, "Competency rallying in electronic markets: implications for open source project success," *Journal of Electron Markets,* vol. 22, p. 11, 2012.

[2] R. Sen, "A strategic analysis of competition between open source and proprietary software," *Journal of Management Information Systems* vol. 24, 2007.

[3] V. Krogh , *et al.*, "Carrots and Rainbows: Motivation and Social Practice In Open Source Development " *Management Information Systems Quarterly (MISQ),* vol. 36 pp. 649-676, 2012.

[4] O. Nov and S. Oreg, "Exploring motivations for contributing open source intiatives " *Journal of Computer In Human Behaviour* vol. 24, pp. 2055-2073, 2007.

[5] K. Crowston, *et al.*, "Information Systems Sucess in Free and Open Source Development : Theory and Measures " *Software Process and Practice,* vol. 11, pp. 123-148, 2006.

[6] V. Midha and P. Palvia, "Factors affecting the sucess of open source software " *The journals of systems and software* vol. 85, pp. 895-905, 2012.

[7] S. Comino and F. M. Manenti, "Government policies supporting open source software for the mass market," *Review of Industrial Organization,* vol. 26, pp. 217-240, 2005.

[8] K. Stewart, *et al.*, "A preliminary analysis of the influences of licensing and organizational sponsorship on success in open source projects.," *In Proceedings of the 38th HICSS,* 2005.

[9] R. Grewal, *et al.*, "Location, location, location: How network embeddness affects project success in open source projects " *Journal of Management Science* vol. 52, pp. 1043-1056, 2006.

[10] S. Y. T. Lee, *et al.*, "Measuring open source software success," *Omega, ,* vol. 37, pp. 426–438, 2009

[11] K. J. Stewart and S. Gosain, "The impact of ideology on effectiveness in open source software development teams," *MIS Quarterly,* vol. 30, p. 24, 2006.

[12] J. Wang, "Survival factors for Free Open Source Software projects: A multi-stage perspective," *European Management Journal* vol. 30, pp. 352-371, 2012.

[13] S.Chen, "Determinants of Survival of Open Source Software: An Empirical Study," *Academy of Information and Management Sciences Journal,* vol. Vol 13, p. 11, 2010.

[14] N. J. Choi and S. Chengalur-Smith, "An Exploratory Study on the Two New Trends in Open Source Software: End-Users and Service " presented at the Proceedings of the 42nd Hawaii International Conference on System Sciences, 2009.

[15] I. Samoladas, *et al.*, "Survival duration on the duration of open source projects " *Journal of Software and Information Technology* vol. 52, pp. 902-922, 2010.

[16] C.-S. a. Sidorova, "Survival of open-source projects: A population ecology perspective. ," in *In the 24th International Conference of Information Systems. Atlanta, GA*, 2002.

[17] J. Lerner and J. Tirole, "The Scope of Open Source Licensing," *Journal of Law, Economics, and Organization,,* vol. 21, pp. 20-56, 2005.

[18] C. Subramaniam, *et al.*, "Determinants of open source software project success: a longitudinal study.," *Decision Support Systems* vol. 2, pp. 576-585, 2009.

[19] E. L. Kaplan and P. Meier, " Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*vol. 53, pp. 457-548, 1958

Service Design for Cloud Services of Dental Clinics

Kang-Lin Peng
ICSE/ISIM
NCTU/MUST
Hsinchu, Taiwan
kp@must.edu.tw

Chia-Yung Lin
Dentistry Department
National Taiwan University Hospital Hsinchu Branch
Hsinchu City, Taiwan
chiayunglin@ntu.edu.tw

Yu-Chee Tseng
Dept. of Computer Science and Engineering
National Chiao Tung University
Hsinchu City, Taiwan
yctseng@cs.nctu.edu.tw

Jui-Yuan Tsai
Department of Information
National Taiwan University Hospital Hsinchu Branch
Hsinchu City, Taiwan
jytsai@hch.gov.tw

Ji Cheng
Dept. of Computer Science and Engineering
National Chiao Tung University
Hsinchu City, Taiwan
cs8mcg8m.cs01g@nctu.edu.tw

***Abstract*—Dental clinic could get new vision and better service through cloud techniques. This project applies the techniques of cloud computing and service design to enhance the convenience and value for dental clinics. A team from the field of Dental Clinics, Computer Science and Service Science cooperate to construct and improve the techniques and quality of cloud services for dental clinics. Through methods of System Developer Life Cycle (SDLC) Model and Service Experience Engineering (SEE), cloud dental clinic system could help dentists' diagnosis and treatment, in the meantime, patients can also get better dental care to fulfill the concept of value co-creation of service science, management, engineering and design (SSMED).**

Keywords- Cloud services; Dental Clinics; Service science; Service design.

I. Introduction

Traditional dental care process is time consuming for both dentist and patients. This research aims to apply the techniques of cloud computing and service design to make an efficient, quality, and valuable dental care system. Purposes of this research are to improve dental clinic and care service system; to promote interdisciplinary collaboration for value co-creation among hospital, patients, and university and other related entities; to enhance techniques of cloud computing and SSMED in dental clinics and research.

II. Literature

An approach for capturing services as an R&D object is presented in the following under the general heading of SSMED, service experience engineering (SEE) and cloud computing. SSMED is the principle, using methods from SEE, then conduct service design to cloud dental clinic systems as a case study for practice. Actor network theory applied as a theory basis in the research which tries to explain how material–semiotic networks come together to act as a whole; the clusters of actors involved in creating meaning are both human and nonhuman[1]. As a part of this it may look at explicit strategies for relating different entities together into a network so that they form an apparently coherent whole. These networks are potentially transient, existing in a constant making and re-making. This means that relations need to be repeatedly "performed" or the network will dissolve. Networks of relations are not intrinsically coherent, and may indeed contain conflicts. Social relations, in other words, are only ever in process, and must be performed continuously.

A. Service Science, Management, Engineering and Design (SSMED)

SSMED is the study of service systems, which aims to create a basis for systematic service innovation [2]. SSMED integrates diverse fields with an interdisciplinary approach aiming to study the service phenomena occurring in human society, and develop service systems toward a better society. Such an approach requires the collaboration of different disciplines, as well as of the government, academia, and enterprises for the purpose of service innovation. At the heart of service science is the transfer and sharing of resources within and among service systems. Four categories of resources have been noted and examined: 1) resources with rights, 2) resources as property, 3) physical entities, and 4) socially constructed entities [2]. This research aims at applying SSMED to services content for cloud dental clinic systems concentrate on dentists' orientation with help of technology instead of focusing on technology factors to build up service system. In perspective of human factors, our systems can be constructed to satisfy the real needs of users [3].

B. *Service Experience Engineering (SEE)*

Engineering customer experiences is an important strategy for establishing and maintaining customer preference for customer relationship management (CRM). It should, therefore, be in the capability portfolio of any firm investing in long-term customer relationships [4]. Along with service sector becoming the largest economic activities in many countries, it is getting important to improve the service quality and stimulate innovation through systematic methods. This research introduces and discusses a methodology called SEE for developing a new service. SEE describes the new service development framework completely from ideas creation to service market-launch. It divides the new service development into 3 phases: 1) FIND, includes consumer demand survey and technology observing research; 2) Innovation Net, includes two research focuses respectively on specific service-related industrial value chain and service modeling; 3) Design Lab, includes proof of concept, proof of service, and proof of business[5]. In fact, SSMED and SEE overlap in service design. SSMED offers a broad concept of interdisciplinary knowledge base, and SEE provides a sequence of methods that can apply to service design for designing services of intelligent vending machine[3].

C. *Service Design*

The service concept plays a key role in service design and development[6]. Service design is the activity of planning and organizing people, infrastructure, communication and material components of a service in order to improve its quality and the interaction between service provider and customers. The purpose of service design methodologies is to design according to the needs of customers or participants, so that the service is user-friendly, competitive and relevant to customers. The backbone of this process is to understand the behavior of customers, their needs and motivations. Service designers draw on the methodologies of fields such as ethnography and journalism to gather customer insights through interviews and by shadowing service users. Many observations are synthesized to generate concepts and ideas that are typically portrayed visually, for example in sketches or service prototypes. Service design may inform changes to an existing service or creation of new services.[7].

D. *Cloud Computing and Services*

A cloud is a type of parallel and distributed system connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service level agreements established through negotiation between the service provider and consumer[8]. Cloud computing is a large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the internet [9] and hand held devices. Cloud computing is a popular topic on software and distributed computing based on Internet, which means users can access storages and applications from remote servers by web browsers or other fixed or mobile terminals. Because the constrained resources of fixed or mobile terminals, cloud computing will provide terminals with powerful complementation resources to acquire complicated services. For cloud dental clinic systems, the cloud computing platform has to provide a long term offsite medical data archive solution [10]. New computational technologies allow users to "see" into and understand human‘s bodies with unprecedented depth and detail. As a result of these advances, biomedical computing help to produce exciting new biomedical scientific discoveries and clinical treatments [11].

From the beginning of the cloud concept till now, cloud services have been connected to concepts such as Software as a Service (Saas), Platform as a Service (Paas), Infrastructure as a Service (Iaas), Xen, Virtual machine (VM), Virtual Data Center (VDC), Virtual Cluster (VC), Microsoft Azure, Acadia-virtual computing environment coalition, etc. This research aims to explore these concepts apply to the dental clinic on the principle of SSMED so as to offer a better solution for dental care. Cloud services have advantages as a service platform such as flexibility in dynamic provisioning through ICT. The normative function of service systems is to connect people, technology and information through value propositions with the aim of co-creating value for the service systems participating in the exchange of resources within and across dental clinic systems.

III. METHODS

A. *System Developer Life Cycle (SDLC) Model for Cloud Dental Clinics Systems*

Dental implant scenario were applied to our case study. Dental implants are risky and time consuming for dentists and patients. In the meantime, implanted care and services are also important for patients. Thus, lower risk and better quality has become issues of dental clinics. Cloud computing and service design are brought together to enhance dental implant efficiency and service quality in this research. Research steps shown as follows: First, system analysis and SEE methods are applied to investigate problems of existed dental clinic. Second, cloud services system prototype was constructed based on investigation results of first step. Third, add-on Apps services are added to mobile devices for dentist and patients to try and adjust systems. Fourth, Trail systems were parallel to add and fix components in order to solve the occurrence of problems and reduce the failure possibility of the systems. Fifth, Interviews and questionnaires are used to identify the efficiency and quality of cloud dental clinic systems. Finally, the complete systems are applied to dental clinics between branches of National Taiwan University Hospital.

B. *Methods of Service Experience Engineering (SEE)*

SEE conducts two stages of processes to design dental services systems that meet dentists’ and patients’ real needs. The first stage is Service Experience Inquiry which contains three sub-processes, Contextual Inquiry, Working Model and Service Requirements Discovery, to discover the insight of users’ needs. Then follow by Service Design of second stage to deploy three sub-processes, Service Product Model, Service Quality Function Deployment (QFD) and Service Resource Model, to construct a cloud dental services system. By doing thess two stages of SEE, service oriented contents of cloud dental care system could be done (Fig.1).

This research is sponsored by joint projects of National Chiao Tung University and National Taiwan University Hsinchu Branch.

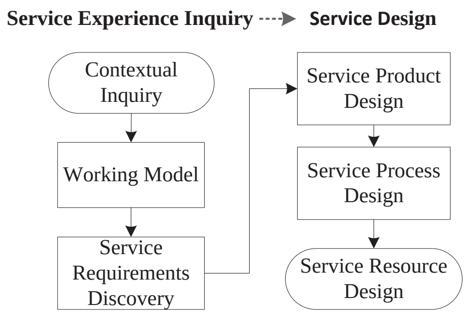


Figure 1. Methods of Service Experience Engineering

1) Service Experience Inquiry

a) Contextual Inquiry

Contextual Inquiry is a user-centered design ethnographic research method. A contextual inquiry interview is usually structured as an one-on-one interaction in which the researcher watches the system users do their normal activities and discusses what they act in systems. Service oriented contents can be assured by contextual inquiry to offer dental services for both dentists and patients, instead of engineering oriented design that can't really satisfy users.

b) Working Model

Working Model is an actual or proposed system that can do on a small scale of work which the system does or is expected to do. Five models are most used in working model; Flow Model is used to understand how system users interact with tangible systems and intangible services in operating process. Sequence Model draws the process of users' experiences in dental care systems. Artifact Model helps to realize users' intention in systems by auxiliary artifacts. Cultural Model is to detect the insight of users' needs in services. Physical Model finds layout or environmental influences to users. Prototype or trial system is essential to let users see or experience whether the functions satisfy their needs or not. Problems can be discovered and solved at this process.

c) Service Opportunities and Requirements Discovery

To develop every possible service opportunity, Brain Storming is one of frequent used method to create new ideas. Following by Affinity Diagram can be adapted to converge on the deep insight of users' needs among big data and creative ideas to capture validate items and service requirements for service design of the next stage of research.

2) Service Design

a) Service Product Design Model

QFD can be used to reveal the relations of service requirements, functions and stakeholders that help communication during the design period. Prototype of cloud dental care systems were constructed as a foundation to reveal users' real needs, then improve prototype gradually to complete cloud dental clinic systems.

b) Service Process Design

A mutual involvement design process can help both dentists and patients experience and adjust service contents to what they really want in systems. Service blueprint is a technique used for service contact. The technique shows processes within the dental clinic, divided into different components which are separated by end users, on stage, back stage and support process to clarify who is responsible for different interfaces of services. Roles of service provider offer the unique functions to satisfy users' needs.

c) Service Resource Design

Service Resources Design was to make services shown in QFD and service processes fulfilled. Resources can be departments, staffs, hardware and software to back up the operating of service design mentioned above.

To sum up, SDLC and SEE were used parallel to construct the cloud dental clinic systems which can offer better solutions for dental care. Systems were built for the real needs of dentists and patients to promote the quality of dental clinics.

IV. RESULTS

A. Service Experience Inquiry

1) Contextual Inquiry

To model dentists' behaviors in using cloud dental clinic systems, a prototype of system was constructed to do adjustments for better solutions of dental clinic. We inquired when, where, what, who, why and how dentists have used systems to deploy dentists' motivations, behaviors and experiences of using systems in a real working environment by interviews and observations, such that we can conduct service design of systems with the following process.

2) Working Model

Model sketch after contextual inquiry help the tangible of dentists' behaviors in using cloud dental clinic systems. For example, sequence model of dental implants in working model help to understand dentist's process in using cloud dental clinic systems as Fig.2. Service failures such as worse situation of dental implants have been found in sequence model, and then we proposed reschedule function for these service failure points to solve dentists' and patients' problems in using cloud dental clinic systems. Moreover, new and creative ideas were brought up to the development of service innovation in dental clinics.

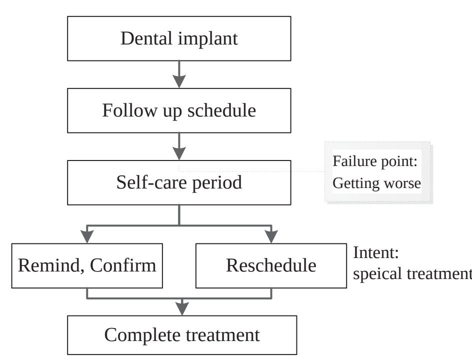


Figure 2. Sequence model in dental implant

3) Service Opportunities and Requirements Deployment

By using Brain Storming, all dentists' data, behaviors, service failure and solutions for the process of dental implant can be converged to requirements deployment, which are reminding patients, dental imaging and upload, rescheduling, image device and mobile app showed in Fig.3. To design cloud dental clinic systems to fulfill dentists' real needs, a better interface between dentists and systems should have more human-base interactions by dentists' instincts. Then more functions can be offered to satisfy dentists' expectations of cloud dental clinic systems for dental implant.

B. Service Design

1) Service Product Design

QFD was used to construct service product design (Fig.3) after first stage of service experience inquiry. QFD revealed the whole dimensions of dentists' requirements, competition analysis, service functions as well as service functions experience threshold which offered a systematic scenario to decide the essential functions of cloud dental clinic systems. Six service requirements are reveled in QFD shown in Figure 3. They are reminding patients for next term treatment, dental imaging and upload to cloud dental system for dentists' reference, rescheduling if emergency case happens, integrating dentists' different schedules for their convenience, imaging device for patients taking pictures for dentists' reference, as well as mobile app for functions mentioned above.

In addition, QFD shows that cloud platform and mobile App are essentials for service providing. Processing speed and user friendly can form service threshold that other dental clinics cannot compete. Compare to traditional clinics, our cloud dental clinic performed better on five service requirements except imaging device item.

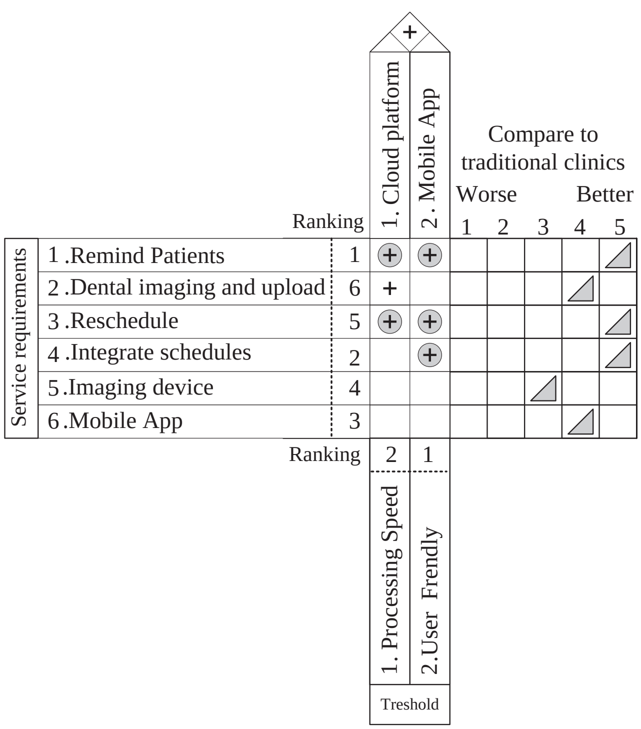


Figure 3. Quality function deployment

2) Service Process Design

Service process design was deployed by service blue print (Fig.4) to show interactions between four layers of dentist, on stage, back stage and support process. Details of service processes are filtered from cloud platform and mobile app with calendar in QDF. Service failure points illustrated in Fig.1 have been solved through the analysis of service blue prints to make systems process more convenient and intuitive for dentists.

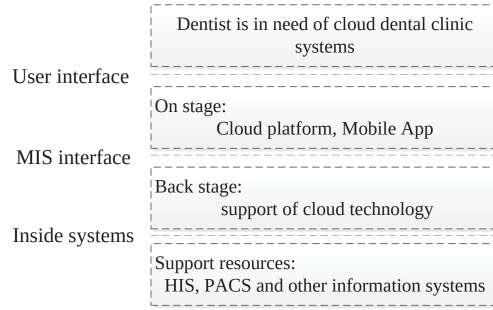


Figure 4. Service blue print of cloud services for dental clinic

3) Service Resource Design

Corresponding to QFD and service process model, service resource design made sure related departments, hardware manufacturing, software designer, researchers, consultants, and maintain staff can play their roles to offer services to dentists as shown in Fig 5.

	1.Dentistry Department	2.Information Management Office	3.Dept. of Medical Images	4.College of Computer Science	5.Platform manager	6.Software engineering	7.Machanical staffs	8.researchers	9.Consultant
1.Remind Patients	+	+				+	+	+	
2.Dental imaging and upload		+	+	+		+	+	+	+
3.Reschedule	+					+	+		
4.Integrate schedules	+			+		+		+	
5.Imaging device			+	+	+	+		+	+
6.Mobile App				+	+	+		+	+

+ means service resources to support service requirements

Figure 5. Service resource design

V. CONCLUSION

Results of our cloud dental care system can not only remind patient to come back his/her dental clinic for dental implant follow up, but also advice capture his/her own dental implant images for further diagnosis. Fig.6 shows a scenario. Firstly, user interface confirms patient's schedule of dental clinic. Secondly, service management server assesses and allocates medical resources for dentists' requirements, will be ready to communicate with picture archiving and communication system (PACS) and database in hospital information system (HIS). Thirdly, access PACS and HIS database through hospital VPN for further request. Fourthly, Service Management Server helps communicate between dentists and patients, if necessary, dentist can choose to inform specific patient to come back earlier for better treatment of dental implants. To sum up, dentists can do diagnosis and treatment through the help of this integrated system, patients can get a better dental care with good quality.

Our cloud techniques that can automate routine but manual and time consuming tasks would benefit dentists. An obvious gap between systems and dentists was closed when service design was conducted after our serials models, methods and tools. Service experience inquiry helped service design for dental image processing systems, which is convenient and efficient than traditional dental clinic services not only in cloud computing technology but also in comprehensive human factors-dentists' real needs. An approach for capturing services as an R&D object is made within this research under principles of SSMED, and methods of SEE to systematize the development of service contents which really meet dentists' requirements. Ways in which the service design for dental clinic systems can really benefit dentists' satisfaction with cloud technology development without neglecting humanity are demonstrated.

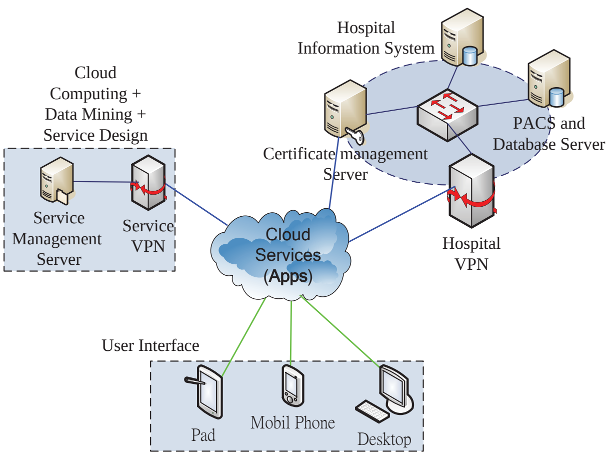


Figure 6. Cloud Dental Care Systems

ACKNOWLEDGMENT

This study is conducted under the collaboration of Nation Chiao-Tung University (NCTU) and National Taiwan University (NTU) Hospital Hsinchu branch which is subsidized by the joint projects of NCTU and NTU.

REFERENCES

[1] Wikipedia. Actor-network theory [Online]. Available: http://en.wikipedia.org/wiki/Actor_network_theory

[2] P. P. Maglio and J. Spohrer, "Fundamentals of Service Science," *Journal of The Academy of Marketing Science*, vol. 36, pp. 18-20, 2008.

[3] C. Y. Lin, K. L. Peng, J. Y. Tsai, Y. C. Tseng, C. C. Wu, M. Chen, *et al.*, "Apply Cloud Services of Medical Image Processing on Dental Field," presented at the International Association for Dental Research Meeting, Seattle, USA., 2013.

[4] L. P. Carbone and S. H. Haeckel. Engineering Customer Experiences [Online]. Available: http://www-935.ibm.com/services/th/igs/

[5] S.-L. Hsiao and H.-L. Yang, "A Service Experience Engineering (SEE) Method for Developing New Services," *International Journal of Management*, vol. 27, 2010.

[6] S. M. Goldstein, R. Johnston, J. Duffy, and J. Rao, "The service concept: the missing link in service design research?," *Journal of Operations Management*, vol. 20, pp. 121-134, 2002.

[7] Wikipedia. Service design [Online]. Available: http://en.wikipedia.org/wiki/Service_design

[8] R. Buyya, C. S. Yeo, and S. Venugopa, "Market-oriented cloud computing: Vision, hype, and reality for delivering it's services as computing utilities," presented at the The 10th IEEE International Conference on High Performance Computing and Communications, Los Alamitos, CA, USA, 2008.

[9] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-degree compared," in *Grid Computing Environments Workshop*, Austin, TX, 2008, pp. 1-10.

[10] W. Zeng, Y. Zhao, and J. Zeng, "Cloud service and service selection algorithm research," in *2009 World Summit on Genetic and Evolutionary Computation*, Shanghai, China, 2009.

[11] C. R. Johnson and D. M. Weinstein, "Biomedical computing and visualization," in *ACSC 2006 Proceedings of the 29th Australasian Computer Science Conference* Darlinghurst, Australia, 2006.

Ontologies in Global Software Development

Rodrigo G. C. Rocha, Ryan R. Azevedo, Sergio Mendonça
Federal Rural University of Pernambuco (UAG)
Garanhuns, Brazil
{rodrigo, ryan, sftom}@uag.ufrpe.br

Alex N. Borges, Catarina Costa, Silvio Meira
Federal University of Pernambuco (Cin - UFPE)
Recife, Brazil
{anbj, csc, srlm}@cin.ufpe.br

The distributed Software Development (DSD) has brought several competitive advantages in software industry, as well as new challenges such as communication and information sharing. In this context, the ontologies can provide benefits such as the definition, standardization and sharing of knowledge involved in the project, allowing a uniform understanding of information and facilitating the collaboration among distributed software development teams. This paper presents a systematic mapping study conducted in order to investigate which ontologies proposed for this context. This work presents evidences from each paper collected and an brief analysis of results reached. The results support the foundation for proposing and developing a feature based on ontologies to support the DSD. The Searches were performed both in manual and automatic way in a set of digital libraries engines and leading conferences in the Software Engineering field. The results support the foundation for proposing and developing a feature based on ontologies to support DSD, besides encouraging further researches that may promote advancements in this area and fostering the adoption these resources by the global software industry.

Empirical Software Engineering, Systematic Mapping Study, Distributed Software Development, Ontology.

I. Introduction

In DSD projects, the teams working are dispersed in different locations [1]. When it comes to projects with scattered teams, the limited communication and lack of sharing and knowledge cause some disadvantages: misinterpreted tasks, lack of collective consciousness relating to the work that has been developed, which issues have been brought up, difficulties following the project's plan and as to take place in a real-time discussion [2].

These challenges have encourage researchers to look for strategies that can help solve these problems, especially, the search for clear, effective information sharing mechanisms, that is essential in this environment. In this context, the use of ontologies as a standard for representing a domain's concept [3] can bring significant benefits to DSD projects, by allowing a simple share of information among dispersed teams.

This scenario motivated the conduction of a research to better understand how ontologies can give support to DSD projects and identify in what way this resource is being applied to this field. In this context, experimentation, which is a kind of empiric study, permits knowledge generation in a systematic, classifiable controlled way, which generates results with greater scientific value [4]. This paper proposes the execution of a systematic mapping study, which is another type of empiric study that is often applied in a research having a broader scope and when it is desired to find as many works as possible in the literature on a certain knowledge field [5]. Therefore, a systematic mapping is conducted to find out which ontologies have been formalized in the context of the software development in a distributed environment.

This article is organized as it follows: there is a description of methodology utilized and results found in this work on Section 2; on Section 3, there is the analysis of the results; and, finally, the concluding remarks.

II. Methodology and Results

In this research, a Systematic Mapping Study was conducted to identify ontologies supporting the DSD. And indirectly to identify tools, techniques, best practices, and models that use ontologies to support this area. An important issue in this process was to search for reviews and accurate analyses on the field, looking for current researches and open challenges related to the use of ontological resources in Distributed Software Development processes. Thus, the following research question were intended to be answered: "Which ontologies have been proposed or adopted in the context of DSD?"

The research questions were defined based on the scope of this mapping. An important issue in this process was to search for reviews and accurate analyses on the field, looking for current researches and open challenges related to the use of ontological resources in Distributed Software Development processes. Thus, the following research questions were intended to be answered:

Which ontologies have been proposed or adopted in the context of DSD?

The first step was to build a search string. Based on the research questions, we identified the main search terms and its synonyms. These definitions were based on other reviews and systematic mappings that involved the search terms: DSD and Ontology. Besides, the definitions were developed under the guidance of experts and researchers. The search string definition involved a testing phase, aiming to refine and obtaining the most appropriate string to the research objectives. The test phase was conducted using different versions of the string and performing automated searches in a few selected digital libraries, such as IEEE Digital Library and Elsevier Scopus. The first step of the test used a most comprehensive search string, composed of several terms related to ontologies and knowledge, such as: knowledge representation, knowledge management, knowledge transfer, conceptualization and concepts formalization. These searches returned several papers, but just a few were related to the ontology topic. Hence all the research terms related with Ontologies have been removed. Using a reduced version of the string, the amount of works returned decreased, however, all the relevant works to the research found in the first search have also been collected. Consequently, the tests performed with the reduced version of the string showed more efficient results. The resulting search string from this process is summarized in Table 1.

TABLE I. SEARCH TERMS

Population	(Ontology or Ontologies)
Intervention	AND (Distributed Software Development OR Global Software Development OR Collaborative Software Development OR Global Software Engineering OR Globally Distributed Work OR Collaborative Software Engineering OR Distributed Development OR Distributed Team OR Distributed Teams OR Global Software Team OR Global Software Teams OR Globally Distributed Development OR Geographically Distributed Software Development OR Offshore Software Development OR Offshoring OR Offshore Outsourcing OR Dispersed Team OR Dispersed Teams OR Distributed Software Project OR Multi-site Software Development OR Distributed Environment of Software OR Outsourced Software Project OR Virtual Team OR Virtual Teams)
Outcome	AND (Technique OR Techniques OR Method OR Methods OR Tool OR Tools OR Software OR Program OR Programs OR System OR Systems OR Model OR Models OR Framework OR Frameworks OR Methodology OR Methodologies OR Good Practice OR Good Practices OR Best Practice OR Best Practices OR Lesson OR Lessons OR Learned OR Success Factor OR Success Factors)

The search strategy used to map the primary studies involves automated searches through well-known digital library search engines. They were chosen based on the relevance for the computer science community, and the availability of papers on this field on the Internet or with libraries, which have partnership with the Federal University of Pernambuco. The search has been performed in the following digital libraries:

- ACM Digital Library (http://dl.acm.org);
- EI Compendex (http://www.engineeringvillage2.org);
- IEEE Digital Library (http://ieeexplore.ieee.org);
- Science Direct (http://www.sciencedirect.com);
- Scopus (http://www.scopus.com/home.url).

The search process also involved manual searches in conference proceedings in the research area. In this stage, the research considered some of the main conferences related to the subject, considered to be more relevant. The conferences defined to perform the manual searches were: International Conference on Global Software Engineering (ICGSE) and Workshop of Distributed Software Development (WDSD). Besides the research in conference proceedings, the manual search involved conversations with experts in DSD, resulting in the inclusion of some articles they found important.

Finally, in the search process, no limitation was imposed regarding to the initial period of publications. The final deadline for the publication of articles was December, 2011, date of the last stage of the search process. Regarding the manual searches, the conferences selected were surveyed from its first year of achievement until the 2011 edition. Studies published later in 2012 have not been considered in the research, in order to produce a more homogeneous result and also to allow a possible future update of this mapping study, which may consider publications from this date.

The searches for the primary studies were conducted according to the research plans defined in the protocol. The search process retrieved 1588 studies from the chosen scientific databases. The Table 2 summarizes the selection process results of the primary studies. The first column presents the digital libraries used in this study and the conferences where manual searches were performed. The second column represents the number of papers retrieved in the search process.

The third column shows the number of selected papers after the first step of selection process, which consisted of evaluating title, abstract and keywords to exclude studies clearly irrelevant to this search. The fourth, fifth, and sixth columns present the number of papers excluded after the second stage (selection process). Finally, the last two columns show, respectively, the number of papers included in the mapping and the percentage of inclusion by research source.

TABLE II. PRIMARY STUDY SELECTION PROCESS RESULTS

Source	**Search Results**	**Relevant Studies**	**Irrelevant**	**Repeated or Duplicated**	**Incomplete**	**Primary Studies**	**% Included**
ACM	236	37	26	7	2	2	5%
EI Compendex	53	21	5	16	0	0	0%
IEEE	550	122	88	9	6	19	50%
Science Direct	225	38	27	7	1	3	8%
Scopus	365	66	23	30	2	11	29%
Manual	157	15	11	3	0	1	3%

Experts	2	2	0	0	0	2	5%
Total	**1588**	**301**	**180**	**72**	**11**	**38**	**100 %**

Analyzing the Table 2, the EI Compendex looks the less efficient digital library, with the lowest number of papers returned. However, the EI Compendex provided a more accurate list of papers in comparison with other research sources. For being the last search performed, the 16 papers found in the EI Compendex base that were considered relevant to the study were excluded because they were duplicates, i.e., they were returned and accounted in other search engines in this research previously.

Other sources that have also been underperforming were the ACM and Science Direct, requiring even more work in the selection of studies. Despite the large amount of studies returned in these two sources, it can be observed the low number of selected articles from them. It is also interesting to notice that the IEEE and Scopus are the libraries that hold the largest number of published papers related to the research topic of this study, representing almost 80% (30 articles) of primary studies included combined. Besides the results of the automatic searches (92% - 35 articles), it can be observed that a few studies (8% - 3 articles) included in the research were found from manual searches or indications of specialists in DSD.

By analyzing Table 2 it is also possible to observe a small number of primary studies returned by search engines of the digital libraries and by manual searches when compared with other mappings and systematic reviews in the field of Software Engineering. This happens especially due the fact that the theme of this research is relatively recent, with many ongoing studies. Furthermore, of the 1588 papers returned in searches, only 38 were included in the research. Therewith is noticeable that the queries presented a considerable level of noise, since only 2,4% of returned studies were really relevant for the search. Many issues may contribute to this result, such as the use of an inappropriate search string or the inefficiency of automated search engines, as discussed by Kitchenham [10].

An analysis of the results of the quality evaluation process, shows that 13 out of 38 primary studies included in this research (34%) have been classified as Excellent, 12 (32%) as Very Good, 10 (26%) as Good e 3 (8%) as Regular, but none being considered Bad. The complete results of the quality evaluation are available at (http://www.rgcrocha.com/ease2013).

This systematic mapping did not restrict the period of publications, although all selected studies were published between 2001 and 2011. This evidences that studies involving the use of ontologies to support DSD are still recent. Hence, most studies (80% - 31 articles) were published between 2006 and 2011, which therefore portraits the relevance this particular topic has been acquiring recently.

The complete protocol is available at (http://www.rgcrocha.com/ms). This question aims to find out which are the ontologies normalized on the DSD context, i.e. To answer this research question, 4 ontologies have been found. The Table 1 presents the proposed ontologies in the distributed context. The first column presents the name and identifier from each ontology. The second column shows a description of each one.

TABLE III. ONTOLOGIES FOR DSD

Models	Description
OFFLOSC [PS10]	This ontology is formalized in the context of open-source software development communities. Its goal is help coordinate activities, management of resources and knowledge sharing. It is composed by 46 classes and describes the concepts related to open-source communities such as actors, artifacts, activities, operations, relationships and resources.
Knowledge Management Ontologies [PS18]	A set of ontologies that formalize structural concepts of DSD environments, directed to knowledge management. It describes concepts of software artifacts, environment problems, interaction among the distributed development teammates, infrastructure, business rules and general information of the project.
Open Source Communities [PS32]	This ontology is also formalized in the context of open-source software development and its main purpose is to compose a project knowledge basis having semantically related, categorized data, which allows the execution of semantic searches and data inferences by smart agents. It is composed of 6 classes that describe concepts of actor's relations, rules, activities, processes, artifacts and tools from open-source communities' projects.
OntoDISEN [PS35]	This ontology is formalized in the DSD Project scenario and is used to aid the establishment of communication between distributed teams. It is integrated to a textual information-spreading model, enabling sharing information in distributed environments to be comprehended by all the software engineers in a clear, homogeneous way. It describes concepts of elements that are represented and shared in a DSD environment, such as users, tools, other environments, activities and processes.

III. DATA ANALYSIS

This section presents an brief analysis of the results found in this study. Besides, it is of interest to notice that numerous another resources were found in this research. Like models, tools and best practices that use ontologies for better their activities. Each resource found is focused in a specific area from Software Engineering aiming to improve this area.

Based on results, it is evident that the development phases that are benefiting from the use of ontologies are: process, management, requirements and design. On the other hand, some important branches have not been fully approached, for example, quality and tests, which involves lots of information management activities, and may have a considerable evolution with the utilization of ontologies as means to standardize, manage and share knowledge.

By answering the research question from this mapping, there have been found four works that propose some ontologies especially developed for distributed software development, according to what is presented previously. Since these ontologies have been designed specifically for distributed teams, they bear the concepts and features required to work in this environment. Noteworthy to mention that two of the four ontologies were developed for open-source software development communities. The free dynamic nature of this environment poses challenges to the coordination of activities and knowledge sharing.

Therefore, the use of ontologies as a support to open-source software development simplifies the management of knowledge resources in the communities. Noticeable that several other works use ontologies to solve or mitigate challenges and in DSD environments, however, these ontologies are not specific for this environment. They are ontologies for Software Engineering, but if applied on DSD projects, they might help. The Table 4 depicts the several ontologies found in this study.

TABLE IV. LIST OF ONTOLOGIES

Field	Primary Study
Software Component	PS01
Business Domain	PS02, PS09, PS16, PS27, PS31
Software Engineering	PS02, PS06, PS07, PS08, PS21, PS28, PS30, PS38
Project Management	PS02
Problems and Solutions	PS02
Collaborative Structure	PS03, PS04, PS14, PS23
Team Division and Role Assignment	PS05, PS13
General Project Data	PS06, PS08, PS19, PS20, PS22, PS25, PS27, PS30, PS31, PS33, PS34
Software Requisits	PS09, PS16, PS24, PS29, PS31
Code and Bugs	PS14
Software Artifacts	PS17
Software Tests	PS26
Software Design and Architecture	PS27, PS37
Linguistics Services	PS34
Constitution of Electronic Contract	PS36

As seen on Table 3, there are numerous tools that utilize nonspecific-to-DSD ontologies only to mitigate challenges and limitations. These tools are distributed and used as support in the various project parts, from actual Software Engineering branches to specific project activities.

With these results, it is clear that there are a lot of advantages in using ontologies to support DSD, specially to generate solutions aiming at mitigating the communication, collaboration, knowledge flow management, coordination of project activities and knowledge, and process management issues.

IV. CONCLUDING REMARKS

The DSD work environments are very complex and there are no mature practices for this context since it is relatively new. In this sense, ontologies can bring benefits such as a shared understanding of information, ease of communication among distributed teams and effectiveness in information management.

This work presents evidences from each paper collected and an analysis of the results reached. The results support the foundation for proposing and developing a feature based on ontologies to support the DSD. This research aimed to identify ontologies formalized in DSD context and resources (models, tools, techniques and best practices) that use ontologies to support the DSD. Most studies have been being published from 2006 to the present. Through results, it is possible to affirm that ontologies were essential for some researches and some teams and projects already use tools based on ontologies aiming to establish information sharing and to improve the software development process as a whole. Is possible to view all the Sistematic Mapping Results in Borge's work [7].

From this work, some research can be developed: through of OOPS! Tool [8] (OntOlogy Pitfall Scanner!), is possible to detect some of the most common pitfalls of four ontologies found. Futhermore, development of an ontology to map all the DSD domain; presentation of solutions to assist project management in such an environment, proposing solutions to test process and software quality in DSD; and to indicate and use tools to support collaboration among distributed teams with the attachment of knowledge through ontologies.

REFERENCES

[1] Carmel, E. 1999. Global Software Teams – Collaborating Across Borders and Time-Zones. Prentice Hall, NJ, EUA.

[2] Wongthongtham, P., Chang, E., Dillon, T. S. 2004. Ontology-based Multi-agent System to Multi-site Software Development. In Proceedings of the Workshop on Quantitative Techniques for Software Agile Process (QUTE-SWAP). (Newport Beach, USA, Oct. 31 - Nov. 05, 2004). ACM, New York, NY, 66-75. DOI= http://doi.acm.org/10.1145/1151433.1151443.

[3] Gruber, T., A. 1993. Translaction Approach to Portable Ontology Specifications. Technical Report, Knowledge Systems Laboratory, Computer Science Department at Stanford University.

[4] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., Wesslén, A. 2000. Experimentation in software engineering: an introduction. Kluwer Academic Publishers, Norwell, MA, USA

[5] Kitchenham, B. 2007. Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report. Keele University at Staffordshire and University of Durham at Durham.

[6] Kitchenham, B. 2004. Procedures for performing systematic reviews. Technical Report. Keele University at Staffordshire.

[7] Borges, Alex., Rocha, Rodrigo., Costa, Catarina., Tomaz, Hilário., Soares, Sérgio., Meira, Silvio. Ontologies Supporting the Distributed Software Development: a Systematic Mapping Study. In Proceedings of the International Conference on Evaluation & Assessment in Software Engineering (EASE). Porto de Galinhas, PE, Brasil. 2013.

[8] Poveda Villalon, Maria and Suárez-Figueroa, Mari Carmen and Gómez-Pérez, A. (2010) A Double Classification of Common Pitfalls in Ontologies. In: Workshop on Ontology Quality (OntoQual 2010), Co-located with EKAW 2010 , October 15, 2010, Lisbon, Portugal.

Manitou: An Open Framework for Multimodal Interaction

Roman Hak, Tomas Zeman
Department of Telecommuniation Engineering, Faculty of Electrical Engineering
Czech Technical University in Prague, Czech Republic
roman.hak@fel.cvut.cz, zeman@fel.cvut.cz

***Abstract*—Mastering of multimodal input processing requires in-depth domain knowledge as well as extensive engineering and development skills. Moreover, time- and resource-demanding implementation of large infrastructure along with other supportive tools is necessary in order to provide desirable background, facilitate rapid prototyping and allow further research in the field. To address these requirements and offer solid basis and foundation for development of multimodal interfaces that is publicly available, we propose a multimodal interaction framework Manitou, which represents a feature rich and flexible system with advanced architecture, pluggable interfaces and extensive scripting capabilities. The framework also provides an additional web browser plugin exposing its functionalities to the web environment and related library for integration with Google Web Toolkit.**

Keywords-multimodal interaction framework; multimodal interaction; multimodal fusion; multimodal web; speech; gestures

I. Introduction

Traditional form of human-computer interaction – GUI and related WIMP paradigm – is far away from full utilization of natural and flexible manner of human communication. Multimodal interaction offers the solution by permitting our highly skilled and coordinated communication behavior to control computer systems in natural and intuitive ways. The key concept behind multimodal interaction is utilization of multiple communication channels working in concert to supply complementary information, increase robustness by partial redundancy, and provide more natural and transparent human-computer interaction [1].

Basic principles of multimodal interaction were first widely introduced and demonstrated in Bolt's seminal work "Put that there" [2] in 1980. Other multimodal systems became emerging in the 1990s. Necessary implementation of complex infrastructure, integration with external systems and devices, and associated in-depth knowledge of individual components makes further research and development in the multimodal interaction domain considerably difficult and demanding. To this end, a number of toolkits and frameworks have been proposed. ICARE [3], for example, presents a component-based platform for design and development of multimodal interfaces. Its direct successor, the OpenInterface framework [4], is trying to provide a tool for rapid development of interfaces with a multimodal input. Individual components are independent and can be chained into the *pipelines* constituting application data flow. Aforementioned systems can be classified to the group of frameworks with data stream-oriented architecture processing raw data sources and performing fusion on event-per-event basis (i.e. on *data level*). Another group comprises frameworks fusing multiple inputs on *decision level* typically consisting of multi-layered architecture. The representatives include the toolkit for rapid prototyping of multimodal interfaces HephaisTK [5] with versatile agent-based architecture allowing to employ various methods of multimodal fusion. The authors utilized that to extend the framework with testbed for benchmarking of fusion engines. The unified multimodal interaction framework Mudra [6] represents other similar solution. It provides three-layered architecture featuring the ability to integrate inputs from different levels of processing.

Many attempts to deliver a voice interaction to the Web were introduced in the past. For example, IBM and Opera Software developed XHTML+Voice[1] in effort to provide multimodal user interfaces. Unfortunately, the voice part – mostly derived from VoiceXML[2] – remains closely coupled with traditional telephony environments making it difficult to use speech in the context of applications with GUI. The WAMI toolkit [7] represents an effort to provide a framework for developing multimodal web interfaces. Components for voice interaction are deployed on a server while a support for input modalities is achieved via Java Applets on a client side. In contrast, the AT&T speech mashup framework [8] proposes several types of distributed web service architectures for speech processing. The data exchange is accomplished via HTTP streaming. Although it presents an interesting solution, a developer is required to have in-depth knowledge of the whole system, protocols and underlying technologies in order to use the framework.

II. Manitou Framework

Multimodal input processing submits a great challenge with complex tasks and multidisciplinary extent. Its mastering requires in-depth domain knowledge as well as extensive engineering and development skills. Moreover, time- and resource-demanding implementation of large infrastructure for individual modalities along with other supportive tools for multimodal input processing is necessary in order to provide desirable background, facilitate rapid prototyping of multimodal interfaces and allow further research in the field of multimodal processing. For these purposes, we propose Manitou representing open, feature rich and flexible framework with advanced architecture for multimodal input processing.

[1] http://www.w3.org/TR/2001/NOTE-xhtml+voice-20011221/
[2] http://www.w3.org/TR/2007/REC-voicexml21-20070619/

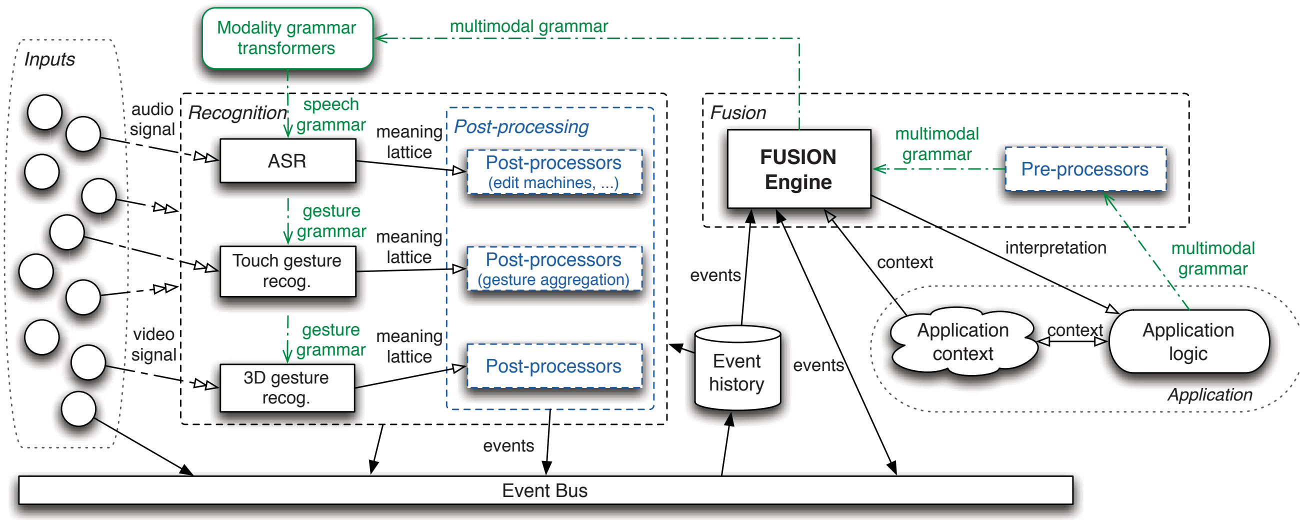


Figure 1. Basic architecture of the Manitou framework from the perspective of multimodal fusion

A. *Architecture*

An analysis of related work and existing techniques of multimodal fusion led us towards the design of advanced and flexible architecture. Fig. 1 illustrates its basic structure from the perspective of multimodal fusion. A detailed description of the most important parts is discussed in the following parts.

Input Processing & Recognition Processing initiates in input devices where diverse signals (acoustic, visual, etc.) and events (e.g. pressed key) are captured. The input data is then transferred to appropriate recognizer for processing. It involves extracting of relevant features and results in an assignment of the most probable meanings. The results are packed and wrapped into a relevant *event* object and sent to the *Event bus* or passed on for additional processing (*post-processing*).

Typical probabilistic character of interpretation assignment process that is performed during a recognition phase negatively influences performance and accuracy of the system. Robustness, on the other hand, depends strongly on precision level of a multimodal grammar and the ability of multimodal fusion subsystem to handle out-of-grammar inputs. Hence, various techniques for increasing robustness and accuracy are employed as a part of multimodal fusion, including for example edit machines or gesture aggregation methods proposed by Bangalore and Johnston in [9]. In order to provide infrastructure for robust-enhancing techniques the architecture of Manitou offers special components called *post-processors*.

Communication & Events A central communication platform across the framework is realized through the *Event bus*. Every important data flow within the system travels over this platform in the form of events. Individual components can observe, monitor and utilize information of events arising in different levels of processing. At the same time, components can create and sent events notifying about new facts obtained within their activity. Events traveling across the bus are monitored, logged and finally stored for their lifetime (i.e. until their expiration time exceeds) by the *Event history*. In concern of communication protocol standardization, the EMMA[3] [10] markup language is used to encode event data.

According to Sharma et al. [11] input data can be classified by the level of processing it arises in (*data*, *feature* and *decision*). The introduction of the central communication bus in combination with the event history brings a key feature of the framework facilitating fusion of input data from various stages of multimodal processing. It enables to process information from inputs with very different characteristics, structure and, the most importantly, data rate. The only other framework providing the corresponding function is Mudra.

Multimodal Fusion The most important stage in the multimodal input processing is a fusion accomplished by the *Fusion engine*. It performs final interpretation and meaning assignment. Within this activity the engine is responsible for determination of temporal relations between sequential and/or parallel inputs and distinguish whether they compose several consecutive unimodal events or one multimodal. In the latter case the inputs need to be appropriately combined.

The fusion engine is directly controlled by the *Application logic*, which provides a set of possible and expected inputs typically in the form of multimodal grammars (e.g. [12, 13, 14]). Various methods of pre-processing can be performed before the fusion engine parses the grammar. A common task of pre-processing is enrichment of the grammar leading towards greater robustness of the system and reduction of grammar complexity resulting in lower demands on its development. An example of such *pre-processor* within the framework is a component that searches the WordNet[4] lexical database for synonyms, which are then inserted to the original grammar as alternative tokens.

A number of techniques for the multimodal fusion process have been proposed. Their fusion strategies are based on one of the following approaches: *(a)* frame-merging [15]; *(b)* unification of feature structures [12]; *(c)* finite-state machines processing multimodal context-free grammars [13]; *(d)* statistical approaches [16]. Manitou – due to its advanced design – allows to implement and employ arbitrary of the methods. Currently, only a finite-state approach is provided out-of-the-box, yet we plan to implement other methods.

[3] http://www.w3.org/TR/2009/REC-emma-20090210/

[4] http://wordnet.princeton.edu/

B. Modalities

Speech In general, the speech processing is a very complex task with many challenges. Hence, Manitou relies on already delivered third-party technologies. Integration is addressed by facilitating an advanced plugin management system, which is described later. The current version provides a support for several external speech technologies, including PocketSphinx for recognition and Loquendo, Festival, Flite and others for synthesis. We have also integrated an MRCP[5] client into the speech subsystem. Thus, any speech technology achievable via this distributed way is ready to be directly used.

Gestures Quickly growing ubiquity of devices equipped with multi-touch screens has made gestures a very popular type of user input. Manitou comprises support for touch gestures as well as 3D gestures captured by novel depth and image sensors facilitating to track human body movement in real-time. In order to employ the sensors in multimodal interfaces the OpenNI[6] framework is integrated into the gesture subsystem. Currently, the framework provides several algorithms for gesture classification. The default and preferred one is based on the $1 recognizer [17], since its accuracy is superior to other classifiers and, more importantly, requires only a single training sample per gesture class, which also makes it a perfect candidate for employing in applications that allow users to define custom gestures. Manitou is supplied with a gesture model containing 58 predefined gesture classes. Moreover, a convenient graphical tool for collection and definition of large training sets has been created and is delivered as a part of the framework.

Geographical Location Geographical location of user (or device) can provide valuable information with a wide range of applications beyond location-based services. A component providing a location is delivered as a part of the framework.

C. Other Features

Plugable Interfaces An integration of existing technologies provided by other parties, which Manitou relies on, comes with practical and legal issues that must be concerned. The practical ones include resolving dependencies on different versions of libraries leading to compilation or linkage errors. The legal issues are associateed with different licensing strategies their incompa-tibilities. In order to address the mentioned obstacles the framework is equipped with an advanced plugin management system. Integration modules (plugins) are realized by imple-mentation of predefined interfaces and compiled into separate binary units in the form of dynamic libraries that are loaded at runtime. Hence, their sources and binaries are completely independent with ability to use different licensing strategies.

Scriptability Necessity to compile, link and deploy application, or some parts of the framework after any change in source codes causes significant slowdowns of development. It is crucial in the phase of rapid prototyping or when debugging, tuning up or slightly modifying an algorithm while observing its effects in runtime. Scripting languages provide a suitable solution since their codes are interpreted and evaluated *on the fly* without the need for time-consuming compilation and redeployment. To this end, scripting cores of the Python and JavaScript language have been integrated into the framework. Thus, individual parts and algorithms of the system can be implemented with scripts, which source codes could be modified at runtime.

Web Browser Plugin & GWT Binding The Web has become a primary and dominant environment for delivering of information and deployment of applications of all kinds. The arrival of HTML 5 delivers many new features and broader multimedia support. However, no significant changes were undertaken in the respect of user interaction. Engagement of novel and alternative modes of interaction in web applications is still extremely problematic. We address deficiencies in this field by exposing all key features of our framework to the web environment through a web browser plugin (implemented with Netscape Plugin API[10]), which enhances the DOM structure with new scriptable objects. Sample code snippet demonstrating the scriptable objects usage is listed in Fig. 2.

Enterprise libraries and tools closely linked to the Java programming language have evolved in a primary domain for creating robust web applications and systems. Thus, we decided to implement an extension library named *gwt-manitou* mapping all functionalities that are provided by the web plugin into Google Web Toolkit (GWT). The extension is mainly realized using JavaScript Native Interface (JSNI), which facilitates native calls from Java to JavaScript. By using *gwt-manitou*, researchers as well as practicioners are empowered with possibilities of new multimodal features in their web applications directly through GWT along with optimized, debugged and type-safe APIs. Moreover, the combination of GWT and Java provides robust debugging and profiling tools that are not available for JavaScript (or at least significantly less advanced and convenient).

1) Format Conversions and Transformations

The framework allows choosing from various formats for synthesis input (e.g plain text, SSML[7]), recognition grammars (e.g. SRGS[8], JSGF[9]) etc. However, underlying speech engines often support only a limited subset of them. To overcome these limitations, special transformations between the related formats were built into the framework. In cases where a format, which is not directly supported by the underlying engine, is used the appropriate transformation is automaticaly performed. Features that have no counterparts or their functionality is rather limited in other formats are completely ignored or converted to the nearest adequate replacement respectively.

```
<script ...>
  function initializePlugin() {
    var plugin = document.getElementById('plugin');
    window.speechManager = plugin.getSpeechManager();
  }
</script>
...
<object id="plugin" type="application/x-manitouplugin">
  <param name="onload" value="initializePlugin">
</object>
...
<script ...>
  var synthesizer = window.speechManager
          .getSynthesizer(voiceId);
  synthesizer.speak('Hello world.', 'text/plain');
</script>
```

Figure 2. HTML/JavaScript code demonstrating synthesis of plain text

[5] http://www.ietf.org/rfc/rfc4463/

[6] http://www.openni.org/

[7] http://www.w3.org/TR/2010/REC-speech-synthesis11-20100907/

[8] http://www.w3.org/TR/2000/NOTE-jsgf-20000605/

[9] http://www.w3.org/TR/2004/REC-speech-grammar-20040316/

[10] http://developer.mozilla.org/en-US/docs/Gecko_Plugin_API_Reference

III. Sample Application (ManiTutor)

Multimodal interaction offers a considerable potential especially in the domain of e-learning, since it employs more senses cooperatively. Various types of students in terms of their preferred mode of learning (auditive, visual, etc.) can be addressed more specifically, which leads to increased efficiency of learning and magnifying an interest of learning materials.

We have created an e-learning web platform ManiTutor for the purposes of multimodal education. The platform allows lecturers and tutors to create learning materials augmented with multimodal interaction in a simple and fast manner. Compared to traditional electronic courses the multimodal ones are enriched from the perspective of both inputs and outputs. The materials can be controlled by speech, touch and 3D gestures. The output is extended with synthesized voice, which is primarily involved in the sense of accompanying commentary.

An extensive statistical module composes another important part of the e-learning platform. Its main objective is to record, analyze and statistically evaluate all events and actions performed by a user or system during the work with learning materials. This evaluation provides indispensable feedback from the perspective of the system, interaction and control design, and learning material quality. Trends in usage of individual modalities and determination of common and divergent elements in users' behavior can be recognized based on the collected data together with identification of problematic parts of the system or materials that should be adjusted, modified, or simplified respectively.

IV. Conclusion

Manitou[12] with its focus, advanced architecture and other features is in many respects comparable to frameworks Mudra and HephaisTK. However, Manitou offers more sophisticated and richer infrastructure, modularity and extensibility in the perspective of support for individual modalities. A remarkable feature differentiating Manitou from other frameworks and toolkits is the exposure of the key features to the Web through the universal web browser plugin and related library for integration with GWT. This unique solution enriches present-day possibilities of the Web. Moreover, extensive scripting capabilities provide significantly efficient development as well as easier and convenient testing and debugging of multimodal algorithms. In combination with the aforementioned web-related features presents an advanced tool for rapid prototyping of multimodal interfaces. Notable attributes are also public availability[12] and multi-platform character of the framework, which allows deployment on traditional desktop environments as well as modern mobile platforms. Eventual incompatibilities with third-party technologies are addressed by the system of plugins offering both their source code and binary separation from the framework.

Our proposed framework provides a solid foundation for development of multimodal systems and opens up unique and widespread possibilities from the perspective of further research in the field. Although the current version is full featured, we are still continuing on its development. In the near future we plan to extend the framework with other implementations of multimodal fusion approaches. Afterwards, we will perform their detailed analysis, mutual comparison of key features and diverse aspects, measurements and evaluation of computational efficiency and memory consumption, robustness, scalability and degradation of accuracy depending on size and complexity of multimodal grammars. In the long term we will focus on further methods for increasing robustness of multimodal system and also on its automatic adaptation related to environment conditions and its changes in runtime.

[12] http://manitou.comtel.cz/

[13] source codes are available under the GNU LGPL 3 license

References

[1] S. Oviatt and P. Cohen, "Perceptual user interfaces: Multimodal interfaces that process what comes naturally," Communications of the ACM, vol. 43, no. 3, pp. 45–53, Mar. 2000.

[2] R. A. Bolt, "'Put-that-there'," in Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '80, 1980, vol. 32, no. 4, pp. 262–270.

[3] J. Bouchet and L. Nigay, "ICARE: a component-based approach for the design and development of multimodal interfaces," in Extended Abstracts of the 2004 Conference on Human Factors and Computing Systems - CHI '04, 2004, pp. 1325–1328.

[4] M. Serrano et al., "The OpenInterface framework : a tool for multimodal interaction," in Extended Abstracts on Human Factors in Computing Systems - CHI '08, 2008, pp. 3501–3506.

[5] B. Dumas, R. Ingold, and D. Lalanne, "Benchmarking fusion engines of multimodal interactive systems," Proceedings of the 2009 International Conference on Multimodal Interfaces, pp. 169–176, 2009.

[6] L. Hoste, B. Dumas, and B. Signer, "Mudra: a unified multimodal interaction framework," in Proceedings of the 13th International Conference on Multimodal Interfaces, 2011, pp. 97–104.

[7] A. Gruenstein, I. McGraw, and I. Badr, "The WAMI toolkit for developing, deploying, and evaluating web-accessible multimodal interfaces," in Proceedings of the 10th International Conference on Multimodal Interfaces, 2008, pp. 141–148.

[8] G. Di Fabbrizio, T. Okken, and J. G. Wilpon, "A speech mashup framework for multimodal mobile services," in Proceedings of the 2009 International Conference on Multimodal Interfaces, 2009, pp. 71–78.

[9] S. Bangalore and M. Johnston, "Robust understanding in multimodal interfaces," Computational Linguistics, vol. 35, no. 3, pp. 345–397, 2009.

[10] M. Johnston, "Building multimodal applications with EMMA," in Proceedings of the 2009 International Conference on Multimodal Interfaces - ICMI-MLMI '09, 2009, pp. 47–54.

[11] R. Sharma, V. I. Pavlovic, and T. S. Huang, "Toward multimodal human-computer interface," Proceedings of the IEEE, vol. 86, no. 5, pp. 853–869, May 1998.

[12] M. Johnston, "Unification-based multimodal parsing," in Proceedings of the 36th Annual Meeting on Association for Computational Linguistics - ACL '98, 1998, vol. 1, pp. 624–630.

[13] M. Johnston and S. Bangalore, "Finite-state multimodal integration and understanding," Natural Language Engineering, vol. 11, no. 2, pp. 159–187, Jun. 2005.

[14] B. Dumas, D. Lalanne, and R. Ingold, "Description languages for multimodal interaction: a set of guidelines and its illustration with SMUIML," Journal on Multimodal User Interfaces, vol. 3, no. 3, pp. 237–247, 2010.

[15] M. T. Vo and C. Wood, "Building an application framework for speech and pen input integration in multimodal learning interfaces," in 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, 1996, vol. 6, pp. 3545–3548.

[16] B. Dumas, B. Signer, and D. Lalanne, "Fusion in multimodal interactive systems: an HMM-based algorithm for user-induced adaptation," in Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems - EICS '12, 2012, pp. 15–24.

[17] J. O. Wobbrock, A. D. Wilson, and Y. Li, "Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes," in Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology - UIST '07, 2007, pp. 159–169.

Playing on large displays to foster children's interest in archaeology

Carmelo Ardito, Giuseppe Desolda, Rosa Lanzilotti
Dipartimento di Informatica
Università degli Studi di Bari "Aldo Moro", Bari, Italy
{name.surname}@uniba.it

***Abstract* — Information and Communication Technology can contribute to fostering a wider appreciation of archaeology. People tend to think that archeologists' work consists in the excavation and in the research of valuable elements without any scientific rigor. This paper presents an educational game for large multi-touch display aiming at developing a keen interest in archaeology in pupils, aged 8-12 years old. Playing with ArcheoGame, pupils discover the complexity of the archaeological investigation and of the overall archaeological research process. A formative evaluation showed that pupils found the game experience fascinating and engaging.**

Keywords: Archaeology, educational game, multi-touch display, educational format

I. Introduction

Since 2003, we have been involved in research projects investigating how Information and Communication Technology (ICT) can foster a wider appreciation of archaeology by offering tools able to increase awareness of its importance in cultural heritage domain. In Italy, school pupils constitute a large proportion of the visitors to archaeological parks. Thus, we have primarily developed educational games on different devices (i.e. mobile devices, large multi-touch displays), which support school pupils learning about ancient history through their active involvement during visits to such sites and follow-up activities at school [1][2][4][9].

The field studies performed to validate the educational games revealed that students enjoyed a lot playing with the systems, which stimulate very much their curiosity about the park and its places of interest. During the studies, students often asked to their teachers how such places were identified, their origin, their use at the ancient time, etc. In some cases, teachers were not able to explain in details what is the process to discover the ruins and archaeologists were not presented during these visits, so that it was not possible to answer to students' questions. As a consequence, we thought it would be useful to illustrate the archaeologists' work. Current multimedia technologies provide several possibilities to support pupils in their learning activities. In [3], we presented a documentary video that, through multimedia resources, such as images, videos, 3D reconstructions, sounds, informs people about the archaeological investigation process.

This paper goes on in that direction and describes a new educational game, called ArcheoGame, to further stimulate pupils' interest in archaeology. Simulating an archaeological investigation process, ArcheoGame allows pupils to discover the scientific value and complexity of the excavation process. Our aim is to combine this game with the above mentioned documentary video in a new educational format we have developed [4]. The educational format is inspired by Bruner's Discovery Learning technique [6]. Bruner states that learning is facilitated by presenting new information through three different types of representation, emphasizing the involvement of pupils in practical activities that allow them to discover as well as apply new concepts.

The paper has the following organization. Section II describes the archaeological investigation process. Section III summarizes the educational format we proposed. Section IV briefly reports the documentary video and Section V presents the ArcheoGame. Section VI illustrates the results of a formative study performed to get feedback by pupils about the design and the usability of ArcheoGame. Section VII concludes the paper.

II. The archaeological investigation

Archaeology is a complex discipline that investigates past human cultures and civilizations and their relationship with the environment, through the collection, documentation and analysis of the traces left during the time. People tend to think that archeologists' work consists in the dig and in the research of valuable elements without any scientific rigor. But the dig is only one of the phases of a complex scientific process, called *archaeological investigation*. It consists of three main phases: 1) archaeological site identification, 2) excavation, 3) historical interpretation. The archaeological site identification phase aims at locating the areas where past human civilizations left their tracks. Archaeologists define areas to look for tracks through appropriate techniques, which allow archaeologists to formulate hypotheses about the nature and appearance of the site going to dig.

The excavation phase represents the central part of the archaeological investigation. In this phase, archaeologists collect and organize the unearthed tracks. Archaeologists provide documentary evidence of the findings and produce a stratigraphic sequence, which graphically shows the development of the archeological stratification.

Finally, during the third phase of the process, in which the historical interpretation of the findings occurs, archaeologists reconstruct the "life" of the site. By adopting specific dating methods, archaeologists understand the history of the findings by referring them to the culture and habits of the ancient civilizations that populated that site.

III. THE EDUCATIONAL FORMAT

Technology can play an important role in the children's learning paths. In particular, games based on ICT can be considered a valid support for effective learning, since they are able to capture pupils' attention and engage them in their learning activities [7][11][13][14][15]. Based on these research results, we have defined a new educational format that, exploiting the educational value of ICT and games, integrates formal learning (traditional classroom lessons) with more informal and technology-based learning.

The educational format has been inspired by the Discovery Learning technique defined by Jerome Bruner in his pedagogical theory of Constructivism [6]. Bruner considers learning an active process involving people in which information goes through three different types of representation, namely *symbolic*, *active*, and *iconic*, based on languages, actions, and images, respectively. In the symbolic representation, information is mostly in the form of words or other symbolic systems (e.g. mathematical, musical symbols, etc.). In the active representation, learning occurs by carrying out physical activities (e.g. manipulating objects, working in laboratory). In the iconic representation, visual, auditory, olfactory, or tactile images are exploited to illustrate information.

In the educational format, we have proposed in [4], the three types of information representation can be enhanced by ICT. Pupils get new information by: 1) attending the lesson(s) by their teacher in the classroom during which teacher can use both traditional and electronic materials (*symbolic phase*), 2) acting in a real/virtual context (*active phase*), and 3) using visual, auditory, olfactory, or tactile images (*iconic phase*).

In order to foster a keen interest in archaeology in school pupils aged 8-12 years old, we have designed and developed two technological educational tools to be used in the three phases of the educational format: a documentary video and an educational game. The documentary video illustrating the archaeologists' modus operandi can be used by teacher in the symbolic phase and by students in the iconic phase. Specifically, during the symbolic phase, pupils get new information by attending lesson(s). The teacher introduces pupils the work of archaeologists by adding to her classical material the documentary video [3]. A LIM or a PC connected to a projector can be employed. During the lesson, teacher and pupils discuss and comment on the notions the documentary video provides. During the symbolic phase performed with the support of the documentary video, the iconic phase also occurs. In fact, pupils acquire information through the multimedia objects (images, sounds) provided by the video.

The educational game, called ArcheoGame, can be used by pupils both in the active and iconic phases. In fact, during the active phase pupils, acting as archaeologists, virtually dig an excavation; for example, they have to define areas to be dug, perform the excavation, reconstruct the unearthed remains. In the iconic phase, ArcheoGame, running on a multi-touch display, allows pupils to interact with visual, auditory, and tactile images. The following two sections describe the documentary video and ArcheoGame, respectively.

IV. THE DOCUMENTARY VIDEO ON THE ARCHAEOLOGISTS' WORK

This section briefly describes the documentary video, which through images, videos, 3D reconstructions, sounds, informs people about the scientific value of the archaeological investigation and the complexity of the overall archaeological research process. Its detailed description is illustrated in a paper published in the proceedings of DET 2010 [3]. The documentary video describes the archaeologists' modus operandi through a dialogue between two avatars, Gaius and Miranda. The choice of this type of presentation is based on the fact that it is well known the advantage of the animated characters in the learning process [12], especially if a team of characters are involved [5].

Gaius is a citizen of the Roman Empire, who lived in Egnathia in the second century AD. In 2013, he comes back to his city. Here, Gaius meets Miranda, a young archaeologist involved in the digs of the Egnathia archaeological park. Miranda drives Gaius through the remains of the ancient Roman city showing what archaeologists have unearthed and illustrating how the process has occurred. Gaius and Miranda virtually visited the interest places of the park. When they reach a place, Miranda illustrates the difficulties of her work in discovering that particular place asking Gaius confirmation of conjectures, which she and her archaeological team have made during the excavation process. Gaius is interested in understanding how archaeologists work and confirms or declines interpretations and reconstructions of the archaeological findings illustrated by Miranda. In this way, Gaius helps Miranda rebuild the life of the ancient Roman city. Finally, a 3D reconstruction of the visited place is shown and contextual sounds reproducing the noises of the ancient daily activities are played e.g., carts running on the paved Trajan Way, lawsuit announcements in the Civil Basilica, cows lowing in the Foro Boario, etc.

V. THE "ARCHEOGAME"

ArcheoGame is an educational game for large multi-touch displays, designed for pupils of 8-12 years old. Its aim is to allow children to understand the complexity of the overall archaeological research process, the work of the archaeologists during the excavation, the tools they use, etc.

The game simulates an archaeological investigation process. It is organized in three main phases: 1) the introduction, 2) the excavation, 3) the composition of the unearthed remains. During the introduction phase, an avatar, representing a professor of archaeology, illustrates the various steps of the archaeological investigation process, e.g the analysis of the historical sources of a geographic area, the observation of aerial photos, the identification of the specific area to be dug. The avatar explains such steps in different screens. To go to the next screen, pupil has to touch the "Forward" ("Avanti" in Fig. 1) arrow, while touching the "Back" ("Indietro" in Fig. 1) arrow, pupil can return to the previous screen. Pupils can read the avatar's explanation in the thought balloon displayed on the screen.

Fig. 1. The archaeological site identification phase.

After the introduction, the excavation phase starts. Initially, pupils have to identify the area to be dug. On the right side of the screen ("Strumenti" in Fig. 2), there are the most important tools archaeologists use during a dig. Pupils start the excavation selecting the radar from the tools menu to scan the virtual map and discover the area to be dug. This area is identified when the portion of the screen below the radar changes its colour from red to green. At this point, pupils have to drag over it an excavator to start the dig. During the excavator movements, the ground changes its aspect giving the feel that the tool is working correctly.

The avatar follows all excavation phases by suggesting pupils the order in which the tools have to be used. For example, after the excavator, pupils have to use a shovel, a pick, a trowel and brush. In fact, if pupils try to use the excavator, when the dig is practically finished, they can damage the unearthed remains. Conversely, if they try to use the brush in the initial phases, he can't quickly remove the ground.

Every time pupils use a tool they have to take a picture to document the excavation, like archaeologists really do. These pictures are collected in the work diary and pupils can view them in every moment.

Fig. 2. The excavation phase.

When all remains have been found, the third phase begins. Pupils have to compose the unearthed remains in a puzzle-based game (Fig. 3). The puzzle is composed of 9 tiles, which pupils have to correctly positioned to discover the object they have discovered during the previous phase. At the end of the puzzle resembling, pupils can choose to start another excavation process or finish the game.

Fig. 3. Pupils are resembling a puzzle during the formative evaluation.

VI. EVALUATING ARCHEOGAME

A formative evaluation was performed in order to receive students' feedback about the design and the usability of the game prototype. The evaluation involved 4 pupils (3 males), aged 8-10 years old. It was carried out in a university laboratory. The ArcheoGame was installed on a 46'' multi-touch display.

A training session, lasting 20 minutes, introduced the participants to the game and its rules. Then, pupils were free to interact with the game in order to familiarize with the new technology (i.e. the multi-touch display) and the game itself. After this session, the four pupils were divided into two groups. Each group independently played with ArcheoGame. Each evaluation session lasted about one hour, in which each group completed the game twice carrying out 4 excavations and resembling 2 puzzles. At the end of this session, pupils were interviewed. The pupils were also given an assignment to write an essay about the experience they have had.

All pupils behaved in a similar manner: they showed great interest and involvement related to both the used technology and the game. Initially, they tangled to establish who among them has to begin to play. Fortunately, they soon developed various strategies to share tasks they have to accomplish to complete the game. They showed a strong sense of collaboration, helping each other.

Three problems emerged. The first one is related to the technology used. In fact, the major interaction difficulty pupils faced with is related to the dragging of objects shown on the display. The screen is not very sensitive either at the margins or in a central strip about two centimeters wide, making it difficult to interact with objects displayed there.

The other two problems are related to the ArcheoGame interface. The instructions communicated by the avatar are visualized at the top left of the display, far from the area where pupils dig. Initially, pupils did not notice the instructions and were disoriented about what to do next. In addition, the radar tool generated some confusion: while in the menu all the radar device is visualized, once it is selected only its display is shown in the digging area.

From the analysis of both the interview performed at the end of the game and the essays, it appeared that the pupils were satisfied by playing on the multi-touch display. The pupils found the game experience fascinating and engaging. The pupils also commented that they would have liked to dig more areas and to compose more puzzles presenting different difficulty levels.

VII. Conclusions

Over recent years, archaeologists have come to realize that the objective of developing a strong consensus on the importance of fostering public interest in archaeology has been only partially met [8]. This paper has presented an educational game, called ArcheoGame, for large multi-touch display to foster pupils' interest in archaeology.

The ArcheoGame has been combined with a documentary video, we proposed at DET 2010 [3], in a new educational format, inspired to Bruner's Discovery Learning technique, in which information goes through three different types of representation, namely symbolic, active, and iconic, based on languages, actions, and images, respectively. The three types of information representation can be enhanced by technology.

A formative evaluation on the educational game revealed that pupils are satisfied in interacting with a multi-touch display. In addition, pupils found the game experience fascinating and engaging. We are now planning a more in-depth study to get more objective data through which validating the learning effectiveness of the educational format and the overall user experience.

Acknowledgment

Partial support for this research was provided by Italian MIUR through grant "L4ALL". We thank Professor Maria Francesca Costabile for her valuable guidance in our research. We are also indebted to all the people involved in the participatory design, in particular the archaeologists of the Department of Ancient History of the University of Bari.

References

[1] C. Ardito, M.F. Costabile, A. De Angeli, and R. Lanzilotti, "Enriching archaeological parks with contextual sounds and mobile technology," ACM TOCHI, 19, 4, Article 29 (December 2012), 2012, 30 pages.

[2] C. Ardito, M.F. Costabile, and R. Lanzilotti, "Gameplay on a multitouch screen to foster learning about historical sites," in Proc. of AVI 2010, ACM Press, New York, NY, USA, 2010, pp. 75-78.

[3] C. Ardito, and Lanzilotti R., "How do archaeologists work? A multimedia digital object to learn their investigation process," in Proc. of DMS 2010, Knowledge Systems Institute, Skokie, Illinois, USA, 2010, pp. 206-210.

[4] C. Ardito, R. Lanzilotti, M.F. Costabile, and G. Desolda, "Integrating traditional learning with educational games on large multi-touch displays," Intern. J. of Educat. Tech. & Soc., 16, 1, 2013, pp. 44–56.

[5] G.H. Bower, et al., "Hierarchical retrieval schemes in recall of categorized word lists," Journal of Verbal Learning and Verbal Behavior, 8, 1969, pp. 323–343.

[6] J. Bruner, "Acts of meaning," Harvard University Press, Cambridge, MA, 1990.

[7] J.S. Cabrera, H.M. Frutos, A.G. et al., "Mystery in the museum: collaborative learning activities using handheld devices," in Proc. of MobileHCI 2005, ACM, New York, NY, USA, 2005, pp. 315-318.

[8] T. Copeland, "Presenting archaeology to the public", Routledge, 2004, pp. 132-144.

[9] M.F. Costabile, A. De Angeli, R. Lanzilotti, C. Ardito, P. Buono, and T. Pederson, "Explore! Possibilities and challenges of mobile learning," in Proc. of CHI'08, ACM, New York, NY, USA, 2008, pp. 145-154.

[10] E. Klopfer, S. Osterweil, and K. Salen, "Moving learning games forward", Massachusetts Institute of Technology, 2009

[11] R. Lanzilotti, and T. Roselli, "An experimental evaluation of Logiocando, an intelligent tutoring hypermedia system," International Journal of Artificial Intelligence in Education, 17, 1, 2007, pp. 41-56.

[12] R. Moreno, R. Mayer, and J. Lester, "The case for social agency in computer-based teaching: do students learn more deeply when they interact with animated pedagogical agents?," Cognition and Instruction, 19, 2, 2001, pp. 177-213.

[13] S.L. Oviatt, "The future of educational interfaces," Routledge Press, 2012.

[14] Y. Rogers, S. Price, C. Randell, et al., "Ubi-learning integrates indoor and outdoor experiences," Communication of ACM, 48, 1, 2005, pp. 55–59.

[15] R. Van Eck, "Digital game-based learning its not just the digital natives who are restless," Educause Review, 41, 2, 2006.

Proceedings
International Workshop on
Distant Education Technology
(DET 2013)

A web-based phonetics tutor using generative CALL

Greg Lessard, Michael Levison, Jonathan Reid, Boyang Fu
Queen's University
Kingston, Canada

***Abstract*—One of the skills required of apprentice linguists is the ability to identify, analyze and represent the sounds of various languages. This includes phonetic transcription, notation of syllable structures, minimal pairs, and intonation, among others. We show how an existing environment for generative Computer Aided Language Learning known as VINCI may be used to aid in teaching these various skills in a web environment, providing both detailed diagnostics and multimedia output.**

Keywords: phonetic transcription; web-based teaching; generative computer-aided language learning; multimedia materials

I. Introduction

One of the more difficult skills to acquire for apprentice linguists is the identification, representation and analysis of phonetic information. In the traditional teaching of phonetics, students are often given words or utterances expressed in the regular alphabet, asked to imagine them being pronounced, and then asked to provide a phonetic transcription or other analysis.[1] Unfortunately, this approach has the weakness of introducing orthography into the equation. It also adds unnecessary variation based on the student's own dialect, and it replaces the important empirical element of a student's listening to sounds produced by others, in various languages, with listening only to oneself.

The advent of the Internet, with its multimedia resources and tools, offers the opportunity to advance beyond this state of affairs. We propose here a means of providing practice materials for phonetic transcription, as well as several other areas of phonetics. The proposed system includes three central elements: audio and visual materials replace a student's own pronunciation; questions, answers and diagnoses are generated by rule; and the use of diagnostic rules permits a detailed analysis of student responses and subsequent production of remedial materials.

II. Some general principles

Over the half-century of its existence, Computer Aided Language Learning (henceforth, CALL) has seen an evolution from a behaviourist to a constructivist approach. The former focused on skill acquisition, individualized instruction in a task-based framework, and a prescriptivist perspective, while the latter has shifted the emphasis to authentic, sometimes complex tasks, the potential for collaborative learning, and construction of knowledge ([1], [2]). At the same time, there has been an increasing emphasis on providing more varied input to learners, including visual, auditory, and spatial ([3]). Thirdly, as Chapelle [4] has shown, there is value in providing learners with enhanced input, including making more important items more salient. To this she adds the importance of interaction, where the learner navigates through learning materials. The model proposed here draws on a number of these principles. In particular: the tasks proposed for learners are multimodal, based on problem solving skills, and can be undertaken either individually or collaboratively; we use parallel sources of output, including aural, textual and graphical; and we have sought to provide an environment in which learners feel that they are exploring a problem rather than moving lockstep through a series of fixed questions.

III. Generative CALL and the VINCI system

CALL materials have tended to adopt one or the other of three general approaches. In the simplest case, a learner is presented with a finite number of questions drawn from a closed set, a fixed set of possible answers is provided, in multiple-choice or other formats, and learner responses are evaluated with respect to this set. At the other end of the spectrum, in the case of what is sometimes referred to as Intelligent CALL (ICALL), learners may be presented with open-ended questions or data and their responses parsed and reacted to in a human-like fashion. (See for example [5], [6], [7].) We propose here an intermediate approach, **generative CALL**. As an alternative to the hand coding of canned exercises, a grammar is developed which generates questions, answers, and potential diagnoses. Such an approach is of particular value in areas like lexical creativity, where predefined lists fail to capture productive rules [8]. At the same time, following principles for intelligent tutoring systems (see, for example, [9]), we make use of an adaptive approach, in which student responses are used to determine which questions should follow (see [10], for example). These approaches are embodied in a system called VINCI. For details, see http://cs.queensu.ca/CompLing.

VINCI is based on the interaction of a number of modules. Attribute classes and values carry grammatical and semantic information used by the system, ranging from gender and number to semantic classes such as 'human', 'physical object' and so on. Terminal classes define parts of speech (noun, verb, etc.). The syntax defines structures to be generated, while transformations build upon these to produce parallel questions,

---

[1] An example with both written and oral output can be found at http://davidbrett.uniss.it/phonology/vowel%20sounds/phonemic%20transcription/transcribing_phrases3/phonemicTranscription.html, but it includes no diagnostics.

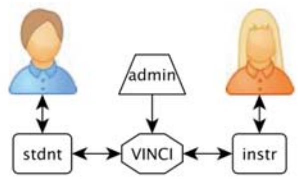


Figure 1 The web-based elements of VINCI

answers, and expected errors. These various specifications are passed through a lexicon, which produces the appropriate words, and these words are inflected by mean of morphology rules. VINCI allows for parallel morphology outputs, so that the same generated utterance may be output as an orthographic string, as a sequence of phonemes to be turned into sounds (see details below), or as an image file. In addition, a set of error analysis routines seek to determine whether incorrect forms produced by a student correspond to forms which might be generated by rule, thus providing a diagnosis of the possible student strategy which led to the error.

VINCI may be used either stand-alone, or in a web-based framework as shown in Fig. 1. In the latter case, VINCI itself does all the work of linguistic production and analysis. Between it and an actual user sits a module called **stdnt** which takes output from VINCI and renders it to the web interface as well as accepting input from the user, in the form of typed text, mouse clicks or other actions. These are preprocessed and then sent to VINCI. Behind VINCI sits another module called **instr**, which chooses materials to be presented to the learner, notes VINCI analyses, and decides on subsequent interactions. It also hands materials to a human instructor, in the form of diagnostics and log files, and accepts instructions from the human instructor in the form of frameworks for choosing which materials to present and how to react to student input.

This system is currently used to produce a variety of exercises including grammar testing, lexical choices based on pictures and text, reordering exercises in which a learner assembles lexical materials into a coherent sentence, and reading exercises, in which a learner is presented with a multi-sentence text (a description of a room, a short fairy tale, etc.) and asked questions about the content of the text. We will not consider these here, but rather turn our attention now to how this environment may be used to teach phonetic transcription and other elements of phonetics. We will focus on the teaching of four skills: phonetic transcription, minimal pairs, syllable structures and intonation.

IV. TEACHING BASIC PHONETIC TRANSCRIPTION

In its most basic form, phonetic transcription is the representation of the sounds of some language using some system of notation like the International Phonetics Alphabet (http://www.langsci.ucl.ac.uk/ipa/ipachart.html). A significant challenge in teaching phonetic transcription is to provide detailed feedback to students. Since student errors are frequently systematic, one should ideally show how the incorrect form provided by the student is related to but yet differs from the expected correct form. To accomplish this, we represent the various traits which distinguish different sounds as rule-based variants. So, for example, the consonants /s/ and /z/ share the traits ‘apico-dental’ and ‘fricative’, but differ in that the first is unvoiced (the vocal cords do not vibrate), while the second is voiced (the vocal cords do vibrate). These various forms may be captured by a set of morphology rules, of which we provide a small sample here:

apico-dental, fricative, voiced : "z";

apico-dental, fricative, unvoiced : "s";

etc.

Each subrule is composed of two elements: on the left a set of attributes which together identify some particular sound, and on the right, in double quotes, the sound corresponding to the attributes, in phonetic transcription. As noted above, VINCI has the ability to check what particular morphology subrule corresponds to input produced by a student. Thus, given the string /s/ erroneously produced by a learner and the expected form /z/, VINCI can calculate the similarities and differences between the traits shown above, and determine that the student has replaced a voiced apico-dental fricative by a voiceless apico-dental fricative. It can then use this to explain to the student where his or her error lies.

Of course, in order to elicit the student's response, VINCI must produce some oral output. This is accomplished by the following series of steps. Based on a terminal file containing two classes (Consonants and Vowels), and an attribute file, which specifies the traits associated with each consonant or vowel, a syntax rule generates the specification for some sequence of phonemes. Thus, for example, the syntax rule:

```
ROOT = C[bilabial, occlusive, voiced]
       V[front, high, unrounded]
%
```

corresponds to the sound sequence /bi/. To produce a question, the VINCI morphology transforms this sequence into a series of lines, where each line corresponds to a phoneme, a length in milliseconds, an initial frequency in Hz, and a final frequency. As an example, the following lines represent the same sequence /bi/.

```
b 83
i  141 150 150
```

This data is passed to the Mbrola synthesis system (http://tcts.fpms.ac.be/synthesis/) along with the specification of the voice to be used (English or French, male or female, and so on). Mbrola generates oral output in the form of a *wav* file. The **stdnt** module sends this to the web page and, using features present in HTML5, presents it as a small playback button. Having heard the oral output, the student is presented with an onscreen keyboard, where each button corresponds to a phonetic symbol. Buttons clicked by the student are transformed by the **stdnt** module into forms suitable for analysis by VINCI and then sent to VINCI. The result of a typical analysis is shown below, for the case where the student was expected to enter the sequence /z I p/, corresponding to the English word *zip*, but has mistakenly entered instead /s I p/.

EXPECTED : z I p
RESPONSE : s I p
ORDEREXACT
C1 S1 z/s MORPH voiced/unvoiced
C2 S2 EXACT
C3 S3 EXACT
***

This diagnosis is generated by the software and has a specific form. So, for example, C1 corresponds to the computer's first generated symbol, while S1 corresponds to the student's first symbol. In this case, they differ by one trait. Since the format of this output is constant, it may be parsed by **instr** and transformed into plain language feedback, in whatever language the learner happens to be using. For example, the specific output might be "Your first phoneme is incorrect. You inserted /s/, which is unvoiced, in place of /z/, which is voiced. Please listen to the recording again."

As the system now stands, errors corresponding to rule-based predictions are captured with some degree of precision. Other classes of errors currently captured by the system include deletion of elements, insertion of spurious elements, and transposition of correct elements.

In addition, to allow students to practice the correlations between phonetic symbols and sounds, right-clicking on the onscreen keyboard synthesizes the sound corresponding to the symbol chosen, thereby allowing a student to explore options before deciding on a particular transcription.

V. Teaching Minimal pairs

The various sounds of a language form a system such that some of them, by their opposition in a context where all other elements remain the same, lead to differences of meaning. Such contexts are called **minimal pairs**, and the two sounds are therefore different **phonemes**. Thus, in English, the vowels /I/ and /æ/ may replace one another in the context /p _ t/ and the resulting forms have two different meanings in English (*pit* versus *pat*). On the other hand, some differences in sound are not always associated with differences in meaning. For example, some speakers of English pronounce the word *economics* as /ɛkənɔmɪks/ and others as /ikənɔmɪks/, but both forms have the same meaning. These are instances of **free variation**. Other differences in pronunciation are conditioned by the context in which a sound occurs. Thus, in initial position in a word, the phoneme /p/ in English tends to be aspirated (followed by a puff of air), whereas no puff of air is produced when the /p/ is preceded by an /s/ (as in /p^hIn/ versus /spIn/. Such differences in pronunciation conditioned by context represent instances of **conditioned variation** and the two sounds are in **complementary distribution**.

The goal in teaching minimal pairs is to encourage students to discover the minimal pairs of a language and the phonemes they illustrate, as well as the role of conditioned and free variation. To make this possible, underlying lists of potential minimal pairs are constructed from online lexicons containing orthographic and phonological forms, processed by means of the Awk pattern-matching language [11]. The result is a set of phonological possibilities divided among various structures such as CV (consonant + vowel), CVC, VC and so on. Different minimal pairs are distinguished by their attributes: thus, the phonetic versions of the French words *seau*, *zoo* and *beau*, which share the ending /o/ in the context CV may be represented by the rules:

s, o, phono : "so";

z, o, phono : "zo";

b, o, phono : "bo";

etc.

Comparable morphology rules exist for words of different combinations of consonants and vowels. Any two lines of a morphology rule which differ by only one trait represent a minimal pair.

In order to explore these systems, a student may proceed in one of two ways. In the first, the student specifies a particular phonological context such as an open syllable (CV). The system selects and presents (in IPA and, if desired, in sounds) a particular sequence of sounds conforming to the pattern. In the present case, this might be /so/. The student must then find a sequence which brings to light a minimal pair, and enter this in IPA. For example, given /so/, the student might enter /zo/. As potential candidates are entered, they are evaluated by the system, which will either show the traits of a correct minimal pair, or the message that no minimal pair exists. In all cases, the student can also see the corresponding orthographic forms.

Alternatively, the student might enter two candidates for the status of phoneme, for example, /a/ and /i/. The system will then choose, by rule or randomly, some syllabic structure, like CV and present it to the student. The student must then find a minimal pair which functions in this context. Importantly, in this case, as in the previous one, the student has the sense of exploring, since he or she must find the items of the problem.

VI. Teaching Syllable structures

Words are not simply sequences of consonants and vowels. Rather, they are formed of syllables which themselves have an internal structure. Each syllable may be divided into an onset and a rhyme, where the onset is typically composed of consonants (so in the sequence /bIn/ 'bin' in English, /b/ is the onset). The onset may be empty in the case of syllables without initial consonants (think of /In/ 'in'). The rhyme represents the remainder of the syllable, where the nucleus, typically a vowel, forms the core and the coda the elements which follow. So, in the previous example, /In/ forms the rhyme, /I/ the nucleus and /n/ the coda. An important exercise for apprentice linguists is to learn how to divide syllables in this way.

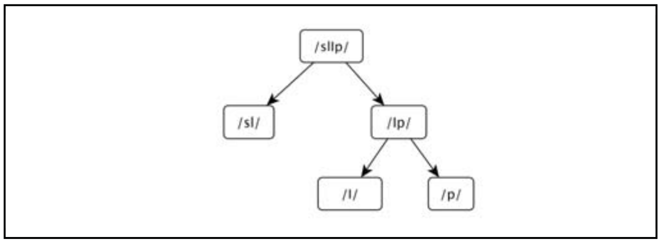


Figure 2 Division of /slIp/ into a syllable structure tree

The teaching of syllabic structures may be accomplished in VINCI in a variety of ways. We will illustrate one here. Assume that each inflected word of the lexicon for some particular language is composed of a set of fields, where the first gives the word form in orthography, the second the part of speech, the third the appropriate attributes, the fourth the frequency, the fifth the morphology rule, and so on. Assume now that there exists an attribute value for each syllabic structure, such that the value s2 means a two-syllable word. Assume as well that later fields of the lexical entry contain, in phonetic transcription or in orthography, the various onsets, nucleii and codas, so that field 7 contains the onset of the first syllable in phonetic transcription, field 8 the nucleus, and field 9 the coda. So, for example, the English word *slip* would appear as:

"slIp"|V|s2, ...|...|...|"sl"|"I"|"p"|...

where vertical bars divide fields of the lexical item. It is then possible to use the VINCI syntax to select a lexical item, present it in phonetic transcription, as an audio output, using Mbrola, or in orthographic form, and ask the student to divide it into its component parts. These may either be entered in phonetic transcription, with some symbol pair like [ ] containing each element or alternatively on the screen within a branching tree. Fig. 2 shows the one-syllable English word *slip* divided into its component parts in this way.

Errors in attribution are captured by the VINCI error analysis routines, which note cases where two items have been transposed, where an item is missing, and so on. Furthermore, since the syntax 'knows' which elements belong to which class, it is possible to provide diagnostic messages of some precision, such as "the syllable is lacking a nucleus; every syllable needs a nucleus". And, as in the case of minimal pairs, students are given the opportunity to explore. Among other things, a student may enter a candidate word, and if it is in the lexicon, the appropriate framework is generated.

VII. Teaching Intonation

In speech, pitch varies over the course of an utterance. Such changes may carry meaning, as in French, where the sequence /ilɛlɑ/ (Il est là?/He is there?) with rising pitch represents a question, whereas the same sequence with falling pitch represents a statement. Fortunately for the teaching of phonetics, there now exist a number of free pieces of software such as Praat (http://www.fon.hum.uva.nl/praat/), which show pitch, intensity, length and other factors in some detail. These allow students to record and analyze authentic segments of speech. There remains however the additional challenge of hearing and representing these changes in pitch. A variety of

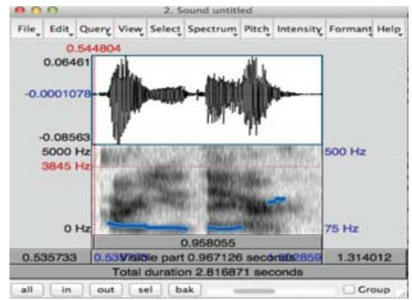


Figure 3 Praat representation of the pitch of "Are you there?"

representation schemes exist for graphically representing changes in intonation with typographical symbols, notably INTSINT [12]. It is also possible to use a system of digits ranging from 0 for lowest pitch, to 4 for highest. (See [13], [14].) So in the case of a sentence like "Are you there?" the first part of the utterance is at level 2, while the last is at level 4.

The web-based version of VINCI provides one means of using this model. The starting point is a set of syntactic patterns corresponding to various structures ranging from noun phrases ("the little boy"), to lists ("the boy, the girl and the dog"), to sentences ("Are you there?"). As in other cases, these may be presented to a learner as phonetic transcription, spoken speech, or orthography. Since VINCI can generate parallel outputs, one of these can contain the pitch levels for each segment. This is sent to Mbrola, as described above, which synthesizes output with the appropriately changing intonation and produces a wav file. Another hidden parallel stream contains the phonetic transcription together with digits for each level of pitch, as described above, where the digit indicates the pitch for the following sequence. In the case of the example shown above, this takes the form:

[2] ɑʁ juʷ [4] ðɛʁ

On hearing the output, the student's task is to assign the appropriate digits to the phonetic string. The VINCI error analysis routine then compares the student's values with those generated by the system and issues the appropriate diagnostics. Such an approach works at a basic level. It must be recognized, however, that the synthesis of speech is exceedingly complex (see [15]) and in particular that detailed contextual information is also important if one seeks to produce natural utterances [16]. This is an empirical question which is beginning to be answered (see [17]). For example, preliminary work on visual feedback on intonation is encouraging (see for example [18], [19]). In our system, for example, a wav file might also be sent to Praat itself to generate visual representations, which may be shown on the screen, although we have not explored this in detail yet.

VIII. Conclusions and Next Steps

The previous examples show that a generative CALL model provides a useful intermediate ground between canned exercises and fully freeform input, at least in the area of teaching phonetics. They show as well that the current web environment makes it possible to present students with various modalities, including sound, phonetic transcription, and visual representations.

As has been noted elsewhere, the generative approach to CALL is resource-intensive [20] and requires distinct areas of expertise. So, for example, the materials discussed here were produced by a team which included a linguist and a computer scientist as well as research assistants in both those fields. Given this disciplinary diversity, one of the overarching goals of the project is to provide for each of the experts a convivial and comprehensible working environment. Thus, the VINCI software is written in C for portability, the web interfaces in HTML, PHP and Expect for maximum flexibility, and the language representations are editable text files using familiar formalisms such as phrase structure grammars and lexical databases.

In the space available here, it has only been possible to show a limited range of facilities. In the coming months, we will be focusing on four areas:

we will apply the system to groups of test subjects to evaluate its effectiveness. As we noted in the introduction, we are not aware of other work which attempts to teach phonetic transcription, so comparison with other systems is not currently projected. That being said, it should be possible to test alternative approaches within the VINCI system;

building on previous work on adaptive testing and on detailed log files, we will develop more robust algorithms for using previous learner responses to focus questions on areas of difficulty;

using the facilities available in HTML5, JavaScript and JQuery for tracking user behaviour such as mouse direction and velocity, we will explore realtime responses before answers are selected;

and as alluded to above, we will explore the interplay of the VINCI system with other pieces of software like Praat to improve visualization of phenomena such as intonation.

IX. References

[1] M. Levy, *Computer-assisted language learning: context and conceptualization*. Oxford: Clarendon Press, 1997.

[2] K. Beatty, *Teaching and researching computer-assisted language learning*, 2nd ed. Harlow, England: Longman, 2010.

[3] J. M. Reid, “The learning style preferences of ESL students,” *TESOL Quarterly*, vol. 21, no. 1, pp. 87–111, 1987.

[4] C. Chapelle, *English language learning and technology: lectures on applied linguistics in the age of information and communication technology*, vol. 7. Amsterdam: John Benjamins Pub., 2003.

[5] V. M. Holland, J. D. Kaplan, and M. R. Sams, *Intelligent language tutors: theory shaping technology*. Mahwah, N.J.: Erlbaum, 1995.

[6] M. L. Swartz and M. Yazdani, *Intelligent tutoring systems for foreign language learning: the bridge to international communication*, vol. 80. Berlin: Springer-Verlag, 1992.

[7] M. Shulze, “Interfaces in intelligent CALL,” *Computer Assisted Language Learning*, vol. 21, no. 4, pp. 301–303, 2008.

[8] G. Lessard and M. Levison, “Lexical creativity in L2 French and natural language generation,” in *French Applied Linguistics*, D. Ayoun, Ed. Amsterdam: John Benjamins Pub., 2007, pp. 299–333.

[9] A. T. Corbett, K. R. Hoedinger, and J. R. Anderson, “Intelligent Tutoring Systems,” in *Handbook of Human-Computer Interaction*, M. Helander, T. K. Landauer, and P. Prabhu, Eds. Amsterdam: Elsevier, 1997, pp. 849–874.

[10] M. Levison, G. Lessard, A. M. Danielson, and D. Merven, “From symptoms to diagnosis,” in *C.A.L.L. - The Challenge of Change: Research and Practice*, K. Cameron, Ed., Exeter: Elm Bank Publications, 2001, pp. 53–59.

[11] A. V. Aho, B. W. Kernighan, and P. J. Weinberger, *The AWK programming language*. Reading, Mass.: Addison-Wesley Pub. Co., 1988.

[12] D. Hirst and A. Di Cristo, *Intonation systems: a survey of twenty languages*. Cambridge, U.K.: Cambridge University Press, 1998.

[13] C. Thomas, “A Prosodic Transcription Mechanism for Natural Language Generation,” Queen’s University, 2002.

[14] C. Thomas, M. Levison, and G. Lessard, “Experiments in Prosody for the Generation of Oral French,” in *NLP and Speech Technologies in Advanced Language Learning Systems*, 2004, pp. 123–126.

[15] T. Dutoit, *An introduction to text-to-speech synthesis*, vol. 3. Dordrecht: Kluwer Academic Publishers, 1997.

[16] B. Grote, E. Hagen, A. Stein, and E. Teich, “Speech Production in Human-Machine Dialogue: A Natural Language Generation Perspective,” in *Dialogue Processing in Spoken Language Systems*, E. Meier, M. Mast, and S. Luperfoy, Eds. Berlin: Springer, 1997, pp. 70–85.

[17] Z. Handley, “Is text-to-speech synthesis ready for use in computer-assisted language learning?,” *Speech communication*, vol. 51, pp. 906–919, 2009.

[18] O. Engwall, “Analysis of and feedback on phonetic features in pronunciation training with a virtual teacher,” *Computer Assisted Language Learning*, vol. 25, no. 1, pp. 37–64, 2012.

[19] R. Hincks and J. Edlund, “Promoting increased pitch variation in oral presentations with transient visual feedback,” *Language Learning & Technology*, vol. 13, no. 3, pp. 32–50, 2009.

[20] D. Galloway and K. Peterson-Bidoshi, “The case for dynamic exercise systems in language learning,” *Computer Assisted Language Learning*, vol. 21, no. 1, pp. 1–8, 2008.

Visualizing Geographic Learning Objects Through a Mobile Learning Application

Vincenzo Del Fatto, Gabriella Dodero, Ashraf ul-Islam Chowdhury, Luca Geraci
IDSE, Faculty of Computer Science
Free University of Bolzano
Bolzano-Bozen, Italy
{vincenzo.delfatto, gabriella.dodero}@unibz.it
{ashrafulislam.chowdhury, luca.geraci }@stud-inf.unibz.it

***Abstract*— In this paper we explore how topics from two different fields of computer science, such as Geographic Information Systems and E-learning Systems, can contribute to the Mobile Learning field. In particular, starting from a new concept, the Geographic Learning Object, we discuss how such Geographic Learning Objects can be used in a Geographic Information System in order to provide information and learning content to the citizens of a territory. A prototype mobile architecture supporting Geographic Learning Objects is also described.**

Mobile Learning; Geographic Information System; Geo-tagging, Augmented Reality.

I. Introduction

In this paper we explore how topics from two different computer science fields, such as Geographic Information Systems (GIS) and E-learning Systems, can be joined, inside Mobile Learning applications, by exploiting the GIS capability of managing and analyzing a territory, together with the E-learning Systems capability of managing learning content. In particular, we tested our hypothesis that most e-learning content, expressed as Learning Objects (LOs), contains hidden geographic information, that can be revealed and used to improve the learning content, its searching and its traceability.

In fact, geoscience studies [1-3] estimated that between 80% and 95% of all information and decisions in the public sector contain some geographic reference. As for the E-Learning component, Downes states in [4] that today's learning concept is closely related to the concept of LOs, which are coherent content, that have been refined and standardized into a rigorous form, together with specifications on how to sequence and organize them into courses. In this sense, we propose the vision of a territory as the object of study, it is hosting a large amount of LOs, available to citizens living in such a territory, in need of techniques and methods for searching and displaying such learning resources.

In a previous work [5] we analyzed several Learning Object Metadata (LOM) Standards and Learning Object Repositories (LOR), getting an overview of the state of the art in terms of de jure and de facto standards, and verifying how geographic context is taken into account within such repositories and metadata. We then introduced a first idea of new entity, the Geographic Learning Object (GLO), and we discussed how such GLOs can be used in a GIS context, in order to provide learning content to citizens in a Smart City. In this paper, starting from the definition of GLOs, we propose and describe a Mobile Learning application, implementing a Visualization system that acts as access point to a territory's learning platform.

The remainder of the paper is organized as follows. Section II introduces the context of the paper, and analyzes systems related to our proposal. In Section III we summarize the state of the art of e-learning standards, presenting GLOs, then discussing how GLOs can be used in a GIS context. Section IV describes the architecture of the Visualization System and the Mobile Learning application we propose. Section V draws the final conclusions.

II. Background and Related Work

This section analyzes systems related to our proposal, which joins two particular computer science fields, GIS and E-learning Systems, and specifically sub-areas of GIS and E-learning Systems such as geo-tagging and mobile learning.

In particular, in this paper we investigate our hypothesis that in most e-learning content, expressed as learning objects, there is hidden geographic information, that can be revealed and used to improve the learning content and its search and traceability. To strengthen our hypothesis, geoscience studies estimate that in the last 20 years about 80% of all information contains some geographic reference [1]. In particular, in [2] it is specified that about 80% of all decisions in the public sector are based on geo-referenced data. Furthermore, Perkins [3] updates the percentage claiming that today 95% is more accurate because of new technology such as cell phones, GPS devices and electronic toll collectors.

As reported in [6], nowadays, this huge variety of mobile devices providing integrated GPS receivers is the reason of the renaissance of location-based mobile applications and in the same time the reason of the widespread use of geo-locating applications such as geo-tagging. This technique, mostly used for images, associates a digital resource with a pair of geographical coordinates. The additional geographic information offers new teaching and learning possibilities, in particular in fields strongly dependent on geo-located data, such as civil engineering, geosciences or archeology. Moreover, as mentioned in [4], the combination of geo-tagging with other Web 2.0 technologies provides a further contribution to e-Learning 2.0. Our proposal builds upon the geo-tagging technique, but to our knowledge this is the first paper which addresses the problem of associating more than a simple pair of geographic coordinates to a digital learning resource. An intuition similar to our idea stimulated other studies [7-9], where micro-blogging services like Twitter are analyzed in order to extract hidden geographic patterns. In this case, differently from our approach, the focus is both to discover language patterns and to extract users' interests starting from the analysis of geo-tagged messages. Our proposal, in fact, is not to present a data mining algorithm which uses existing geographic coordinates, but to provide a flexible structure which allows the user to easily associate a geographic context to a learning content.

As for the E-learning, we started from the concept of LO, which as Downes claims [4], today are closely related to the learning concept. They are coherent content, that have been refined and standardized into a rigorous form together with specifications on how to sequence and organize them into courses.

As combination of GIS and E-learning, the Mobile learning (m-Learning) area is another area related to our topic. It is focused on e-Learning using mobile devices and, as reported in [10], it deals with applications that support learning anywhere, anytime. For instance, the Handheld-Centric Classroom approach, presented in [11], uses mobile devices as an integral part of a learning activity. As reported in [12], a main characteristic of mobile learning is the possibility of ongoing assessment and feedback. An interesting example of using m-Learning in higher education is the EU research project RAFT (Remote Accessible Field Trips), which was conducted from 2002 to 2005 [13]. The goal of RAFT project was the support of classes with virtual excursions, using portable Internet-conferencing tools. A common feature between these applications and our approach is related to study topics which rely on education in-the-field, for which m-Learning is particularly interesting. An important difference with respect to our approach is that the learning content we propose could be accessed through m-Learning applications, but they are not limited to it. Finally, an approach that mixes mobile learning and augmented reality has been presented in [14], where a location aware augmented reality system based on mobile technology, named ACCampus, has been proposed. Although this work concerns augmented reality and mobile learning, it differs from our proposal because is designed for indoor mobile users, by using QR codes and wi-fi connection to detect user localization. Our approach is designed for outdoor mobile users, and that totally changes the issues involved and the implementation choices.

III. From Learning Object Repositories and Metadata to GLOs

In a previous work [5] we analyzed different LOR and LOM standards in order to

- get an overview of the state of the art, in terms of de jure and de facto e-learning standards, and to
- verify whether and how a geographic context is taken into account within such repositories and metadata.

It is worth noting that LO research is less and less active in recent years. Nevertheless, in many years of studies and projects a huge number of LOs have been created. Our aim is to propose a method to facilitate reuse and improvement of existing LOs.

In particular, the GLOBE (Global Learning Objects Brokering Exchange) project was analyzed, which contains about 1,2 million learning objects, and uses the well-known IEEE LOM as learning standard. In [15] an interesting large-scale study is reported, about use and quality of LOM instances in GLOBE. Among the many conclusions of [15], two were interesting for our work:

1. Only 20 out of the 50 data elements are used more than 60% of the time, thus suggesting that just a core of LOM elements have been used. Specifically, just 2 elements are used 100% of the times, and just 7 elements are used more than 80% of the times, thus suggesting that LOM is not fully used by Globe providers.
2. GLOBE providers used 2 extended geographic data elements, which are not present in LOM standard. The need to enrich LOM standard clearly suggests that it does not contain enough elements to associate a significant and usable geographic context to LOs

More attention to geographic aspect has been given in the MACE (Metadata for Architectural Contents in Europe) [16] project. MACE is a project which connects several repositories of architectural knowledge, and enriches their contents with new metadata, to support different learning scenarios. Although MACE provides means to associate a basic geographic context to a learning object, we argue that associating only one geo-location (a pair of coordinates) to a learning object can severely limit the use of that resource. Also the popular geo-tagging technique presents this kind of limitation, that is, just a pair of geographic coordinates can be associated to a digital resource. In order to fill this gap, in [5], we proposed the Geographic Learning Object concept, as an extension of a LO which embeds information about geographic contexts where the LO is valid and/or applicable, as shown in Fig. 1.

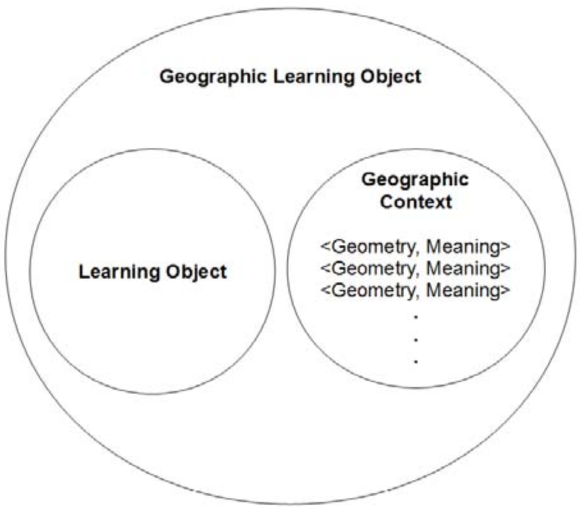


Figure 1. The Geographic Learning Object Structure.

A Geographic Context is a sequence of n pairs <Geometry, Meaning>, where

- Geometry is a generic geographic information, which can take a variety of shapes in agreement with the geometry class of the Open Geospatial Consortium (OGC) Standard for Geographic information - Simple feature[1], and
- Meaning is the meaning associated to each Geometry instance.

Such a solution overcomes the limit of associating a single pair of coordinates to a GLO, and it allows to fully explicit all geographic information that is often hidden inside a learning content.

As an example, a Marine Biology study could be associated with a Multi-polygon (collection of polygons) corresponding to the areas where the studied species live, and also with a point, corresponding to the research center where the study was carried out, and also with a Multipoint (collection of points) representing locations where experimental data was collected. As a second example, a biography of a historical figure could be associated with a Multipoint, representing the birthplace, different places where he lived, and the place of death. As a last example, we report in Table 1 a LO taken from the Learning Resource Exchange Repository, showing how different geographic contexts could be associated to such LO, by turning it into a GLO.

[1] http://www.opengeospatial.org/standards/sfa

TABLE I. "SCHOOL AT HOME" LEARNING OBJECT.

Title	"School at home". "Look what we do!" Woodland and forest fires (Verge del Tallat Primary School in Blancafort)
Description	With the help of forestry agents in the Conca de Barbera area, pupils at Verge del Tallat Primary School in Blancafort get to know the characteristics of the Mediterranean woodland, with special emphasis on the disastrous ecological effects of forest fires.
User's Tags	Not Available
Descriptors	environmental education environmental protection forest
Keywords	Not Available
Age range	18-99
Resource type	Website-Video
Available in	ca
License	See License
Provider	XTEC, Spain
Read about in	ca de en es fr it pt

As the reader can see, no geographical feature is used in the set of metadata associated to the original LO, although there are many geographical references in the title and in the description. By using the GLO structure, such geographic information could be explicated, and stored as different geometries.

So, a GLO derived from the LO in Table 1 should contain:

- a Polygon representing the Blancafort Municipality,
- a Point representing the address of the Verge del Tallat Primary School, and
- a Polygon representing the Conca de Barbera area

In the "School at home" website, as part of the LO, a video shows an excursion made by children of the Verge del Tallat Primary School in the Conca de Barbera area, thus additional related geometries could be added to the GLO:

- a Linestring representing the route taken, and
- a MultiPoint representing trees planted by children.

All this additional geographic information can be used to improve search functions among LOs, and to reuse the same LO, e.g. by following the same route in subsequent years by other Primary School classes, checking the status of planted trees.

As the examples highlight, GLOs can be used in a GIS context in order to provide learning content to citizens in a territory. In particular, GLOs can be used to access LOs, by using the territory both as the starting point, and as the filter of a content learning search.

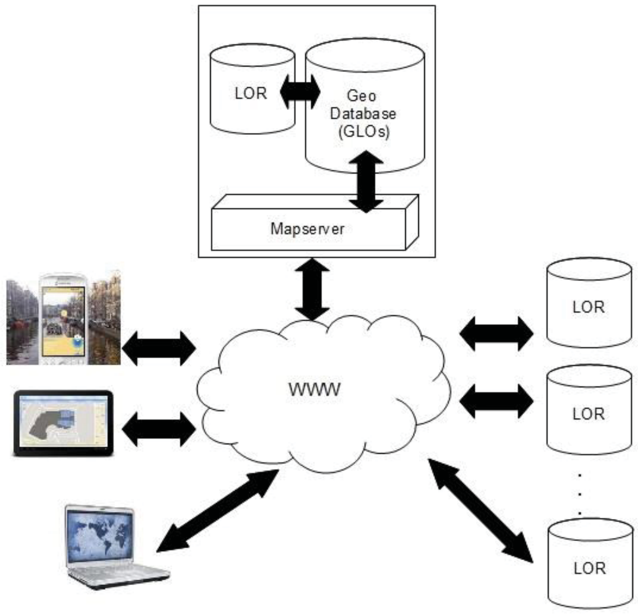


Figure 2. GLO System Architecture.

IV. GLO APPLICATION PROTOTYPE

A first prototype system, based on a client-server architecture, has been implemented at the Free University of Bolzano-Bozen. The current implementation consists of a server side, a very light (possibly mobile) client side, and the communication system between client and server. As shown in Fig. 2, the core component of the architecture is represented by a Geographic DBMS containing GLOs, in a standard format within a standard service, ready to be used in different display modes, such as maps on PCs, tablets and smart-phones, or as augmented reality on mobile devices. GLOs in the Geographic DBMS are either linked to already existing LOs (external LORs) or to newly developed GLOs (internal LOR). In particular, the server side contains:

- PostgreSQL[2] as DBMS, with PostGIS[3] component as geospatial extension, in order to store GLOs, and
- OSGeo[4] MapServer as mapserver, an Open Source geographic data rendering engine, in order to publish data through the standard web service (WFS[5]) in the Geography Markup Language (GML[6]) format.

The very light client has been implemented as an Android (version 4.0.3) mobile application, able to show GLOs in two different display modes, 2D map and augmented reality. The mobile application extends the Mix Augmented Reality Engine (Mixare[7]), which is a free open source geo-based augmented reality application with basic functions, published under the GPLv3 license, and thus easy to extend and improve. The chosen software meets the following criteria: (1) be geo-based, (2) be free of charge, and (3) be as complete and functional as possible.

The original Mixare application shows only Points of Interest (POIs), extracted from default data sources (Geonames and Wikipedia) in GeoJSON and XML format, on both 2D map and augmented reality view.

Extensions have been made to the original Mixare code. In particular, a new data source, the GLO server, has been added to the existing data sources, Geonames and Wikipedia. Now, the extended application queries the Geographic DBMS containing GLOs on the server, and parses GLOs, obtained from the GLO server in GML format, which was not supported by Mixare. Then, visualization as desired is performed. The Map View has been extended to visualize all kinds of geometries in addition to points. In particular, we extended 2D Map View with Google API version 3, that allows to show different types of geographic objects, such as lines and polygons, on the map.

Fig. 3 shows GLOs in Bolzano-Bozen, Italy, in the 2D Map View. On the right side of the figure some points in the historical center of Bolzano-Bozen are displayed, and on the left side a polygon that represents the GLO of the airport of Bolzano-Bozen is displayed. The user can interact with the individual map object, and access the contents of the corresponding GLOs.

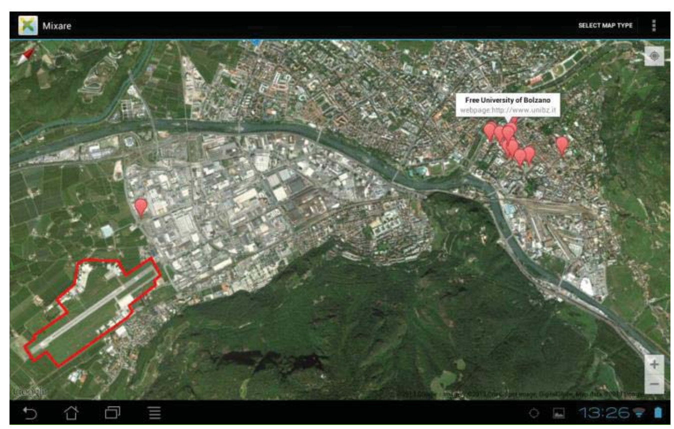


Figure 3. GLOs displayed on the 2D Map View.

Also the type of interaction between client and server has been changed. In the original Mixare version, the application downloads all POIs from the database on the server and then, it is the client application that filters the number of POIs based on the distance set by the user. We modified the code in order to download only those GLOs that fall within the distance set by the user. This change has been introduced to improve client performance, focusing its computational resources on the

[2] http://www.postgresql.org/
[3] http://postgis.net/
[4] http://mapserver.org
[5] http://www.opengeospatial.org/standards/wfs
[6] http://www.opengeospatial.org/standards/gml

[7] http://www.mixare.org/

display mode, and to minimize data transmission. Fig. 4 shows a GLO in the Augmented Reality View near the Computer Science Faculty of Free University of Bolzano-Bozen. Also in this modality, the user can interact with the individual GLO and access its content.

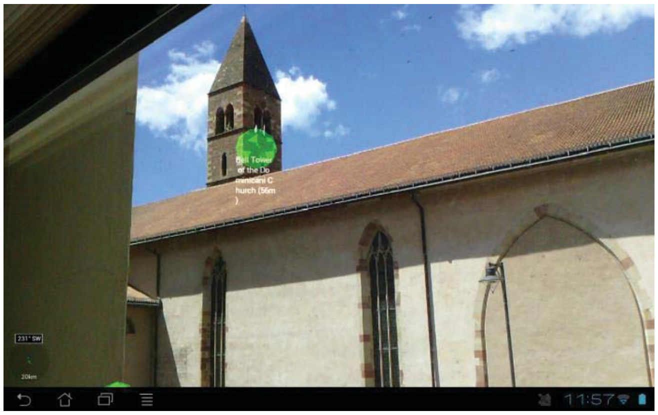


Figure 4. A GLO displayed on the Augmented Reality View.

At present the augmented reality and the 2D map modes are both being experienced with students. More significant extensions have been planned and are being implemented at the time of writing this work, concerning visualization of different kinds of shapes in the Augmented Reality View, by integrating an OpenGL[8] module, an environment for developing portable, interactive 2D and 3D graphic applications, in order to add, different types of geographic objects.

V. CONCLUSIONS

In this paper we proposed a new architecture, and its prototype implementation, managing a new entity, the Geographic Learning Object (GLO). A GLO is an extension of a Learning Object (LO), which embeds information about a geographic context, where the Learning Object is valid and/or applicable. GLOs derive from two well established fields, Geographic Information Systems and E-learning Systems, by exploiting the GIS capability of managing and analyzing a territory, and the E-learning Systems capability of managing learning content. In particular, we showed some examples of e-learning content, expressed as learning objects, where hidden geographic information can be revealed, and used to improve the learning content, its search and its traceability. A possible application of the GLO concept would be in tomorrow's Smart Cities. In fact, as defined, a set of GLOs concerning the area of a Smart City would be the information layer in different types of Geographic Information Systems, which thus become the access point of a city-wide learning platform.

As future work, we aim to compete the implementation of the prototype, so to easily allow users to associate a geographic context to a learning object, as defined in the GLO structure. A second extension would be to improve GLO structure by adding temporal context information. Moreover, we aim to investigate how to automatically or semi-automatically extract hidden geographic information from existing LOs. Tests on the effectiveness and usability of our prototype system are also planned, by undertaking a usability study with potential users in the educational domain (school teachers and pupils).

[8] http://www.opengl.org/

REFERENCES

[1] S. Hahmann, D. Burghardt, and B. Weber, "80% of All Information is Geospatially Referenced"??? Towards a Research Framework: Using the Semantic Web for (In)Validating this Famous Geo Assertion, in Proceedings of the 14th AGILE Conference on GIScience, AGILE, Utrecht (2011)

[2] J. Riecken, "The improvement of the access to public geospatial data of cadastral andsurveying and mapping as a part of the development of a NSDI in Northrhine-Westfalia,Germany", in Proceedings of the 4th AGILE Conference on GIScience, AGILE, Brno pp. 215– 221. (2001)

[3] B. Perkins, "Have you mapped your data today" (2010) http://www.computerworld.com/s/article/350588/Have_You_Mapped_Your_Data_Today_.

[4] S. Downes, "E-Learning 2.0", in eLearn Magazine October 16, 2005.

[5] V. Del Fatto, and G. Dodero, "Geographic Learning Objects in Smart cities Context", Interaction Design and Architecture(s) Journal - IxD&A, N. 16, 2013, to appear.

[6] C. Safran, V.M. Garcia-Barrios, and M. Ebner, "The Benefits of Geo Tagging and Microblogging in m-Learning: a Use Case", in Proceedings of ACM Academic MindTrek 2009 - Everyday Life in the UbiquitousArea, MindTrek 2009, Tampere, Finland, p. 135-141, 2009

[7] L. Hong, A. Ahmed, S. Gurumurthy, A. J. Smola, and K. Tsioutsiouliklis, "Discovering geographical topics in the twitter stream". WWW 2012: 769-778

[8] Q. Mei, C. Liu, H. Su, and C. A. Zhai, "Probabilistic approach to spatiotemporal theme pattern mining on weblogs". In Proceedings of WWW 2006, pages 533–542, New York, NY, USA. ACM.

[9] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks". In Proceedings of KDD 2011, pages 1082–1090, New York, NY, USA. ACM.

[10] D. Tatar, J. Roschelle, P. Vahey, and W. R. Penuel, "Handhelds Go to School: Lessons Learned". Computer, 36, 9(2003), 30-37.

[11] C. Norris, and E. Soloway, "Envisioning the Handheld-Centric Classroom". Journal of Educational Computing Research, 30, 4 (2004), 281-294.

[12] R. Klamma, M. A. Chatti, E. Duval, H. Hummel, E. T. Hvannberg, M. Kravcik, E. Law, A. Naeve, and P. Scott, "Social software for life-long learning". Educational Technology & Society, 10, 3 (2007), 72-83.

[13] M. Kravcik, M. Specht, A. Kaibel, and L. Terrenghi, "Collecting data on field trips - RAFT approach". ICALT 2003, 2003,478.

[14] A. De Lucia, R. Francese, I. Passero, and G. Tortora. 2012. "A Collaborative Augmented Campus Based on Location-Aware Mobile Technology". Int. J. Distance Educ. Technol. 10, 1 (January 2012), 55-73. DOI = 10.4018/jdet.2012010104 http://dx.doi.org/10.4018/jdet.2012010104

[15] X. Ochoa, J. Klerkx, B. Vandeputte, and E. Duval, "On the use of learning object metadata: the GLOBE experience". In *Proceedings of* EC-TEL'11, Carlos Delgado Kloos, Denis Gillet, Raquel M. Crespo García, Fridolin Wild, and Martin Wolpers (Eds.). Springer-Verlag, Berlin, Heidelberg, 271-284. (2011).

[16] E. Arlati, E. Bogani, and A. Cammarata, "MACE – Metadata For Architectural Contents in Europe". In: ICERI2008 - Proceedings CD. ICERI 2008. Madrid (Spain). 17th-19th November 2008. ISBN/ISSN: 978-84-612-5091-2. Valencia: IATED (SPAIN). (2008).

Design and Development of ePortfolio Using Gamification to Increase Engagement

Yuki TERAWAKI[1], Taketo SHIMOHIGOSHI[2], Makoto WATANABE[2],
Toshiyuki KAMADA[3] ,Yasushi KODAMA[4], Kazuo YANA[1]
[1]Research Center for Computing and Multimedia Studies, Hosei University, Tokyo, Japan
[2]Hosei University Faculty of Engineering & Design, Department of Architecture, Tokyo, Japan
[3] Department of Technology Education, Aichi University of Education, Tokyo, Japan
[4]Hosei University Faculty of Business Administration Department of Business Strategy, Tokyo, Japan

Abstract— **Intrinsic motivation plays an important role in learning. The number of students learning in digitized educational environments is increasing, but online learning systems have limitations in sustainability due to a lack of motivation on the part of the learner. To create intrinsic motivation, we focus on "gamification". This paper outlines the major trends in gamification, and explains the design and development of an e-Portfolio that has adopted gamification. We expect this design to enhancing emotional engagement in learning, and thereby improve educational achievement.**

Keywords—**e-Portfolio, e-Learning, Gamification, Game design elements, Engagement**

I. INTRODUCTION

Younger generations of students are immersed in a digital environment. These students, who are very familiar with digital learning environments, are called "digital natives" [1]. They have access to the Internet at all times, and instant access to information for learning; whether it is a discussion on a Learning Management System (LMS), video lectures, or other information on the Internet, there are many possible ways to learn. This creates a unique situation that requires unique attention.

Recently, e-Learning has become a popular educational platform that facilitates a richer educational experience. However, many conventional e-Learning styles are implemented in the teacher's perspective, without consideration for the perspective of digital natives. In addition, teachers have many good real-world classroom educational devices that have not been properly ported to e-Learning environment.

The research on motivation for learning has been carried out to bring about the desire in the learner. In educational psychology, the quest for an explanation of study motivation has resulted in two contrasting theories: "Extrinsic Motivation" and "Intrinsic Motivation". However, because these theories are highly generalized, they are criticized for not being of much use in the improvement of education in practice. Ichikawa [2] redefined the structure of study motivation and explained that those learners who start study motivation and explained that learners who start studying with content-involved motivation have a greater difficult for content involved motication to get learners to start studying and so it is important to bring about the desire in the learner by the relationship they have with their educator and from their peers (e.g. my friend is studying hard so I should do the same) and to lead them so they become aware of the significance of what they have studied.

In real-world learning situations, enhanced emotional engagement invokes a positive attitude towards learning. This is also true in e-Learning environments. Therefore, the enhancement of emotional engagement should be one of the main goals for e-Learning systems. By evoking emotional engagement, we can help students achieve a sustainable learning motivation.

Gamification is a very important way of enhancing emotional engagement in e-Learning. The idea of using game design elements in non-game contexts to motivate and increase user activity and retention has rapidly gained traction in interaction design and digital marketing. The effectiveness of gamification in enhancing emotional engagement has already been demonstrated in business. Even though these successes have been focused on economic success the same principle can be applied in education if we think more deeply about the meaning of engagement in education. Consequently, in this paper, we focus on enhancing motivation, and append a function comprising game design elements to our existing e-Portfolio.

The remainder of this paper is organized as follows. Section2 describes our ePortfololio, named IAE (Integrated Archiving Environment). Section3 outlines gamification, and reviews some related works in order to consider for appropriateness of these techniques. Section4 presents design and function of ePortlfolio which adapted the gamification. Section5 discusses our experimental usage of the function. Section 6 concludes this paper and outlines some of the key challenges to be tackled in future work.

Ⅱ. Integrated Archiving Environment (IAE)

Portfolios serve as showcases and have been used to evaluate people's achievements and learning objectives. Portfolios were originally used in art- and architecture-related departments and they are very familiar in the Department of Architecture.

e-Portfolios are electronic versions of portfolios. Most architectural works produced by students in the Department of Architecture, such as architectural models, drawings, and presentation boards to explain design concepts, were not electronic creations originally. Therefore, in order to accumulate all the results of learning activities in the e-portfolio system, methods to computerize originally non-electronic deliverables had to be devised.

We previously developed an e-Portfolio system to save all learning outcomes and educational guidance records in the Department of Architecture electronically [3]. This e-Portfolio system is named the Integrated Archiving Environment (IAE), has been utilized in learning activities in all departments since 2010.

The IAE [4] is an e-Portfolio system that is used to construct methods to computerize originally non-electronic achievements and to realize seamless connections with these methods. Figure 2 shows a conceptual image of the IAE. The IAE consists of three tools: a tool for management of outcomes and evidences (store and accept), a digitizing tool that utilizes Quick Response (QR) codes (see Fig. 1), and a sequential uploading tool for the management of photographs. The IAE has been utilized in various types of classes. For instance, it has been used in courses involving practical training to learn a serial design processes such as construction and materialization of space, as well as in review meetings for architectural works at the end of the courses. In addition, the IAE has a function to save picture data at the same time as the pictures are taken. We utilize this function, when we hold discussions to project pictures of architectural models that were instantly saved in the IAE when the pictures were shot. Further, when students hand in their learning outcomes written by hand, a QR code seal is applied to the report. The report to which put the QR code seal is affixed is read with a high-speed scanner, and subsequently digitized and stored in a file identified by student number. All such digital files are then uploaded as a group and automatically saved in an appropriate folder in term of class name and student ID.

Since the introduction of the service in 2010, the IAE has been running smoothly, without any unexpected system interruptions and has gained a certain amount of appreciation from faculty members and students in the Department of Architecture. The IAE has since become essential for the effective conduct of learning activities in the Department of Architecture.

Ⅲ. Gamification

The idea of using game design elements in non-game contexts to motivate and increase usage has rapidly gained Popularity in digital marketing. Marketers and product managers frequently use gamification to engage customers and to influence desirable usage behaviors, Online application platforms such as LinkedIn (Linkedin.com), and eBay (ebay.com) are popular examples of increased user engagement driven by game design elements. Zichermann and Cunningham state that engagement metrics such as time spent on site increased by 20% after application of gamification techniques [5].

With the growing popularity of digital marketing, the game design elements draw human-computer interaction (HCI) researcher's attention. For instance, researchers have begun to study concepts such "motivational affordances [6]", "pleasurable products [7]", and "science of enjoyable technology [8]". As part of this movement, multiple researchers have explored "playfulness" as a desirable user experience or mode of interaction, and how to design for it. Thus, interest from researchers in gamification is on the increase.

As stated above, the idea of inserting game dynamics into web interactions to enhance engagement is referred to as gamification. However, the word "gamification" is still a contested term. Since 2008 different approaches have viewed "gamification" from different angles, resulting in many definitions of the concept of "gamification". Recently, many researchers have shown that the game design elements can be an effective medium for deeppeing learning and increasing motivation [9] [10]. It is now evident that gamification is now being applied in "non-game contexts". The term " non-game context" can be defined as follows: "Gamification uses games for other purposes than their normaly expected use for entertainment (asserting that entertainment constitutes the prevalent expected use of games)". Therefore, we can assume that the game design elements which are powerful educational methods, can be used as media for guided learning.

Fig. 1 QR code label used in the IAE

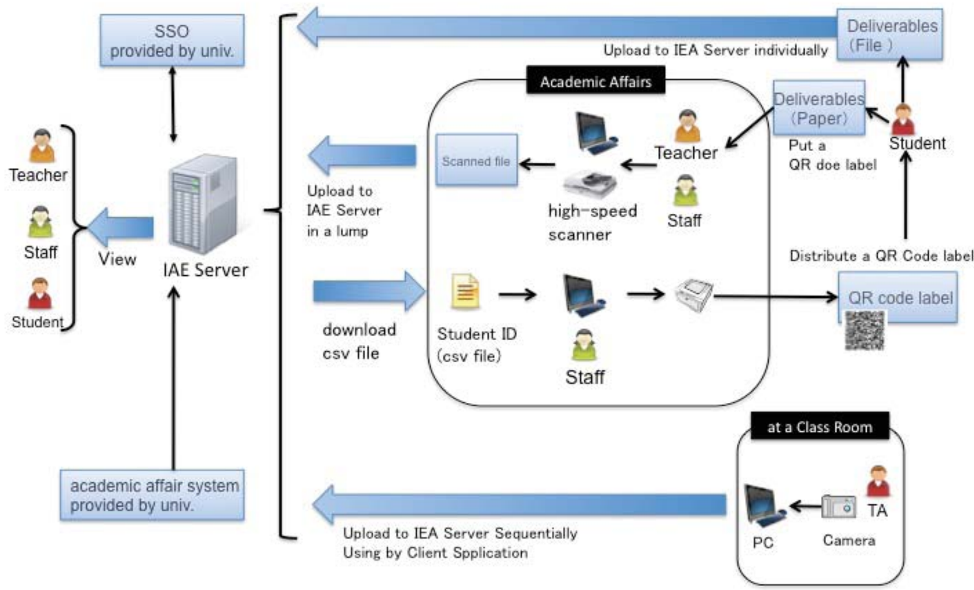


Fig. 2 Conceptual diagram of the IAE

Ⅳ. Design and Implementation of Gamification in the IAE

Several elements of gamification have been in common use. Elements such as Points, Badges, On Boarding, and Loyalty have been contributed to enhance engagement and generate extra revenue for commercial ventures. However, the use of gamification education requires careful consideration because gamification can be both positive and negative [11]. Therefore, we have to consider the fundamental traits of gamification.

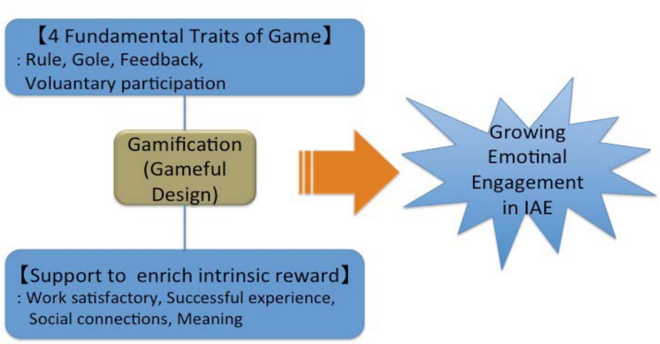


Fig. 3. Design Concept

A. *The Fundamental Traits of Gamification*

Gamification has four fundamental traits; goals, rules, feedback system, and voluntary participation [12]. We have added gamification elements to our existing e-portfolio. Fig.3 shows design concept of our e-Portfolio. The goal is for students to create and publish an e-Portfolio from learning outcomes. The rule is that students must receive comments from as many people as possible on e-portfolios that they have published. The feedback is realized by visualization of how many comments they received and how many people viewed the portfolio.

"Voluntary participation" is an important issue in recent gamification research. For example, findings from positive psychology have been adapted to the issue. McGonigal [12], state that games are essentially self-purpose activities that enrich intrinsic reward rather than activities to stimulate demand for extrinsic reward. The work also describes how games help human beings to experience the intrinsic reward that they seek, and summarizes the intrinsic elements into the following four categories: "work satisfactory," "successful experiences," "social connections," and "meaning (wanting to be a part of something bigger than their own)." We have decided to adopt gamification so that our e-portfolio can be a tool for students to use in accordance with their intrinsic motivation, i.e., for them to use actively. The four categories of intrinsic reward are very informative and guide the evolution of our ePortfolio IAE.

However, Gamification is a new concept, and there is no definitive definition in the game field itself to use as a guide in the realization of "voluntary participation." Our attempt to adopt gamification to e-portfolios (or adopting gamification to software systems for educational purposes) is also a new challenge. For this reason, we plan to adopt the four characteristics of gamification to e-portfolios in stages. First, we plan to develop functions to realize "goals," "rules," and "feedback system." Then, observe how they are utilized by students and analyze the results to see what advantages and disadvantages exist. Next, we plan to examine the changes that take place in the learners by adopting the three characteristics, and realize functions for appropriate "voluntary participation" in our portfolio. In this paper, we describe functions created to realize "goals," "rules," and "feedback system."

B. *Goals*

Our goal is to create an e-portfolio from learning results and publish it. The IAE has all the learning outcomes stored. Thus, we created two areas: an area where all the learning outcomes of students are stored and the e-Portfolio area. Students create e-Portfolios by selecting works that they like from among the works in their storage area and moving them to the e-portfolio area. Students can also publish their e-Portfolios on a page-by-page basis when moving works to the e-portfolio area. In reality, only three pages of five-page works created using PowerPoint can be published as an e-Portfolio. The IAE's e-Portfolio areas are connected to SlideShare. Works published as an e-portfolio can therefore also be viewed from SlideShare. As such, e-Portfolios published by students can be seen not only by IAE users but also by people around the world.

Figure 4 is a screenshot of the new functions. As previously mentioned, the IAE is divided into the storage area and the e-Portfolio area. "My Files" is a place for storing materials. In the creation of an e-Portfolio, materials are first selected from "My Files" and the portfolio created in "My Portfolios." If a student would like other people to comment on the portfolios he/she created, he/she moves the relevant file(s) to "Request for Comments." "PinBoard" is a place where a student can store other people's portfolios that he/she likes. The number of portfolios stored here is counted and can be published to all IAE users by ranking (see Fig. 6).

C. *Rules*

The rule is that students must receive comments from as many people as possible on e-portfolios that they have published. The Request for Comments (RFC) function is used to post comments on e-Portfolios. RFC is the IAE's original microblogging service (see Fig. 5). RFC allows a user of the IAE to post comments on published portfolios. Comments can be posted on a page-by-page basis and a student can bookmark pages on which he/she commented and archive his/her favorite works. In addition, comments left by other users can be downloaded in CSV format.

D. *Feedback system*

Feedback is realized by visualizing how many comments were received and how many people viewed the portfolio. The feedback system function has three features: First, it notifies students by email of comments posted on their portfolio and replies from the authors of portfolios on which a student posted comments. Next, IAE users are notified by email of lists of newly published portfolios and portfolios on which comments have just been posted. Finally, the portfolios bookmarked by other people are ranked according to number of bookmarks (the ranking of heavily-bookmarked portfolios).

Fig. 4. Screenshot of IAE

Fig. 5. Screenshot of RFC

VI. Experimental Usage

We utilized four classes in our experimental usage. New functions were tried in classes taught by teachers who were engaged in the development of IAE, especially in those that involved training and students' presentations. From April 2012 through March 2013, 163 portfolios were published. Although some only aimed to be published, the number of comments posted by students on the published e-portfolios was 354. Typical opinions from students were as follows:

- Easily searching other people's reports is of particular value in studying.
- If I can, I'd like to post to many friends' portfolios.
- Finding an excellent report has given me a creative urge.
- This makes posting comments to other students extremely simple.
- In the beginning, I could not figure out how to use the system.
- Uploading a new portfolio takes a long time.

VI. Conclusion and Future Works

In this paper, we outlined previous research on gamification, and presented our new functions developed using game design elements. We added gamification functionality to our ePortfolio system to develop it into a more advanced system that encourages students to actively use it. While we have achieved a partial formative evaluation based on responses from the target students who participated in the classroom trials, the results failed to show how the new function affected students' use of ePortfolios or what kinds of effects can be expected. "Voluntary participation" is known to be one of four basic factors of gamification," but further research is necessary to evaluate its effect on active students' participation. The new function, however, improved the educational environment by providing a means of sharing report materials among participants (i.e., instructors and students) in classes that require students' presentations, as well as a means of allowing a large number of students to express their opinions in the classroom.

However, there is room for improvement. For example, the current system takes a long time to upload relevant files, and there are some technical issues regarding its use. Although students received guidance on how to use e-Portfolios, and technical support is available, there is a concern that the slow upload speed during guidance sessions may reduce the motivation of students. Upon the release of e-Portfolios with gamification functionality, we plan to introduce simpler methods that could lead to an increase in its usage. Going forward, we plan to solve the technical problems encountered, and conduct more detailed evaluations in order to find ways to engage students in their learning activities through our ePortfolio.

Fig. 6. Screenshot of Top Page (Ranking)

REFERENCES

[1] M.Prensky, "Digital natives, digital immigrants Part1", the Horizon 2001.

[2] Shin'ichi Ichikawa, "The structure of leaning motives and its relations to the concepts of learning.", Proceedings of the 37'th Annual Conference of Japanese Educational Psychology Association, pp. 177, 1995 (In Japanese).

[3] Y. Terawaki et al., "Introducing ePortfolio to Architectural Course: the Integrated Archiving Environment", Proceedings of the 17th International Conference on Distributed Multimedia Systems, pp. 250-253, 2011.

[4] IAE 2011 Server, Retrieved April 8, 2011, from https://iae.hosei.ac.jp, May 2013.

[5] G. Zichermann and C. Cunningham, "Gamification by design: Implementing Game Mechanics in Web and Mobile Apps", OReilly, 2011.

[6] P. Zhang, "Motivational Affordances: Reasons for ICT Design and Use", Computers in Human Behavior 18(4), pp. 369-288, 2002.

[7] P.W. Jordan, "Designing Pleasurable Products: An Introduction to the New Human Factors", Taylor & Francis, London, New York, 2002.

[8] M., Blythe et al., "Introduction: Beyond Fun", Interactions 11(5), pp. 36-37, 2004.

[9] J. P. Gee, "What video games have to teach us about learning and literacy", Computers in Entertainment, 1(1), pp. 20, 2003.

[10] A. Graesser et al., "Deep Learning and Emotion in Serious Games", In Serious Games Mechanisms and Effect, New York, USA Routledge, Tayloy & Francis, pp. 83-102, 2009.

[11] J. J. Lee, and J. Hammer, "Gamification in Education: What, How, Why bother?" Academic Exchange Quarterly 15(2), pp. 146-150, 2011.

[12] J. Mcgonigal, "Reality is Broken: Why Games Make Us Better and How They Can Change the World", Penguin Press HC, 2011.

Address of the corresponding author:
Yuki TERAWAKI
Research Center for Computing and Multimedia Studies,
Hosei University
3-7-2 Kajino, Koganei City Tokyo 184-8584
Tel: +81-(0) 423-87-6128
Email: yuki.terawaki.dc@k.hosei.ac.jp

Educational MMORPG for Computer Science: DeBugger, a Virtual Lab on PC and Smart Phones

Ilmi Yoon[1], Arno Puder[1], Gary Ng[1], Manori Thakur[1],
Hunvil Rodrigues[1], and Jae H. Paik[2]
[1]Computer Science, SFSU, [2]Psychology, SFSU
San Francisco, CA, USA

Eun-Young Kang
Computer Science, CSULA
Los Angeles, CA, USA

***Abstract*— There are two important observed facts for this research. One is that computer games with social interactions have become strong cultures to young students. The other is that introducing a computer programming language to majority of students is challenging as programming paradigm and concepts are completely alien to most of them. In this research, we introduce a Multiplayer Online Role Playing Game (MMORPG) named "DeBugger". It is developed to utilize the MMORPG online culture to introduce computer science education to improve learning outcome. Considering the increase in awareness and early exposure to programming at high school or even at middle school, developing publicly available multiplayer online educational games for CS students is a timely and sought effort. The DeBugger aims to build a collection of educational mini games within a virtual community of learners where players fight with bugs by solving problems, interact with other players and continue to play on smart phones during mobile gaming. The DeBugger is designed to take advantage of social interactions, mobile games and community to retain players longer, promote players to solve more quests, and encourage players to discuss and learn from each other more actively, which encourages peer learning and tutoring. A positive efficacy testing of the DeBugger on learning and social interaction among students will be discussed in paper.**

Keywords*— *Multiplayer Online Role Playing Game; Video Game; Java; Programming; Computer Education; Mobile Games

I. Introduction

Recent computer science education strongly recognized the significance of abundant exposure at an early age in improving one's computational thinking and problem solving approach [1,2]. Similar to Algebra, for example, where concept of variables in algebraic expressions takes some exposure time to get familiar to most 5-7th grade students, concept of variables in programming language also takes some time to become familiar. Moreover, learning computational approach usually takes sufficiently iterative practices for core concepts (variables, data types, logical flow, conditional statements, iteration, objects, methods, etc…) for most students whether young or not. To shorten the learning curve and to entertain the iterative learning process, game-based learning tools can be an effective aid for computer education.

MOOC (Massive Open Online Course) is one of noticeable approach in education. Among many computer science courses, introductory computer science course (CS1) has become pilot course for MOOC and diverse approaches are being explored. Majority of these approaches focuses on lecturing mechanism and interactions between the instructor and students even though peer interactions have served important roles for learning explicitly or implicitly [3,4,5]. As much as effective instructor-student interactions are explored for MOOC, similar effort should be made to explore and develop means to improve peer interactions. Here, we argue that virtual community where students can meet each other and play collaboratively or competitively via educational game challenges will serve as virtual labs and provide an atmosphere for desirable peer interaction. In addition, DeBugger game can be continuously played on Android and iPhone while students are away from computer.

Section 2 discusses related work; using multiplayer game for educational purposes and other game-like approach for CS education. Section 3 discusses the principles of designing the DeBugger game and section 4 discusses implementations of Android version and cross-compilation of the game for iPhone version, community features within the DeBugger game, and game architect. Section 5 presents results of user trial and on-going efficacy evaluation and implementations. Section 6 concludes the paper.

II. Related Work

Game-based learning has been an active area of educational research with hopes that playing game would effectively aid students' learning. Large number of studies showed the educational impact of the game-based approach [6].

Educational games (also called edutainment, serious game, gamification) have been adapted and used in diverse areas such as elementary math to professional medical trainings [7,8]. Recent study shows that game-based education reaches beyond original objective of educational and motivates players to achieve much more [9]. Cooper [9] utilized multiplayer online game as social interactions serve well to motivate players. Other studies also showed significance of social interactions in the game to retain players longer. Figure 1 shows the general play time of stand-alone game vs. MMORPG game; games with social interactions keep players longer over time [10]. We observed that popular MMORPG games are retaining their players for very long period time (months to years) and invite friends to play together (spread through human network). Recently, Farmville, a relatively simple farm nurturing (crops or farm animals) Facebook game made a nice show case of how quickly a game can spread through social interactions (80 million active players) and keep players playing repeatedly.

Hours per week	MMORPG players	Non-MMORPG players
0–1 h	1%	11%
1–2 h	5%	38%
3–6 h	10%	35%
7–10 h	13%	7%
11–20 h	25%	2%
21–40 h	34%	4%
40 or more h	11%	2%

Figure 1. MMORPG vs. non-MMORPG games, hours played; Source: Addiction to the Internet and Online Gaming.

Popular MMORPG games like World of Warcraft has several millions players. Lots of research has been done to analyze the main driving factors [11,12]. These studies pointed out that players especially enjoyed the fact that there is a community of audiences to give compliments when player achieved high level or completed a quest. It is an important finding which can significantly increase educational games' efficacy. Educational games that are well designed for good educational impact often suffer from the fact that players do not play voluntarily over and over. Another strong trend of young generation is continuous usage of smart phone at all times. Mobile DeBugger game is created to take advantage of this mobile trend, to encourage students to connect to other players, and seamlessly continue playing DeBugger game while students are away from computer.

The MMORPG game, "DeBugger" was designed and developed to make use of these findings to assist Introductory Computer Science course. It is widely accepted that programming is very difficult to learn[13]. For example Bergin and Reilly noted that it is well known in the Computer Science Education (CSE) community that students have difficulty with programming courses which can lead to high drop-out and failure rates [14]. The concepts in CSE are non-intuitive or/and overwhelming. Students need lots of exercises to digest the concepts, with close interactions, guidance and help when they are lost. Average pass rate for CSE course at colleges across US is around 60% [15]. Considering that significant shortage of workforce in CS [16], it is imperative to develop and provide publicly available educational game for CS students to assist students as well as instructors.

III. DEBUGGER GAME

The DeBugger is a MMORPG game where there is a persistent virtual world that continues to exist and progress even after a player exits the world (i.e. stop playing the game). In that world, players are represented by their characters with status properties like game level, virtual money, health, list of friends, game items like weapons or tools.

Design principles of the DeBugger game are; (1) Make use of players attachment to the character. In commercial MMORPG games, players have shown strong attachments to their online status (or their own characters). To be able to level up their characters, players can enjoy irksome process –known as boring, time consuming and repetitive tasks [17]. Players happily spend hours and hours to succeed in a quest. While there is pleasure of succeeding in quests itself, players are strongly motivated by the fact that their successes in quest result in level-up or acquisition of special awards and recognition in the community. Therefore, to level up characters in DeBugger, players are often observed to play DeBugger longer with focus to keep their rank high. (2) Studies have shown that students learn from peers as much as from teachers [18,19,20,21]. Students tend to ask questions for clarifications among peers first and then ask questions to teacher if no answer can be found among peers. In addition, peer pressure pushes students not to be left behind. Online game communities have shown that experienced players take great pleasure helping novice players. These advanced players can be further motivated to help new players like a TA giving close individual interactions that teachers may not give in real world. Frustration from the learning can be alleviated by sharing the similar troubles of individual with peers. The peers can be connected in the virtual world, available when needed, and increase educational impact. (3) For timid students, they can hide behind their virtual character, to avoid the fear of failure. (4) Importantly, typical MMORPG games advance and adjust depending on the needs or demands of the player community unlike console games that have pre-fixed settings once manufactured. The DeBugger game can advance and adjust the same way. In addition, the DeBugger can collect through the game server, a massive amount of educationally meaningful data based on the players' activities. These data can be useful for iterative design cycle (design, development, user test) of the DeBugger game for various users to maximize their learning outcome.

One thing to note is that the DeBugger game is not intended to replace traditional classes, but to be used as an effective support tool like a virtual lab with a TA that provides repetitive practices (lab) and feedback and clarifications from other players (virtual TA), so students can master the core pedagogical components to then progress smoothly.

In the DeBugger game, there are many mini games that players can play with other players or against bugs; the DeBugger title was inspired by the origin of the word (removing a bug from computer to fix errors). These mini games (games inside of DeBugger game) are intended for players to develop competence in computer science concepts and fundamental computational approaches when played repeatedly. In the DeBugger game, players 1) fight with bugs by solving flashcard games - chapter review questions after reading text book, 2) play "CodeGame" to solve a given mission by dragging and dropping the proper code segments in correct sequence, 3) play "Variable" games to practice the variables, data types and operations in Tetris style game, 4) play board game with other players by solving quiz style questions, and 5) interact with other players via chatting, friends list, reputation list and other features.

IV. IMPLEMENTATION

A. Mini Games

Students who take first Computer Programming courses are introduced to the core concepts of problem solving with pseudo-code, variables, data types and operators, flow controls (selection and iteration), methods, strings, and arrays in sequential order with slight variations. Each mini game has a specific focus so it is designed to work for one concept or one group of a few related concepts.

Figure 2. "CodeGame" Interface: a mission is displayed at upper left corner and the answer segments to drag are displayed below the mission. A segment can be dragged and dropped individually.

Figure 3. "Variable" game. Player is supposed to fill in the left column where variable slots are, according to instructions that appear one at one time on the right column.

The mini game called "CodeGame" is for learning and practicing program structure, importance of syntax rules, and simple program flow. When the mini game starts, a player will receive a simple mission that is randomly assigned according to the player's current level with a list of possible answer segments displayed on the left panel as shown in figure 2. To solve the mission, the player should pick up correct answer segments and then arrange them on the right answer panel in proper sequence by drag-and-drop. The goal of this mini game is to enable the player to comprehend the functionality of given answer segments and then focus on the program logic. The answer segments can be either actual program codes or abstract pseudo code as shown in figure 2. Abstract high level code helps students learn problem solving using top-down or divide-conquer approach without thinking of the syntax details. As for the actual program code answers, the mini game presents both grammatically correct and incorrect answers. In this case, to solve the mission, the player must know the precise language syntax and this allows the player to review important language syntax. Screen capture of playing this mini game is recorded and available on YouTube (http://www.youtube.com/watch?v=wfFW7mfd1M4).

"Variable" game is designed to teach concepts of variable, data type and operations associated with data type. For example, the minigame shows instructions related to a variable 'i' type of int, such as i++, then the player should calculate the value for the variable and enter the value to a variable slot for 'i' on the left column within a given time. An instruction can involve more than one variable such as j = i/3. In this case, the player should read value from i slot and perform integer division. When answer is wrong or the player has reached the time limit, then the row turns to red and stays red, while the right answer turns the line green and clears out. Wrong answers stack up in red rows and the player loses when the stack reaches the top (see the figure 3). Also, the screen capture of this mini game being played is recorded and available on YouTube (http://www.youtube.com/watch?v=vtcVQG-n8e0).

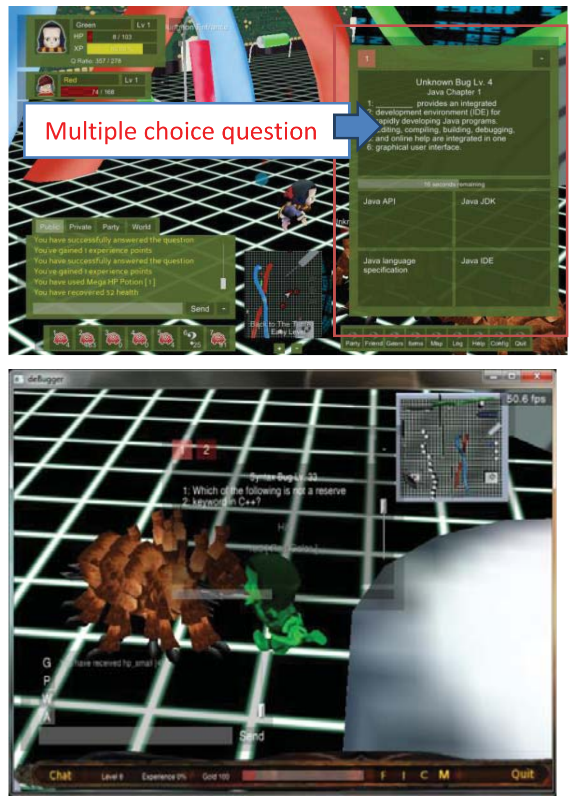


Figure 4a (top) & 4b (bottom). A player is fighting with a bug. Bug asks a question and the player has to answer. Health decreases as time lapses. Player's health, level, and gold are shown at the upper left corner. Lower left corner panel serves as a message panel and chat panel.

Based on the designed game architecture, many more mini games can be easily added to allow players to be exposed to further concepts continuously with fun (Figure 4a & 4b).

Players can seamlessly continue to play the game while they are away from the PC. Mobile DeBugger game allows for mini games play and maintain the player's inventory, score and all other belongings since they share the same account as the game server and DB server. Currently players can solve multiple choice questions that enable collection of stones for correct answers, so players can directly fight with bugs with stones as weapon. Mobile DeBugger makes use of touch pad for throwing stones in natural pattern, GPS information for finding friends nearby, and accelerometer for shaking off bugs or wrong options during the play.

Figure 5. Screen capture of Mobile Debugger. Players can throw stones to the bugs using the touch pad screen,

B. Social Activity

There are many social activity features within the DeBugger game. Players can chat with other players in a few different ways. Public chat is for chatting with anyone in the same virtual space (room) and private chat is for chatting with a specific person that a player wants to chat with. Also, it is possible to chat with friends who are online. A player can manage a friends list. Player can choose to show their level and other performance to their friends or the public

Player's achievement can be measured by levels, number of questions played and the accuracy ratio. Achievement board displays the top 10 high score achievers to encourage players to play longer and better (Figure 6).

Figure 6. – Player's performance is measured in diverse ways, and players can see other players' performance unless blocked by the player.

Social gaming and MMORPG are relatively recent game genres as they gain popularity over the availability of Internet. Social interactions within the game generates excitement of the game which allows popular games to easily develop communities of multimillion players. Once community of learners is created, students can learn not only from educational game components but also from their peers through discussion of problems with them. The DeBugger game was created with the vision of nurturing such a community of learners playing collection of games together, helping each other and inviting more friends to join.

The DeBugger is not only for students. The tool assesses students' learning outcomes automatically and provides useful information (e.g. grading of homework) back to teacher. This is very useful and efficient in that teachers save time for grading and students receive feedback immediately. Figure 7 shows the data available for teachers.

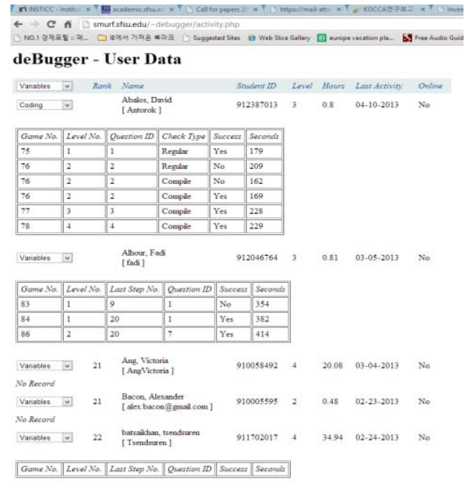


Figure 7. Website where teachers can see all of the students' performance in DeBugger.

C. Game Architecture

For DeBugger game, we need to create and support a virtual world that exists continuously on the game server. The game server runs at all time, is used by all clients to play the game, and updates the Database Server. A game client establishes a connection to the game server when a player wants to play the game, receive the latest status of the virtual world that has changed since the last time player logged out, and allows the player to interact with other players who are online. The database server keeps all the data of individual players (health, level, money, game items, friend's list, performance, and etc) and other data to maintain the game world continuously. An efficient game protocol between the server and clients is developed to make the communication effective. The DeBugger game also has a bug server that maintains all the bug characters – how often they appear (spawn), what kind of game item they drop, how aggressively attack players, etc. The bug server also controls bugs' behaviors, such as wandering and reacting to collisions, with simple AI (Artificial Intelligence) to improve the player's fun experience with the game.

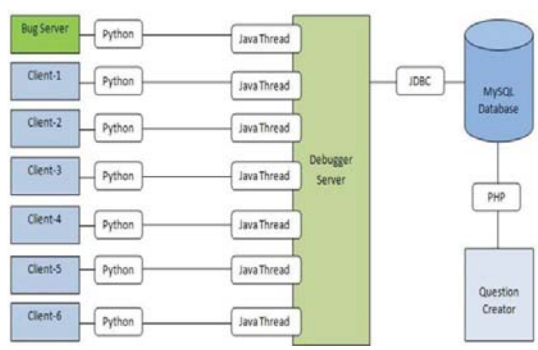


Figure 8. DeBugger Architect. It shows the relationship between a Client and the Server

Figure 8 shows an architect of the DeBugger game, depicting each component and their connections to each other. The game server was developed using JAVA and MySQL as DB server. Bug server was extended from game client that already included collision handling. The game client was developed utilizing Panda3D, an open source game engine, and python scripts. Figure 9 shows the further development of mobile DeBugger connecting game server directly, supporting seamless game play for players. We also employ XMLVM cross compilation technology for effortless porting from Android to iPhone version.

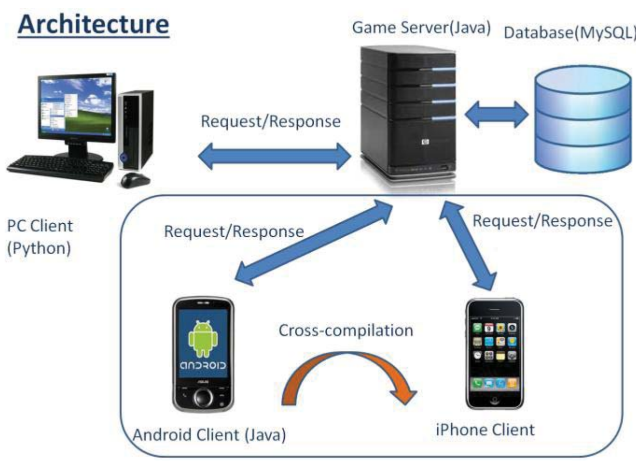


Figure 9. DeBugger Architect including Mobile Debugger. This shows the relationship between a PC Client, Android Client, iPhone Client and the Server

Stable support of the DeBugger game server is critically important for MMORPG game. We evaluated the scalability utilizing Clouds Computing acting as individual clients and the DeBugger server can stably support 100 concurrent players at more than 100 heartbeat per second performance. Figure 10 shows a screen capture showing all the clients running in the Amazon Clouds in a same room [22].

Figure 10. Testing Server Scalability using Amazon Clouds.

V. Results

A. Project Efficacy Testing

The DeBugger game has been used at SFSU CSc 210 Introduction to Programming using JAVA during Spring 2012 (pilot test) & Spring 2013 (official efficacy testing). Usability and learning impact of the DeBugger is continuously being evaluated. During the Spring 2013, 29 student participants were recruited from 4 introductory Computer Science programming classes. Seventeen students were introduced to DeBugger and were asked to play at the minimum 1 hour a week for the first 2 months as a part of their learning (Experimental Group) while 12 students were not given such exposure to DeBugger (Control Group). All students were given a pre-test that examined their general knowledge about JAVA programming (e.g., Data types, Variables, Syntax flow) in order to set the baseline performance scores. Mixture of forced choice (i.e., true/false and multiple choice) and open-ended problem solving questions were included. Once pre-test was completed, students in the Experimental Group were given the access to DeBugger for the next 8 weeks. To examine the effectiveness of playing DeBugger while taking CSc 210, all students were also given a post-test at the end of the study. The questions on the post-test were identical to the pre-test and students' improved scores from pre- and post-test across two groups were compared. We are currently in the process of analyzing the data. However, our preliminary data analyses are already showing positive impact of playing DeBugger on student learning outcomes.

Figure 11 illustrates participants' mean percent correct on the pre- and post-tests. It was found that overall, students' general knowledge about JAVA programming improved significantly from pre-test to post-test, $F(1,27) = 7.49$, $p = 0.01$, $\eta 2=0.22$. However, such significant improvement was driven mostly by the Experimental Group as alluded by the significant interaction between the group and testing phase, $F(1,27) = 4.6$, $p = 0.04$, $\eta 2=0.15$. Indeed, follow-up post-hoc analysis confirmed that students' performance improved significantly from pre- to post-test only in the Experimental Group where DeBugger was integrated into their learning experience (pre-test: $M = 44.12$; $SD = 13.25$; post-test: $M = 64.71$; $SD = 16.25$), $t(16) = 4.42$, $p < 0.001$. For the Control Group, there was no significant improvement in students' scores from pre- to post-test (pre-test: $M = 54.17$; $SD = 11.65$; post-test: $M = 56.68$; $SD = 20.6$), $t(11) = 0.33$.

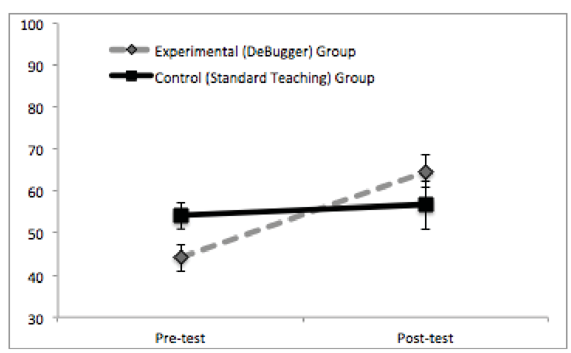


Figure 11. Students' Mean Percentage Scores on Pre- and Post-tests.

B. Student Learning Experience

The ffollowing comments from students in the Experimental Group provide noteworthy social cooperative and interactive nature of the DeBugger experience that seems highly conducive for acquiring knowledge while enjoying the process of learning:

"We played the board game for pretty much the whole hour. It was fun. I explained as much as I could to John, and we had a great time. Debugger is a great place to talk about coding and Java and just hang out. "

"There were two other players from CSc 210 that played Debugger today. We played the bugs and board game for about 30 - 40 minutes. It was really fun, and we talked a lot about programming and school and stuff. When we both played the board game at level three, and it's more fun at a harder level because both players can talk about the problem and help each other out with the solution."

"I was able to explain to the other player about short circuiting, compile/run time errors, and other stuff. It was time well spent. I really enjoyed playing Debugger today."

Based on the present research findings, we advocate for educational studies that thoroughly investigates the learning impact of individual mini games for pedagogical practices and consequential student learning experience (e.g., peer interactions in virtual community).

C. On-Going Development

Further developments are in progress. First, additional mini game to add is one for practicing selection flow control concept. It will utilize idea of racing game and player's route will be selected depending on the value in the car and the condition at the branch. To be able to finish the lap in given time, player should choose the value wisely, to utilize the conditions at the branch, and to stay at desired routes. This game will be extended to repetition flow control by running multiple laps. For the beginners, the car move very slowly to give enough response time and the car will accelerate as player progresses well. Second game to add is for practicing repetition flow control, specifically nested repetition flow control using pattern printing style arcade game. The third addition is a game for practicing array and repetition flow control.

VI. Conclusion

We introduced a MMORPG game, "DeBugger", that can be used as a primary assistance tool to improve students' learning outcome for Introductory Computer Science course through inherent properties of the MMORPG: addictive and fun game play for repetitive activities and social interaction which results in longer retention and active peer interaction/tutoring. The evaluation study shows that the impact of the game is effective and socially cooperative as a virtual lab and space for peer instruction.

Acknowledgment

This project is being partially funded by National Science Foundation NSF DBI- 0543614 and NSF TUES 1140939. CSc 631/831 Multiplayer Game Design and Development course students at fall 2009 contributed to majority of game concept, art work, DB, game contents, client and server implementation.

References

[1] V. Barr and C. Stephenson, Bringing computational thinking to K-12: what is Involved and what is the role of the computer science education community?, ACM Inroads, Vol. 2 No. 1, March 2011.

[2] Lee, Irene, et al., Computational Thinking for Youth in Practice, ACM Inroads, Vol. 2, No. 1, March 2011.

[3] C. Greenwood, J. Delquadri, and R. V. Hall, Longitudinal effects of classwide peer tutoring, Journal of Educational Psychology, Vol 81(3), Sep 1989, 371-383.

[4] A. King, A. Staffieri, and A. Adelgais, Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning, Journal of Educational Psychology, Vol 90(1), Mar 1998.

[5] A. Kolb, The evolution of a conversational learning space, In A. C. Baker, et al. (Eds.) Conversational learning: An experimental approach to knowledge creation. Westport, CT: Quorum Books, 2002.

[6] Papstergiou, M., Digital Game-Based Learning in High School Computer Science Education: Impact on Educational Effectiveness and Student Motivation, Computers & Education, v52 n1 p1-12 Jan 2009.

[7] GREEN, C S, AND BAVELIER, D. 2003. Action video game modifies visual selective attention. Nature 423 (2003), 534-537, Letters to Nature.

[8] PRENSKY, M. 2001. Digital Game-Based Learning. McGraw-Hill, New York. ISBN 0-07-136344-0.

[9] Cooper, et al, Predicting protein structures with a multiplayer online game, Nature 466, pp 756-760, August, 2010.

[10] Ng, Brian D. and Peter Wiemer-Hastings, Addiction to the Internet and Online Gaming, Cyber Psychology & Behavior 8.2 (2005)

[11] Seungkeun Song, Joohyeon Lee (2007), Key factors of heuristic evaluation for game design: Towards massively multi-player online role-playing game, International Journal of Human-Computer Studies, no. 65, 709-723

[12] Yee, N. (2002) Understanding MMORPG Addiction

[13] T. Jenkins. On the Difficulty of Learning tp Program. In Proceedings for the 3rd Annual conference of the LTSN Centre for Information and Computer Sciences , Loughborough, UK August 27 - 29, 2002.

[14] S. Bergin and R. Reilly. The influence of motivation and comfort-level on learning to program. In Proceedings of the 17th Annual Workshop of the Psychology of Programming Interest Group pages 293-304, University of Sussex, Brighton UK 2005.

[15] Jens Bennedsen and Michael E. Caspersen, Failure rates in introductory programming,
SIGCSE 2007, pp 32-36.

[16] Computing Education for 21st Century (CE21) Program Solicitation NSF 12-527 http://www.nsf.gov/pubs/2012/nsf12527/nsf12527.htm

[17] S. H. Hsu, , M. H. Wen, and M. Wu, Exploring user experiences as predictors of MMORPG addiction, Elsevier, Computers & Education, Vol. 53, No. 3, pp. 990–999, November 2009.

[18] Christudason, A. Successful Peer Learning, http://www.cdtl.nus.edu.sg/success/sl37.htm

[19] Falchikov, N., Blythman, M., Learning together: peer tutoring in higher education.

[20] Johnson, D.W.; Johnson, R.T.; & Holubec, E.J. (1993). Circles of Learning. Edina, MI: Interaction Book Company.

[21] Kaufman, D.B.; Felder, R.M.; & Fuller, H. 'Peer Ratings in Cooperative Learning Teams'. Proceedings of the 1999 Annual ASEE Meeting, ASEE, Session 1430'.

[22] Athalye, A., "Extending the DeBugger Game by Adding 3D Models and Testing Server Scalability" SFSU CS Master Project, 2011, http://cs.sfsu.edu/techreports/11/CE-11.07.htm

Scenes extraction from videos of telementored surgeries

Paolo Buono, Giuseppe Desolda, Rosa Lanzilotti
Dipartimento di Informatica
Università degli Studi di Bari "Aldo Moro"
via Orabona, 4, 70125 - Bari, Italy
{paolo.buono, giuseppe.desolda, rosa.lanzilotti}@uniba.it

***Abstract* - The huge amount of videos, available for various purposes, makes video editing software very important and popular among people. One of the uses of video in medicine is to store surgical operations for educational or legal purposes. In particular, in telemedicine, the exchange of audio and video plays a very important role. In most cases, surgeons are inexpert in video editing; moreover, the user interface of such software tools is often very complex. This paper presents a tool to extract important scenes from surgery videos. The goal is to enable surgeons to easily and quickly extract scenes of interest.**

Keywords: Telemedicine, video editing, scene extraction

I. INTRODUCTION

Video editing software is extremely popular today. Both, commercial [11] and open source [8] software present many useful features that allow their users to cut scenes, merge scenes, apply transitions as well as add soundtracks, video and audio effects, etc. The user interface of video editing software is often not easy to use for occasional users.

In medicine, and especially in laparoscopy surgery, videos are becoming common in surgical practices. However, many surgeons do not have adequate skills for video editing. Laparoscopic surgery, also called minimally invasive surgery (MIS), is a modern surgical technique in which the surgeon performs several small incisions (typically 3 or 4 incisions 0.5–1.5 cm in length) in the abdomen of a patient through which tools and the endoscopic camera are inserted. During the surgery, images are displayed on large monitors that magnify the area of interest. The whole surgery video is stored in order to be used for educational and/or legal purposes.

Surgery videos are usually long (at least 2 hours) and often only a few scenes are relevant. Moreover, typical functions of video editing software, like video filters, transition effects between scenes, advanced export functions (e.g. publishing on YouTube, Facebook, etc.), multi-audio/video tracks, etc. are not needed in the case of the surgery videos and may only confuse surgeons. We are involved in a research project aimed at developing a telemedicine system for supporting surgeons learning new surgical techniques. The system will provide different tools to support the surgeons work. Such tools are designed in order to accommodate the needs of their users. In particular, the aim is to reduce the complexity of the software tools by providing surgeons only those functions that they need. In this paper, we describe a tool that allows surgeons to easily and quickly perform the main task they are interested in, i.e. extract scenes of interest from a surgery video.

In order to create a tool able to satisfy needs and requirements of the specific category of surgeons, we adopted a participatory design approach [5]. A contextual enquiry was performed at the "Perrino" Hospital in Brindisi (Italy), which is actually a partner of our project. We carried out interviews and focus groups with several surgeons. Naturalistic observation of surgeons while performing laparoscopic surgeries or using video to teach to younger surgeons was also performed.

These studies showed the difficulties surgeons had in using software for video editing and provided hints for the design of a tool that can support surgeons in analyzing video and retrieving scenes of interest.

Next section presents a brief state of the art in video editing. Afterwards, we illustrate two alternative interfaces of the tool for extracting scenes. Then, the results of a formative evaluation in which these two alternative interfaces have been compared are described. The running version, implementing the prototype that resulted more effective and usable is later presented. Last section concludes the paper.

II. RELATED WORK

Video manipulation and, in particular, video summarization are gaining increasing interest due to the proliferation of digital camcorders, online video databases (e.g. YouTube), videos collected on large storage device, etc. Video summarization aims at extracting, from a long video, scenes that are more relevant for a certain purpose.

Some commercial tools for video summarization are already available. Examples are Windows Movie Maker, Pinnacle Studio, and Adobe Premiere. Such tools provide many functions, whose complexity confuses users that, occasionally, use them. Research on video summarization presents many different approaches. We briefly report here some of them.

AVST (Automatic Video Summarizing Tool) utilizes MPEG-7 visual descriptors to generate video thumbnails to search for similar scenes and cluster scenes [15]. AVST splits scenes also according to abrupt and/or gradual transitions. It works well for videos characterized by scene changes, like in sports or movies.

Jang et al. propose an algorithm that uses visual and audio content to automatically generate improved video summaries [13]. The system performs audio segmentation and classification according to audio and visual diversity, face quality, and overall image quality. These characteristics are gathered from users, who provide feedbacks on summarized videos. The system has been applied in the contexts of birthdays, weddings, shows, and parades.

Bailer et al. propose TRECVID, an interactive video browsing tool based on a multimedia content abstraction model [1]. The tool clusters scenes according to: camera motion, visual activity, audio volume, face occurrences, global color similarity, repeated takes and relations in multi-view content, in order to reduce the content to a manageable number of scenes.

Novel visual techniques in the field of video surveillance have been explored, an example is in [2] and [3] in which, in order to speed-up the selection process of interesting scenes, an interactive image of movements is created.

The previously mentioned approaches do not specifically address laparoscopic surgery videos, in which the camera is not stationary, videos are characterized by very similar scenes, the audio is produced by surgeons talking among them, and no face, or people characteristics are present in the videos.

Among several works carried out for summarizing videos in laparoscopic surgery, Leszczuk and Duplaga present a prototype that creates summaries of bronchoscopy video recordings [16]. The summarization algorithm removes poor quality frames due to blurry images. Such frames are unavoidable due to the relatively tight endobronchial space, rapid movements of the respiratory tract, and secretions that occur commonly in the bronchial tube, especially in suffering patients. During a classification phase, the algorithm identifies non-informative frames, which are discarded in the summarized video.

According to the opinions of surgeons we have worked with, shortening videos, by removing blurry or uninteresting scenes, can be useful but is not their primary goal. The problem addressed in our work is to choose a few relevant scenes that are representative for a video. It is worth remarking that the approach we propose does not rely on automatic algorithms only, but requires human intervention in order to reduce the possibility of errors and make sure that the final scenes identification are those required by the surgeons.

III. THE SCENE EXTRACTION TOOL

The tool, described in this paper, supports the extraction of scenes from video produced by a telementoring system that we have been developing in the last few months. Telementoring is gaining momentum. It is useful for surgeon training and can be used for consultancy requests to mentors working in different hospitals, cities or even continents [14].

Our telementoring system allows surgeons (learner) to be assisted by experienced remote surgeons (tutor) during a laparoscopic surgery. The system main components are the *Learner* and the *Tutor* (we distinguish Learner and Tutor devices from learner and tutor surgeons by capitalizing the first letter when referring to the devices). The Learner is a small device installed in the surgery room that sends video signals produced by the endoscopic camera and the audio of the surgery room to the Tutor. The Tutor is a different device available at the remote tutor surgeon location having the functionalities of audio and video I/O and a pointing feature (mouse, pen, touch, etc.). The Tutor sends audio, telestration and images to the surgery room. A high-level architecture is shown in Figure 1.

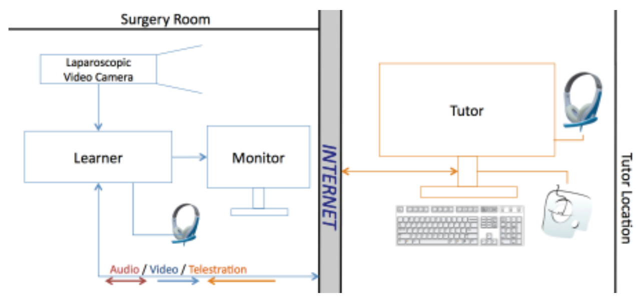


Figure 1: System architecture

In this paper, we refer to telestration [9] as the activity of drawing, sending images or applying a still image on a monitor from a remote location. Specifically, we use telestration for improving the remote communication between the learner and tutor.

During the discussions with surgeons in our participatory design team, important requirements emerged that were implemented in our telementoring system. Surgeons asked for the possibility to mark interesting moments of the surgery, like adding a bookmark while reading a book, in order to easily retrieve such moments that they would like to show later. To comply with this need, we provided the Tutor interface with a marking function that allows the remote surgeon to indicate relevant moments by pressing a button in the Tutor user interface.

Since surgeons, in specific moments, needed higher precision, provided the Tutor with a pause feature, which consists of visualizing, at the same moment, a high-definition still image of the learner camera to both tutor and learner monitors, on which the telestration is displayed.

Another feature, we added after a specific request of the surgeons, is the possibility to send images. Often, experienced tutors also teach, so they need to send images to improve their explanations. This picture is displayed like a paused image.

Each telementored laparoscopy surgery is video recorded by the telementoring system, which records also the audio interaction between learner and tutor. Surgeons often need to extrapolate a few meaningful scenes in order to teach, provide consultancy on a specific topic, show a novel technique at a conference, etc. Beside the recorded videos, our telementoring system stores XML files that contain all details of telestration data, marks, pauses and images sent. In order to support scenes extraction, our approach exploits data available in the XML file.

During the laparoscopic surgery, the tutor performs different actions, such as marking a certain time, pausing the video, performing a telestration, sending an image to the learner. A telestration can be a free-hand drawing or an arrow. For the tutor, such actions are performed in specific moments. Thus, in our tool, the identification of such actions is the starting point for retrieving scenes of interest and extracting them. Our approach is based on the visualization of actions in a compact form, which aims at helping the user to identify scenes of interest and save them in a personal area for future uses.

During the participatory design, several alternative designs and several low fidelity prototypes were proposed. In particular, two alternatives were better investigated: 1) visualize actions on a fixed area and show the details into another timeline; 2) visualize the different types of actions occurred in a given moment using a zoomable timeline. Furthermore, these two alternative prototypes present two different scene selection modalities, due to the different timeline visualization.

The two running prototypes were developed using Axure, a software for creating high fidelity interactive prototypes [10]. Figure 2 and Figure 3 show the two prototypes. In both cases, the main area is devoted to the video play and actions are visualized in a timeline. The panel on the right side is used to collect interesting scenes and see their previews.

A. Prototype 1

As shown in Figure 2, the interface of the Prototype 1 is composed of three main areas: the *play* area at the center of the screen; the *timeline*, at the bottom of the screen; the *preview* area, on the right side of the screen containing the selected videos.

The video is played in the *play* area. The play/pause button is located at the left-bottom corner of the *play* area.

The *timeline* was inspired by [6], which uses two timelines: one showing the overview of the video actions and the other one focusing on a small time span of the video. Indeed, as shown in detail in Figure 5, this area is divided into two parts: at the bottom there is an *overview timeline* visualizing data of the whole video, and on top of it, a *details timeline* containing details of the selection performed in the *overview timeline*. The blue pins in the *overview timeline* indicate all actions performed by the tutor. When the user clicks somewhere on the *overview timeline*, the details of about 1 minute video, centered at the clicked time point, is visualized in the *details timeline*. In the example of Figure 5, the user has clicked at minute 30:00. The system visualizes the details of the video from minute 29:30 to minute 30:42 in *details timeline*. Three icons represent three actions made by a tutor, i.e. a free hand draw, a pause action and an arrow, respectively. Moving the mouse pointer over these icons, a balloon shows a preview; by clicking on these icons, the corresponding part of video is played in the *play* area. To help the user to understand which part of the *overview timeline* is shown into the *details timeline*, the selection has a green border, like the *detailed area* border.

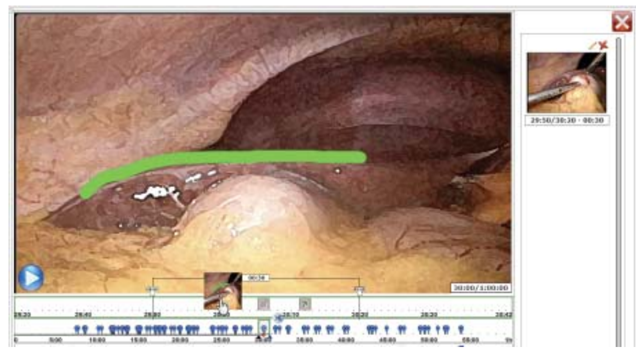


Figure 2: Screenshot of Prototype 1

One of the main goals of this paper is to bring out an interaction modality to easily extract scenes. The steps that the user has to perform with this prototype are:

1. Clicking on *overview timeline* to visualize the details;
2. Clicking the button represented by the scissor icon to visualize the selection function;
3. Resizing, if needed, the selection by using the handles;
4. Accepting or discarding the selection by clicking on SAVE button or X button respectively.

The *preview* area, on the right side of the user interface shown in Figure 2, contains thumbnails of saved scenes. At the top-right corner of each preview, the X button allows users to delete the selection, while the *pencil* button permits to change the video interval. By clicking on the *pencil* button, the previously saved selection appears again on the timeline in order to allow the user to change the begin/end of the video.

B. Prototype 2

As shown in Figure 3, the interface of the Prototype 2 is very similar to the first one, except for the timeline. In this prototype actions are visualized like in [4], in which the timeline shows different rows, each representing a type of action. Actions are organized into invisible tracks over the timeline, with each action allocated to a given segment of its track.

Moreover, these actions are vertically grouped into columns according to the concept of *indication*. We consider an *indication* as a group of actions performed by the tutor, useful to give suggestions to the learner about surgical procedures. An *indication* starts when the tutor draws something (free hands or arrow), pause the video or sends an image on which drawing something; the *indication* ends when the user deletes all actions or plays the video (if it was in pause or if picture was sent). Color, size and actions priority has been considered. We used the ColorBrewer online tool [7] to choose a set of 5 quantitative colors in order to associate the color to the type of action.

Figure 3: Screenshot of Prototypes 2

The size of actions is problematic. The columns width can be too tight. Typically, actions last in average for less than 20 seconds. On a common monitor with a screen resolution of 1024x768 pixels, the timeline is composed of about 800 pixels. In a video that lasts 2 hours (the average length in the surgical domain), 20 seconds can be represented in about 2 pixels wide. In order to make the information visible, we adopted two solutions. The first one is to start the video editing with a default timeline zoom that visualizes an interval of 30 minutes. In this way, the columns have an average width of about 10 pixels.

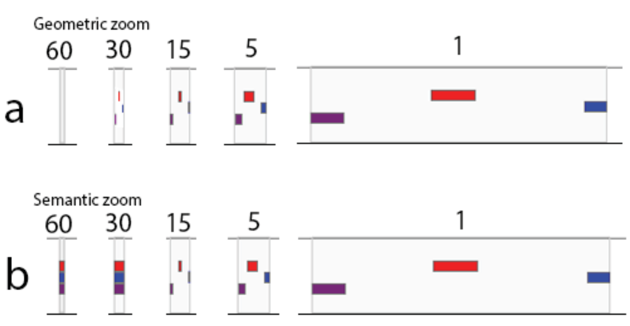


Figure 4: Comparison between action visualizations with (a) geometric and semantic (b) zoom

The second solution implemented a semantic zooming, which visualizes different levels of detail in a view when zooming in and out. If the system implements a geometric zoom (case A of Figure 4), when the user zooms, out over a certain level, the visual information about the actions disappears.

In the adopted semantic zoom, when the user zooms out and the visual information about the actions is too small to be visible, the actions have the same width of the containing column (e.g. Figure 4, case B, see details at 60 and 30 minutes). In any case, the group of actions is at least 2 pixels wide.

Finally, we ordered the action track according to an importance criteria expressed by interviewed surgeon. They considered the *mark* as the most important action, since it is the only one that the tutor performs explicitly. After, they considered the *image* as the next important, then *pause* function, lastly, *free hand* and *arrow*.

Similarly to Prototype 1, we designed a feature to select a scene. The green rectangle in Figure 6 starts at minute 30:00 and contains a scene with two bars composed of 3 and 2 actions, respectively. The green rectangle appears when the user clicks on the timeline and the 20 pixel default time span of the selection can be modified acting on the handles on both sides of the rectangle. A click on the floppy stores the selected scene and a corresponding thumbnail appears in the area at the right of the user interface. The user can zoom into the timeline and a preview of the scene is visible as a popup when the mouse pointer is moved over a specific action.

IV. FORMATIVE EVALUATION

As a part of formative evaluation during the early development phase, a user study was performed in February 2013 to get feedback from the intended users about which one of the two different prototypes is more usable and more appropriate for their main tasks. The study involved 6 surgeons and was performed in the field, i.e. in the surgeons' office at the Perrino hospital in Brindisi.

The 6 surgeons (5 males) were tested separately. The thinking aloud technique was used to evaluate the prototypes. Each test consisted of two phases, each one for analyzing a prototype. In order to avoid the learning effect, 3 surgeons first interacted with the Prototype 1 and, then, with the Prototype 2; the other 3 surgeons used the two prototypes in reverse order. At the beginning of each phase, the surgeon was given a brief introduction on the prototype to be used and its main functions. After this, the surgeon performed five predefined tasks and, finally, s/he was interviewed to collect data about her/his opinions on the used prototype. At this point, the other phase, in which the surgeon had to interact with the other prototype, started. This latter phase followed the same procedure of the former one: surgeon had to perform the same tasks, but with the support of the second prototype.

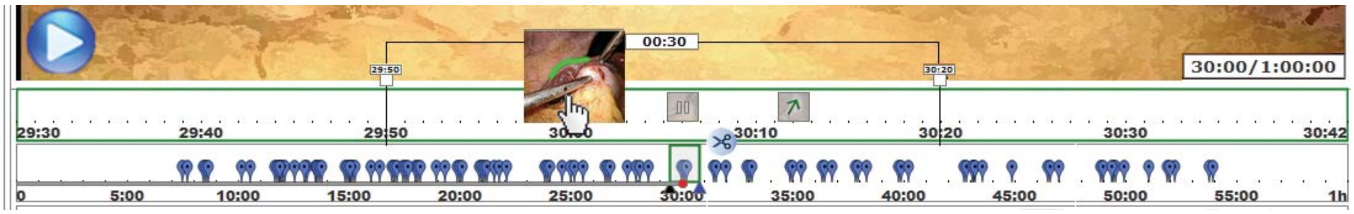


Figure 5: Detail of Prototype 1 with two timelines

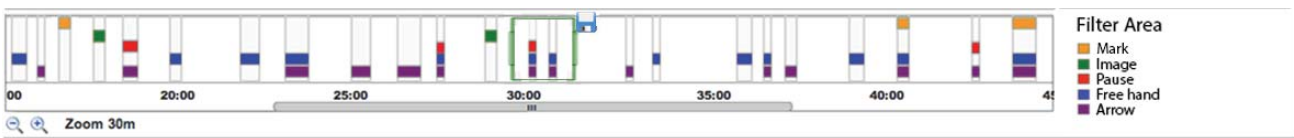


Figure 6: The timeline in Prototype 2 with semantic zoom

The surgeon was observed by two HCI evaluators and was videotaped. A video analysis was performed to collect data on the number of tasks successfully completed. Each test lasted about 30 minutes.

The five tasks were defined in order to allow surgeon to use the functions implemented in each prototype (i.e. play/pause, select, modify, save, delete). They were of different complexity. In order to analyze the ease of learning of the two prototypes, two tasks were very similar. In order to accomplish a task, surgeons have to do more than 2 steps.

The success rate was calculated for all tasks performed with the two prototypes [18]. Specifically, it resulted 54% for the Prototype 1 and 58% for the Prototype 2. Generally, no significant difference emerged. However, it is worthwhile to highlight that Prototype 2 better supports its users in performing the selection of a scene from the telementored video. In fact, only one user successfully completed this task with the Prototype 1, 3 users partially accomplished it and 2 users did not finish it. Regarding the Prototype 2, 4 users successfully completed the task and 2 users did not able to finish it.

Another important result concerns the ease of learning that was analyzed by the difference between the success rates of the two similar tasks for each prototype. Specifically, the success rate increased of the 8% for the Prototype 1 and of the 42% for the Prototype 2. This showed that Prototype 2 seems to be more easily learnable than the Prototype 1.

The thinking aloud was instrumental to identify usability problems of both prototypes. Specifically, all surgeons did not understand in what the two timelines visualized in the Prototype 1 differs. On the other side, it was not so clear to the surgeons that the labels, on the right side of the Prototype 2, were not only a legend, but they also were filters, which permit to refine the scene search. Other interaction difficulties surgeons concerned limits of the rapid prototyping software. In other words, if surgeons did not tightly follow the time indications given in the task definition (for example, “Select a scene starting at 29.50 min and ending at 30:20) they did not able to accomplish the specific task.

The interviews were useful to collect opinions and, especially, suggestions to improve the prototypes. Surgeons were agreed that the Prototype 2 was easier to use than the Prototype 1. They said that the Prototype 2, differently from the Prototype 1, not only allowed its users to accomplish the tasks without serious difficulties, but also to have at a glance an idea of actions available in the video.

Two surgeons required a scrubbing mechanism to facilitate the scene detection. In fact, being inspired by the Prototype 1, they would like that the Prototype 2 provided a scene preview, which appeared in a small popup window when the user goes through the timeline. In this way, the video analysis could become more rapid.

Another surgeons’ request concerns the way in which the video can be annotated. They explicitly said that, during a classical laparoscopic surgery without the telementoring support, they would prefer to have a vocal command to annotate a scene of the video. For example pronouncing: “System, mark now!” and the system stores a vocal mark.

V. THE FINAL PROTOTYPE

Starting from the results of the user test, we developed a final prototype in Java. We used JavaCV framework to visualize the video and export the final summary. The input of the application is a video of telementoring and its XML file produced by the telementoring software. Figure 7 shows the interface, composed by different areas: (1) video player, (2) previews; (3) filter by action type; (4) Timeline. The software is similar to Prototype 2, since it has been shown to be more adequate for surgeons.

Another evaluation of the application user interface was performed with three surgeons chosen among those that participated to the prototype user testing. The surgeons carried out the same five tasks performed to evaluate the first two prototypes. All tasks were correctly accomplished without any problem.

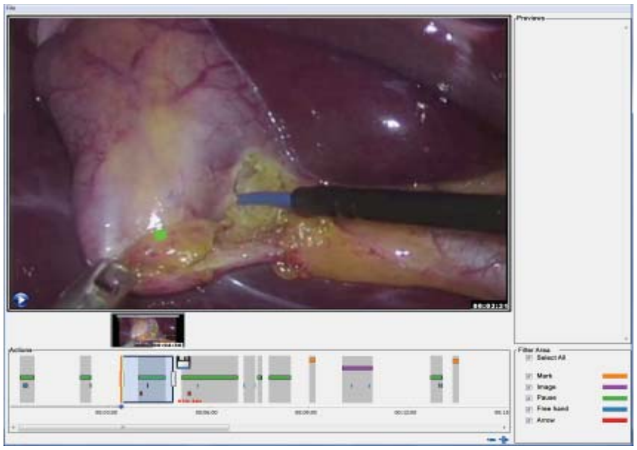


Figure 7: The current version of the video-editing tool

VI. CONCLUSIONS

In medicine and, especially, in laparoscopy surgery, surgeons have to edit surgery videos both for educational and legal purposes. It often happens that surgeons have difficulties in using video editing software given to its functional complexity. In fact, video editing software provides several functions, such as transition effects, advanced export function, etc., which surgeons do not use.

This work has presented the development of a tool that allows surgeons to extract scenes from a surgery video. Two different prototypes have been implemented and evaluated with end users in order to identify which one of them better supports the work of surgeons in extracting the important scenes.

The Prototype 1 provides two timelines: one visualizes data of the whole video and contains blue pins indicating actions performed by tutor; the other one shows details of a selected scene. The Prototype 2 shows a timeline containing different rows that represent the tutor's actions. Such actions are vertically grouped into columns, which provide actions performed by the tutor useful to give suggestion to the learner about the surgical procedures.

The performed usability testing revealed that end users preferred the Prototype 2; also some improvements to be implemented in the new version of the tool were suggested. For example, the use of advanced video scrubbing techniques [17] will be investigated to enhance the detection of interesting scenes. We are planning to implement in the telementoring system a speech recognition module [12] to give surgeons the possibility to vocally mark the videos.

In the immediate future, we will preform a comparison study to investigate which one of the two visualization techniques implemented in the two prototypes is more efficient to detect interesting moment in a video.

The analysis of a corpus of telementored surgery videos could be conducted to reveal actions patterns that allow the system to provide suggestions for the selection of important scenes.

Another improvement could be to export the selected videos as SCORM learning objects, in order to allow surgeons to easily integrate interesting videos into e-learning systems.

ACKNOWLEDGMENTS

We thank Professor Maria Francesca Costabile for her valuable guidance in our research.

REFERENCES

[1] Bailer, W., Weiss, W., Schober, C. and Thallinger, G. 2012. A video browsing tool for content management in media post-production. In Proceedings of the Proceedings of the 18th international conference on Advances in Multimedia Modeling (Klagenfurt, Austria, 2012). Springer-Verlag, 2189122, 658-659.

[2] Buono, P. 2011. Analyzing video produced by a stationary surveillance camera. In Proceedings of the DMS (2011). Knowledge Systems Institute, conf/dms/Buono11, 140-145.

[3] Buono, P. and Simeone, A. L. 2010. Video abstraction and detection of anomalies by tracking movements. In Proceedings of the Proceedings of the International Conference on Advanced Visual Interfaces (Roma, Italy, 2010). ACM, 1843036, 249-252.

[4] Costa, M., Correia, N. and Guimarães, N. 2002. Annotations as multiple perspectives of video content. In Proceedings of the Proceedings of the tenth ACM international conference on Multimedia (Juan-les-Pins, France, 2002). ACM, 641065, 283-286.

[5] Douglas, S. and Aki, N. Participatory Design: Principles and Practices, 1993.

[6] Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., Merkl, H. and Pankanti, S. 2005. Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. Signal Processing Magazine, IEEE, 22, 2 (2005), 38-51.

[7] Harrower, M. and Brewer, C. A. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. In The Map Reader (2011). John Wiley & Sons, Ltd.

[8] Hsiang-An, W., Yen-Chun, L., Hua-Ting, L. and Shang-Te, L. Open Source for Web-Based Video Editing. Institute of Information Science, Academia sinica, Taipei, Taiwan, Province de Chine, 2012.

[9] http://en.wikipedia.org/wiki/Telestrator. Telestrator(

[10] http://www.axure.com. City.

[11] http://www.toptenreviews.com. Video Editing Software Review(

[12] Jacob, M. G., Li, Y.-T., Akingba, G. A. and Wachs, J. P. 2013. Collaboration with a robotic scrub nurse. Commun. ACM, 56, 5 (2013), 68-75.

[13] Jiang, W., Cotton, C. and Loui, A. C. 2011. Automatic consumer video summarization by audio and visual analysis. In Proceedings of the Proceedings of the 2011 IEEE International Conference on Multimedia and Expo (2011). IEEE Computer Society, 2193850, 1-6.

[14] Latifi, R., Peck, K., Satava, R. and Anvari, M. 2004. Telepresence and telementoring in surgery. Stud Health Technol Inform(2004), 200--206.

[15] Lee, J.-H., Lee, G.-G. and Kim, W.-Y. 2003. Automatic video summarizing tool using MPEG-7 descriptors for personal video recorder. IEEE Transaction on Consumer Electronics, 49, 3 (2003), 742-749.

[16] Leszczuk, M. I. and Duplaga, M. 2011. Algorithm for video summarization of bronchoscopy procedures. Biomed Eng Online, 10(2011), 110.

[17] Matejka, J., Grossman, T. and Fitzmaurice, G. 2013. Swifter: improved online video scrubbing. In Proceedings of the Proceedings of the 2013 ACM annual conference on Human factors in computing systems (Paris, France, 2013). ACM, 2466149, 1159-1168.

[18] Nielsen, J. Usability Engineering. Morgan Kaufmann Publishers Inc., 1993.

An Investigation of the Attitudes of Instructors and Students to On-line Assessment in Mathematical Subjects

Elizabeth Acosta-Gonzaga [1,2]
Niels R. Walet [2]
[1]National Polytechnic Institute
Mexico City, Mexico
[2] School of Physics and Astronomy
The University of Manchester
Oxford Road, Manchester, M13 9PL, UK
elizabeth.acostagonzaga@postgrad.manchester.ac.uk
niels.walet@manchester.ac.uk

***Abstract*—We study factors that facilitate or hinder the adoption and use of Information and Communication Technologies in the assessment process in STEM subjects involving mathematics. To that end, we investigate both the behavioural intentions of instructors and the attitudes of students. We construct two models, which incorporating aspects of established models of acceptance of information technology, such as the Technology Acceptance Model, the Theory of Planned Behaviour, and Social Cognitive Theory. Apart from the variables from these traditional models, in our model for instructors we propose a new empirical variable, "perceived usefulness to students". In the students' model three new empirical variables--"perceived reliability", "perceived suitability" and "feedback"--are added. The models are analysed with data collected from staff and students at nine STEM schools at the University of Manchester using two on-line questionnaires. A Generalized Linear Model is then used to analyse both datasets. For the instructors we conclude that "perceived usefulness" and "perceived usefulness to students" have a significant positive effect on the use of web-based assessment. "Attitude" has a negative effect. For students the evidence indicates that "perceived usefulness" also has a direct effect on students' attitude.**

Keywords—Educational technology adoption; e-assessment, web-based assessment.

I. INTRODUCTION

The influence of Information and Communication Technologies (ICT) on the learning and teaching process is significant. It is increasingly seen as a way to fulfil the requirements for learning in a modern society and this has created great demand from business and all stages of education (Sun et al. 2008). Its impact has widely expanded in all fields, but we are especially interested in teaching mathematics and realted subjects. The work of (Gunasekaran et al. 2002) gives some evidence of the effectiveness of using ICT to do this. As a particular example, they explain how (Larson & Bruning 1996) used a qualitative research methodology to explore perceptions in an interactive collaborative mathematics course; they found that a distance learning format gave teachers access to more resources, is useful for under-achieving students, and can be an effective way to implement national curriculum and instruction standards.

It seems less research has been conducted concentrating on the assessment process. Since it can be said that assessment is the culmination of the interaction and feedback of teaching and learning, it is a core component for effective learning (Gikandi et al. 2011).

The specific process of assessment using ICT has come to be known as e-assessment. This includes the entire assessment process from designing assignments to storing the results using ICT (Stödberg 2011). ICT can also play an important role building a useful link between the assessment, teaching and learning processes creating new opportunities and approaches. As (Bennett 1998) states, computer-based assessment offers prospects for innovation in testing and assessment.

This study tries to contribute to our understanding of the factors that influence students' and instructors' technology use. We shall concentrate on assessment in mathematics and allied subjects in a Higher Education setting. We want to understand the factors that influence adoption of successful e-assessment from both a lecturer and student perspective.

So we shall look at attitudes of students to the use of web-based assessment. Although students are very familiar with the use of the technology, they are nevertheless worried about the security of testing (Cassady & Gridley 2005), the possibilities of cheating (King et al. 2009) and the fairness of question banks (Dermo 2009). If students do not have sufficient confidence in a test, they can be affected in their levels of engagement and cooperation (Domino & Domino 2006).

Although both students and instructors are important stakeholders determining the success or failure of a system, the role that instructors play appears to be the most important (Selim 2007). We also investigate what motivates the instructors.

II. RESEARCH MODELS

Our work is both influenced by and based on previous papers which have considered the role of several key concepts, which we shall call "variables" since they will become variables in a model, for the study of a person's acceptance, perception, feelings and attitudes towards the use of ICT.

In the case of instructors' behavioural intention to use technology the research model and hypotheses were broadly based on the Technology Acceptance Model

(TAM), the Theory of Planned Behaviour (TPB), and Social Cognitive Theory (SCT). These are useful theoretical frameworks in their own right.

(Davis, F. D. 1989) introduced the TAM as a theoretical extension of the Theory of Reasoned Action (TRA). The TAM assumes that "perceived usefulness" and "perceived ease of use" are the primary drivers for technology acceptance. Here "perceived usefulness" is defined as "the degree to which a person believes that using a particular system would enhance his/her job performance" whereas "perceived ease of use" is defined as "the degree to which a person believes that using a particular system would be free of physical and mental effort" (Davis, F. D. 1989). The TAM states that "perceived usefulness" has a direct effect on a user's behavioural intention, and "perceived ease of use" affects behavioural intention indirectly through perceived usefulness (Davis, F. D. 1989). Several empirical studies have found support for this idea, see for instance (Venkatesh & Davis 2000) and (Wu & Chen 2005). Similarly, in the work by (Pynoo et al. 2012) the main drivers for instructors to accept technology are found to be "perceived usefulness" and "attitude".

We hypothesize that "perceived usefulness" can also be strongly affected by attitudinal beliefs. Previous studies found that there are important attitudinal factors that impact the intention to use technology. For example, (Benson Soong et al. 2001) found that instructors and students' technical competency, instructors' and students' mindset about on-line learning, level of collaboration, and Information Technology (IT) infrastructure are all crucial factors for successful applications on-line resources a course. Likewise, in the research of (Volery & Lord 2000) it is concluded that technological factors (ease of access, support, interaction, design, etc.), instructors' characteristics (attitude toward students, teaching style, technical competence, encouraging interaction, etc.) and students' characteristics all affect effectiveness of on-line delivery. In addition, the study by (Webster & Hackley 1997) mentions other factors for instructors such as teaching style, attitude, and control of technology. All of these affect a person's attitude towards using technology. These attitudes determine behavioural intentions, which in turn lead to actual system use.

On the other hand, several studies have extended TAM by adding other variables. For instance, in the work of (Ilias, Razak, and Yasoa 2009), the concepts of "perceived credibility", "information system quality", and "information quality" were investigated in the case of ICT acceptance amongst taxpayers in Malaysia.

Following this line of thought, we include "perceived usefulness to students" (PUS), the instructors' belief that students can improve their learning by using technology, as a empirical variable. We hypothesise that "perceived usefulness to students" will have a positive effect on instructors' behavioural intentions in the context of web based-assessment.

The TPB proposed by (Ajzen 1991) is another improvement on the TRA which holds that attitude, subjective norms (social influence), and perceived behavioural control are direct determinants of intentions that influence behaviour. Social influence (SI) refers to "the perceived social pressure to perform or not to perform the behaviour" (Ajzen 1991). The effects of social influence on behavioural intention are expected to be direct (Teo et al. 2008). Attitude (AT) refers to ''the degree of a person's favourable or unfavourable evaluation or appraisal of the behaviour in question'' (Fishbein M. and Ajzen, I. 1975). According to the TPB, attitude impacts users' behavioural intention. We also use attitude as a predictor of the behavioural intention to use technology.

Finally, the SCT (Bandura 1986) is an empirically validated model for understanding and predicting human behaviour and identifying ways in which behaviour can be changed. It has been applied as a framework to predict and explain an individual's behaviour in Information Systems (IS) settings. SCT proposes two cognitive factors, "performance expectations" and "self-efficacy", that influence individual behaviour. Self-efficacy is defined "as one's judgements and beliefs of his/her confidence and capability to perform a specific behaviour" (Bandura 1986). In order to understand the impact that self-efficacy has on instructors' behavioural intention we decide to include it as a variable in our model.

Not all of these factors are relevant for the student model. In this case we extended the TAM by including perceived suitability (PS), perceived reliability (PR) and feedback (FE). We define PS as "the belief that the technology used is appropriate to evaluate understanding of a specific topic". PR is considered as "students' belief that the technology assesses their work fairly and accurately". Finally, FE is "the students' belief that the support and feedback they receive from instructors and the software will improve their learning". Based on these theoretical frameworks we propose the following models for the instructors and students.

III. RESEARCH MODELS AND HYPOTHESIS PROPOSED

A. *Instructor model*

The model consists of eight variables: attitude (AT), perceived usefulness (PU), perceived usefulness to the student (PUS), perceived ease of use (PEU), behavioural intention (BI), perceived system satisfaction (PSS), computer self-efficacy (CSE) and social influence (SI) see Figure 1. We hypothesize (letter *a* for instructors) that:

H1a. PUS will positively influence instructors' behavioural intention to use web-based assessment.
H2a. PU will positively influence instructors' behavioural intention to use web-based assessment.
H3a. PUE will positively relate to instructors' behavioural intention to web-based assessment.
H4a. PSS will positively influence instructors' behavioural intention to use web-based assessment.
H5a. CSE will have a positive effect on instructors' behavioural

intention to use web-based assessment.
H6a. SI will positively influence instructors' behavioural intention web-based assessment.
H7a. AT towards use will positively relate to instructors' behavioural intention to use web-based assessment.

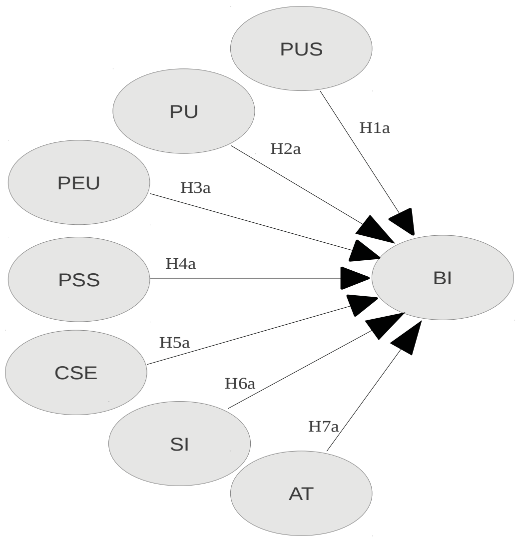


Figure 1. The instructor model. BI is the dependent variable, the other ellipses represent the independent variables. The labelled arrows are the hypotheses

B. Student model

For the student model five variables were used: perceived usefulness (PU), perceived suitability (PS), perceived reliability (PR), feedback (FE) and attitude (AT). In this model AT was considered as a dependent dependant while PU, PS, PR and FE are the independent variables as shown Figure 2. To establish the students' attitude towards the use web-based assessment and taking into account the variables proposed the following hypotheses (letter *b* for students) are proposed:

H1b. PU will positively influence students' attitude towards using web-based assessment.
H2b. PS will positively influence students' attitude towards using web-based assessment.
H3b. FE received will positively influence students' attitude towards using web-based assessment.
H4b. PS will positively influence students' attitude towards using web-based assessment.

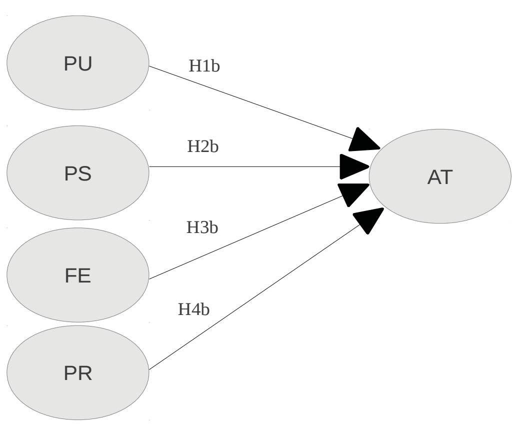


Figure 2. The student model. AT is the dependent variable, the other ellipses represent the independent variables. The labelled arrows are the hypotheses.

IV. RESEARCH METHODOLOGY

We used two separate on-line questionnaires to obtain data to test the empirical models for two groups. These questionnaires were applied during the same semester. The students' questionnaire consists of 15 items whereas that instructors' questionnaire consists of 32 items. Each item corresponding to each variable was measured using a five-point Likert scale with answer choices raging from strongly disagree (1) to strongly agree (5).

V. DATA COLLECTION AND INSTRUMENTS

34 responses were obtained from the instructors' questionnaire. Of these 2 were removed as incomplete in resulting in a dataset of 32 (9 female, 23 male). For students' questionnaire a total of 127 responses were obtained, 6 were removed because these were partially filled. We thus has a sample of 71 male and 50 female students. According to the student survey, the responders came from 9 academic schools at the Manchester University, with most of studying in the fields of physics and astronomy, mathematics, chemistry, mechanical, aerospace, and civil engineering. For the instructor survey the distribution by school is most from physics and astronomy, chemistry, earth, atmospheric, and environmental, chemical engineering and analytic science, electrical and electronic engineering. The survey shows that 5.88% have less than 1 year experience in teaching, followed bye 8.82% with 1-5 years. Also 14.71% have the 6 to 10 years, 5.88% have the 11 to 15 years and 61.76% with more than 15 years experience. These results show that most of the instructors surveyed has a very experienced. This factor is significant since we want to determinate the willingness to use technology, and we expect resistance to change from (some) experienced instructors.

VI. DATA ANALYSIS

A descriptive analysis and correlation tests were calculated for both questionnaires. The results are shown in Table 1 for instructors and Table 2 for students. The reliability of our variables is established through Cronbach's alpha. We use the rule of thumb. The accepted rule of thumb that for $\alpha \geq 0.9$ is excellent internal agreement, $0.9 > \alpha \geq 0.8$ is good, $0.8 > \alpha \geq 0.7$ is acceptable, $0.7 > \alpha \geq 0.6$ is questionable, $0.6 > \alpha \geq 0.5$ is poor, $0.5 > \alpha$ is very poor.

Table 1. Descriptive statistics for the instructors' dataset

variable	mean	standard deviation
AT	12.6	3.5
BI	11.0	3.3
CSE	11.6	4.4
PEU	11.3	3.4
PSS	10.7	2.7
PU	12.5	3.9
PUS	12.4	3.0
SI	11.9	2.9

Values of AT and PU above the mean indicate that instructors' attitude is in general favourable and that they believe that they can improve their teaching and performance by using technology. Results show that

PUS, BI, CSE, PEU, PSS and SI are not important factors affecting instructors' intentions to use technology.

Table 2. Instructors' dataset: Pearson correlation values (lower triangle), scale reliability (Cronbach alpha) on the diagonal in parentheses.

	AT	BI	CSE	PEU	PSS	PU	PUS	SI
AT	(0.86)							
BI	0.62	(0.76)						
CSE	0.68	0.54	(0.89)					
PEU	0.50	0.65	0.49	(0.71)				
PSS	0.70	0.67	0.67	0.70	(0.53)			
PU	0.82	0.72	0.72	0.47	0.59	(0.84)		
PUS	0.83	0.75	0.48	0.47	0.58	0.78	(0.78)	
SI	0.65	0.44	0.44	0.35	0.55	0.57	0.53	(0.70)

The values for Cronbach alpha indicate that CSE, AT and PU have a good internal reliability indicating that the questions included seem to measure a single variable. We can also say the values of PU and AT reveal that instructors' attitude is favourable, and also they see web-based assessment as useful. The reliability scores for BI, PEU, PUS and SI show that they acceptably measure a single variable. The reliability of the entire dataset is considered as excellent (0.91).

The correlation analyses shows that BI has a strong positive correlation with PUS and PU. It has a slight correlation with CSE and a low correlation with SI. This reveals that both PUS and PU are important factors to predict instructors' intentions to use web-based assessment. If instructors perceive usefulness using technology, they are more willing to use it. The value of CSE does not determine totally BI.

Table 3. Descriptive statistics for the students' dataset

variable	mean	standard deviation
AT	8.6	2.6
FE	14.2	2.7
PR	4.9	1.6
PS	8.6	2.5
PU	10.5	2.4

According to the students responses, FE is the most significant variable to the model; this means that getting feedback is an important factor to them. On the other hand, the results reveals that PU, AT, PS and PR are not important determinants for students' attitude towards the adoption of technology.

Table 4. Student' dataset: Pearson correlation values (lower triangle), scale reliability (Cronbach alpha) on the diagonal in parentheses

	AT	FE	PR	PS	PU
AT	(0.62)				
FE	0.38	(0.69)			
PR	0.10	0.25	(0.39)		
PS	0.44	0.56	0.33	(0.59)	
PU	0.59	0.46	0.21	0.54	(0.56)

As we can see in Table 4, The internal consistency of the entire students' dataset is considered as acceptable (0.79), although the result for each variable shows only limited internal consistency.

Scores show that "perceived usefulness" has a effect on the entire model. This variable is positively related to attitude and perceived suitability. This reveals that if students believe that the technology is appropriate or useful for their learning, they are more willing to using it. "Perceived reliability" has the lowest correlation coefficient which states that reliability is not an important factor to determine students' technology use.

VII. HYPOTHESIS TESTING

Results for both models were obtained by GLM regression analyses. For the instructor model we consider BI as dependent variable while AT, PU, PUS, PEU, CSE, PSS and SI are independent variables. Results are shown in Table 5. For student model we consider AT as dependent variable whereas PU, PS, FE, and PR are considered as independent variables. Results are shown in Table 6.

Table 5. Results of a GLM regression analysis of the instructor dataset. The βs are the coefficients in the regression. The model is considered significant when $p<0.05$

Hypothesis	Independent variable		Dependent variable: Behaviour Intention
		β	p-value
H1a	PUS	0.65	0.01
H2a	PU	0.37	0.05
H3a	PEU	0.25	0.08
H4a	PSS	0.34	0.15
H5a	CSE	0.00	0.99
H6a	SI	0.00	0.99
H7a	AT	-0.50	0.04

For the instructor model, hypothesis H1a is strongly supported (p=0.01) which means that the relationship between PUS and BI is the most significant one in the model. H2a is also supported (p= 0.05). Hypotheses H3a, H4a, H5a and H6a are not supported. These variables appear do not determine BI. Hypothesis H7a is not supported, but its small p-value shows that instructors' attitude actually has a significant negative effect on the behavioural intention towards using web-based assessment.

Table 6. Results of a GLM regression analysis of students' dataset. The βs are the regression coefficients

Hypothesis	Independent variable		Dependent variable: attitude
		β	p-value
H1b	PU	0.53	$1.34x10^{-6}$
H2b	PS	0.15	0.17
H3b	FE	0.09	0.32
H4b	PR	-0.12	0.39

Hypothesis H1b is very strongly supported; it shows that PU has a positive influence on students' attitude. H2b, H3b and H4b are not supported. Thus only perceived usefulness is useful for predicting students' attitude to web-based assessment.

VIII. DISCUSSION AND CONCLUSION

Our results reveal that the most significant factor determining instructors' behavioural intention to use web-based assessment is given by *"perceived usefulness to students"*. In the literature we can find studies which have demonstrated that teachers' opinions strongly influence on the students' attitude regarding the adoption of technology. For instance, (Gray et al. 2010) shows that students tend to adopt technology recommended by their teachers and peers**.** Likewise, (Margaryan et al. 2011) discusses that teachers are the main source of ideas on technology use for learning by UK university students. However, we show by analysing data for instructors there is a reverse effect, where the instructor considers the benefit to the students, and this then leads to adoption of technology. Since it is also clear that instructors' suggestions have an impact on students' attitude, so we can also enhance students' appreciation through stimulating instructors use of web-based assessment. This can be an important matter for policy-makers, who can potentially develop useful strategies to convince instructors to adopt technology by demonstrating students' improvements of their learning.

Also, our findings show that *"perceived usefulness"* has an important impact on instructors' behavioural intention to use web-based assessment. This finding is largely consistent with other studies. For instance, (Andersson 2006; Zhao 2007; Rogers and Finlayson 2004; Teo 2009; Pynoo et al. 2012) have found perceived usefulness as the mayor determinant to instructors' acceptance to use technology. Therefore, we can conclude that whether instructors perceive web-based assessment as a useful and where using it also increases their productivity, their intentions to use it will be significant increased.

Interestingly, our results regarding instructors show that perceived usefulness to the student has more significance than perceived usefulness. We can draw that instructors are more interested in that their students can improve their learning and performance by using web-based assessment than they can obtain in their teaching, that is in a nice sense of altruism.

There is another side to this, since we see a rather sceptical attitude to the usefulness of e-assessment in mathematics. This may explain why, contrary to what is generally found in the literature, attitude is predicted negatively. Even though instructors perceive that it is useful to use technology during their assessment practices and test and that their students can also enrich their learning, they are not totally convinced to adopt it as the attitude variable shows. This is can be shown in a detailed analysis, since it is mainly driven by a few questions that concentrate on the assessment of mathematical skills. This could mean that instructors still keep some attitudes that it is more effective to do mathematical exercises and exams by hand, rather than using technology. Also, it seems relatively few instructors are aware of the improvements that they can achieve to their teaching practices, and even less are prepared to learn how to use technology to perform assessments.

Furthermore, the results show that "perceived ease of use", "computer self-efficacy", "perceived system satisfaction" and "social influence" are not major determinants to predict instructors' behavioural intention to use web-based assessment. In other words, their computer skills or whether the system is easy to use are not important. The pressure that instructors get from colleagues or other University staff to use technology also has little effect.

On the other hand, the results show that *"perceived usefulness"* is the major factor impacting students' attitude to use web-based assessment. This result is consistent with the work of (Davis, F. D. 1989) and with current research by (Teo 2011). Thus, in order to obtain a favourable students' attitude to use web-based assessment it is necessary to increase perceived usefulness. It is matter of the educational managers to encourage students by providing more useful technological tools that raise students' expectations of achieving a better performance on their learning. If that is so, they will be more willing to use technology.

Likewise, our empirical model reveals that perceived suitability, perceived feedback and perceived reliability are not important factors determining students' attitude. It is surprising that issues that are a common point for complaint, such as the lack of fairness of assessments, play such a small role.

It is interesting to point out that the results of both models show that *"perceived usefulness"* is the most important factor to predict the use of technology. It seems that the main driver for adoption of such a technology is the user's belief that using technology will enrich his/her performance.

Although the amount of educational research into factors affecting learners' adoption electronic learning has increased in the last years (Wang 2003) on (Liaw & Huang 2013). Our evidence shows that it is necessary to develop strategies in order to stimulate the use of technology particularly in the assessment process. First at all, there is a need to built some to effective schemes to convince to instructors who have the power to persuade to students to use more broadly technology. Students as digital natives are always more willing to use technological tools as long as they think that these are useful and enjoyable.

We are particularly interested in fomenting and enhancing mathematical activities and task involved in the assessment process by making them more efficient and effective by using of technology. We believe that the use of technology can enhance in great way the performance of students by making more effective the learning activities and consequently in the way in which they are assessed. Besides, instructors can enrich considerably their teaching practices with the use of helpful technological resources.

We believe that we have gained some interesting insights regarding to the adoption of technology considering perceptions and opinions of instructors as of students for the assessment of mathematics subjects. We consider that these insights can be useful for policy-makers, instructors and students.

Likewise, these results can be very helpful as a way to fulfil necessities of more and better educational processes in developing countries.

In addition, we believe that these insights can contribute to lead an successful implementation of web-based assessment systems in a context of Higher Education.

IX. REFERENCES

Ajzen, I., 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), pp.179–211.

Andersson, S.B., 2006. Newly qualified teachers' learning related to their use of information and communication technology; a Swedish perspective. *British Journal of Educational Technology*, 37(5), pp.665–682.

Bandura, A., 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*, Englewood Cliffs, NJ: Prentice Hall,.

Bennett, R.E., 1998. *Reinventing Assessment. Speculations on the Future of Large-Scale Educational Testing. A Policy Information Perspective.*, Policy Information Center, Mail Stop 04-R, Educational Testing Service, Rosedale Road, Princeton, NJ 08541-0001; Web site: http://www.ets.org; Available at: http://www.eric.ed.gov/ERICWebPortal/detailaccno=ED424254

Benson Soong, M.. et al., 2001. Critical success factors for on-line course resources. *Computers & Education*, 36(2), pp.101–120.

Cassady, J.C. & Gridley, B.E., 2005. The effects of online formative and summative assessment on test anxiety and performance. *The Journal of Technology, Learning and Assessment*, 4(1).

Davis, F. D., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology,. *MIS Quarterly (13:3)*, pp.pp. 319–339.

Dermo, J., 2009. eAssessment and the student learning experience. A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), pp.203–214.

Domino, G. & Domino, M.L., 2006. *Psychological testing: An introduction*, Cambridge University Press Cambridge.

Fishbein M. & Ajzen, I., 1975. Belief, attitude, intension and behavior: An introduction to theory and research. In MA: Addison Wesley, Reading,.

Gikandi, J.W., Morrow, D. & Davis, N.E., 2011. Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), pp.2333–2351.

Gray, K., Chang, S. & Kennedy, G., 2010. Use of social web technologies by international and domestic undergraduate students: implications for internationalising learning and teaching in Australian universities. *Technology, Pedagogy and Education*, 19(1), pp.31–46.

Gunasekaran, A., McNeil, R.D. & Shaul, D., 2002. E-learning: research and applications. *Industrial and Commercial Training*, 34(2), pp.44–53.

Ilias, A., Razak, M.Z.A. & Yasoa, M.R., 2009. Taxpayers' attitude in using e-Filing system: Is there any significant difference among demographic factors? *Journal of Internet Banking and Commerce*, 14(1).

King, C.G., Guyette, R. & Piotrowski, C., 2009. Online exams and cheating: An empirical analysis of business students' views. *The Journal of Educators Online*, 6(1), pp.1–11.

Larson, M.R. & Bruning, R., 1996. Participant perceptions of a collaborative satellite-based mathematics course. *American Journal of Distance Education*, 10(1), pp.6–22.

Liaw, S.-S. & Huang, H.-M., 2013. Perceived satisfaction, perceived usefulness and interactive learning environments as predictors to self-regulation in e-learning environments. *Computers & Education*, 60(1), pp.14–24.

Margaryan, A., Littlejohn, A. & Vojt, G., 2011. Are digital natives a myth or reality? University students' use of digital technologies. *Computers & Education*, 56(2), pp.429–440.

Pynoo, B. et al., 2012. Teachers' acceptance and use of an educational portal. *Computers & Education*, 58(4), pp.1308–1317.

Rogers, L. & Finlayson, H., 2004. Developing successful pedagogy with information and communications technology: how are science teachers meeting the challenge? Available at: http://hdl.handle.net/2381/1048 [Accessed March 18, 2013].

Selim, H.M., 2007. Critical success factors for e-learning acceptance: Confirmatory factor models. *Computers & Education*, 49(2), pp.396–413.

Stödberg, U., 2011. A research review of e-assessment. *Assessment & Evaluation in Higher Education*, 99999(1), pp.1–14.

Sun, P.-C. et al., 2008. What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, 50(4), pp.1183–1202.

Teo, T., 2011. Factors influencing teachers' intention to use technology. Model development and test. *Computers & Education*, 57(4), pp.2432–2440.

Teo, T., 2009. Modelling technology acceptance in education: A study of pre-service teachers. *Computers & Education*, 52(2), pp.302–312.

Teo, T., Lee, C. b. & Chai, C. s., 2008. Understanding pre-service teachers' computer attitudes: applying and extending the technology acceptance model. *Journal of Computer Assisted Learning*, 24(2), pp.128–143.

Venkatesh, V. & Davis, F.D., 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), pp.186–204.

Volery, T. & Lord, D., 2000. Critical success factors in online education. *International Journal of Educational Management*, 14(5), pp.216–223.

Wang, Y.-S., 2003. Assessment of learner satisfaction with asynchronous electronic learning systems. *Information & Management*, 41(1), pp.75–86.

Webster, J. & Hackley, P., 1997. Teaching Effectiveness in Technology-Mediated Distance Learning. *The Academy of Management Journal*, 40(6), pp.1282–1309.

Wu, I.-L. & Chen, J.-L., 2005. An extension of Trust and TAM model with TPB in the initial adoption of on-line tax: An empirical study. *International Journal of Human-Computer Studies*, 62(6), pp.784–808.

Zhao, Y., 2007. Social Studies Teachers' Perspectives of Technology Integration. *Journal of Technology and Teacher Education*, 15(3), pp.311–333.

Methodology and Software Support for Evaluating Teams' level of collaboration

Luigi Romano
IDSE, Faculty of Computer Science
Free University of Bolzano-Bozen
Bolzano, Italy
e-mail: luigi.romano@unibz.it

***Abstract*—In a perspective of future working life, students must possess both good team-working skills and good knowledge about topics. This paper describes a methodology and presents a software prototype, to support tutors in assessing collaborative activities and level of knowledge reached by students, while working on E-Learning platform and in presence. Results from an extensive experiment performed with 30 high school students are provided as well.**

Keywords: Social Network Analysis, Analysis of Collaboration Patterns, Assessment, Communities of Practice, E-Learning

I. INTRODUCTION

In a modern educational approach, students are required to master a set of competencies, related to the disciplines in their curriculum, but this is not enough. Students should also learn various soft skills, including team working on real problems: soft skills and disciplinary competencies are both needed in their future working life. Consequently, tutors should teach, and assess, soft skills [1][5], and to this aim, an assessment based on traditional disciplinary tests is no longer sufficient. The importance of mastering team working, among soft skills, has been stated by both OECD [15], and the European Community in its Lifelong Learning Programme [13]. To achieve it, adoption of technological artifacts of various kinds has been considered beneficial: for example, using an E-Learning platform helps in increasing team activities, by working during standard classroom hours, and also from home.

Tutor assessments should take into account results gathered in all educational environments, by evaluating students from different points of view. Assessment thus integrates data collected with different aims, both on the level of understanding of the discipline, and on the level of proficiency gained in team-working, in IT use and, more generally, in soft skills. This data is collected from the e-Learning platform, and during traditional classroom activities.

To assess the quality of team working, we rely on studies from Wenger [27], who highlighted that a Community of Practice (CoP) is the highest level of team-working, because it helps and supports the creation of new (and deeper) level of knowledge. As a consequence, it is fundamental for a tutor to investigate the relationships between students' collaboration paths, examined with the support of the Social Network Analysis (SNA), looking for the creation of a CoP, and consequently, building of new knowledge by them. This analysis is not easy for a tutor, who is responsible for a class, especially for the overwhelming amount of data to be analyzed within course time constraints. Considering a class of 30 high school students, a thorough assessment process for just one discipline over a semester may result in approximately 350 working hours.

The purpose of this paper is to describe a methodology and a software artifact named TASSE (Tutor Assessment Support with SNA Engine) for supporting tutors in both aspects of the evaluation process, i.e. in assessing student collaboration patterns, and their disciplinary test results. TASSE provides tutors with an interface, allowing them to have at a glance a feedback on class situation, as well as the possibility for careful and detailed analysis of selected data sets, by focusing on individual students and on teams.

The structure of this paper is as follows: Section II presents the theoretical scenario behind this research project; Section III describes the architecture of TASSE; Section IV presents the experience, Section V analyses the results, and in Section VI conclusions and future developments of this project are presented.

II. RATIONALE AND BACKGROUND LITERATURE

Examination of collaboration patterns among students has often been performed with Social Network Analysis (SNA). Theories related to SNA are well established: this kind of analysis deals with communication relationships inside a group of people, and it allows understanding how a "network" develops inside a community. For example, SNA may show who is the "leader" of a team, or highlight who is "not well involved" in team activities. Papers from Bonacich [2][3], Borgatti [4], Freeman [8], Granovetter [10], Hanneman [11], Nadel [16], Radcliffe-Brown [18], Scott [25] and Wasserman [26] provided a clear analysis of SNA and its possibilities.

Studies from Reuven et al. [21][22], Reffay et al. [19][20] stated that it is possible to recognize if a team could create a CoP, again by adopting the SNA techniques. This analysis is based on the concept of clique, which is a set of actors (usually a subset of a larger set) strictly interconnected among each other. Clearly, the presence of a clique in a team is not implying that a CoP is indeed active, because actors might be strictly related among one another, without sharing any common educational achievement. The long term goal of this research activity is to experience a methodology for:

1. Evaluating collaboration paths inside the class.
2. Assessing results gathered by students in disciplinary tests.
3. From activities described at points 1 and 2, derive a comprehensive assessment of students´ performance.

The present paper is based on previous work [23][24], and builds upon it by describing the architecture of TASSE, the software artifact implementing the methodology, and by deepening on the experimental results. TASSE supports tutors in decision making, by displaying collected data on a dashboard (see Fig. 5 later on for an example of TASSE dashboard). The author believes that a fully automated system would be unable to substitute tutor's judgment, since an automated assessment would be at risk of missing significant information.

Metrics listed in TABLE I. have been used to express student results, as defined in Romano [23][24]:

TABLE I. METRICS FOR EVALUATING STUDENTS

a.	LOCW: Level of Collaborative Work
b.	LOR: Level of Result
c.	CCOP: Creation of a CoP (inside each team)
d.	GCCOP: Global Creation of a CoP (inside the whole class)

Metrics in TABLE I. have the following meaning:

a. LOCW (Level Of Collaborative Work) assesses the contribution of each student to group activities inside his/her own team.
b. LOR (Level Of Results) measures the level of understanding of each student in the relevant discipline
c. CCOP (Creation of CoP) is monitoring the creation of a CoP (Community of Practice) inside each team of students
d. GCCOP (Global Creation of CoP) monitors the creation of one CoP inside the whole class

LOR is measured w.r.t. results gathered by students in tests performed in class and in interviews. LOCW is a metric deduced from Polya [17], who described a series of steps for supporting students in performing heuristic analysis.

In this paper we will deepen point (c) in TABLE I. and the assessment of its related metric CCOP. Analysis of point d and of GCCOP will be developed in a further paper.

III. TASSE ARCHITECTURE

Data on students' performance and collaboration patterns are collected from various sources. One such source is the log of student on-line activities, for example exchanges of messages, comments and so on. They are automatically collected by an E-Learning platform called DIEL [6][7]. DIEL is based on the popular platform Moodle [14] (An example of DIEL screenshot is shown in Fig. 1). It provides a 2D and 3D graphical interface, plus all Moodle typical features, including activity logs.

Figure 1. DIEL screenshot

Additional data related to collaboration paths is provided by observations and interviews with students during classroom activities. This is measured by the LOCW metric (see TABLE I. Information related to students' understanding of proposed topics is gathered by the tutor by periodical class tests (LOR metric in TABLE I. LOR and LOCW results are stored manually by the tutor into TASSE.

TASSE uses the Cytoscape [12] engine for analyzing logs from DIEL and building the SNA, identifying clique(s) inside teams. As stated above, the presence of clique(s) is a necessary (but not sufficient) condition for the creation of a CoP inside a group of people. So, when the tutor considers results shown in the dashboard, he/she needs to merge information collected by SNA together with LOR and LOCW. These metrics all together support tutors to understand what teams of students have a high probability of having built a CoP. This result is measured by CCOP, the third metric from TABLE I. .

It is important to clarify that LOCW and SNA measure a different kind of information about collaboration within teams. LOCW collects tutor observations on the level of information sharing, and participation of each student to group activities, as undertaken in classroom. SNA on the other hand, yields a graph of on-line collaboration paths among students. In principles, an automatic tool like TASSE might also be used to measure LOCW from DIEL logs, distinguishing (with content analysis) whether information exchanges are relevant to study goals or not. Such an analysis is at present beyond the scope of our implementation. Any way, it would not eliminate the need for manual input, in what concerns the assessment of classroom activities by the tutor.

In previous works [23][24], a dashboard has been introduced, as an integrated model for displaying such data. The dashboard gives the tutor an overview of daily results, both about disciplinary level of understanding, and about the level of collaboration, providing at a glance the full scenario of the whole class, and the possibility of filtering results from each team or student.

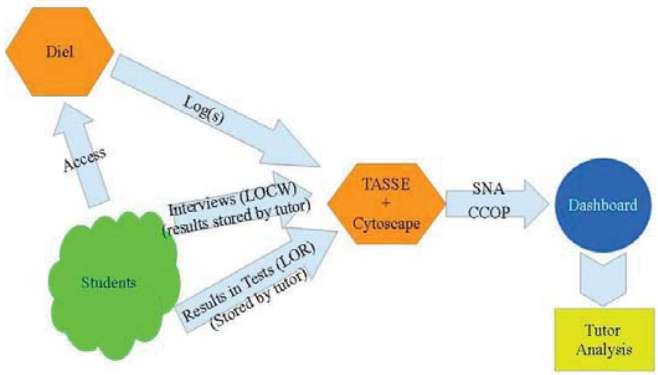


Figure 2. TASSE Architecture

IV. Description of the Experience

The experience was conducted on a class in a Liceo Scientifico (Italian high school with specific reference to Mathematics and Physics). The sample was made by 30 teenage students, that have been monitored over three terms, starting in February 2010 (at the beginning of their second term of 2nd year), till June 2011, when they concluded their 3rd year of study. Students were divided in 8 teams, and the subject area involved was Mathematics. They initially worked on Euclidean Geometry, and later on Analytical Geometry. Mathematics is often perceived by students as being too abstract, and poor performance is rather commonplace. Math tests were based on problems strictly connected with the real world; some examples are listed in [24]. Solutions to those problems were proposed by teams, and each team solution was assessed. Students also took periodically individual Maths tests, aimed at understanding the level of proficiency in Maths for each of them. Term assessment combined results achieved both by team work and by individual tests.

V. Result Analysis

The analysis of data gathered during activities made in class and on DIEL by our sample, showed that all teams communicated enough so to build a clique among members. Assessing CCOP produced these results i.e. teams' results typically scored between two extremes:

1. Teamwork activity of students was satisfactory, and their collaboration paths showed the existence of a clique; they achieved a good and quite homogeneous LOR and LOCW, both in tests and in interviews. As a consequence, they built a CoP and they consolidated their methodology of team-working and they got a high score of CCOP.

2. SNA reveals the existence of a clique, yet all students scored a very poor LOR; interviews showed a superficial (insufficient) analysis of proposed issues i.e. a low LOCW. Then, the team was unable (for reasons that should be investigated by the tutor) to work properly and consequently a CoP was not built. This wrong (and immature) approach should be timely and properly dealt with by the tutor, since, if not corrected, it shall cause problems in the future, both in further studies and in working life. Obviously the CCOP score in this case was really poor.

Situations described at points 1 and 2 represent the extreme team situations, all the experienced team situations were somewhere between them. Figures 3 to 5 show screenshots of the dashboard for Team 3 and for Team 4.

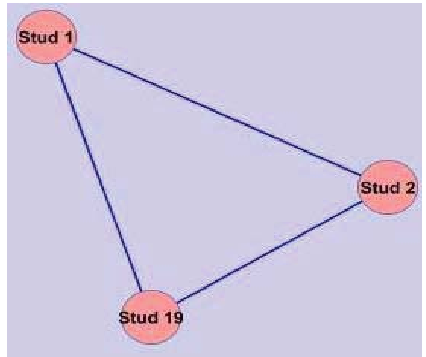


Figure 3. Team 3 SNA

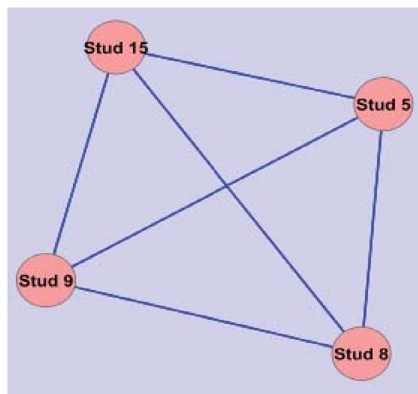


Figure 4. Team 4 SNA

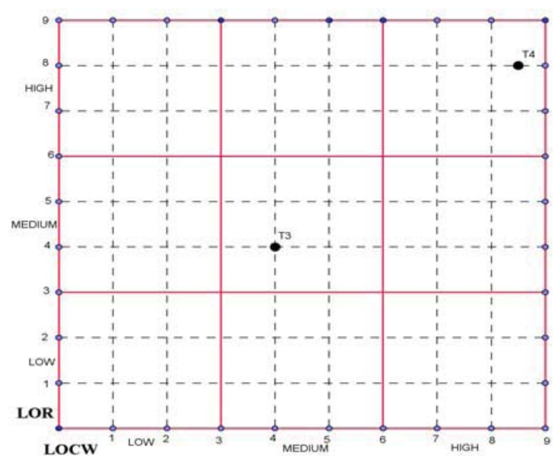


Figure 5. LOR LOCW Diagram for Team 3 and Team 4

Figure 5 shows the class dashboard, filtering just results on LOR and LOCW by Team 3 and Team 4. Teams placement in the diagram area depends on their results, gathered in tests and interviews, related both to LOR and LOCW. A position in the upper-right corner means high level performance both in LOR and LOCW. A position in the left-bottom corner means very poor performance both in LOR and LOCW. Combined observation of Fig. 3 and Fig. 5 show that even if Team 3 built a clique, they didn't work properly on the proposed topics (medium-to-low results in LOR and LOCW), so Team 3 is an instance of the above described situation #2 with a poor result of CCOP. Fig. 4 and 5 show instead that Team 4 built a clique and worked proficiently (good results in LOR and LOCW); so they are an example of situation #1, and they built a CoP in their team, and consequently a good CCOP.

The added value of the methodology and of TASSE results, summarized in CCOP metric, is supporting tutor in monitoring and assessing the activities performed in a class, especially w.r.t. the level of proficient collaboration and sharing of information inside a team.

VI. Conclusions and Future Works

This paper describes a methodology and a software artifact named TASSE, and its use in supporting a tutor in his/her daily assessment of students. TASSE combines results from Social Network Analysis with tutor’s observation, and helps the tutor in discovering whether students' teams built a CoP, the educational goal of team collaboration. Our experience with TASSE involved 8 teams of teenage students studying Maths, both in their classroom activities as well as in collaborative activities on an e-learning platform. TASSE supports a tutor in performing a complete and detailed assessment of students: its dashboard and filters allow a quick, daily check, as well as a detailed term assessment. Results from Romano [24] show that the use of TASSE saves approximately 33% of tutor time in managing assessment data, making the tutor timely aware of problematic situations of students with severe deficiencies, and leaving the tutor with more time to work on learning process improvement.

Currently the author is working on improving the user interface of TASSE, being it still a prototype, to achieve a seamless flow of data between DIEL and Cytoscape. This further step of data integration will allow assessing the fourth metric, named GCCOP, foreseen in our methodology.

References

[1] J. Andrews, H. Higson, Graduate Employability, ‘Soft Skills’ versus ‘Hard’ Business Knowledge: a European Study, Higher Education in Europe, 33:4, 411-422, 2008.

[2] S. Bonacich, Power and Centrality, American Journal of Sociology, 5, 1170 – 1182, 1987.

[3] S. Bonacich, Factoring and Weighting Approaches to Status Scores and Clique Identification, Journal of Mathematical Sociology, 2, 113 – 120, 1972.

[4] S. P. Borgatti, A. Mehra, D.J. Brass, G. Labianca, Network analysis in the social sciences science, 323, 5916, 892 – 895, 2009.

[5] R. Crosbie., Learning the soft skills of leadership, Industrial and Commercial Training, 1, 37, 45-51, 2005.

[6] Di Cerbo F., Dodero G., Papaleo L., Integrating a Web3D interface into an e-learning platform. In Proceedings of the ACM, International Conference on Web 3D Technology, pp. 83 - 92, ACM NY USA. (2010)

[7] F. Di Cerbo, G. Dodero, P. Forcheri, V. Gianuzzi, M. G. Ierardi, Hybrid Learning Experiences with a Collaborative Open Source Environment, Proceedings of ICHL 2nd Int. Conf. On Hybrid Learning 2009, F.L. Wang, J. Fong, V.K.S. Lee (eds.), Lecture Notes in Computer Science, vol. 5685, Springer 2009, 45-54.

[8] L. C. Freeman, Centrality in Social Networks Conceptual Clarification, Social Networks 1, 215 – 239, Elsevier Sequoia (eds.), 1978.

[9] N. E. Friedkin, Theoretical foundations for centrality measures, American Journal of Sociology, 96, 1478 – 1504.

[10] M. S. Granovetter, The strength of weak ties, American Journal of Sociology, 78, 1360-1380, 1973.

[11] R. A. Hanneman, M. Riddle, Introduction to Social Network Methods, University of California, Riverside, 2005.

[12] http://cytoscape.org

[13] http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11090_en.htm

[14] http://moodle.org/

[15] http://www.oecd.org/pisa/pisaproducts/46961598.pdf

[16] S. F. Nadel, the theory of the social structure, Free Press, Glencoe, IL, 1957.

[17] G. Polya, How To Solve It, Princeton University Press, 1945.

[18] A. R. Radcliffe-Brown, On Social Structure, Journal of the Royal Anthropological Institute, 70, 1 - 12, 1940.

[19] C. Reffay, T. Chanier, Social Network Analysis Used for Modeling Collaboration in Distance Learning Groups, S.A. Cerri, G. Gouardères, and F. Paraguaçu (Eds.): ITS 2002, LNCS 2363, 31–40, Springer-Verlag, Berlin, Heidelberg, 2002.

[20] C. Reffay, T. Chanier, How social network analysis can help to measure cohesion in collaborative distance-learning, In Proceeding of Computer Supported Collaborative Learning conference (CSCL'2003) - Computer Supported Collaborative Learning, Norway, 2003.

[21] A. Reuven, E. Zippy, R. Gilad, Cohesion and Roles: Network Analysis of CSCL Communities, Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies (ICALT’03), 2003.

[22] A. Reuven, E. Zippy, R. Gilad,, A. Geva, Network Analysis of Knowledge Construction in asynchronous Learning networks, JALN Volume 7, Issue 3, 2003.

[23] L. Romano, Tool and Method for Evaluating Students Working on E-Learning Platforms, DET Conference 2011, Florence, 200, 203.

[24] L. Romano,, An Artifact for Assessing Students' collaborative and Individual Activities, ATEE 2013 Winter Conference, Genoa, Italy.

[25] J. Scott, Social Network Analysis: A Handbook, 2nd ed., London, Sage (eds.), 2001.

[26] S. Wasserman, K. Faust, Social Network Analysis, Cambridge, MA: Cambridge University Press, 1994.

[27] E. Wenger, Communities of practice: learning, meaning, and identity, Cambridge University Press, 1998.

Modeling Multiple Common Learning Goals in an ETC*plus* Educational Project

Angela Guercio
Dept. of Computer Science
Kent State Univ. at Stark
Canton, OH
aguercio@kent.edu

Paolo Maresca
Dept. of Ingegneria Elettrica e Tecnologie dell' Informazione
Univ. Napoli *Federico II*, Italy
paomares@unina.it

Lidia Stanganelli
Dept. of Ingegneria Elettrica e Tecnologie dell' Informazione
Univ. Napoli *Federico II*, Italy
ldistn@gmail.com

Abstract. *The ETCplus project tackles the challenges of developing a sequence of learning activities, topics and assessments that improve the learning skills of the modern students. The ETCplus portal connects faculty and students from different institutions to define learning goals and discipline-related skills and collaborate in a distribute environment. The validation of the cooperation process model of the ETCplus project has been performed in an academic environment with a creation of a community of practice that joins students and teachers of two selected academic courses: one in Italy, and the other in the United States. In this paper we describe the ETCplus project model, its design issues and its problems. The collaboration is designed on top of a shared IBM Jazz platform, which in union with Moodle and Eclipse have been used to generate the documentation, the analysis and the monitoring of the experiment.*

Keywords: ***Collaboration, E-learning, learning models, community of practice, cooperation.***

1. INTRODUCTION

Role exchanges, programming camps, and discussion on a competition are some of the many collaboration activities that can be created to foster cooperative learning. *Cooperative learning* is a paradigm of collaboration aimed to reach a common goal [11]. It differs from *individual learning*, or *competitive* learning where individuals must reach the same goal but cannot collaborate, rather compete, with other people. Competitive learning generates a winner and many losers a point that is not as significant as gaining knowledge over time. Cooperative learning is designed to support and reinforce this particular aspect. In this learning paradigm students learn from each other, the learning process moves faster and it produces better results [3]. The approach is student-centered since the teacher's figure is freed of its hierarchical role and is moved to a more peripheral role such as the role of a facilitator, of a coach, or of a counselor. In this new role the teacher can better communicate with the students who see the teacher as an active participant in the learning process, learns from him/her, and mimics him/her while tutoring other participants. This boosts and speeds the overall learning process of the whole community. Collaborative activities in software developments are often performed in Programming Communities of Practice (CoP) whose objective is the software development or the code generation. When distributed communities share a set of learning goals, a well-designed CoP is the place where a modern and mobile community can find the best and broadest learning support for collaboration and knowledge exchange. A CoP is defined by Wenger [17, 19] as "a network of people who share a common interest in a specific area of knowledge (and) are willing to work and learn together over a period of time to develop and share that knowledge", and is the result of the union of competence, experience and engagement of its members [18].

The model of cooperation proposed in this paper is called ETC*plus* and extends the learning paradigm of the ETC[1](Enforcing Team Cooperation) project [4, 5, 12]. The project is supported by IBM in the area of Academic Initiative and fosters the creation of CoPs among people of academic communities on a joint platform (the IBM Jazz-Hub platform). ETC*plus* uses a dynamic process model which encourages dynamicity at a broader level not only between the single instructor and the students, but also among all the participating instructors who must be open to accept and incorporate new knowledge that comes from other communities. The coordination model and the steps for the creation of a learning environment used in ETC*plus* generate an *open innovation network* which fosters the creation of an "intelligent" community. To validate the model and the process of ETCplus a CoP around an academic environment has been created. The CoP joins two universities, one in Italy, one in the US, around two selected academic courses and creates a set of collaborative activities The students cooperated on collaborative projects and

[1] ETC is a project supported by IBM in the area of Academic Initiative. The project received the IBM Academic Award 2011 (http://www.ibm.com/developerworks/university/facultyawards/) and an IBM Rational Champion 2012 award was assigned to Prof. Maresca (https://www.ibm.com/developerworks/champion/) for his work on the project. ETC has been nominated IBM Best Practice 2011 at the IBM Innovate 2012 [10], received a special mention among the worldwide academic institutions [7], and was cited as an Internal Rational success story [8]. Finally ETC has been accepted for publication in the IBM conference *Innovate 2013* to be held in Orlando 2-6 June 2013.

the result of this cooperation is used to validate the learning benefits of ECT*plus* [14, 15] and to identify the points of weakness and strength of the model.

The paper is organized as follows. In Section 2 we present the Coordination Model Process required for the creation of an open academic CoP. In Section 3 we present an application of the ETC*plus* project and we analyze the common objectives and the cooperative activities of practice. In Section 4 we discuss the resources required to foster the learning activities of the community, and in Section 5 we discuss the results. Finally, in Section 6 we present the state of the art and the future research.

2. CoPs and a Model of Cooperation

A CoP consists of a group of people who share a craft, a profession, or a passion and are a cornerstone in the design and innovation process of knowledge acquisition and knowledge transfer [13, 17]. Often their goal is gaining and sharing knowledge related to their field. In the area of Internet CoPs include people from different countries, languages and traditions who find common ways of communicating and learning over the network. Academic CoPs are of great interest in the educational challenge of this century since their participants will share not only what they need to learn specifically and in depth for their own academic courses, but will acquire additional knowledge as well as the necessary techniques to conquer that information. We believe that this approach contributes to the creation of a T-shaped people, so highly aimed in industry but so hard to generate in academia, which requires a deep knowledge generation in some area and a more shallow and broad knowledge in other areas [9]. Preparing the young generation to face the challenges of this new century is one of the greatest challenges. While in some countries the educational system is adapting quite well to the challenges, in other countries the lack of resources and old preconceptions have slowed down the innovation process. In open innovation networks every student of any country can be part of this worldwide learning process that can be generated from the development of academic CoPs.

In this section we describe the ETC*plus* model of cooperation used to create virtual academic CoPs that join academic participants who share common goals. This model creates a concept of an *innovative university*, which is an *open innovative network* of people where:

- the students learn from people or students of other universities or industry;
- relations among students and future employers are stimulated;
- T-shaped knowledge so strongly aimed in the job market of this century is built;
- instructors are helped and supported in their endeavors;
- researchers contribute in the maintenance and the update of the curricula to provide the latest innovative knowledge.

The process of the cooperation model used for the generation of an open CoP is depicted in Fig. 1. Before the model is applied, the identifications of the participating entities must be completed. Then, the first step identifies the areas that may benefit from the cooperative learning, and analyzes the common objectives, while maintaining the constraints required by each entity. In the second step a set of cooperative learning activities of practice that achieve the common objectives are identified. The activities are designed in a joint fashion. Clear and traceable objectives, prerequisites, team formation, and grading are also identified and their coordination is designed. This complex activity requires an extensive collaboration of the entities before the project's deployment in order to implement a successful CoP.

Fig. 1. The ETC*plus* Cooperation Process Model

It is in this step that the customization of the needs and the constraints of each entity must be properly evaluated. When this analysis is completed, the model proceeds with the identification of the resources required for the deployment of the process. This produces a Work Breakdown Structure (WBS) which identifies and plans the distribution of the organizational tasks.

3. An ETC*plus* Application

In order to validate the ECT*plus* Cooperation Process Model, an application has been applied in Fall 2012. An experimental virtual CoP was formed. The CoP involved two academic institutions: the Faculty of Engineering of the University of Naples "Federico II" (UNFII), and the Department of Computer Science of the Kent State University at Stark, Ohio, US (KSU). During the first step of the coordination process model (the analysis of common objectives) two programming courses which had multiple overlapping goals were chosen. The selected courses were: “*Programmazione I”* (Programming I) at UNFII and *Computer Science II* at KSU. Learning goals and objectives were clearly highlighted in this stage before moving to the next step. Since the selected courses were delivered in a face-to-face fashion, the entities were required to preserve the characteristics of a regular academic course. However they were free to incorporate a set of distance learning activities to foster cooperative learning and to connect the participating entities. In this phase the policies, the rules and regulations required by each entity, and the synchronization of the activities were analyzed. For example the courses involved in the validation were held in the fall semester, but the participating universities started and ended at different time. The number and the length of the lessons of each course were different in term of total number of hours per week and number of weeks. A synchronization plan beneficial to both institutions was devised. When this analysis phase was concluded, the entities started identifying teaching activities to boost students learning abilities and to empower them with tools that allow them to operate as independent learners, cooperating team leaders, and tutors for other students. Such activities included the deployment of the course material, the virtual laboratory, and the students’ assessments. The collaborative activities identified were: programming camps (a collaborative activity in which participants write algorithms, identify or vote for the best posted solutions), discussions on a competition (a collaborative activity in which the participants are invited to discuss and analyze problems and artifacts - see [9]), and general team related activities.

During the analysis of cooperative activities of practice, a set of activities for the design and organization of the course were identified. This included the generation of a joint syllabus, the installation, testing and revision plan of the common platform, the set-up plan of the shared virtual laboratory, and the layout of the tentative outline. Dropbox was used to share documents and perform asynchronous activities in this phase. Additional customized activities were inserted in the joint part of the course by each institution to meet their needs. Deadlines were identified for each course with the constraint to maintain joint deadlines for joint activities. The activities of cooperation identified in the previous stage, were designed. For example, the students of KSU were selected to tutor the students of UNIFII since their course started 4 weeks before the UNIFII course and those students were assumed to be ready to share part of the course material that the students of UNIFII were just starting to learn.

4. The Resources

The course laboratory was implemented in a virtual environment accessible via either the native client, or a browser, or the Eclipse plugin as shown in Fig 3.

Fig. 3. The Eclipse ETC*plus* virtual laboratory.

The virtual laboratory was designed on top of the IBM Rational Jazz platform, which includes the Rational Requirement Composer (RRC), the Rational Quality Manager (RQM), the Rational Team Concert (RTC), and the Rational Asset Manager (RAM). The RTC provides tools to build software in an Agile manner, while the RRC application has the ability to handle the requirements. The RQM is able to manage the testing activity, while the RAM enables to track versioned artifacts back to the assets that were used to create them. This platform provides the ability to handle the whole project in a cooperative fashion, to share documentation and to align modifications performed during the development process. The RAM tools support cooperative activities such as “discussion on a competition” for the selection of the most preferred resources stored in the RAM repositories.

Moodle was used as a content management system to coordinate the teaching activities.

Finally, the WBS for the application was completed and the work breakdown for the preparation of the teaching material, the setup of the virtual laboratory, and the preparation of the documentation of the virtual laboratory was established. Where possible, the activities were carried out in parallel.

5. The ETC*plus* Validation

The experiment between UNFII and KSU lasted for an entire semester and involved around 100 at the UNFII and around 20 at KSU. In addition to producing tangible deliverables, which consisted in a collection of several artifacts and the development of a large entire project, it was an opportunity for the evaluation of the training process delivered to the joint classes. The evaluation was performed at the end of the experiment by inviting the students to answer to a questionnaire. The answers were traced to a 5 scale values metrics (expressed in terms of relation between adjectives and numbers) which were used to measure the results and analyze them objectively. Appendix A shows the ETC*plus* Experience Survey consisting of 22 questions. The questions were designed to analyze the features chosen for the implementation process of the experiment. Table 1 shows the correspondence between a feature of the process and its corresponding question(s). There are many points of interest to test in a complex experiment as this one, however we kept the number of questions around 20 to encourage serene and truthful answers, and avoid survey dismissal that often occurs when large requests are submitted. In a future work we would like to emphasize the quality of the training process by extracting, where possible, appropriate data by measuring the process directly from the platform used during the ETC*plus* experiment.

Table 1. ETC*plus* features associated with the survey

	Feature(s)	Questions
1.	ETCplus preprocess - precondition assessment	1
2.	ETCplus postprocess - learning assessment	2-3
3.	ETCplus - cooperation assessment	4-5
4.	ETCplus - cooperation barriers assessment	6-7
5.	ETCplus - time spent assessment	8-10
6.	ETCplus - quality of educational material & organization	11-16
7.	ETCplus - process & tools evaluation	17-23

Fig. 4 represents each question with its mean value on a star at 22 branches of a Kiviat diagram. This diagram is useful to highlight the strengths vs. weaknesses of our project. The results of the survey have also been plotted on a graph depicted in Fig. 5. The graph shows the standard deviation, the mean and the median of each question. The standard deviation, with its index of dispersion around an expected value, is used to represent the precision with which the measurement has been carried out. Low values indicate high precision.

The first question, with its mean value of 1.88, shows that the groups had little experience of e-learning environment before the ETC*plus* project. This may be an advantage, since our students sample can be seen as a "blank slate" on which the rules of the process act. This value should not surprise us and, in our opinion, has contributed to reach a sufficient level of cooperation between the groups and good learning results of each individual. The learning process is assessed by using questions 2 and 3. The graph of Fig. 5 shows that the students were sufficiently satisfied in acquiring knowledge on software design and implementation using the ETC*plus* paradigm. The third feature of table 1, which corresponds to questions 4 and 5, is oriented towards the co-operation between international students and shows a nearly sufficient level of cooperation. The fourth feature of table 1, which uses questions 6 and 7, is used to detect the barriers encountered during the cooperation phase. As the answer to question 6 shows with its mean value of 1.77, students of different languages, with different teachers, cultural paths and different logic, naturally have barriers of communication and of knowledge exchange. A separate discussion should be done on actions aimed at removing the above mentioned barriers in order to converge faster toward common objectives.

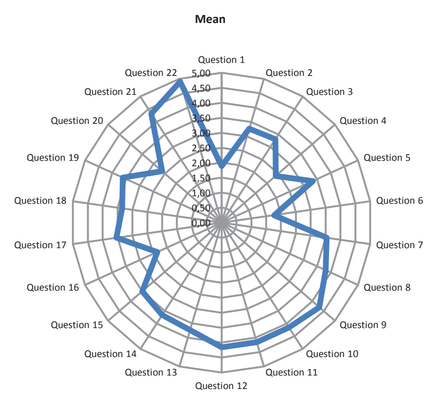


Figure 4. The Kiviat graph of the ETC*plus* process.

The language is still one of the barriers and this problem was detected especially by the Italian students who had to interact with Americans in their mother tongue. On the other side, from question 7 we observe that the instructions given in the process were very clear and this has allowed them to better approach the problems resolution and compensate in part the language barrier. Two additional fundamental aspects of this project were the time spent by students to participate in the ETC*plus* project, and the way they strategically used it to acquire knowledge in the context of their cultural improvement. The time dedicated to acquire the know-how required for both the Eclipse and the RTC platform was between 1 and 2 months. This is not a very long time if we consider the fact that the majority of the participating students worked either full or part time, and if we consider that the average time spent by an individual who does not participate in any collaborative learning activity, ranges between 2 to 3 months to acquire the know-how of the tools used in ETC*plus*. The quality of the teaching material used as well as the quality of the organization perceived by the student was is a key point of the project and it has reported the best result (see questions 11-16) and the full approval of the students' population who participated in the project.

Finally after analyzing the assessment process and the tools used for designing and writing code (see questions 17-22) we observe that the students have appreciated the many efforts made to connect different and distant universities for the development of a joint course by crediting the goodness and the quality of the process as well as the validity of the experiment. However the same experiment received some complaints from the fact that some collaborative activities were time consuming during the development and cooperation with other groups. This, in fact, required them to be engaged in additional activities such as conference call, video call, and so on, usually not required in a non-cooperative environment.

At the end of the survey the students had the opportunity to provide suggestions to incorporate in future project developments. The quality of the work developed by the groups is the tangible aspect of the success of the process development. Each group has been engaged in the development of an application of an ATM teller machine. We were struck by the high quality of the code produced in such a short period of time.

6. Conclusions and Future Work

In this paper we describe ETC*plus*, a project designed to create virtual CoPs to generate a collaborative learning environment. An application that connected the students and the teachers of two international academic communities is also presented. The model of the process of collaboration used in the application was designed on top of a shared platform. The model of the process was validated in the experimental community that joined two universities in the formation of a CoP around two academic courses. The experimental community focused on creating collaborative activities by sharing teaching materials, and supporting the educational formation of the participating students. We are aware of existing experimental CoPs and this shows the great interest in this approach. For example, experimental projects such as eXtension, [16] an initiative designed to foster collaborative learning across land-grant universities in the area of sustainable development; the ACUI (Association of College Unions International) Late Night Programs CoP [2] which provides programming options for registered students who like to works during the nighttime; a CoP in Education at the Rotterdam University [1]. These experiences provide large amount of resources and communication tools to its members to engage in online discussion. An experiment to teach Intro to Programming in a virtual CoP by using Second Life is also described in [6]. What makes our experience unique is the fact that beyond the set of available resources and communication tools available to engage online discussion, there are several coaches over the CoP that help the members of the community and that stir and spin the dynamicity of the learning process when it slows down. The connection of the two academic courses across the ocean enlarges the classrooms' physical limitations and opens the students to a broader group of selected members that face the same challenges while still remaining committed to the requirements of the single classroom. The opening of the community is important both to the participating students in the project as well as to the teachers who can learn new teaching techniques used by other peers. A common shared platform must support the virtual community and provide the software tools for the development of the work required for the course participation. The advantages of this extended community is to provide the virtualization of an environment, such as a course laboratory, that favors students' distance learning, that provides study flexibility, and that minimizes the set-up and maintenance time for the instructor. The results of the experimentation show the strength of the ETC*plus* project as well as the weakness that need to be addressed. The learning assessment, the cooperation, and the knowledge acquisition are positive. However language barriers are still in place. The project was well received; however additional time might have improved some results that were produced under stress. The overall quality of the artifact was very satisfactory. In a future work we would like to emphasize the quality of the training process by measuring the process directly from the platform used during the ETC*plus* experiment.

References

1. A Community of Practice in Education at the Rotterdam University.http://www.zylstra.org/blog/archives/2008/06/community_of_pr.html. Retrieved on Aug. 2012.
2. ACUI (Association of College Unions International). Late Night Programs Community of Practice. Retrieved on Aug. 2012. http://www.acui.org/groups/group.aspx?id=12329
3. Bermeio S., "Cooperative Electronic Learning in Virtual Laboratories through Forums", *IEEE Transactions on Education*, vol.48 (1), pp.140-149, IEEE Press (2005)
4. Coccoli, M., Maresca, P., Stanganelli, L.,"Enforcing Team Cooperation: an Example of Computer Supported Collaborative Learning in Software Engineering", Procs. *of 16th International Conference on Distributed Multimedia Systems* (DMS2010) – *Workshop on Distance Education Technologies (DET*), pp. 189-192, ISBN 1-891706-28-4, Oak Brook Illinois, USA (2011)

5. Coccoli, M., Maresca, P., Stanganelli, L., "Enforcing Team Cooperation Using Rational Software Tools into Software Engineering Academic Projects", *Proc. of the 5th Workshop of the Eclipse Italian Community* (Eclipse-It 10), pp. 90-103, ISBN: 9788890438813, Savona, IT(2010)
6. Esteves, M., Antunes, R., Fonseca, B., Morgado, L., Martins, P., "Using Second Life in Programming's Communities of Practice" Briggs, R.O., Antunes, P., DeVreede, G.J. and Read, A.S. (eds): *CRIWG 2008, LNCS 5411*, pp. 99-106, Springer-Verlag, Berlin, Heidelberg, (2008)
7. ETC-HUB 2012, IBM Offers New Tools to Help University Students, Professionals Prepare for Information Technology Top Jobs, http://www.sysmannews.com/SearchResult/35610
8. ETC success 2011, Internal Rational Success Story - ref: W. Batten (Wendy J Batten/Lexington/IBM)
9. IBM Jam, https://www.collaborationjam.com/
10. IBM Innovate 2011, J. Shildt, "Announcing JazzHub - NEW Cloud Based Software Development Resources for Academia", *IBM Innovate*, June 2011, Orlando, FL, USA, http://ibm.co/O87OKw
11. Johnson, D.W., Johnson, R.T., Holubec, E., *Apprendimento cooperativo in classe*, Erickson (1994).
12. Maresca, P., Stanganelli, L., Gorga, F., "ETC: Stato dell'Arte e Prospettive Future", *Didamatica*, pp. 1-13, ISBN: 978-88-905406-7-7, Taranto (2012).
13. Maresca, P., Stanganelli, L., "Il progetto ETC: i Risultati della Sperimentazione in un Corso di Programmazione", *Didamatica*, pp.1-13, ISBN: 978-88-905406-7-7, (2012)
14. P. Maresca, A. Guercio, L. Stanganelli, "ETC*plus*: a Project for the Creation of Innovative Universities", *Procs. of the 7th Workshop of the Italian Eclipse Community* (Eclipse-IT 2012), Napoli, Italy, Sept 20-21, 2012, ISBN 978-88-904388-3-7.
15. P. Maresca, A. Guercio, L. Stanganelli, "Building Wider Team Cooperation Projects from Lessons Learned in Open Communities of Practice", *Proc. of the 18th Intl. Conf. on Distributed Multimedia Systems (DMS 2012) – Intl. Workshop on Distance Educations Technologies (DET 2012)*, Miami Beach, FLA, USA, Aug 9-11, pp. 144-149, 2012.
16. NASULGC, eXtension: Tapping the Power of Cooperative Extension.http://www.landgrant.org/docs/FY2006/BAC/2006_5.pdf, (2006)
17. Wenger, E., "Communities of Practice: Learning as a Social System", *Systems Thinker*, June 1998. http://www.co-i-l.com/coil/knowledge-garden/cop/lss.shtml,
18. Wenger E., "Communities of Practice and Social Learning Systems". *Organization*, vol. 7, n. 2, pp. 225-246, (2000)
19. Wenger, E., *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, (1998)

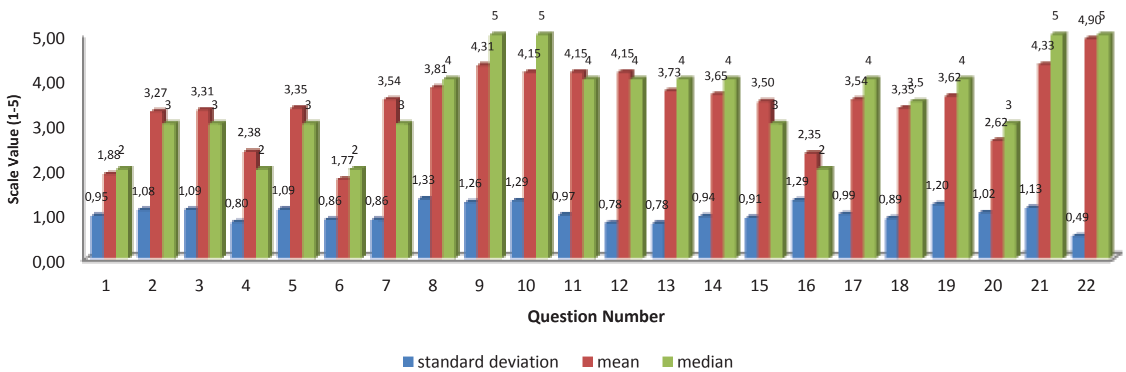


Fig. 5. The assessment graph of the ETC*plus* process.

Appendix A - ETC Experience Survey

LEGEND: wks=weeks

1. Did you have previous experiences of Communities of Learning? *0: Nothing 1: A little 2: Enough 3: More than enough 4: A Lot*	2. After the experiment did you learn something new about designing and implementing software in a collaborative environment? *0: Nothing 1: A little 2: Enough 3: More than enough 4: A Lot*
3. Overall, how much did you learn in the ETC experience? *0: Nothing 1: A little 2: Enough 3: More than enough 4: A Lot*	4. What is the level of cooperation that you achieved in your group? *0: No cooperation 1: A little 2: Enough 3: More than enough 4: A Lot*
5. What did you learn about software development in cooperation? *0: Nothing 1: A little 2: Enough 3: More than enough 4: A Lot*	6. Did you encounter language barriers in your group? *0: Not at all 1: Very little 2: Some 3: Quite a bit 4: A Lot*
7. How clear where the instructions of the project you had to solve in the ETC experiment? *0:Totally insufficient 1:Below Sufficient 2:Average 3:Good 4:Excellent*	8. During the experiment you had to use some software applications such as Eclipse and the Jazz platform. What was the personal amount required to learn the above mentioned applications? *0: I do not know 1: ≥ 2 months 2: 1-2 months 3: 3-4 wks 4: 1-2 wks*
9. What was the personal amount of time required to learn only the Eclipse application? *0: I do not know 1: ≥ 2 months 2: 1-2 months 3: 3-4 wks 4: 1-2 wks*	10. What was the personal amount of time required to learn only the Jazz application? *0: I do not know 1: ≥ 2 months 2: 1-2 months 3: 3-4 wks 4: 1-2 wks*
11. How much would you evaluate the quality of the material provided to instruct you on the installation and use of the Eclipse application (i.e. video, files, etc.)? *0:Totally insufficient 1:Below Sufficient 2:Average 3:Good 4:Excellent*	12. How much would you evaluate the quantity of the material provided to instruct you on the installation and use of the Eclipse application (i.e. video, files, etc.)? *0:Totally insufficient 1:Below Sufficient 2: Average 3:Good 4:Excellent*
13. How much would you evaluate the quality material provided to instruct you on the installation and use of the Jazz application (i.e. video, files, etc.)? *0:Totally insufficient 1:Below Sufficient 2:Average 3:Good 4:Excellent*	14. How much would you evaluate the quantity of the material provided to instruct you on the installation and use of the Jazz application (i.e. video, files, etc.)? *0:Totally insufficient 1:Below Sufficient 2:Average 3:Good 4:Excellent*
15. How much would you evaluate the organization and the use of the material and the tools offered in Moodle? *0:Totally insufficient 1:Below Sufficient 2:Average 3:Good 4:Excellent*	16. Besides the material provided by the teacher, how much personal search did you need to understand the use of the software used in the ETC experiment? *0: Nothing 1: A little 2: Enough 3: More than enough 4: A Lot*
17. How quick was the help provided to solve the software issues that you have encountered? *0:Totally insufficient 1:Below Sufficient 2:Average 3:Good 4:Excellent*	18. How would you evaluate the whole ETC experiment? *0:Totally insufficient 1:Below Sufficient 2:Average 3:Good 4:Excellent*
19. How useful, in your opinion, is this type of experiment in a course of CS 2? *0: Not useful at all 1: A little 2: Enough 3: More than enough 4: A Lot*	20. Did the ETC experiment satisfy your expectations? *0: No 1: A little 2: Enough 3: More than enough 4: A Lot*
21. Did you know how to use Eclipse before this course? *Yes/ No*	22. Did you know how to use the Jazz application before this course? *Yes/ No*

Closing the Distance in ODL - e-Portfolio Assessment

Petra le Roux

School of Computing
University of South Africa, (UNISA)
lrouxp@unisa.ac.za

Abstract

In outcomes-based education (OBE), outcomes, not content, must be assessed. Formative assessment takes place during the teaching and learning process, whereas summative assessment refers to assessment that is used for making a judgment about the achievement of an outcome. Authentic assessment mirrors the real world and is integrated with learning as learners are viewed as active participants in the learning process. Integrated assessment assesses a number of outcomes simultaneously, using a combination of assessment methods and instruments. One appropriate method that can be used to include summative, authentic and integrated assessment in a qualification would be the inclusion of a capstone module, to be taken in the final year of study. The purpose of such a module is to integrate the learning that has taken place throughout the programme and assess if the exit outcomes have been achieved. A well-designed portfolio will meet the teaching goals of engaging the students in the higher order cognitive of activities of reflection and creativity. However, distance in ODL has different implications in different environments. In a third world environment where infrastructure cannot be taken for granted different options of the presentation of the portfolio must be offered, which include non-venue-based assessment; a driving force in open distance learning (ODL) assessment. Furthermore, to strive to preserve the environment and to improve service to students all assessment is done electronically (green assessment).

I. INTRODUCTION

In OBE, with each module or qualification, the outcomes and associated assessment criteria are clearly stated so that students understand, in advance, what they have to do to achieve these outcomes and assessors can use the criteria to assess the outcomes with reasonable objectivity and reliability. With OBE, the outcomes, not content, must be assessed [8].

Knowledge or content of modules is no longer the principal focus but instead the focus is on the *application* of that knowledge and the demonstration of the required skills and values within specific contexts. Furthermore, outcomes based assessment is criterion referenced as there is a complete shift away from the tendency to use a normal curve as the basis for making decisions about individual performance. In contrast to *norm* referenced testing which is designed for the purpose of comparing students with one another, *criterion* referenced assessment is a form of assessment in which judgements are made about learners by measuring their work against set criteria that are independent of the work of other learners [6].

Outcomes based assessment also includes integrated assessment which provides evidence that the purposes of a module or qualification as a whole have been achieved.

Formative assessment supports teaching and learning during the teaching and learning process, whereas summative assessment refers to assessment that is used for making a judgment about the achievement of the outcome. Summative assessment may take a form other than an examination, such as the submission of a portfolio of evidence or reports from workplace mentors [1, 8].

Furthermore, authentic assessment mirrors the real world and is integrated with learning [7]. Assessment is a process whereby the learners are active participants and the criteria are open and negotiable. The goal is to engage learners in the assessment as well as the learning process. This engagement will assist the learner to develop better learning and self-evaluation skills that are vital, as there is a move to a society where life-long learning must be the norm. Portfolio assessment is one form of authentic assessment.

Integrated assessment refers inter alia to the assessment of a number of outcomes simultaneously, using a combination of assessment methods and instruments for outcomes.

One method that can be used to include integrated assessment in a programme would be the inclusion of a capstone module [8]. A capstone module is a programme-specific module taken in the final year of study. The purpose of such a module is to integrate the learning that has taken place throughout the programme through the use of portfolios, which will be assessed to ascertain whether the exit outcomes have been achieved [2, 4] .

II. E-PORTFOLIOS

A portfolio is a purposeful collection of a person's work. Portfolio assessment has traditionally been used in creative fields such as fine art and music. More recently it has become common in a much wider range of educational settings. A well-designed portfolio will meet the teaching goals of engaging the students in the higher order cognitive of activities of reflection and creativity [3, 5]. According to Plimmer [7], portfolios can, at best, "act as a silent mentor becoming an instrument of learning as well as a repository and students become a responsible partner in documenting and evaluating their own learning".

The University of South Africa (UNISA) offers an *ICT Diploma in Information Technology* qualification. After scrutinizing the rules, structure and purpose of this qualification, the necessity of a capstone module was identified and, due to the nature of the outcomes, an e-portfolio was identified as assessment method.

III. ASSESSMENT PROCESS

In the distance based education environment at hand, the submission of assignments via the internet entailed the printing of these submissions. This process consumes a lot of paper and huge bottlenecks due to the volume of assignments that require processing. This has a negative effect on the quality of service. However, faster turnaround times, with that amount of paper, are simply not possible and students do not receive important feedback on their assignments in time to prepare for their examinations. Furthermore, the process of printing out submitted assignments does not contribute to the preservation of the environment. Green assessment addresses these issues and entails onscreen marking.

The onscreen marking process routes the assignment from the student to the lecturer or marker and then back to the student while it updates the student system with a mark. The marking tools consist of two components. The first component allows for the adding of marking symbols (^, √ or x) to the text; the allocation of marks as well as calculating a final mark. The second component is a commenting tool to comment on the typical errors students make. The commenting tool and marking tools combined can be used as a marking rubric.

Due to the numbers of assignments external markers are contracted to help with the marking. The marking tools are also available to the external markers.

IV. MODULES

As mentioned earlier, the University of South Africa offers an *ICT Diploma in Information Technology* qualification. Two third year modules, which form part of the qualification, were earmarked as capstone modules with the first, a pre-requisite for the second. In the first, the newly qualified ICT professional is provided with the experience in the implementation and evaluation of an ICT project by analysing a problem scenario in a fictional environment, designing an ICT solution underpinned by theoretical arguments, and presenting a proposal for an ICT solution. In the second module the student has to implement, evaluate and present the ICT project.

Both modules are assessed in two assignments and a final e-portfolio. The assessment is cumulative in that the comments on one assignment lead to changes in the next assignment. The e-portfolio would be the culmination of the work that has been done in the semester. All assignments are compulsory and, except for the e-portfolio of the last module, will have to be electronically submitted and will be marked onscreen with a marking schema.

The three assignments of the first module are structured as follows: The first assignment consists of the preparation of a project brief. The project brief describes the proposed system that the student will be working on. The outcome of the second assignment is a draft project proposal. This assignment is an opportunity for the student to refine the project proposal and to provide detailed planning of the system that is being developed. The final assignment, the complete and final project proposal, is a repeat of the second assignment, but the student has to incorporate all the feedback given in the second assignment.

The outcome of the first assignment of the second module is a detailed project plan that will show how the student intends to complete the implementation of the proposed solution. The outcome of the second assignment is a set of analysis and design documentation. The final assignment of the second module is an e-portfolio where the students have to implement the proposed solution. Students can utilise any development environment with a production quality deliverable to provide a concrete, real-world experience. This will be assessed and presented during an interview.

V. E-PORTFOLIO ASSESSMENT

The next step is the final evaluation of the project. The student is responsible for arranging an interview to demonstrate the project implementation through the "sign-up" tool on *myUnisa*, a portal that allows for the creation and posting of meetings. *myUnisa* is the course management system of Unisa. Students are also required to download the documentation of the project in a "drop box" tool available on *myUnisa* the day before the demonstration. This will allow the assessors to be prepared. A rubric (marking scheme) is used during the summative evaluation and is given to the students in advance. The rubric refers to the outcomes of the module. Each outcome is broken down into smaller parts and a rating scale of 0 to 5 is used.

Due to the existing internet infrastructure in South Africa (e.g. the lack of affordable bandwidth) as well as the economic realities, there are four options available for the live demonstration.

Non-venue based assessment via *myUnisa*. The *myUnisa* site incorporates an open source web conferencing system developed primarily for distance education. It supports multiple audio and video sharing, presentations with extended whiteboard capabilities – such as a pointer, zooming and drawing – public and private chat, desktop sharing, and integrated VoIP. Moreover, users may enter the conference in one of two roles: viewer or moderator. As a viewer, a user may join the voice conference, share their webcam, raise their hand, and chat with others. As a moderator, a user may mute/unmute others, eject any user from the session, and make any user the current presenter. The presenter may upload slides and control the presentation. Once a meeting is in progress, the student is assigned the role of moderator and demonstrates the project via desktop sharing. Even though this is the preferred option it not always viable, due to a lack of bandwidth.

As the majority of students reside near the physical location of the department, an **interview** is scheduled and the demonstration is done in an office setting. One hour slots are allocated to each student. The rationale behind one hour slots benefits both student and assessor. For the student, it allows enough time to setup their equipment, create the environment if needed and feel at ease. For the assessors it allows for discussion of outcomes and the finalization of necessary administration requirements.

Students residing near a Unisa regional office e-portfolios will be assessed via the **video conferencing** (VC) facility within the university. The student is made the presenter, which makes screen, desktop and/or application sharing possible, thus allowing the student to demonstrate a project. As this is an established VC facility within the university, no technical problems is experienced.

The last option, but least desirable, is to submit the e-portfolio via an **external storage device** (e.g. CD or flash drive) or e-mail (if size permits). The assessors then conduct the demonstration with the help of the documentation. This option can be very challenging for several reasons. The major being the environment the student used to develop the application. This environment is to be re-created in order to evaluate the system and it is not always possible. For this reason, this option is only allowed if the other options are not executable due to infrastructure.

VI. EXPERIENCE OF STUDENT AND STAFF

Though this research is work in progress, we found that it is feasible to have the capstone project within the *ICT Diploma in Information Technology* and are encouraged by early results and student perceptions. Based on informal feedback, the students' reaction to the e-portfolio were positive, with the better students making the best use of the feedback they were given on the earlier submissions to improve their e-portfolio.

By creating the capstone module, the students experienced that they gained professional skills that used cutting edge technologies to provide real-world understanding whereas the department gained insight into teaching strategies that will provide students with the skills that industry need.

The assessment procedure facilitated face-to-face engagement between student and lecturer in the ODL environment. This was observed as a positive experience by both students and lecturers. From the assessor's point of view, onscreen marking improved the turnaround time to less than a week and provided for comments on how the students might improve their work. The server also allowed comparisons to the student's previous submissions.

By designing the appropriate assessment, it contributed to the process of teaching and helped students to develop lifelong learning skills as well.

REFERENCES

1. Box, I. 2002. *Assessing the Assessment: An empirical study of an Information Systems development subject.* Australasian Computing Education Conference. Adelaide, Australia.

2. Bruhn, R & Camp, J. 2004. *Capstone course create useful business products and corporate-ready students.* Inroads - The SIGCSE Bulletin , 36 (2), 87-92.

3. Cassiday, S & Schwitter, R. 2007. *An evaluation of portfolio assessment in an undergraduate web technology unit.* UniServ Science 2007 .

4. Keefe, K & Dick, M. 2004. *Using extreme programming in a capstone project.* Sixth Australasian Computing Education Conference, (pp. 151-160).

5. Klenowski, V, Askew, S & Carnell, E. 2006. *Portfolios for learning, assessment and professional development in higher education.* Assessment and Evaluation in Higher Education, (pp. 267-286).

6. Nightingale, P., Te Wiata, I. *et al*. (eds), 1996 *Assessing Learning in Universities*, Professional Development Centre, University of New South Wales, Sydney.

7. Plimmer, B. 2000. *A case study of portfolio assessment in a computer programming course.* Proceedings of the NACCQ, (pp. 279-283).

8. Unisa. 2005. Assessment Policy. Retrieved from http://www.unisa.ac.za/cmsys/staff/contents/departments/tuition_policies/docs/AssessmentPolicy_CouncilFinal_271005.pdf

msMLO:A Novel Approach for Selecting and Fusing Learning Objects

Aldo Ramirez-Arellano, Elizabeth Acosta-Gonzaga
IPN
Mexico City, Mexico
aramirezar@ipn.mx, eacostag@ipn.mx

***Abstract*— Retrieving Learning Objects (LO) that fulfills user's requirements, is still an issue. For instance, a user interested in linear regression topics (theoretical background) and examples (problems solved using a spreadsheet), might find two different resources covering these topics. Merging the two LOs as one would be a more appropriate solution rather than browsing them separately. This paper presents a novel approach for automatic selection of source LOs in order to construct a more appropriate one based on the users query. The combination of two or more LOs gathers not only knowledge but also tends to rank the new LO in a higher position than the source ones. This is accomplished by driving the merging process using the query. Our algorithm avoids including similar items in the new LO.**

Keywords—Learning object; Recommendation system; Content Authoring; Algorithm.

I. INTRODUCTION

Nowadays, Learning Management Systems (LMS) are broadly used in universities, institutes etc. They are able to retrieve learning resources in an easy way; most of the time the user provides a query then LMS searches and ranks learning resources according to it. The large number of ranked learning resources could render the process of choosing the best one tedious because the user has to browse them one at time. In addition to this, if none of the learning resources in the set matches the majority of requirements, it is necessary to build a new query taking into account the user´s requirements. For instance, a user interested in linear regression topics (theoretical background) and examples (problems solved using a spreadsheet), might find two different resources covering these topics. However merging the two LOs as one would be a more appropriate solution rather than browsing them separately.

Rebuilding manually the LOs is impractical and it represents a hard workload for a human instructor hence relying on automatic reuse of resources is our goal. In this paper the Management System for Merging Learning Objects (msMLO) is introduced, a LMS that picks out and merges Learning Objects (LOs) using novel techniques. msMLO works without any additional information added to the Learning Object (LO) such as domain ontology [1] or attribute structure [2]; instead, msMLO converts a LO, packed using the SCORM [3] or IMS [4] standard, to a hierarchy [5]. A hierarchical representation [6] allows us to search for similar topics and avoid duplication [7], as well as to perform the merging stage [5]. Our approach does not target individual multimedia resources such as images, videos, pdf documents etc. ; however it focuses on items of LO, defined by the SCORM [3] standard.

This paper is organized as follows. Section II, describes the process of selecting LOs to be merged. The architecture of msMLO is presented in section III. The results of our experiments are shown in section IV. We talk about the related work to this paper in section V. Finally, our conclusions and further work are presented in section VI.

II. SELECTING AND MERGING LO

LOs stored in LMS belong to different topics or even different fields of knowledge so the first goal of the users query is to restrict the number of LOs in the search set. The second goal is to drive the merging process.

First, msMLO ranks the LOs according to the query provided, this process is based on the Vector Space Model (VSM) [8] technique. The new LO resulting from fusing the two top ranked LOs, is assigned in its respective position according to its score. It could be at the top of the list. In this case the new LO becomes the pivot and msMLO will merge it with others LOs, except its parents or ancestors. msMLO will repeat this stage until no more top ranked LOs are found or until no more LOs are added to the search set.

A. Relationship Detection

A genealogical tree depicts the ancestors and descendants of a given LO. It allows tackling the issue of knowing which LOs are the ancestors of the pivot, our approach encodes the genealogical tree in a single sequence of characters. Figure 1 shows a brief example.

Note that the relationship detection between “F” and “B” requires a single traversal of the encoded genealogical tree, EGT for short. The EGT is stored into the index. Putting the EGT into the index, does not alter the LO so an extension of the standard is not necessary. Also it is the recipe for rebuilding any merged LO provided that the source LOs are still stored in msMLO.

The merge [6] function produces the same LOs no matter how the algorithm merges “A” and “B”. In other words, it could merge “B” and “A” or vice versa. The resulting LO would contain the same items. This is an important property to reduce the number of merging processes and design a one-pass function to detect the relationship.

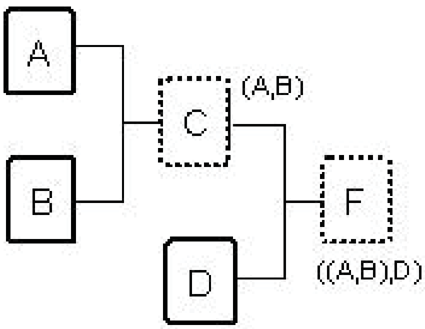


Fig. 1.The genealogical tree shows the ancestors of F, which are encoded as a sequence of characters ((A,B),D). A dashed square denotes a merged LO. On the other hand a solid square is a source LO stored by msMLO.

B. Algorithm for selecting and merging LOs

The rationale of our algorithm is to search for the best LO, performing the minimum number of merging processes, using the following relationship detection algorithm. The algorithm avoids merging LOs that have a common ancestor. For example AB, see Figure 2, will not be merged with A or B. Similarly AB will not be merged with BC or AC to produce ABC. The algorithm produces the combination ABC using two merging processes at most. It could merge A and B then C.

We describe the algorithm as follows:

Let P_0, LO_i and LO_f be hierarchical representations [6] of LOs where P_0 is the pivot, LO_i is the candidate to be merged and LO_f is the outcome of the merging process.

The set S has LOs filtered by the users query. Each $LO_i \in S$ was ranked according to its score. At the first stage, the pivot is the top ranked LO_i. $|LO|$ means the score of the LO assigned by msMLO.

Relationship-based Algorithm (RA)

P_0=top ranked LO
do
  For each $LO_i \in S$
    If(not $relationshipdetection(P_0, LO_i)$ &
    $sim(P_0, LO_i)$ & $P_0 \neq LO_i$)
      LO_f=$merge(P_0, LO_i)$
      If($|LO_f|>|P_0|$)
        $P_0 = LO_f$
      End if
      $S = S \cup \{LO_f\}$
    End If
  End for
While(S has been changed)
Return P_0

The function $sim(P_0, LO_i)$ is essential for merging LOs strongly linked to each other, since it searches for the root item from P_0 similar to the root item from LO_i; it avoids merging LOs whose degree of matching is below a threshold. The threshold is the fraction of the number of children of LO_i coinciding with corresponding children of P_0. Since the outcome of *sim* is in the interval [0, 1], a threshold of 0.8 was used. Using a low value such as 0.4 could produce a merged LO which includes items from different topics. Also the comparison of two root items takes into account the title, the metadata and the content. A depth explanation of *sim* can be found here [7]. The function *merge* [6] compares each item from P_0, to each item from LO_i. If they are similar, it evaluates the two items using the user´s query and picks out the top ranked one then it will be added to LO_f. The LO_f is temporally stored in the index so its score is computed using the user´s query and the VSM [8] technique. If it has a higher score than P_0, it will become into the new pivot else it will be removed from the index.

If the merging process of RA produces a new pivot for each iteration, the function *relationshipdetection* will discard most of the combination of the current level of the lattice, see Figure 2. For instance, let A be the pivot at the first level of the lattice then the first merging process could produce the pivot (A, B) at the second level of the lattice. Now S= {AB, A, B, C}. Since the function *relationshipdetection* avoids merging the pivot with A or B, the remaining combination is ABC. It is clear that RA performs only a merging process for each level of the lattice no matter what combination is the current pivot hence RA performs at most d-1 merging processes. Where d is the number of the source LOs at the top of the lattice.

C. Comprehensive algorithm for merging LOs

In order to compare the effectiveness of RA, we designed a second approach based on the lattice structure [9], as shown in Figure 2. The structure is tree-like with each node representing a LO in the search set.

At the first stage, the structure has only source LOs at the top; therefore, at least the algorithm has to perform 2^d-d-1 merging processes to obtain the remaining LOs. Since the complexity of the algorithm increases exponentially, it must traverse the entire lattice, doing the minimum number of merging processes. For example the node ABC could be built merging A and B then C. It could be written using EGT notation as ((A, B), C). The other path is ((A, B), (B, C)). Note that the first path performs two merging processes so it is faster than the second one.

The algorithm avoids keeping the lattice structure in memory. This is accomplished by assigning a binary value to each source LO. The first LO is labeled with 2^0 the second with

2^1 and so forth for the all other cases. To discover the remaining nodes in lattice, the algorithm adds 1 to the current node until it reaches 2^d, see Figure 2.

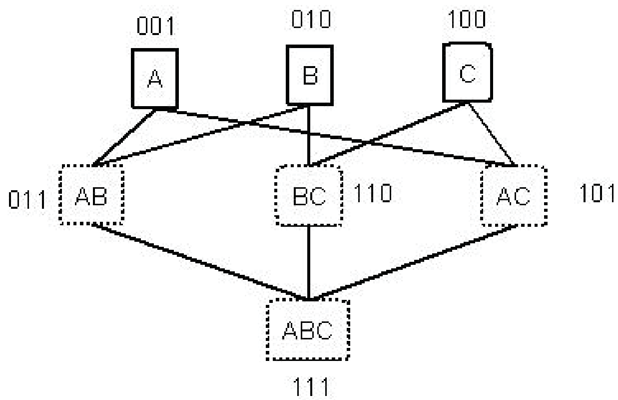


Fig. 2. Lattice structure with d=3. Each node represents a LO.

The lattice algorithm is as follows:

The set S has LOs filtered by the users query. Let M be the set to store the previous merged LOs but $S \cap M = \emptyset$. Let P_0, LO_a, LO_b and LO_f be hierarchical representations of LO. P_0 stores the outcome of the algorithm.

Lattice-based algorithm (LA)

1. P_0= top ranked *LO*
2. For (i =1;i<$2^{|S|}$;i++)
3. If (combination *i* does not exist in $S \cup M$)
4. i_a =*firstcombination(i)*
5. i_b = *secondcombination(i)*
6. LO_a=*load*(i_a, $S \cup M$)
7. LO_b=*load*(i_b, $S \cup M$)
8. LO_f =*merge*(LO_a, LO_b)
9. $M = M \cup \{LO_f\}$
10. If ($|LO_f| > |P_0|$)
11. $P_0 = LO_f$
12. End if
13. End if
14. End for
15. Return P_0

A search set with three source LOs, labeled with 001_2, 010_2 and 100_2 respectively, needs four merging processes to discover all combinations of LOs, see Figure 2. For example, if the current node is 001_2 (A), the next is 011_2 (BA) which means that we have to merge 010_2 (B) and 001_2 (A). Note that LA skips the node labeled as 010_2. The node 111_2 (ABC) requires at most a merging process because 110_2 (BC) was previously computed in the step 6. As a result, each merging process of the algorithm builds a new LO that is a part of the LOs at the bottom of the lattice. The function *firstcombination* searches for the combination with the longest length so any LO in lattice can be built with a merging process. In our example the outcome of *firstcombination* and *secondcombination* is 110 and 001 respectively. In other words the algorithm has merged "A" and "B" then the outcome will be merged with "C" saving a merging process. Note that the order of traversal does not matter because we aim to compute the entire lattice.

III. ARCHITECTURE OF MSMLO

The architecture of msMLO, shown in Figure 3, is composed of four modules (Insertion/Retrieval, Preview, LO File Builder and LO Parser). msMLO uses three repositories, the first one stores a packaged version of the LOs, the second one keeps a preview version of the LOs and the last repository is a term-based index.

With a web-based approach, msMLO becomes not only a flexible and platform-independent LMS, but friendly to the user as well. The merging process of the LOs is accomplished by using core functionalities that are described below.

The LO parser is responsible for getting a hierarchy from the packaged LO. Besides it extracts the metadata, text and multimedia resources such as audio and video files, web pages, images etc. Before a new object has been added to the index, the text from the items, content and metadata are filtered by the Insertion /Retrieval module to delete stop words such as "a", "into", "so". Also a stemmer is used to derive roots from words, for example "runs", "running", "run" all map to "run". then it will put the text in a Lucene [10] document. The document typically consists of several separately named fields with values. For example "content" is used to label the text from the items of the LO. The labels of the metadata fields can be found in LOM standard [11]. Finally, the document is stored in the Lucene [10] index.

The merging process gets the ranked LOs from the Insertion /Retrieval module. If it finds the best merged LO, it will be exported to SCORM or IMS package. This task is carried out by the LO Parser and LO file builder.

Finally the Preview unpacks each LO and copies its content into Preview Repository. Given the multiuser nature of msMLO, keeping LOs stored apart from unpacked ones avoids losing information.

IV. EXPERIMENTAL EVALUATION

We are interested in finding out the scalability of our algorithm and the usefulness of the resulting LO. The next subsections first present the scalability of the two algorithms described previously. Next we will show the evaluation of the merged LO.

We conducted our experiment on a server with two 2.4 Ghz CPUs, 4GB RAM and an HD of 465 GB. Fifty LOs agglutinated by four topics were stored in the repository of the msMLO.

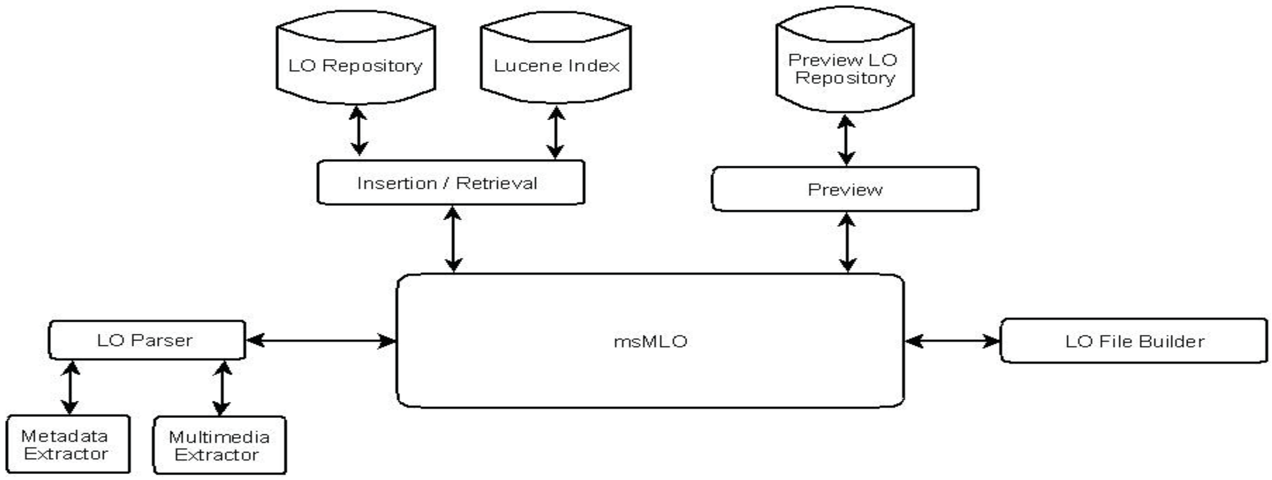
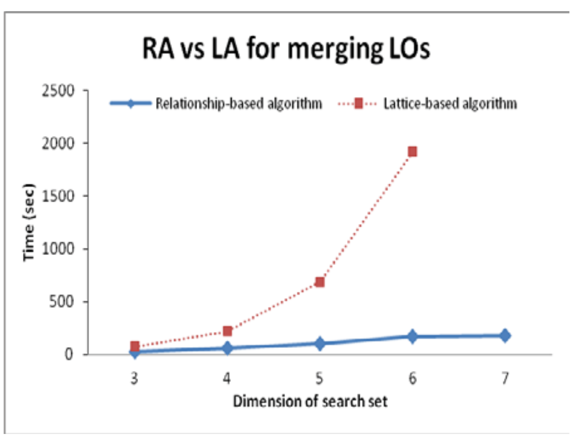


Fig. 3. The architecture of msMLO.

A. *Performance for merging LOs*

In this section, we will compare the time it takes the relationship-based algorithm (RA) and the lattice-based algorithm (LA) to find the best merged LO. Since RA and LA both evaluate the merged LO obtained in each stage using the user´s query and the VSM [8] technique, always the outcome of both algorithms is the top-ranked LO. If it does not belong to the set of LOs initially filtered by the users query, it will be considered the best merged LO. Our experiments were carried out using ten different users queries and the search set was restricted from d=3 to d=7 LOs. Because of the time consuming nature of the *merge* function, LA was tested with a maximum value of d=6.



Fig, 4. Scalability of Relationship-based Algorithm (RA) and Lattice-based Algorithm (LA).

Whereas LA is affected exponentially by the number of LOs in the search set, RA has a slower growth rate than LA, see Figure 4. There are two main reasons for this behavior; firstly RA merges only the LOs which are not related to each other (by EGT analysis) hence we prune the lattice structure. In the best case, in which each merging process makes a top LO, RA will carry out a depth-first traversal, doing at most d merging process. On the other hand, any other case will not exceed this boundary. Secondly, RA tests if the two root items of LOs are similar using the *sim* function. Carrying out the two tests is less expensive than getting all possible merging process.

Table 1 shows the average execution time of each algorithm. We can see that the difference between computation time of RA and LA increases exponentially. LA becomes impractical as d increases.

TABLE I. RELATIONSHIP-BASED ALGORITHM (RA) VS LATTICE-BASED ALGORITHM (LA) VARYNG THE NUMBER OF LOS IN THE SEARCH SET (D)

D	RA (sec)	LA (sec)
3	29.2	77.7
4	64	217.8
5	106.9	690.9
6	168.8	1921.7
7	177.1	*

* Means that the time was too long to get a result.

B. *Usefulness of the top ranked LO*

We compared the LOs found by the two algorithms using d=6. The weakness of the RA approach is that we are not certain of having found the best LO, or if a combination of the LOs can accomplish this goal, see Table 2. Despite the fact that RA is not able to retrieve a top ranked LO for queries 4 and 10, the score of RA and LA for query 3, 8 and 9 are identical. We concluded that the LOs found are the same.

For queries, 6 and 7, the score of LOs using RA are closer to those of LA; therefore RA gets an admissible outcome without sacrificing time. The LA spent too much time on obtaining that there was not a viable merging process to build the best fused LO instead, RA was faster. Both algorithms concluded that it is not possible to obtain a better LO for queries 1, 2 and 5. RA fails to do so for queries 4 and 10. To find out if *sim* affects the result of the two previous queries, we ran RA without the *sim* test. The RA got a LO for query 4 which was evaluated by the author of the query. The opinion was similar to those expressed for queries 4, 6 and 10, described below.

TABLE II. LOs OBTAINED FROM RELATIONSHIP-BASED ALGORITHM (RA) AND LATTICE-BASED ALGORITHM (LA)

QUERY	RA SCORE	LA SCORE	RA TIME (sec)	LA TIME (sec)	# LO MERGED
1	--	--	91	2042	0
2	--	--	58	1276	0
3	0.03161	0.03161	310	2442	2
4	--	0.18338	57	1633	2
5	--	--	20	1908	0
6	0.09549	0.09823	229	1967	3
7	0.02854	0.03676	308	1895	6
8	0.14579	0.14579	285	2032	3
9	0.05676	0.05676	179	2235	2
10	--	0.06019	151	1787	3

-- denote that the algorithm have not found the best merged LO

Finally we asked the students and teachers to browse the LOs merged by RA and LA; then they made a comment about coherence and usefulness of the LOs. The comments about queries 4, 6 and 10 using LA, highlighted the fact that the LOs have items from different topics. For example, linear regression is moderately related to multiple regression and as a consequence they should not be merged by LA. This behavior is avoided by RA because of the test of the *sim* function. Others students pointed out that the resulting LO from both RA and LA tended to be more complex than the original one. The merging process adds the items that fulfill the users query and the non-repeated items to the new LO so the outcome could be a LO that has more items than the source LOs. For instance, LO_1 contains the following items: introduction, least-square regression and applications. LO_2 contains the items: introduction, estimation methods and example. The merging process builds a new LO_3 made of least-square regression, applications from LO_1 and introduction, estimation methods, example from LO_2. Despite the fact that LOs, obtained using RA, are complex. The students opined that the LOs are useful because they contain the topics that they were looking for.

V. RELATED WORK

We propose an efficient algorithm for selecting and merging LOs which does not need additional information. The research of [12-14] and [15] rely on an ontology to perform the combination of LOs. This leads to spending time on build an ontology for every LO. Instead, our approach exploits the hierarchical structure and metadata of the SCORM and IMS standards.

The authors in [2] and [1] focus on combining isolated learning resources such as images and text, whereas msMLO picks out items which could be made of several resources including multimedia files. Using the SCORM standard makes portable LOs which can be published in several numbers of LMS or even exported to different formats automatically. The disadvantage of using only text and images restricts the kind of LO which can be handled. Moreover, ErauzOnt [1] relies on language processing techniques, so it is coupled with the language for which it was designed. It was exploited on basic pdf textbooks for primary school, most of them related to cosmology, geology and anatomy. This research does not report an evaluation of the LOs.

A learning design finder can be found in [16] this framework helps the instructor to build a new document. The system retrieves the documents matching the instructor´s needs then the wizard will suggest the way of combining such documents. The process of building is carried out manually by the instructor therefore it is impractical for a large number of documents.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have presented two algorithms to select and merge LOs packed using SCORM and IMS standards. One of the advantages of using standards is that the merged LOs can be shared between different LMS. Our goal was to evaluate the scalability of the two algorithms and the usefulness of the LOs obtained. To do so, we compared the scalability of RA and LA, and then we asked the authors of the queries to evaluate the merged LOs. RA is faster than LA so it obtains an admissible LO without sacrificing time. Also most of the students thought that the LOs obtained using RA were useful.

The key idea of RA is to detect relationship between LOs to prune the number of merging processes in the lattice structure and try to obtain the best LO. In contrast LA is a comprehensive algorithm so the number of merging processes increases exponentially. Both algorithms were affected by the number of LOs in the search set (d). RA obtained the merged LOs in a reasonable time because it performs d-1 merging processes at most. However, RA does not ensure that the resulting LO is the best or even if there are a combination of LOs to produce it. This handicap is made up for the less time spent on searching.

Although LA tests all possible combinations and lets us know whether a combination of source LOs can produce a top ranked LO, it is impractical for values of $d>4$. The remarkable strength of LA is that it performs only a single merging process no matter how many source LOs are involved in the EGT to get a LO.

We are aware of the fact that the average time of RA is acceptable but it could be speeded up using an ad-hoc index. This index should store not only the metadata and content of the LOs but also their hierarchies. Currently we are working on distributing the LO across different LMS focused on sharing only the description of the LO to reduce the net traffic.

REFERENCES

[1] I. C. Mikel Larrañaga, Jon A. Elorriaga, Ana Arruarte Lasa, Katrien Verbert, Erik Duval, "ErauzOnt: A Framework for Gathering Learning

Objects from Electronic Documents," in *IEEE International Conference on Advanced Learning Technologies*, 2011, pp. 656-658.
[2] A. D. Stefanutti Luca, Hockemeyer Cord, "Structuring and Merging Distributed Content," in *the 2005 conference on Towards the Learning Grid:Advances in Human Learning Services*, 2005, pp. 113-118.
[3] "Advanced Distributed Learning (ADL) initiative, SCORM Specifications Version 2004," ed, 2004.
[4] (2003, *Instruccional Management System (IMS)*
[5] A. G.-A. Serguei Levachkine, "Hierarchies Measuring Qualitative Variables," in *CICLing*, 2004, pp. 262-274.
[6] E. A.-G. Aldo Ramírez-Arellano, "Merging Learning Objects Automatically," in *International Conference on Education, Training and Informatics*, Orlando, Florida USA, 2012.
[7] O.-C. J. M. Guzman-Arenas Adolfo, "Measuring the understanding between two agents through concept similarity," *Expert Syst. Appl.*, vol. 30, 2006.
[8] R.-N. B. Baeza-Yates Ricardo A., *Modern Information Retrieval*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.
[9] O. C. Chen Zhibo, "Efficient OLAP with UDFs," in *the ACM 11th international workshop on Data warehousing and OLAP*, Napa Valley, California, USA, 2008, pp. 41-48.
[10] M. Michael, *et al.*, *Lucene in Action, Second Edition: Covers Apache Lucene 3.0*: Manning Publications Co., 2010.
[11] "Learning Object Metadata (LOM)," in *LOM standard, document IEEE 1484.12.1-2002*, ed, 2002.
[12] M. J. Doan AnHai, Dhamankar Robin, Domingos Pedro, Halevy Alon, "Learning to match ontologies on the Semantic Web," *The VLDB Journal*, vol. 12, pp. 303-319, 2003.
[13] G. A. V. Konstantinos Kotis, Konstantinos Stergiou, "Towards automatic merging of domain ontologies: The HCONE-merge approach," *J. Web Sem*, vol. 4, pp. 60-79, 2006.
[14] J.-Z. L. Jie Tang, Bangyong Liang, Xiaotong Huang, Yi Li, Kehong Wang, "Using Bayesian decision for ontology mapping," *J. Web Sem*, vol. 4, pp. 243-262, 2006.
[15] C.-S. L. Ching-Chieh Kiu, "Learning Objects Reusability and Retrieval through Ontological Sharing: A Hybrid Unsupervised Data Mining Approach," in *IEEE International Conference on Advanced Learning Technologies*, 2007, pp. 548-550.
[16] A. C. N. Chang, Chen, Irene Y. L., "Enhancing Learning Resources Reusability with a New Learning Design Framework," in *the Fifth IEEE International Conference on Advanced Learning Technologies*, 2005, pp. 414-418.

Virtual learning communities: yet another LMS?

Luigi Colazzo[1], Andrea Molinari[2]
[1]*Department of Industrial Engineering – University of Trento - Via Sommarive, 2 (Trento 38122)*
[2]*Department of Economics and Management – University of Trento - Via Inama, 5 (Trento 38122)*
{ luigi.colazzo, andrea.molinari}@unitn.it

Keywords: Virtual learning environment, e-learning, learning management systems

Abstract: After a long experimentation in institutions devoted to deliver educational projects, like universities, public bodies, private companies etc., the platform "Online Communities" created in 1998 by our research development team is ready to jump into the market. In the process of moving a self-made virtual learning platform in the market of free open source software, a number of issues and reflections have been made regarding the real differences between the two approaches, and most of all if we really need yet another learning management system. Many educational institutions are using different solutions for delivering e-learning, taken from the open source community (like Moodle) or from the proprietary market (like BlackBoard™). This paper presents the results of these reflections regarding the real differences between a virtual community approach and what traditionally is present in LMS, and why in our opinion yet another LMS is worth.

1 INTRODUCTION

This paper summarizes our experience in moving a typical self-made e-learning platform towards the market of e-learning platform, trying to differentiate from the several proposals, coming both from the market of open source and proprietary solutions.

We are seeing, on the one hand the development of training tools provided and managed via the web, and on the other side to the diffusion of virtual communities, typical of Web 2.0, some of which also possess the typical characteristics of communities of practice and learning. This rapid growth of e-learning applications certainly puts in evidence the strengths of this learning strategy, but does not hide, however, some of its limitations or weaknesses. Users of a Learning Management System (LMS) are sometimes forced into cramped virtual spaces called "virtual classrooms", or "virtual learning environments", lacking of effective tools for an exchange of views and mutual knowledge exchange. The current LMSs have put at the center of their purpose by far more the teachers rathen than the students. The teacher, however, is free to create, edit and shape the virtual learning environments in which it operates. The learners, in contrast, "suffer" the didactic approach conditioned by the software tool available, without being able to customize their own space. In the case in which a LMS was used as a tool accessory within a path-based learning method, it also presents the difficulty, not to be underestimated, linked to the lack of tools related with evaluating, developing and monitoring skills.

This is one of the directions in which we developed "Online Community" for complete market strategies broader than what has been done with the current LMS, considering that e-learning, especially informal learning, presents a series of undeniable strengths. In this process, it is the person in first to take control of the learning process, using the tools and services offered by the platform to the user in order to share and create knowledge. We can identify a series of conditions that favor the choice of such a learning strategy, namely the possibility of:

- Self-assessment of their skills and training gaps;
- Development of organizational skills to structure study time effectively and the breakdown of the planned activities;
- Flexibility in the simultaneous management of courses of study and work;
- Use of multiple tools and different methods of learning;
- Identification of a customized training.

2 ONLINE COMMUNITIES: THE PLATFORM

A Virtual Community is not the result of a process of social networking, but a virtual space shared by a group of people who have a common purpose. The virtual space of a community can be simple or complex, such as contain within it more virtual communities, thus establishing a hierarchical relationship "father-son". The (virtual) community can be an open space where anyone can enter, as it can be an enclosed space where entry is restricted only to a few people approved by an administrator of the community. Users can play different roles, with different rights and obligations associated with the use of space and collaboration services activated in a virtual community. The system maintains the consistency of the entire social landscape made up of virtual communities active at any time, providing users a choice of community's services that can be activated on request and can be used according to the associated permissions and roles.

This architecture allows us to easily model the organizational structure of a training establishment. For example, talking about a university environment, communities of the courses are components of larger communities called degree courses, and these in turn are part of the community faculty. The Community School on the other hand also includes the Community Faculty Council (a gated community) and other communities very heterogeneous, as the circle of Chess or the Commonwealth of first-year students, with its organizational structures, promoted by the Presidency of the Faculty and used to acclimatize quickly to new students. In each community, based on the purpose for which it exists, can be activated by different services. Community On Line incorporates many services, or features that every user has the option to use during his formative experience. The system is able to provide services of:

- e-Learning "traditional" asynchronous (upload and download educational materials, news groups, forums, bulletin board, calendar teaching, classroom management, course catalog management, user management, etc..) And synchronous (chat, streaming audio / video services "customized" closer to aspects of life-long learning and training on the job: tutoring, training on demand, contextual search tools to the problem, FAQs, etc..).
- integration with external information systems, such as the system of the student administration (register of lessons).
- use for courses in "offline" mode, ie courses already taken by the teacher recorded, digitized and made available to the user community (with the ability to sync movies with slides, podcasts, webcasts, SCORM modules, etc..).
- creation and use of self-assessment tests of learning, questionnaires, surveys, collection of opinions
- statistical analysis of user behavior (Community On Line in a data warehouse collects data on actions taken by users of the system).

All services are configurable, so that the rights of use and the resulting control can be varied depending on the role the user holds within a specific community. Each service is seen as structurally composed of at least three parts.

- The service in the strict sense;
- The collaborative environment in which to run the service, or the management of permissions for users
- The measurement system of its use by users

Although the platform has reached a good level of maturity with respect to technologies and methodologies of e-learning, it is subject to continuous developments and changes, both for the emergence of new standards in the market, which in general due to changing experiences interaction of living users of digital media. It is a certainty that the future evolution of e-learning platforms will be strongly determined by the answers that developers will be able to give to the needs of students, teachers, tutors, administrators and all others who use these systems for both a more traditional and structured learning takes place in formal training, both for feeding opportunities for exchange-involving peers in informal settings.

3 ONLINE COMMUNITIES: ANOTHER LMS FOR THE MARKET?

The market provides many solutions for the delivery of services through e-learning software platforms, proprietary and open. Our decision in order to enter the main stream of e-learning applications has followed a typical differentiation strategy, trying to propose something different from the rest, improving what already provided by competitors, and adding features not offered.

The decision to extend the functionality of an existing platform, developed from a university research group, with a significant number of deployed installations and international experience, is not derived solely from the favorable economic factors that today, in our opinion, are available for e-learning market

The reasons why we decided to adopt a different solution from (yet another) adaptation of existing platform are also strategic, technological, functional. They are in fact known limitations of e-learning platforms, which are summarized below and have been a cause of in-depth discussion in our team.

- Conceptual: the metaphor of the "class" concentrate the whole system around the idea of a "lesson", and this is not suitable for situations where more structured training is only one component of the interaction, such as collaboration among colleagues, working groups, secretariats, research groups, study groups, etc.. In these situations, the traditional metaphor of a training classroom needs to be adjusted, and it is certainly not ideal. In our opinion, much more flexibility and suitability is guaranteed by the metaphor of "virtual community", where the scope of the community can be generalized to any form of collaboration, not only learning. Changing the metaphor means to change not only the vocabulary of the system, but also the services that are provided, the users that could be assigned to the platform, the applications inside the company etc. It is just another LMS re-branding, but completely re-thinking the approach to services delivered to the users
- The difficulties in expanding with new features existing platforms, considering the know-how to build from scratch and the size of the platforms. The amount of functionalities in the major open source platforms (Moodle, DoceboLMS, Sakai, Atutor etc..), and the impressive amount of contributions, additions, different versions more or less certified that characterizes them is a serious obstacle for those who want to create a strong personalization of platform itself. This is particularly true when you want to distinguish your solution respect to the market and provide customized training services to its customers. The university team who have written completely from scratch, fifteen years ago, a platform devoted to collaboration and used in e-learning settings ensures on the one hand the total mastery of the platform in miminal details, and then a time- to-market customizations needed for various clients much faster than other solutions. Secondly, the features of platform devoted to collaboration (including e-learning that is an important declination, but certainly not the only) guarantee the extension of markets and customers reachable with different offers and services compared respect to training traditional services
- Given the known situation of the market for e-learning platforms, what we saw in educational institution is the domination of technically very good solutions, but most of the time used in a logic of "click and install". This is mainly due to funds and competences lack, but this brings educational institutions, or those who provide related services, to a "use what's available" strategy, rather than extending and customizing the existing training model. This latter conforms to what is offered by the platform because "it's free", and brings a further disadvantage, the play on the downside, not wanting to differentiate the educational services provided to the audience, and then flatten out on the solutions offered "as-is" from the platform. Something similar to what happened in the world of accounting systems, where the domination of big players with ERP, all-in-one solutions forced the customer to adapt their business processes to the software and not vice versa.
- The required services to a platform for e-learning inevitably tend to leave the confines of the training, also mediated by technology. Share a file for a lesson is not very different from the share to a secretariat, or to a meeting, or a working group. Use a blog to deliver training is not different from using a blog to provide a virtual space of interaction with our customers. At a time when the services of a traditional platform of e-learning are "open" to other uses, the characteristics of the training platform emerge with clarity and prove inadequate to areas more oriented towards collaboration, meaning they are adapted but certainly are not suitable. Projects are underway to use our platform "Online Communities" for Enterprise 2.0, or social enterprise, an application area where the

platform provides services in addition to those required for training, thus providing other markets for distributing the platform. Our enterprise 2.0 projects are carried out in collaboration with the local development agency, and open up a market of tools and collaboration platforms that will certainly see e-learning as a key asset of this new way of doing business, but certainly not the only one. This use of software platforms is certainly out of the domain of the use of traditional e-learning platforms.

- From a technical point of view, some structural issues of e-learning platforms for which, for applications or uses at Enterprise, substantially heavily customize the platform, often distorting it and then losing or compromising compatibility with future releases, or spending a considerable effort to be able to maintain. Many things, with significant numbers of users have started with the open source platforms, but then had to heavily customize the same, with the risk of losing compatibility with the mainstream software updates, and certainly facing huge costs to maintain the same line.
- Integration with the rest of the information system: the majority of e-learning platforms do not conceive, except by specific (and expensive) customization, integration with parts of the information system of the organization, as if the platform e-learning was a separate island that does not communicate, if not maybe just to the issues of single-sign-on, with the rest of the company. These deficiencies feel mainly with applications to manage human resources, accounting, with the part of logistics management, with ticketing systems with synchronous communications platforms perhaps already present in the company. If we speak of institutions with core-based business training (schools, universities, etc..), often the most important integration is avoided, or integration with the Students management system, and then all issues related to the master data, the votes, to the results and certificates. The two worlds are clearly separated, they live in two different environments very loosely integrated to avoid huge costs of passing data or data duplication.
- The presence of too many modules to do some operations, often in competition with each other, or activities that overlap with other activities in other contexts, and sometimes makes it difficult once again the "downward" integration and evolution of the platform e-learning: we are satisfied with minimal interventions not to touch a platform which, if not taken up and customized on a technical level, appears to be an object not to be touched
- The management of multi-organizations within the same platform is normally very difficult, if not impossible except through the installation of different instances of the e-learning platform, with apparent explosion of redundancies and associated costs. On the contrary, the main idea of "Online Communities", the nested virtual community to which a "parent" community (i.e. the company) may have more community children (the various branches and various partners) which in turn may have their own sub-communities, greatly simplifies these steps, providing a single installation to complex multi-corporate realities, subsidiaries, associated companies, consortia, etc.., each with its own separate organization, visual image, but with a ' unique software platform to manage everything.
- The addition of modules developed by various parties not aware of features developed by others creates a "visual" interface that often makes heavily customized e-learning platforms rather cumbersome to use. This often happens in situations with many individuals who manage the e-learning platform, such as school situations, with results that are often difficult to understand by the end user.
- Having only as support for collaborative activities related to teaching a platform for e-learning, we tend to "e-learning-ize" all these activities, masquerading educational activities with collaboration tasks that have little to do with, such as the management of internal competition, agenda scheduling, project management tasks, public tender and open innovation contexts, discussion of a joint working group etc..
- Finally, some typical concepts in the world of collaboration, long implemented within "Online Communities" and used by the users, are not available in major platforms, such as tools for programmable management training courses, project management tools and workflow management, managing the scheduling of events

and appointments, forms for evaluating the performance of the participants, the training of statistics available in a logic of multi-dimensional analysis and not just reporting, and many other implementation details that naturally "Online Community" presents in a logic of differentiation compared to the other platforms.

Regarding this last point, we want to briefly present the ideas that we are implementing in the “Online Communities” platform in order to follow this differentiation strategy, trying to avoid to present yet another LMS to the market

4 EVOLVING ONLINE COMMUNITIES: NOT ANOTHER LMS FOR THE MARKET

The objectives of the project we have presented in the paper include the expansion of our established collaboration platform based on virtual communities, used mostly for distance education, through the addition of services and unique features that allow us to attack the growing market related to online training, with innovative service offerings, differentiated and from the competitors. The objectives can be summarized into 5 areas:

1- Application of semantic technologies to a collaboration platform for educational components
2- Study of how to manage the collaboration between members of a virtual community
3- Research and development of tools for overcoming the limitations of the SCORM standard, through a greater integration and its generalization in structured educational paths
4- Study of the integrated mode of mobile learning tools geo-referenced (situated-learning) and with proximity services, to meet the needs of users on the move
5- Research and development of a strategy for the publication and management of the source code of the application in the Free / open source software (FOSS), through the creation of business models that ensure the creation of a community of developers for the autonomous development of platform, and consolidate the basis of revenues for installations by customers

These objectives will be reached starting from the consolidated basis of "Online Communities", a platform founded in 1998 in the Laboratory of Maieutics University of Trento, and developed over the years through research and development projects carried out by the institutional partners within the same group of local and international collaboration.

The platform now has a number of installations and uses by public and private clients which has made on the one hand the success of the metaphor used (that of "community" instead of "class"), and the other showed the need for its industrialization that allows, through significant applied research, to innovate in this area where many players exist but, in our opinion, too flattened solution are proposed, thus ending up providing the same educational services to all. It is this differentiation from a very aligned on the same, inexpensive, and “take-as-is" solutions that constitutes an opportunity not only to make a technology transfer from the world of applied research to the market, but attack the same with new ideas

The field of e-learning is certainly an area of unexplored market, in fact, it is well guarded by two main groups of LMS platforms (also known as Virtual Learning Environments-VLE):

- "closed" solutions, which in the past were based on large player (WebCT Blackboard, Thinking Cap LMS, SharepointLMS, JoomlaLMS, AvanzoLMS, GlobalScholer, Coggno, Desire2Learn, CCNet etc..) and now mostly proprietary solutions are often developed according to specific needs by large customers
- open-source and free platforms, (Moodle, Atutor, Sakai, DoceboLMS etc.), then readily available tools, to practically zero cost, with the availability of the source code, and suitable for any training situation, but often require expert knowledge for customization in case they are needed special training aspects, or if what is provided out-of-the-box is not sufficient to meet the needs of trainers.

To do this, the effort of our research team will be directed towards the evolution of the platform towards the above objectives, which aim at a glance

- Add advanced components from the concrete experiences of the world of research, such as the addition of components related to semantic technologies. The addition of components derived from research on semantic technologies and the semantic web to a platform for e-learning that is not an experimental platform, is quite unique in the world

- Create an open source community around the product, developing and distributing a free and open version of the platform in the traditional circuit of the portals for the distribution of open source software (SourceForge, CodePlex, etc..), Run by a community of users / developers on the model of what already exists, in a niche market niche that sees the services of collaboration / Enterprise 2.0 and not just e-learning. Similar to this, in order to guarantee revenues, the idea is to develop and deploy an enterprise version also on which to build "premium" services to sell to advanced users with personalization features of the teaching model, the method of use and integration with its own information system (accounting, office staff, classroom management, shared calendars, document management etc.). It must be noted that this approach is common to most of the LMS available on the market.
- Look for the details and the results of research to be able to abstract from the dependence on the SCORM standard and especially the specific SCORM engines, while maintaining compatibility with the certification requirements of the laws (such as the requirements of the Digital Agenda for Italy), and create a software layer generalized to be able to hook up to any present or future engine that will be indicated by the regulations or by the market as the standard of reference
- To study and develop software components that do definitely turn the platform and its services to the world of mobile learning, exploiting the huge potential of mobility and geo-referencing provided by modern mobile devices
- To study the degree of integration of a platform for e-learning with organizational and administrative aspects to complete the cycle dispensing active training (payments, ticketing, voice teaching, etc..), Providing tools for integration with the host computer system to facilitate this step and then provide a competitive advantage on the enterprise market.

5 REFERENCES

- Colazzo L., Conte F., Molinari A., Villa N. (2007), *Real communities vs. Virtual communities: structural adaptation of a Learning Management System*. In: *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2007*, Chesapeake (VA): AACE, 2007. p. 2416-2423. "E-Learn 2007 Conference", Quebec City, October 2007
- Schonfield E., Yen Y.W. (2007), *It's a Web, Web, Web 2.0 World*. Business 2.0, February 2007
- Wellman B. (2005), *Community: from neighborhood to network*. Commun ACM 48(10):53-55
- Johnson G. J., Ambrose P.J. (2006) *Neo-tribes: the power and potential of online communities in health care*. Commun ACM 49(1):107-113
- Parker K., Chao J. (2007) Wiki as a Teaching Tool, Interdisciplinary Journal of Knowledge and Learning Objects, Vol. 3, 2007
- Boyd D. (2006) *Friends, friendsters and mySpace Top 8: writing community into being on social network sites*. First Monday 11(2), December. http://www.firstmonday.org/ issues/issue11_12/boyd. Accessed October 2008
- O'Reilly T. (2005) What is Web 2.0?, OReillynet.com. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09 /30/what-is-web-20.html. Accessed October 2008
- Anderson C. (2006), The long Tail: how endless choice is creating unlimited demand. Random House Business Books, N.Y.
- Rheingold H., (1993) The Virtual Community, 1993 (italian translation Comunità virtuali, Sperling & Kupfer, Milano, 1994).
- Strother, J., "An assessment of the effectiveness of e-learning in corporate training programmes", International Review of Research in Open and Distance Learning, 3 (1), pp. 1–17, 2002.
- Villa N., Colazzo L., Conte F., Molinari A. (2007), *Real communities vs. Virtual communities: structural adaptation of a Learning Management System*, in Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2007, Chesapeake (VA): AACE, 2007, p. 2416-2423. Atti di: E-Learn 2007 Conference, Quebec City, 15th-19th October 2007.

Proceedings
International Workshop on
Visual Languages and Computing
(VLC 2013)

A Fast Intersection Detection Algorithm For Qualitative Spatial Reasoning

Chaman L. Sabharwal
Computer Science Department
Missouri University of S&T
Rolla, MO – 65409, USA
chaman@mst.edu

Jennifer L. Leopold
Computer Science Department
Missouri University of S&T
Rolla, MO – 65409, USA
leopoldj@mst.edu

Abstract. In this paper we present a framework for determining the intersection of geometric objects based on the 9-Intersection model used in qualitative spatial reasoning. Triangle-triangle intersection is the computational basis for interaction between 2D/3D objects; hence a fast, robust intersection detection algorithm is beneficial. Additionally, we present efficient predicates for classification of the intersections. This work is applicable for most region connection calculi, particularly, VRCC-3D+, which detects intersections in 3D as well as projections in 2D for occlusion detection.

Keywords: Intersection Detection, Classification Predicates, Spatial Reasoning.

I. INTRODUCTION

Detecting possible intersection between pairs of objects and classifying the type of intersection is important in many areas such as geographic information systems [1], CAD/CAM geometric modeling [2], real-time rendering [3], geology [4], and qualitative spatial reasoning [5]. One of the major tasks in modeling and automated reasoning is to decompose the objects and use relevant data structures for efficient processing. Thus triangulation is the first step in the process of intersection detection. In order to detect intersection between objects, we actually need to detect and classify the intersection between the constituent pairs of triangles. In fact, in all the modeling computations, the triangle-triangle intersections are the integral part. Attempts to speed up the intersection process are still being explored to improve various applications [6, 7].

Brute force pairwise intersection of two objects is very inefficient. The quad tree approach is most effective in partitioning the objects before the actual triangle-triangle intersection takes place. There are several approaches to quad tree construction. Axis Aligned Bounding Boxes (AABB) [8] can be used for pre-processing the objects. The overlap test between two AABBs is far more cost effective than the triangle-triangle intersection test. This approach is fast, but can be slow in some cases where objects are slim and elongated; they cannot fit some triangles (e.g., long-thin oriented triangles) tightly. In such cases, non-axis Oriented Bounding Boxes (OBB) perform more efficiently. The minimal ellipsoids and OBBs provide tight fits. However, detecting even the existence of overlap between them is relatively expensive [2, 9].

For some applications it may be beneficial to employ a hybrid approach, Axis-aligned-Oriented Bounding Boxes (AOBB). In this case, the AABB algorithm would be applied to some parts and the OBB algorithm would be applied to other parts of an object selectively. Since there is a large computation cost overhead in such an implementation, this approach has not received much attention; a comprehensive discussion is beyond the scope of this paper. One particular aspect of the larger problem is the decomposition criteria of detecting the leaf nodes in the quad tree, which depends on the application. For example, for the surface/surface intersection algorithm used in CAD/CAM, acceptable surface flatness is the stopping criteria for decomposition [8], whereas in VRCC-3D+ the surface of each object is already triangulated, so a count of triangles is the stopping criteria.

This paper is organized as follows: Section II briefly reviews the background and related work. Section III discusses motivation and terminology, and develops the main algorithm for triangle-triangle intersection. Section IV describes predicates for classifying the intersection between pairs of triangles. Section V discusses the applications, and Section VI concludes, followed by references.

II. BACKGROUND

In QSR, spatial relations between regions are defined axiomatically using first order logic or the 9-Intersection model. Using the latter, spatial relations are defined based on the intersections of the interior, boundary, and exterior of one region with those of a second region. Consequently, the intersection of regions results in the 9-Intersections. In fact, it has been shown that it is sufficient to define the spatial relations by computing 4-Intersections, (namely, Interior – Interior (IntInt), Boundary – Boundary (BndBnd), Interior – Boundary (IntBnd), Boundary – Interior (BndInt)) instead of 9-Intersections [10].

We briefly review three approaches for determining intersection detection that are relevant to our discussion herein, starting with Moller's algorithm [11]. For two triangles T_1 and T_2, two planes P_1 and P_2 that support the triangles are determined. Then it is determined whether triangle T_1 and plane P_2 overlap, and triangle T_2 and plane P_1 overlap. If all vertices of triangle T_1 lie on the same side of plane P_2, and no vertex of triangle T_1 lies in plane P_2, then triangle T_1 and plane P_2 do not overlap; hence the triangles do not intersect. The same test is repeated for triangle T_2 and plane P_1. If the two tests succeed, then the algorithm computes the intersection segment S_{12} of triangle T_1 and plane P_2, as well as segment S_{21} of triangle T_2 and plane P_1. Further, a line-line 3D intersection algorithm is used to test whether the two segments S_{12} and S_{21} overlap; see Fig. 1. If they do intersect, the triangles are guaranteed to cross intersect. It should be noted that this algorithm does not address additional computational questions related to the classification of intersection points, information that would be useful for spatial reasoning.

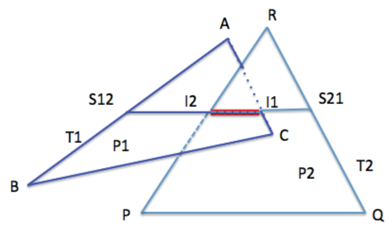


Figure 1. S_{12} is the intersection of T_1 and P_2, S_{21} is the intersection of T_2 and P_1; intersection of S_{12} and S_{21} is the intersection of T_1 and T_2.

The second approach [12] reviewed here starts out similar to the above approach for solving the triangle-triangle intersection problem. If the above preliminary test succeeds, now instead of computing segment S_{21} of the intersection between triangle T_2 and plane P_1, the solution is approached slightly differently. It will be determined whether intersection segment S_{12} of triangle T_1 and plane P_2 overlap. If segment S_{12} and triangle T_2 overlap, the intersection algorithm succeeds, and the determination is made that the two triangles intersect; see Fig. 2. Unfortunately this approach has the same "classification" problems as the previous algorithm.

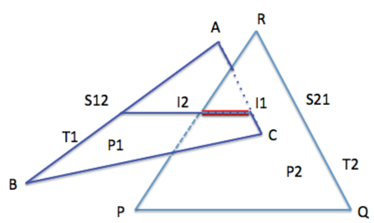


Figure 2. S_{12} is the intersection T_1 and P_2, S_{21} is the intersection of T_2 and S_{12}.

A third approach [6] differs in its strategy from the previous two algorithms: if the edges of one triangle intersect the surface of the other triangle, it can be concluded that the triangles intersect. In this case, six such intersection tests are performed: three to test triangle ABC against the edges of triangle PQR, and three tests where the roles of ABC and PQR are reversed. This technique solves the six sets of linear equations associated with the problem and exploits the relations between these sets to speed up their solution. Like the other algorithms, it does not answer the related intersection classification questions. Notably, Badouel's algorithm [13] uses a vector approach in preference to Cartesian coordinates, yet still performs the same tests as the aforementioned algorithms.

Our approach differs from all of the above. These algorithms are not suitable for spatial reasoning where a qualitative answer of yes or no is sufficient instead of quantitative calculation of the intersection segment. At the root of our algorithm is vector analysis, which is in its own way beautiful, elegant, and is the foundation for making the complex calculations simple.

III. THE ALGORITHM DESCRIPTION

Many papers [11, 12, 13, 14] have concentrated on the fast implementation of similar intersection computation algorithms. In contrast to other approaches, our algorithm relies more on logical tests rather than computational arithmetic tests. Furthermore, our algorithm not only determines whether or not an intersection exists, but also classifies an intersection segment as "degenerate" (a single point) or "non-degenerate" (a segment of positive length).

A. TRIANGLE-TRIANGLE CROSS INTERSECTION

A.1 *Triangles: Representation and Intersection*

Our algorithm is different from all other algorithms mentioned in the above background section. By judiciously formulating the first step in the algorithm, the second step in those algorithms is not needed in our approach for intersection detection. We start with a vector parametric representation of two triangles, equate the two forms, eliminate calculation for one triangle, and carry our calculations in only one triangle to determine if the two triangles intersect. Cross intersection may be differentiated into seven types: vertex-vertex, vertex-edgeInterior, vertex-triangleInterior, edge-edgeCross, edge-edgeCollinear, edge-triangleInterior, triangleInterior-triangleInterior intersection; see Fig. 3. In QSR, these can be used to determine the predicates involving interior, exterior, and boundary intersections (i.e., IntInt, IntBnd, BndInt, BndBnd) that subsequently distinguish the RCC8 base relations DC, EC, PO, EQ, TPP, TPPc, NTPP, and NTPPc [15].

A.2 *The Algorithm Step 1 Triangle-Triangle Intersection Algorithm*

In this section we present the cross intersection algorithm and formulate the intersection detection decision in one step, whereas other algorithms use an additional step before making a decision when an intersection exists.

Boolean triTriCrossInt1(tr1 = ABC, tr2 = PQR)
Input: two triangles ABC and PQR
Output: Boolean value true if there is a cross intersection; otherwise, false value is returned.

The equations of planes supporting the triangles ABC and PQR are

$X = A + u\,U + v\,V$, where $U = B - A$, $V = C - A$,
$0 \le u, v, u + v \le 1$

$X = P + s\,S + t\,T$, where $S = Q - P$, $T = R - P$,
$0 \le s, t, s + t \le 1$

Let N_1 and N_2 be normals to the surfaces of the triangles. The normals are defined by

$$N_1 = U \times V, \quad N_2 = S \times T$$

If the triangles intersect, no matter what the special case may be, the intersection points correspond to some parameter values. We equate the two triangles' parameterization and solve for the parameters simultaneously. Then parameters u and v; p and q must satisfy the vector equation

$$A + u\,U + v\,V = X = P + s\,S + t\,T$$

If the triangles cross, then the normals to the triangles cannot be parallel, so $N_1 x N_2 \neq 0$. Rearrange the equation to get

$$u\,U + v\,V = P - A + s\,S + t\,T \qquad \text{or}$$

$$u\,U + v\,V = AP + s\,S + t\,T$$

Since S and T are orthogonal to N_2, take the dot product of this equation with N_2 to eliminate S and T from the above equation to get

$$u\,U \bullet N_2 + v\,V \bullet N_2 = AP \bullet N_2$$

This is the equation of a line in the uv - plane for real u, v. Using this information, the non-parametric equation of the line is transformed to a parametric vector equation using real parameter t.

The vector equation for real parameter t becomes

$$(u, v) = AP \cdot N_2 \left(\frac{(U \bullet N_2, V \bullet N_2)}{U \bullet {N_2}^2 + V \bullet {N_2}^2}\right) + \lambda (V \cdot N_2, -U \cdot N_2)$$

Then parameter values u, v are explicitly written as

$$u = AP \cdot N_2 \left(\frac{U \cdot N_2}{U \bullet {N_2}^2 + V \bullet {N_2}^2}\right) + \lambda\, V \cdot N_2 \qquad (1)$$

$$v = AP \cdot N_2 \left(\frac{V \cdot N_2}{U \bullet {N_2}^2 + V \bullet {N_2}^2}\right) - \lambda\, U \cdot N_2 \qquad (2)$$

From this point on, our approach is different from all the previous algorithms. Now we write another equation

$$u + v = AP \cdot N_2 \left(\frac{(U \bullet N_2 + V \bullet N_2)}{U \bullet {N_2}^2 + V \bullet {N_2}^2}\right) + \lambda (V \cdot N_2 - U \cdot N_2) \quad (3)$$

and use the barycentric constraint on parameter λ.

The parametric bound for t satisfying equations (1)-(3) are denoted by λ_m and λ_M. The following step of the algorithm shows how the parameter values λ_m, λ_M are used to detect whether or not intersection exists instead of actually computing the intersection points as in the previous methods. Also in order to determine a Boolean answer of yes or no for whether intersection exists, we do not need the second test to find the intersection of the second triangle T_2 with plane P_1 of the first triangle to determine intersection as done in [11], or the intersection of computed segment S_{12} with the second triangle T_2 as seen in [12]. For spatial reasoning, the following algorithm is sufficient and optimally efficient. This algorithm is much faster if a quick rejection test is used at each step while computing λ_m and λ_M. No doubt, these λ_m, λ_M can be used to obtain the 3D intersection points used in those algorithms.

A.3 *The Algorithm Step 2, Triangle-Triangle Intersection Algorithm*

```
Boolean triTriCrossInt2(λm, λM)
    if (λm > λM) // no intersection
        return false,
    else
        if (λm < λ M ) //proper segment intersection
            return true
        else
            (λm == λM) //single point Intersection
            return true
        endif
    endif
```

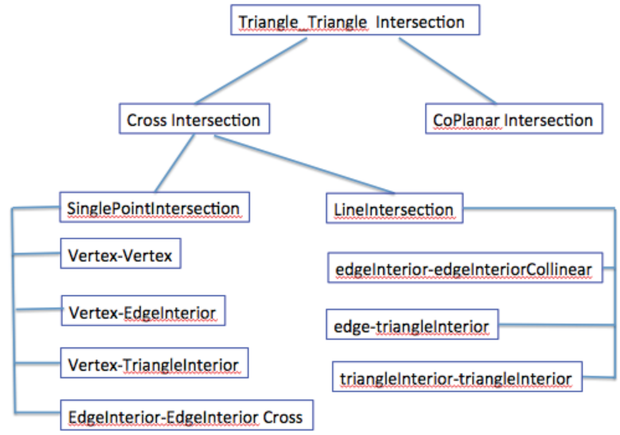


Figure 3. Hierarchy of triangle-triangle cross intersection predicates; coplanar predicates not shown.

If there is no intersection, the algorithm may quickly stop at any of the three inequalities resolution. When there is an intersection, the application may need to know the actual intersection. Using these t parameter values for the intersection line constrained to the uv-triangle, we can calculate the (u, v) coordinates for 3D points. The triangle parametric coordinates (u_m, v_m), (u_M, v_M) corresponding to $\lambda_m, \lambda_{\square}$ are computed from equations (1)-(3).

$$u_m = AP \cdot N_2 \left(\frac{U \cdot N_2}{U \bullet {N_2}^2 + V \bullet {N_2}^2}\right) + \lambda_m\, V \cdot N_2$$

$$u_M = AP \cdot N_2 \left(\frac{U \cdot N_2}{U \bullet {N_2}^2 + V \bullet {N_2}^2}\right) + \lambda_M\, V \cdot N_2$$

$$v_m = AP \cdot N_2 \left(\frac{V \cdot N_2}{U \bullet {N_2}^2 + V \bullet {N_2}^2}\right) - \lambda_m\, U \cdot N_2$$

$$v_M = AP \cdot N_2 \left(\frac{V \cdot N_2}{U \bullet {N_2}^2 + V \bullet {N_2}^2}\right) - \lambda_M\, U \cdot N_2$$

These parameter values (u_m, v_m), (u_M, v_M) corresponding to the intersection points are determined by λ_m and $\lambda_{\square}$ where $\lambda_m \leq \lambda_{\square}$. Note that there is no such ordering as, $\lambda_m \leq \lambda_{\square}$, on u and v in (u_m, v_m), (u_M, v_M). For example, $(u_m, v_m) = (0.2, 0.3)$ and $(u_M, v_M) = (0.3, 0.2)$ correspond to a valid intersection segment, whereas in this case $u_m < u_M$ but $v_m > v_M$.

From now on we use the symbols $A_m = (u_m, v_m)$, $A_M = (u_M, v_M)$. When $\lambda_m \leq \lambda_{\square}$, this guarantees that at least one point of PQR lies on the segment A_mA_M. This could not be claimed from the previous algorithms even when the first step intersection segment existed. To determine the complete intersection, we similarly calculate s_m, s_M, t_m, t_M, P_m, P_M for triangle PQR.

We update them with the common segment as in [11, 12]. Also we update their parametric coordinates with respect to triangles ABC and PQR. So the intersection segment A_m to A_M becomes P_m to P_M *or* P_M to P_m. Note that in all the calculations, we will use the parametric coordinates instead of 3D coordinates as computed and used in other algorithms. When there is no ambiguity, we will continue to use interchangeably A_m for (u_m, v_m), A_M for (u_M, v_M), P_m for (s_m, t_m), *and* P_M for (s_M, t_M).

In QSR, we also need qualitative information about the type of intersection as used in the 9-Intersection model [5]. For this reason, we proceed to the next step to classify the intersection. Remember that the computational part is complete. From now on, unlike other algorithms, we will simply make logical tests on Boolean expressions to categorize the type of intersection.

The algorithm is implemented in Python. In an actual application there can be thousands of such intersections. In the previous algorithms there two steps of multiple operations, whereas here there is a single step of such operations. Without going through the exact number of operations, it is apparent that computational efficiency of even one multiplication can make a difference in user interaction. Thus our algorithm is at least 50% more efficient.

B. *Intersection of Co-Planar Triangles*

Similar to cross intersection, the analysis has been completed for coplanar intersections for pairs of triangles. Due to space consideration, the algorithm, results, and corresponding predicates are not included here. Accordingly, there are predicates for the cases of vertex-vertex, vertex-edgeInterior, edge-edgeCollinear, and triangleInterior-triangeInterior intersection.

IV. INTERSECTION CLASSIFICATION PREDICATES

For completeness, we first define the following predicates relevant to any triangle. For classification of intersection, we apply them to intersections between pairs of triangles and create the integrated composite predicates for use in qualitative spatial reasoning. Let (u, v) and (s, t) refer to the parameter values of 3D points as they relate to each triangle represented parametrically. Here we characterize the triangle in terms of parametric coordinates independent of the 3D positions. No computations or comparisons of 3D points are used.

A. *Position Predicates*

The predicates are overloaded with one argument and two arguments; first we describe the predicates with one argument, then use them to categorize predicates with two arguments. The points in a triangle are denoted by $X = (u, v)$, $X_1 = (u_1, v_1)$, and $X_2 = (u_2, v_2)$ with their parametric coordinates.

A.1 *Taxonomy of Predicates With One Argument*

There are three categories of predicates for a point $X = (u, v)$ relative to a triangle. These predicates are defined below as: *vertex*(X), *edgeInterior*(X), and *triangleInterior*(X). The *edge*(X) encompasses *vertex*(X) and *edgeInterior*(X); the triangle(X) encompasses *edge*(X) and *triangleInterior*(X);

For *vertex* point, one of the corners of the triangle,
vertex(X) = $(u, v) == (0, 0)$ or $(u, v) == (1, 0)$ or $(u, v) == (0, 1)$

For *edgeInterior* point, non-vertex edge point,
edgeInterior(X) = $(u == 0$ and $0 < v < 1)$ or $(v == 0$ and $0 < u < 1)$ or $(u > 0$ and $v > 0$ and $u + v == 1)$

For *triangleInterior* point, non-edge triangle point,
triangleInterior(X) = $(0< u <1)$ and $(0< v <1)$ and $(0< u + v <1)$

A.2 *Taxonomy of Predicates With Two Arguments*

For a pair of points, there are seven categories of predicates. These predicates are defined below as: *vertex*(X_1, X_2), *edge*(X_1, X_2), *edgeInterior*(X_1, X_2), *edge0Interior*(X_1, X_2), *edge1Interior*(X_1, X_2), *edge2Interior*(X_1, X_2), and *triangleInterior*(X_1, X_2).

Two points X_1 and X_2 are the *same* vertex of a triangle if and only if $(u_1, v_1)=(u_2, v_2)$ and X_1 is a vertex.
vertex(X_1, X_2) = $X_1 == X_2$ and *vertex*(X_1)

Two points X_1 and X_2 are distinct edge points on the *same* edge if and only if they have same u coordinates, or same v coordinates, or same $u+v$ values
edge(X_1, X_2) = X_1 != X_2 and *edge*(X_1) and *edge*(X_2) and $(u_1==u_2==0$ or $v_1==v_2==0$ or $u_1+v_1 == u_2+v_2 ==1)$

Two points X_1 and X_2 are on the same edge with at least one *edgeInterior* if and only if they are edge points, and at *least* one of them is an interior point,
edgeInterior(X_1, X_2) = *edge*(X_1, X_2) and (*edgeInterior*(X_1) or (*edgeInterior*(X_2))

Two points X_1 and X_2 such that (1) they lie on the same edge, (2) none is an interior point of the edge are defined as
edge0Interior(X_1, X_2) = *edge*(X_1, X_2) and *vertex*(X_1) and *vertex*(X_2)

Two points X_1 and X_2 such that (1) they lie on the same edge, (2) exactly one point is an interior point of the edge are defined as
edge1Interior(X_1, X_2) = *edge*(X_1, X_2) and (*edgeInterior*(X_1) and *vertex*(X_2)) or (*vertex* (X_1) and *edgeInterior* (X_2))

Two points X_1 and X_2 such that (1) they lie on the same edge, (2) both points are interior points of the edge are defined as
edge2Interior(X_1, X_2) = *edge*(X_1, X_2) and *edgeInterior*(X_1) and *edgeInterior* (X_2)

With parameter values (u_1, v_1), (u_2, v_2), the segment X_1X_2 overlaps with interior of triangle if X_1 and X_2 are in the same triangle and both do not lie on the same edge.
triangleInterior(X_1, X_2) = *triangle*(X_1) and *triangle*(X_2) *and* $(u_1 \neq u_2)$ and $(v_1 \neq v_2)$ and $(u_1+v_1 \neq u_2+v_2)$)

V. APPLICATIONS AND FUTURE WORK

For spatial reasoning, the relations between convex objects in VRCC-3D+ are the RCC8 base relations: DC, EC, PO, EQ, TPP, TPPc, NTPP, and NTPPc [15]. Each relation is in turn defined with 4-Intersection predicates [10], determined from the triangle-triangle intersection predicates defined in Section IV. An AABB tree can be constructed for possible intersection of triangle pairs. For VRCC-3D+, most of the time the entire AABB tree is not necessary for determination of relations, and thus not constructed. As soon as an intersection is found, the choice of 4-Intersection predicates is resolved, and AABB tree construction is terminated. This results in considerable efficiency in the determination of spatial relations among objects in 3D. In case some predicate is not resolved or all triangle pair intersections are false, the complete tree is used in conjunction with ray-object intersection using the odd parity rule to determine if one object is part of the other object. The worst case cost for intersection detection is given by the equation

$$T= N_b x C_b + N_t x C_t$$

where T is the total cost function for intersection detection, N_b is the number of bounding volume pairs overlap tests, C_b is the cost of testing one pair of bounding volumes for overlap, N_t is the number of triangle pairs tested for intersection, and C_t is the cost of testing a pair of triangles for intersection. In an application there are likely thousands of such intersections. In the previous algorithms there two steps of multiple operations, whereas here there is a single step of such operations. Without going through the exact number of operations, it is apparent that computational efficiency of even one multiplication can make a difference in user interaction. Thus our algorithm (which has been implemented in Python) is at least 50% more efficient.

VI. CONCLUSIONS

Herein we presented a new framework for determining and characterizing the intersection of geometric objects based on the 9-Intersection model used in qualitative spatial reasoning. In contrast to other algorithms, our approach relies more on logical tests rather than computational arithmetic tests. Furthermore, our work differs by not only determining whether or not an intersection exists, but also classifies an intersection segment as "degenerate or "non-degenerate." Currently we are applying these methods to the qualitative spatial reasoning system, VRCC-3D+. We hope these ideas will be useful in other related applications that employ triangle-triangle intersection.

VII. REFERENCES

[1] M. J. Egenhofer, R.G. Golledge, Spatial and Temporal Reasoning in Geographic Information Systems, Oxford University Press, USA, 1998.

[2] Houghton E.G., Emnett R.F., Factor J.D. and Sabharwal C.L., Implementation of A Divide and Conquer Method For Surface Intersections; Computer Aided Geometric Design Vol.2, pp. 173-183, 1985.

[3] Oren Tropp, Ayellet Tal, Ilan Shimshoni, A fast triangle to triangle intersection test for collision detection, Computer Animation and Virtual Worlds, Vol17(50, pp.527-535, 2006.

[4] Caumon G, Collon-Drouaillet P, Le Carlier de Veslud C, Viseur S, Sausse J (2009) Surface-based 3D modeling of geological structures. Math Geosci 41:927–945, 2009.

[5] Max J. Egenhofer, R. Franzosa, *Point-Set topological Relations*, International Journal of Geographical Information Systems 5(2), pp.161-174, 1991.

[6] Ahmed H. Elsheikh, Mustafa Elsheikh, A reliable triangular mesh intersection algorithm and its application in geological modeling, Engineering with Computers, pp.1-15, 2012

[7] N. Eloe, J. Leopold, C. Sabharwal, and Z. Yin, “Efficient Computation of Boundary Intersection and Error Tolerance in VRCC-3D+”, Proceedings of the 18h International Conference on Distributed Multimedia Systems (DMS’12), Miami, FL, Aug. 9-11, 2012, pp. 67-70.

[8] Sabharwal C.L., and Factor J.D., Cross Intersection Between Any Two C^0 Parametric Surfaces. Proceedings of The International Australian Computer Graphics Conference, Ausgraph'88, Melbourne, Australia, 1988, 37-42.

[9] S. Gottschalk, M. C. Lin, D. Manocha, OBBTree: A Hierarchical Structure for Rapid Interference Detection, *Proc. of ACM Siggraph'96, pp. 12.*

[10] C. Sabharwal and J. L. Leopold, “Reducing 9-Intersection to 4-Intersection for identifying relations in region connection calculus, ” in The 24th International Conference on Computer Applications in Industry and Engineering, 2011, pp. 118–123.

[11] Moller T. A fast triangle-triangle intersection test. Journal of Graphics Tools, 1997; 2(2): 25–30.

[12] Held M. ERIT a collection of efficient and reliable intersection tests. Journal of Graphics Tools 1997; 2(4): 25–44.

[13] Badouel Didier, "An Efficient Ray-Polygon Intersection," *Graphics Gems* (Andrew S. Glassner, ed.), Academic Press, pp. 390-393, 1990.

[14] C. Sabharwal, “Survey of implementations of cross intersection between triangular surfaces, ” MDC Report Q0909 (Now Boeing at St. Louis, MO-USA), 1987.

[15] C. Sabharwal, J. Leopold, and N. Eloe, “A more expressive 3D region connection calculus,” in Proceedings of the 2011 International Workshop on Visual Languages and Computing (in conjunction with the 17th International Conference on Distributed Multimedia Systems (DMS’11)), 2011, pp. 307–311.

Visual Thinking Design Patterns

Colin Ware
Center for Coastal and Ocean Mapping
University of New Hampshire
Durham, NH, USA
cware@ccom.unh.edu

William Wright
Oculus Info Inc.
Toronto
Ontario, Canada
bill.wright@occulusinfo.com

Nicholas J. Pioch
Systems & Technology Research LLC, Suite 5850, 400 W. Cummings Park, Woburn, MA 01801
npioch@STresearch.com

***Abstract*—In visual analytics, interactive data visualizations provide a bridge between analytic computations, often involving "big data", and computations in the brain of the user. Visualization provides a high bandwidth channel from the computer to the user by means of the visual display, with interactions including brushing, dynamic queries, and generalized fisheye views designed to select and control what is shown. In this paper we introduce Visual Thinking Design Patterns (VTDPs) as part of a methodology for producing cognitively efficient designs. We describe their main components, including epistemic actions (actions to seek knowledge) and visual queries (pattern searches that provide a whole or partial solution to a problem). We summarize the set of 20 VTDPs we have identified so far and show how they can be used in a design methodology.**

Keywords—design patterns, visual thinking, data visualization, visual analytics.

I. Introduction

Visualizations are tools for reasoning about data and to be effective they must support the activities of visual thinking. Part of this is ensuring that data is mapped to the display in such a way that informative patterns are available to resolve visual queries concerning the cognitive task. This requires matching the graphic representation with the capabilities of human visualization. For example, correlations between variables should be visually easy to see and commonly searched for symbols should be more distinctive than those that are rarely sought out. In addition, interactions must be designed to support an efficient visual thinking process. Visual analytics is an example of distributed cognition and cognitively efficient interactions require that perceptual and cognitive processes in the brain of the analyst must be efficiently linked to computational processes in a computer. For example, data points representing companies can be shown simultaneously in a map view and in a scatter plot view; the technique of brushing can be applied so that points on the map, when selected, are highlighted in both views. This can support reasoning about the growth of industries related to geographic regions, but to be cognitively efficient the brushing effect should ideally appear in less than a tenth of a second.

In this paper we introduce the concept of Visual Thinking Design Patterns (VTDPs) as a tool to help with the construction of cognitively efficient visualization designs. VTDPs are based partly on a prior construct developed by Ware [1] and called visual thinking algorithms (VTAs). VTDPs represent a broadening of this original concept with a change in emphasis. VTDPs are a method for describing the combined human-machine cognitive processes that are executed when interactive data visualizations are used as cognitive tools.

First we describe the characteristics of VTDPs followed by a brief description of the set of 20 we have identified to date. Two of the VTDPs are described in somewhat greater detail to show how they combine machine computation with perceptual and cognitive processes. Finally we show how VTDPs can be used in an agile design process.

VTDPs take their inspiration from Alexander's design patterns [2] intended for architects as well as designed patterns as used by software engineer [3]. Although considerable research has shown that perceptual and cognitive principles can be applied usefully to the design of interactive visualization, this knowledge is only applied in practice if a particular designer has taken an interest in the relevant research. VTDPs are intended to provide an accessible structured method for combining knowledge about interaction methods and visualization designs together with cognitive and perceptual principles.

- Like their precedents, VTDPs are intended to describe best practice example solutions to design problems where interactive visualization is an intended component.
- VTDPs provide a method for taking into account perceptual and cognitive issues especially key bottlenecks in the visual thinking process, such as limited visual working memory capacity. They also provide a way of reasoning about semiotic issues in perceptual terms via the concept of the visual query.
- VTDPs incorporate the common set of interactive techniques used in visualization and suggest how they may be used separately or in combination.

This is not to say that there are no prior methodologies for incorporating cognitive principles into design. About three decades ago the GOMS (Goals, Operators, Methods and Selection Rules) model [4] was introduced and more sophisticated approaches have followed in the form of the

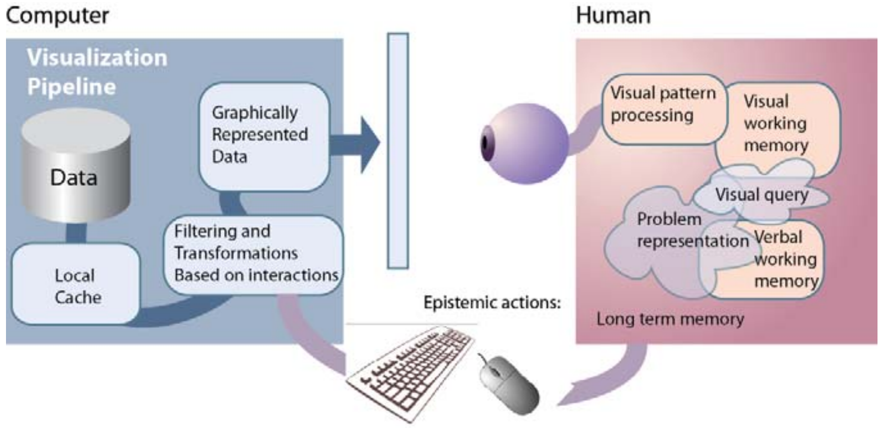


Fig. 1. Key elements of the VTDP virtual machine

ACT-R [5] model and SOAR cognitive modeling systems. These modeling systems provide executable cognitive models, containing timings for cognitive computations and as well as for common interactions such as mouse movements. But to use these models as part of a design process a proposed interface must be designed in detail and executed in a simulation that includes both simulations of cognitive operations and simulations of how the computer application will behave. This is beyond the capabilities of all but a few specialized centers. Set against this is the need for agile design based on rapid prototyping.

VTDPs should be understandable with a modest amount of training, and it should be possible to incorporate them into common design practices. VTDPs are potentially applicable to all visualization design problems, including design of presentation materials, interactive training materials and the design of analytic tools. In the following section we will concentrate on a single application domain — the design of visualization tools used in visual analytics.

Like software engineering and architectural design patterns, VTDPs are not modules and not re-usable. The demands of analysis are almost infinitely varied due to the enormous variety of data types and analytic problems. As a result, any modularization of VTDPs would necessarily restrict the domain. Nevertheless, implementations of VTDPs can take advantage of modular components. For example, map display software may allow for extra magnifying windows which can be used to support the pattern comparisons in a large information VTDP space.

II. THE COMPONENTS OF A VTDP

A VTDP begins with a statement of the problem it is designed to solve, together with one or more examples. In most cases it will contain a pseudo-code description of the combined cognitive process and this will typically incorporate many the following components.

A. Perceptual and Cognitive Operations

These include converting some part of a problem into a visual query, mentally adding both imagery and attributes to a perceived symbol or other feature. Cognitive operations include decisions such as terminating a visual search when an item is found. A key bottleneck in cognitive processing is the capacity of visual working memory, so situations where visual working memory limits can lead to cognitive inefficiently are elucidated, and solutions described.

B. Visual Queries

A visual query is cognitive transformation of some aspect of a problem so that progress can be made using a visual pattern search. For example, finding a relationship in a social network diagram will involve visually searching for a line linking two nodes. Visual queries provide a method for reasoning about key problems of data representation..

C. Visual Pattern Processing

A visual pattern is the target of a visual query. The goal of efficient graphic design is to ensure that data is mapped into graphical form in such a way that all probable visual queries can be efficiently executed.

D. Working memory load

Working memory capacity is assumed to be at most 4 simple patterns or shapes if the patterns are previously unknown. If patterns are well known, a skilled analyst can hold more complex patterns in working memory.

E. Epistemic Actions

These are any actions designed to seek information. They include eye movements to focus on different parts of a display, and mouse movements to select data objects or navigate through a data space.

F. Externalizing

These are instances where someone saves some knowledge gained by putting it out into the world, for example, by adding annotations to a visualization, or checking boxes to indicate that certain information is deemed important or irrelevant.

G. Interaction Computation

This includes all parts of a visual thinking algorithm that are executed in a computer. Of particular relevance to VTDPs are computations involved in rapidly changing how information is displayed. This can be as simple as zooming or changing the range of the data that is displayed because of user interaction via a time-slider.

H. Display budget

The display budget is the amount of visual information that can be usefully displayed at a given time. For example, there is no point in displaying a graph with 5000 nodes if very few of them can be visually resolved.

I. The VTDP virtual machine

When thinking about interactive systems, it is useful to consider the simple virtual machine, illustrated in Figure 1. This contains a simple visualization pipeline as well as the key components of perceptual and cognitive processing. Very often, a gain in cognitive efficiency can be achieved by shifting part of the workload from visual working memory by means of simple computations and interactive methods.

III. A SUMMARY LIST OF VTDPS

We have identified a set of 20 VTDPs to date and the following section gives a brief summary of 11 of them, listing the remainder. In two cases, Drill Down with Aggregation, and Lateral Exploration, we provide somewhat more detail in order to show how they combine perceptual operations with machine operations. The first two design patterns, Visual Query and Reasoning with a Hybrid of a Visual Display and Mental Imagery are components of almost all visualizations. These are followed by two high level patterns, Analytic Framework and Visual Monitoring. The remaining patterns are mid-level. VTDPs that have already appeared as visual thinking algorithms in [1] are indicated.

A. VTDP: Visual Query

Most visualization involves a cognitive operation where some aspect of the problem is turned into a visual search for a pattern that can provide part of the solution. For example, someone may wish to know how a piece of information got from Jack to Jane and they have a social network diagram showing communications as lines connecting nodes representing individuals. The visual query is to find a path of lines between the node representing Jack and the node representing Jane.

B. VTDP: Reasoning with a Hybrid of a Visual Display and Mental Imagery (IV3 CH11:A3)

Almost all visual thinking is a process or reasoning where external imagery is combined with mental imagery. Mental imagery is used to represent alternative interpretations or possible additions to an external visualization. Visual queries are executed on the combined external/internal image.

C. VTDP: Analytic Framework

This provides a broad framework for analytic tasks (as opposed to monitoring tasks or routine processing tasks)[6,12]. In some cases a framework can be configured and the user can set up a system to suit a particular set of tasks and a particular data set. The analytic framework has the overall goal of ‘sense-making’ and as such includes a set of activities such gathering and organizing information. Most of the VTDPs in this list can be part of the analytic framework.

D. VTDP: Drill Down-Close Out with Hierarchical Aggregation

In many big data applications it is not possible to show all entities at once. A common approach for dealing with this is to aggregate items hierarchically to provide a visual overview [7]. The basic interaction is clicking on an object of interest whereupon it opens revealing the visual representations of the objects that it stands for. A reverse operation closes the object. For aggregated objects to be meaningfully used they must visually portray sufficient information scent for decision making. This is done by causing key information to propagate up the hierarchy using task relevant rules, such as a statistical summary (sum, median average), exceptions (anomalies, faults), or temporal changes A screen budget combined with the number of data objects can be used to calculate the degree of aggregation necessary. Drill down-close out works best with balanced hierarchies. For example, a ten way tree requires eight clicks to get to leaf nodes with 100 million items in the data base, but only if the tree is balanced. A 40 way tree only requires 5 clicks and a 450-way tree only requires 3 clicks to access the same amount of data. If the visual cue being searched for is pre-attentive then the time to click may be close to constant, but if it is not then a slow serial search will occur. There can be a large burden placed on working memory if multiple drill down operations are to be compared.

Process: Drill Down, Close Out with Aggregation

Display Environment: Symbols representing aggregations of data.

1. Based on a screen budget and the quantity and architecture of the data, the computer creates a hierarchy of entities, each having a visual representation that reveals some aspect of its constituent parts.

2. The visualization begins with a display of the top level object.

3. The analyst conducts a visual search for informative objects based on some aspect of their visual appearance (visual scent).

4. The analyst clicks on the object judged to have the highest probability of yielding useful information.

5. If useful information is acquired, the analyst saves the object in human or machine memory.

6. The analyst closes the object.

7. Repeat from 3.

E. VTDP: Visual Monitoring (IV3 CH11:A10)

This VTDP is applied in situations where analysts must monitor a set of measured values or instruments. Usually monitoring is interspersed with other work. The basic visual thinking algorithm involves setting up a schedule of interrupts, according to which users will periodically stop what they are doing and conduct a visual scan of a set of displays looking for anomalous patterns that may require action. Key patterns will depend on the monitoring goals. A key decision is whether interrupts should be cognitively or system generated.

F. VTDP: Cognitive Reconstruction

Most analysts will have their work frequently interrupted. They also often need to extend or redo prior analytic work as new data becomes available. Successfully resuming a prior analysis requires that the entire cognitive system be reconstructed and this involves rebuilding both the machine state and the operator state. All long term memory is more a process of reconstruction than a process of retrieval. Any full featured analytic support system will contain features to support cognitive reconstruction such as the ability to add annotations and many packages support the resurrection of a prior analytic system state through the use of scripts.

G. VTDP: Lateral Exploration

Often analysts start with a piece of information and follow links outward [8]. For example, in social networks we may follow a chain of social relationships. Or we may trace linkages between people through an organization. Lateral exploration is an alternative to the 'overview first/details on demand' approach advocated by Shneiderman. Perhaps in most cases, analysts begin by finding a lead, in the form of a piece of information that seems relevant, then follows links laterally to find related information, assembling what is relevant and discarding what is not. One of the simplest instances of this is a search for related information within a large network.

H. VTDP: Find Local Patterns using Degree-of-Relevance Highlighting (IV3 CH11:A7)

Sometimes information objects in a display are interrelated in ways that are highly task relevant. Degree-of-relevance highlighting can be useful when it is possible to display a substantial amount information on the screen at once but because of its density it cannot all be made legible. A simple interaction solves the problem; touching an object causes both it and other task relevant data objects to be highlighted. The highlighted objects may also reveal addition detail. As with the hierarchical aggregation, graphical information scent is needed to provide a starting point for visual search. As a first order approximation, degree-of-relevance highlighting is useful for between 30 and 500 graphical symbols representing data.

I. VTDP: Pattern integration across views using brushing (IV3 CH11:A5)

Brushing [9] is often useful to represent data in several different views. In brushing, selecting a data object in any one of the views causes those same data objects to be highlighted wherever they appear in all of the other views, thereby visually linking them. For brushing to be effective is it important that the highlighting technique used takes advantage of pre-attentive visual cues.

J. VTDP: Pattern Comparisons in a Large Information Space (IV3 CH11:A6)

A need to represent detailed information in a larger context is a common problem for data visualization. The most common example is a map display, where we want to compare small scale features on the map. The same problem occurs with more abstract data, for example, in large network diagrams. Any pattern comparison involves loading some aspect of one pattern into visual working memory, to be later compared to some other pattern. Pattern comparisons are far more efficient if the transfer of attention between one pattern and another can be made using eye movements, because in this case the information only needs to be held for a fraction of a second. The working memory burden is far greater when other techniques, such as zooming, are required for the comparison. Requiring more than a few simple shapes to be held in working memory will result in high error rates. Adding extra magnifying windows is a common way of ensuring that the task can be carried out using eye movements.)

One way of dealing with the pattern comparison problem is to use a method called the Generalized Fisheye View [10]. This method relies on a degree of interest function whereby the computer attempts to show only task relevant information and hides or shrinks other information. The success of this method depends entirely on the predictability of related information.

K. VTDP: Multidimensional Dynamic Queries (IV3 III CH11:A9)

With multidimensional discrete data all entities have the same set of attributes. The attributes define the data dimensions and each entity can be thought of as a point in a multidimensional space. A set of sliders is provided that can narrow the range on each of many attributes. Each slider adjustment is an epistemic action, narrowing the range of what is displayed. Ideally, feedback is very rapid (<100 msec) [11] If we assume that each dynamic query slider can be used to reduce the range selected to 10% of the original, then the number of objects that be interactively queried is ~ 10^d where d is the number of dimensions. The method has most often been used with scatter plots and to a lesser extent time series plots, but it can work with ranges displayed on maps and with node link diagrams.

L. VTDP: Table Data - Sort and Compare

Tables of data can contain glyphs instead of just numbers, in which case they become interactive visualizations. A common analytic strategy is to use sorting to bring out certain kinds of relationships visually. Sorting on one value can be useful in revealing correlated anomalies. For example, suppose we have a data table with system failures represented by a glyph in one of the columns. If failures tend to occur when a particular variable is low, sorting from low to high on that variable will bring marks representing faults to the top of the table where they can be visually compared to values in other columns of the table.

Other VTDPs are the following.

- VTDP: Pathfinding on a Map or Diagram (IV3 CH11:A2)
- VTDP: Discovering Novel Temporal Patterns
- VTDP: Table Data: Compute, Chart and Find Patterns
- VTDP: Model-Based Interactive Planning
- VTDP: Design Sketching (IV3 III CH11:A4)
- VTDP: Task List
- VTDP: Query by Example
- VTDP: Presentation linking images and words.

IV. VTDPs in the Design Process

Visual thinking design patterns are intended as tools for the designer to use in reasoning about visualization design. In the following paragraphs we sketch out the design process as a series of steps. In most cases these steps should be part of a spiral design methodology, with multiple iterations.

A. *Step 1: High level cognitive task analysis*

At the start of the design process the team members must establish, in a general way, the problem to be solved by the final product. Initially the description should not specify the implementation method so as not to prejudge the solution. Refinement will come later.

B. *Step 2: Data inventory*

It is important to know early on in the process what data is available that bears on the cognitive goals, what is readily accessible, and what is likely to be accessible only with time delays. Sometimes confidentiality can be a stumbling block. The structure of the data will be important in defining the visualization and so will the semantics. Data can be enormously varied in its properties, but the following list contains some of the attributes that are likely to appear in any inventory.

- Quantity: Very big data requires different approaches beyond what applies to medium or small data.
- Structure: E.g. hierarchical, map layers, network, multi-dimensional discreet.
- Time to access: Sometimes real-time or near real-time is critical.
- Ease of access: Some data is readily accessible; some data may require new infrastructure; some data has explicit costs; some data has security issues.
- Interrelationships and interdependencies.
- Quality and reliability.
- Potential for derived data products – processing infrastructure.

The data inventory process will almost certainly start to blend into step 3, task refinement, because data is only important as it relates to task requirements.

C. *Step 3: Refinement of cognitive task requirements*

As the data is understood, the set of tasks can be refined. Knowing the specifics of particular data objects will suggest additional questions that may be addressed. As a generally strategy it is useful to work top down, breaking down the overarching goal into subgoals that relate to specific data sets. It is also important to establish interdependencies between tasks, such as the order in which they must be performed. At some point specific computer based tools may be invoked, but this should be held off as long as possible.

Only part of an analytic problem is likely to be amenable to solutions that use visualization and it is important to divide the problem into aspects that can use visual thinking and aspects that use other forms of thinking or computation. The following of aspects of a problem suggest a visualization solution.

- There must be a way of transforming the problem so that solutions are discoverable through a visual pattern search.
- The task should not be so repetitive and standardized that an automatic computer pattern search is more appropriate.

Once a set of subtasks that is amenable to visual thinking solutions has been identified, they can be matched to the VTDPs and their associated visualizations.

D. *Step 4: Identification of VDTPs and visualizations that can bear on the task*

This step is the key creative stage in the design process. It is best led by an experienced designer who knows both the types and nature of visualization and understands the VTDPs that can be used to make them cognitively efficient.

The rapid design and development of visualizations would be almost impossible if it were necessary to come up with a radical new design and a radically new interaction method for every problem. Fortunately, there are only a small number of basic types of visualizations that have widespread use in practice, namely charts, maps, node link diagrams and tables. These are can be provided in component libraries. The key development issue is that the components should support the relevant set of VTDPs.

E. *Step 5: Design decision rules*

Since VTDPs incorporate rules regarding when they are most effective, it is possible to construct a set of rules for visualization problems that provide guidelines for when the different methods can be applied. For example, we have constructed the system of rules below regarding techniques that can be applied to the visualization of graphs as node link diagrams.

- Small graph (<30 nodes) static representation
- Medium graph (>30,<600) degree of relevance VTDP

- Large Graph (>600 < 5m) use dynamic queries VTDP attributes to reduce graph size to get to degree of relevance VTDP (<600).
- Very large graph (>5m) Query by example VTDP may be needed. Start with example, computer finds similar – user selects to refine query.

F. Step 6: Prototype development.

In order to test the cognitive affordances of design alternatives, it is necessary to have some form of prototype. This can be design sketches of key screens, or a rapidly developed prototype. The purpose is to provide a basis for reasoning about cognitive efficiency when the visualization is applied to the cognitive tasks identified in step 3.

G. Step 7: Evaluation through cognitive walkthrough

The cognitive walkthrough should focus on critical or frequent tasks to ensure that the cognitive execution will be efficient using the VTDPs. Small groups should talk through the steps, attempting to identify bottlenecks, such as unreasonable memory load, or instances where repetitive work can be offloaded to the computer.

It is likely that new cognitive tasks will emerge during the process and the list of cognitive tasks can be refined; links between cognitive tasks may be revealed.

H. Step 8: Alpha and beta products.

The final stages of design should apply the common techniques of developing alpha and beta product versions that can be tested and refined through use in problem solving.

V. CONCLUSION

It is often said that designers work intuitively. But they still bring all the knowledge and skills they have acquired to a design problem. Using intuition simply means that the creative reasoning process is largely unconscious and not explicitly analytical. Designers are explicitly analytical at certain stages in the design process, such as when they critique design solutions. VTDPs provide a way of training designers in the way interactive visualization techniques work in the visual thinking process, together with the limitations and strengths of each method. If the designer is familiar with them, VTDPs can become part of the mental landscape of the designer, providing a starting point for intuitive design solutions.

VTDPs are not intended to be rigidly prescriptive; instead they provide a framework for thinking about key parts of the interaction design in terms of their perceptual and cognitive efficiency. Design is an optimization process. The goal is to maximize the value of whatever is produced and to minimize the effort required to produce it. Complete information about the set of tasks is rarely available, and software tools are rarely used in isolation.

We have introduced the concept of VTDPs in this paper and described how they can be applied in a design context. The set described in the present paper is not complete or definitive and will be extended as our work progresses.

The ideal context for the VTDPs to be applied is in an agile design environment such as the one being developed around Oculus Aperture studio [13]. Aperture provides an open and extensible Web 2.0 visualization framework; it has an extensive library of customizable components, including maps, node-link diagrams and charts of various kinds such as scatter plots, bar charts and time series plots. Aperture already provides support for many of the VTDPs listed in this paper and as it develops more will be added.

ACKNOWLEDGMENTS

This study was supported by the DARPA XDATA Program under Contract No. FA8750-12-C-0317 to Oculus Info Inc.; also under STTR Contract No. W31P4Q-12-C-0207 to Systems Technology Research. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA. The ideas presented here have benefitted from discussions with many people, especially the Oculus design team, as well as Tamara Munzner, and Sheelagh Carpendale.

REFERENCES

[1] C. Ware, Information Visualization, Perception for Design, Third Edition. Morgan Kaufman. 2012.

[2] C. Alexander, S. Ishikawa, and M. Silverstein, A pattern language. Oxford University Press. 1977.

[3] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley. 1995.

[4] S.K.. Card, T.P Moran, and A. Newell, The Psychology of Human-Computer Interaction, London: Lawrence Erlbaum Associates. 1983.

[5] J. R Anderson, M. Matessa, and C. Lebiere,. ACT-R: A theory of higher level cognition and its relation to visual attention. Human-Computer Interaction, 12(4), 439-462. 1997.

[6] P. Pirolli, and S. Card, The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In Proceedings of International Conference on Intelligence Analysis (Vol. 5). 2005.

[7] N. Elmqvist, and J.D. Fekete, Hierarchical Aggregation for Information Visualization: Overview, Techniques,and Design Guidelines. IEEE Transactions on Visualization and Computer Graphics. 16(3) 439-454. 2010.

[8] F. van Ham, , and A. Perer. Search, Show Context, Expand on Demand": Supporting Large Graph Exploration with Degree-of-Interest. Visualization and Computer Graphics, IEEE Transactions on 15, no. 6 953-960. 2009.

[9] R. Becker, and W.S. Cleveland, Brushing scatterplots. Technometrics, 29, 127-42.

[10] G.W. Furnas,. Generalized fisheye views. Vol. 17, no. 4. ACM, 1986.

[11] C. Ahlberg, and B. Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays." In Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, pp. 313-317. ACM, 1994.

[12] R. Amar and J. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on, pp. 143-150. IEEE, 2004.

[13] D. Jonker, S. Langevin, N. Bozowsky and W. Wright, Aperture: An Open Web 2.0 Visualization Framework, 46th Hawaii International Conference on System Sciences, 201.

Multi-level Visualization of Concurrent and Distributed Computation in Erlang

Robert Baker
School of Computing
University of Kent, UK
rb440@kent.ac.uk

Peter Rodgers
School of Computing
University of Kent, UK
P.J.Rodgers@kent.ac.uk

Simon Thompson
School of Computing
University of Kent, UK
S.J.Thompson@kent.ac.uk

Huiqing Li
School of Computing
University of Kent, UK
H.Li@kent.ac.uk

***Abstract*—This paper describes a prototype visualization system for concurrent and distributed applications programmed using Erlang, providing two levels of granularity of view. Both visualizations are animated to show the dynamics of aspects of the computation.**

At the low level, we show the concurrent behaviour of the Erlang schedulers on a single instance of the Erlang virtual machine, which we call an Erlang node. Typically there will be one scheduler per core on a multicore system. Each scheduler maintains a run queue of processes to execute, and we visualize the migration of Erlang concurrent processes from one run queue to another as work is redistributed to fully exploit the hardware. The schedulers are shown as a graph with a circular layout. Next to each scheduler we draw a variable length bar indicating the current size of the run queue for the scheduler.

At the high level, we visualize the distributed aspects of the system, showing interactions between Erlang nodes as a dynamic graph drawn with a force model. Specifically we show message passing between nodes as edges and lay out nodes according to their current connections. In addition, we also show the grouping of nodes into "s_groups" using an Euler diagram drawn with circles.

I. Introduction

Erlang applications are typically concurrent and distributed, and run on multicore platforms. By visualizing the concurrent processes involved we can convey how individual processes are performing and communicating. This means the programmer can evaluate performance, verify program correctness and gain insight into program behaviour [10]. This paper describes a prototype visualization system for Erlang applications, giving two (low and high level) views of the system.

Erlang [2], [6] is a programming language with built-in support for concurrency and distribution. Erlang models the world as sets of parallel processes that can interact only by message passing, that is without any shared memory. Erlang concurrency is directly supported in the implementation of the Erlang Virtual Machine (VM), rather than indirectly by operating system threads.

Messages between processes can consist of data of any Erlang type: as well as the basic types (integers, booleans, strings and bit strings, or binaries) these can be tuples (consisting of a number of values of different types), lists or functions. Process identifiers and symbolic names can also be communicated as (parts of) messages.

Erlang supports multi-core programming "out of the box". At startup, the Erlang VM automatically detects the CPU topology on which it is running and creates a process scheduler for each available CPU core. Each process scheduler maintains its own run queue of processes to execute. In order to balance the workload between schedulers, a process migration (work stealing) mechanism is applied periodically, so that schedulers with less work pull processes from schedulers with more work, while schedulers with more work push processes to others.

A distributed Erlang system consists of a number of Erlang VMs communicating with each other. Each VM is called an Erlang *node*. Message passing between processes on different nodes is indistinguishable from communication on the same node (at least when process identifiers (*pids*) are used for message addressing). Whilst different Erlang nodes in a distributed Erlang system can run on physically separated computers, it is also possible for multiple nodes to run on the same computer.

The default model for an Erlang distributed system is for all nodes to be connected to all other nodes, but this has negative consequences for the scalability of systems. Erlang nodes in a distributed Erlang system can be partitioned into groups, where every element in a group is connected to all others, but the fact that the groups are disjoint means that certain kinds of architectures are not possible.

One of the outcomes of the ongoing FP7 RELEASE project [3] is the notion of *s_groups*. These extend Erlang's original partition-based grouping in a number of ways. As far as the work described in this paper is concerned, the major difference between s_groups and groups is that s_groups are not necessarily disjoint, and so one Erlang node may be a member of multiple s_groups, or indeed a member of none.

This paper describes a prototype visualization system for concurrent and distributed Erlang applications. In the system we provide two granularities of view: low level and high level.

The low level view shows the concurrent behaviour of Erlang processes on a single Erlang node, focussing in particular on scheduling aspects. The schedulers are laid out in a circle: we indicate the size of their run queues by the length of bars which are drawn next to the schedulers. We also visualize the migration of Erlang concurrent processes from one scheduler to another, as work is redistributed to fully exploit the hardware: this migration is represented by animated edges between schedulers.

The high level view visualizes the distributed aspects of the system, with each graph node representing an Erlang node. We show message passing between nodes as edges, laying nodes out according to their current connections. The grouping of nodes into s_groups is shown by surrounding each s_group by an interlinking circle. Dynamic communication behaviour is animated using a force model [7].

The high level layout is produced by first drawing an Euler diagram for the intersecting s_groups of Erlang nodes [19]. The nodes are then placed in the appropriate zone of the diagram, and a force-directed layout is applied to the communication graph. The forces in the layout include the standard spring embedder forces [7], with an additional repulsion from the edges of the circles to keep the nodes within the correct regions [16]. Whilst the number of nodes and their grouping into s_groups is fixed and does not change during execution (at least in this prototype), the communication between nodes does change over time, and the use of a force based approach allows the layout of the graph to alter dynamically in an animated way.

The research contribution of this paper is to describe the first visualization for concurrent and distributed Erlang execution. The visualization has novel features:

- The use of Euler diagrams to show groups of virtual machine nodes in concurrent programs is new.
- Whilst inspired by previous approaches [16], [7], our method of integrating Euler diagrams with graphs represents a new combination of forces.
- The two different levels of view with discrete visualizations has not been seen, to our knowledge, in previous systems for concurrent and distributed visualization.
- The combination of bar charts and circular graph layout appears to be new in the context of concurrent process visualization.

In the remainder of this paper, we first discuss related work in Section II. Section III then provides an overview of the low level visualization and animation method, whilst Section IV discusses the high level method. Finally, Section V presents our conclusions and outlines possible future work.

II. Related Work

Circular layout has been used previously for processes on a core [10]. However, in this earlier work, communication was shown by dots moving from one process to another. This requires the users attention during the movement of the dot to identify the start and ending processes, which is particularly problematic when several processes are communicating at the same time. We improve on this by using curved edges to show the communication between two processes. These edges fade over time, to prevent the display from becoming overly cluttered. We note that the use of curved edges follow the Gestalt principle of good continuation [17], and so can be justified by perceptual theories. Justification for grouping processes by surrounding groups with contours – which overlap in the high level view, and do not overlap in the low level view – is provided by the preattentive principle of closure [20].

Visualization of communication graphs in concurrent processes has been prevalent for some time. Visputer [22] has a multi-level approach to visualizing such graphs, where a single node can contain either a process or a communication subgraph. This consistent approach to the different levels, where lower levels have the same syntax as higher levels, contrasts with our approach where the visual display of the two levels is tuned to the different diagram semantics. The Meander project [21] not only visualized communication graphs, it used them as the basis for a visual concurrent programming language. The use of graph visualization for the presentation of communication graphs is widespread, see for instance a 1992 survey [11]. More recent work has applied a variety of graph drawing methods to automatically generate layouts for various graphs associated with concurrent and distributed programs [4], [14], [18].

Another approach to visualising Erlang programs is provided by Percept2 [12]. Percept2 is a tool for offline visualisation of Erlang application level concurrency and identifying concurrency bottlenecks. It utilises Erlang's built-in support for tracing to monitor events from process states. Trace events are collected and stored in a file, which can then be analysed offline. Once the analysis is done, the data can be viewed in chart form through a web-based interface.

Percept2 gives a picture of application-level parallelism, as well as how much time processes spend waiting for messages. It also exposes scheduler-related activities to end-users, and provides finer-grained profiling information about the existing parallelism of an application. It is also designed to scale well to parallel applications running on multiple cores.

VampirTrace/Vampir [15] is a tool set and a runtime library for instrumentation and tracing of software applications. It allows a variety of detailed performance properties to be collected and recorded during runtime, including function enter and leave events, MPI communication, OpenMP [1] events, and performance counters. VampirTrace provides the input data for the Vampir analysis and visualization tool. Scalasca [8] is another representative trace-based tool that supports performance optimization of parallel programs by measuring and analysing their runtime behaviours. The analysis identifies potential performance bottlenecks, in particular those concerning communication and synchronization, and offers guidance in exploring their causes. Finally, ThreadScope [9] is used for performance profiling of parallel Haskell programs. ThreadScope reads GHC-generated tracing events from a log file, and displays the thread profile information in a web-based graphical viewer.

III. Low Level Visualization

The low level visualization shows the concurrent behaviour of Erlang processes on a single Erlang node. We visualize the migration of Erlang concurrent processes from one scheduler to another and the size of the run queue on each scheduler.

The visualization system (both low level and high level components) is developed in JavaScript and rendered in SVG using the Data-Driven Documents library (D3) [5]. The visualization is re-drawn on every iteration and this gives the effect of animating the state of the schedulers to reflect the current

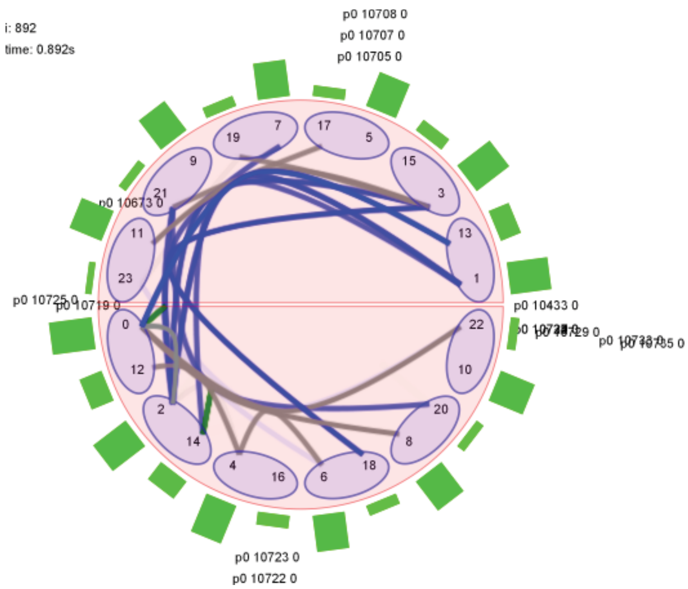


Fig. 1. Low Level Visualization

situation. This constant re-drawing is required as the system is dynamic: processes can be created, terminate or migrate between schedulers at any time.

The data used in this paper was generated from running Dialyzer [13], the Erlang static type analysis tool, on a selection of Erlang applications and observing the computation. During the running of the application, every process migration between schedulers is recorded using Erlang's built-in support for tracing, and the size of run queues is sampled every millisecond. The trace data are recorded in an Erlang file, with each trace datum represented by an Erlang tuple (that is collection of Erlang values), and data analysis is performed by an Erlang program.

The application was run on an Intel machine, running CentOS 6, with two processors, each having six physical CPU cores. Through hyper-threading technology, the OS is able to address two logical cores for each physical CPU core, therefore there are 24 logical cores in total for an Erlang node to exploit.

An example frame in the low level visualization can be seen in Figure 1. The parallel machine is split into two six-core processors, represented by the two pink semicircles. Erlang schedulers are shown as a number that represents the unique ID of the related threads. Pairs of Erlang schedulers that run on a single real core are shown grouped by the blue ellipses.

Each Erlang scheduler has an individual run queue, and this is represented by a bar radiating from the visualization. The height and colour are used to indicate the size of the run queue: larger queues (i.e. busier Erlang schedulers), are shown with longer and redder bars, whereas smaller queues are represented by shorter and greener bars. The run queue size can be displayed on screen as a value by selecting the relevant option in the interface, or by hovering a cursor over the bar or Erlang scheduler in question. It is possible that the queues can be so busy that they are not fully displayed on screen, if this is the case, the user is able to take advantage of the scalability of SVG graphics and alter the zoom level on their browser.

Processes get spawned on a particular scheduler. The list of numbers radiating from a Erlang scheduler shows the unique ID for each process spawned on that scheduler. As can be seen in Figure 1, Erlang scheduler 5 has just spawned a number of new processes.

Processes may migrate from one Erlang scheduler to another and this is shown by edges between various schedulers. These are curved to avoid occlusion with other schedulers (e.g. a migration from 10 to 18 would cross schedulers 20 and 8). The edges are coloured to indicate the three types of migration: within a single core (green), between different cores on the same processor (grey), and from a core on one processor to a core on the other (blue), although these migrations can also be deduced from the start and end position of the edge. These edges fade away after a few seconds to prevent the visualization from becoming occluded.

The entire visualization displays the current state of the trace and is animated such that a new time frame (iteration) is shown every 200 milliseconds and each time frame is 1 millisecond sample of the original trace, that is, the animation runs 200 times more slowly than the original application. This allows the user to focus on the changing state without being overwhelmed with the amount of animation. In addition, it allows enough processing time for the animations to display correctly on slower machines. The animation can be paused and resumed using the interface. A slider allows the selection of a specific time frame and permits the animation of a sequence of time frames in either direction.

Figure 2 shows three snapshots of the low level view at various stages of execution of an Erlang program. The top diagram shows the state at iteration 890, where the none of the schedulers are particularly busy, and the amount of communication between different schedulers is correspondingly low. The middle diagram, at iteration 967, shows the Erlang node in a much more active state , here the run queues are extremely uneven. The bottom diagram shows the state at iteration 3804, where the disparity in run queue size is now extreme, with three outliers that have full run queues, and others with relatively short queues.

A number of interesting phenomena are indicated by the visualizations.

- A clear consequence of hyper-threading is that there is an asymmetry between the two schedulers on a single physical core. This is evident in the visualisation in the different lengths of the two run queues on a single core (e.g. cores 0 and 12), resulting in the alternating pattern seen in the figure.
- Setting aside this intra-core issue, migration does not ensure that run queues have the same size: however this is not necessary, since all that is needed is that each core has enough work to ensure that it is fully occupied.

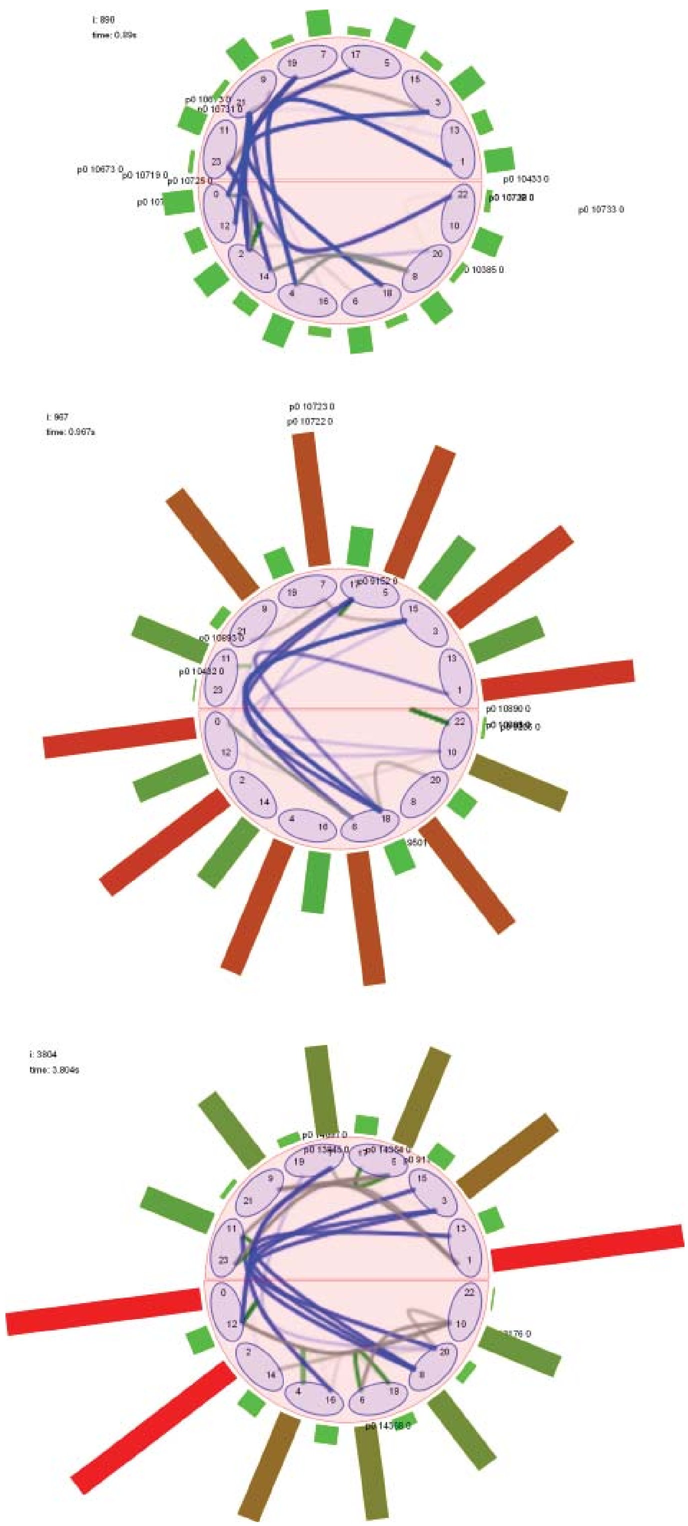


Fig. 2. Three steps in the low level view

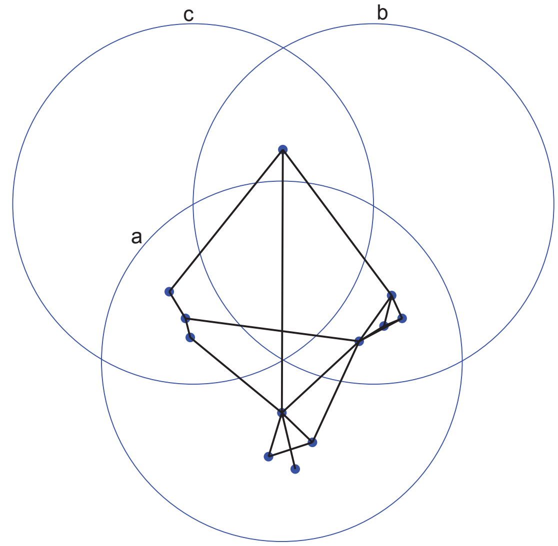


Fig. 3. High level view.

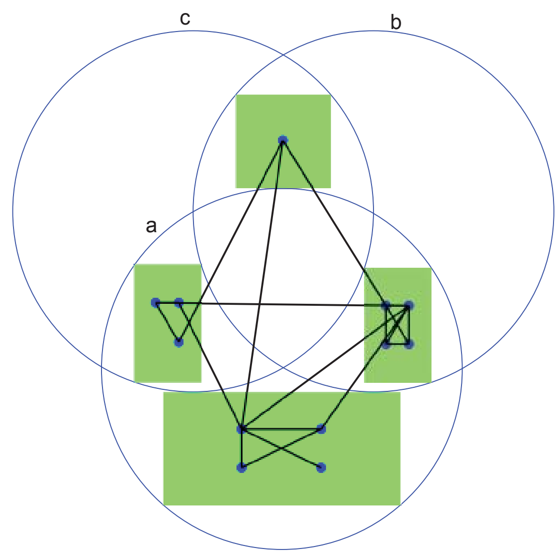


Fig. 4. Initial placement of Erlang nodes before force application.

IV. HIGH LEVEL VISUALIZATION

The high level visualization shows the grouping of Erlang nodes within particular s_groups and edges visualizing the communication between the nodes. Each node on this high level visualization corresponds to an instance of the low level visualization.

An example of the high level visualization is shown in Figure 3. The s_groups are drawn first, with each s_group being represented by a circle. As a node may be the member of more than one s_group, the circles typically intersect. The s_group membership is derived and is passed to code described by [19] which returns the circle centres and radii that will then be drawn in SVG. The intersecting circles form connected regions and the largest rectangle within each region is calculated (shown in green in Figure 4). Nodes that are contained in the relevant s_groups are drawn in a grid pattern within the rectangle. The method for drawing Euler diagrams often creates extra regions, in which no node is placed. In some visualization systems (particularly those using Venn diagrams) these empty regions can be shaded to indicate that they have no elements inside. However, we regard it as self-evident that there are no elements in empty regions and so we do not add the extra syntactic complexity of shading.

A force-directed layout is then applied to the nodes to improve the layout. The forces are the standard spring embedder attraction and repulsion forces [7], with an additional force. This third force is applied to a node to keep it away from the borders of the region, and is a simplified version of the force introduced in [16]. It is proportional to the node's distance from the circumference every circle, pushing the node toward the centre of circles in which it is inside, whilst pushing the node away from the centre of circles which it is outside of. Additionally, a test is made to ensure that nodes do not leave the region to which they belong by disregarding any node movement that moves it outside the correct region (although nodes are allowed to move outside their starting rectangle).

Communication between nodes is represented by edges and these may change during the program's execution. The shading of the edges is related to the frequency of messages within the timeframe. The darker the edge, the more frequently messages are sent. Edges are removed if no communication exists between them and new edges are introduced when communication is initiated between nodes. The use of a force-directed layout method allows the visualization to be dynamic, so that the repositioning of the nodes can be shown in an animated manner as the diagram reaches a new equilibrium.

The data used in this paper was generated by running a distributed Erlang benchmark application on a collection of Erlang nodes, which were configured to form the s_groups (these s_groups were manually defined, as no s_group data is currently present in the trace data). The total wallclock runtime of the application was 222 seconds. On each participating Erlang node, Erlang's built-in support for tracing was used to record message sending events. Each such event recorded contains the name of the sending node, the name of the receiving node as well as the actual message sent. Trace data from each node were written to a file, and the data files were then collected, analyzed and synthesized to provide the visualization tool with the accumulated communication data on each iteration, which occurs every 200 milliseconds.

Figure 5 shows some steps in the animation process. The top diagram shows the visualization after 100 iterations since the start of the trace. At this point there are relatively few communication edges between Erlang nodes. In the middle diagram, an iteration later, the number of communication edges suddenly increases dramatically, indicating that the scheduling has initiated a number of communications between nodes. In terms of the visualization, the middle diagram has not had time for the forces to be applied to the nodes, and so the nodes are shown in the same position as in the top diagram. After another 99 iterations, the forces find an new equilibrium,

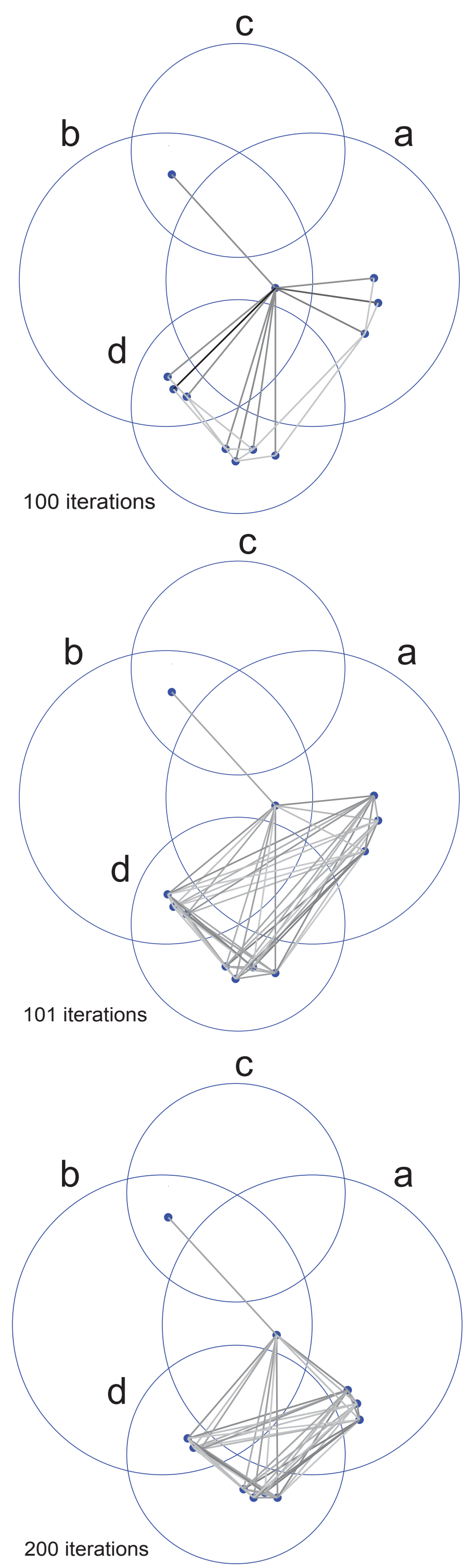


Fig. 5. Three steps in the animation of the high level view.

with the nodes in the positions as shown in the lower diagram. Here, the nodes have moved further away from the region centres, so indicating a stronger affinity to nodes outside their regions. However, the only node in s_group "c" (top circle) has been relatively unaffected by this, as the number of nodes it is communicating with has not changed.

V. Conclusion

This paper has described a two level visualization for Erlang distributed and concurrent processes. The complexity of such systems requires a mixed approach where concurrency and distribution are visualized by distinct methods.

Concurrency in Erlang nodes has been represented in a low level visualization by animating communication edges between Erlang schedulers which are shown in a circular layout. How busy each Erlang scheduler is can be seen by the length of the bar associated with that scheduler.

The distributed nature of Erlang computation is shown in a high level visualization using a dynamic graph of communication between Erlang nodes. Nodes are placed in s_groups which is shown by drawing the nodes in regions of a Euler diagram drawn with circles. The changes in communication between nodes is animated by a force-directed layout, which also maintains the nodes in their s_groups.

Informal feedback from RELEASE project members has been solicited. This has resulted in broadly positive reactions, with suggestions for alternative data that might be represented by the edges between processes or nodes. This leads to the first aspect of further work: once the system is more stable, there is potential for formal evaluation of the system to see if it is successful in its aims of:

- assisting system performance evaluation;
- aiding the verification of program correctness; and
- improving insight into program behaviour.

Given the limited number of distributed Erlang experts, and the considerable amount of time that would need to be devoted to using the visualizations to get useful information, a longitudinal study with a handful of participants might be the best mechanism for this. Before such an evaluation, the user interface will have to be improved, and the connection between the two levels of visualization needs to integrated more closely.

Other further work relates to modifying the system to deal with an increasing amount of distribution. The current high level visualization shows each Erlang node as a graphical node. Whilst this has the advantage of showing all Erlang nodes, it can become unwieldy when their number grows. To allow the system to scale we expect to cluster Erlang nodes, using mechanisms similar to those developed in [22]. When applied to distributed Erlang, a graphical node might represent an s_group, and edge affinity could be a composite measure of overlap and communication bandwidth between groups.

The system so far supports *offline* visualisation, but the same graphical approach can be used to provide an *online* – that is, real time – view. The challenge of online visualisation in this case is to provide streams of suitable aggregate data without adversely affecting the performance of the system under observation. These streams of data could be provided by the built-in Erlang tracing mechanism, or using a Unix teaching framework such as DTrace / SystemTap.

Acknowledgements This research is supported by EU FP7 collaborative project RELEASE (http://www.release-project.eu/), grant number 287510.

References

[1] The OpenMP API specification for parallel programming. http://openmp.org/wp/.

[2] Erlang/OTP. http://www.erlang.org, May 2013.

[3] RELEASE. http://www.release-project.eu/, May 2013.

[4] Benjamin Albrecht, Philip Effinger, Markus Held, Michael Kaufmann, and Stephan Kottler. Visualization of complex bpel models. In *Graph Drawing*, pages 421–423. Springer, 2010.

[5] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011.

[6] Francesco Cesarini and Simon Thompson. *Erlang Programming*. O'Reilly Media, Inc., 2009.

[7] Peter Eades. A heuristic for graph drawing. *Congressus Numerantium*, 42:149–160, 1984.

[8] Markus Geimer, Felix Wolf, Brian J. N. Wylie, Erika brahm, Daniel Becker, Bernd Mohr, and Forschungszentrum Jlich. The SCALASCA performance toolset architecture. *Concurr. Comput. : Pract. Exper.*, 22(6):702–719, April 2010.

[9] Don Jones Jr., Simon Marlow, and Satnam Singh. Parallel performance tuning for Haskell. In *Haskell Symposium 2009*, Edinburgh, Scotland, September 2009. ACM Press.

[10] Eileen Kraemer and John T Stasko. Creating an accurate portrayal of concurrent executions. *Concurrency, IEEE*, 6(1):36–46, 1998.

[11] Eileen T Kraemer and John T Stasko. The visualization of parallel systems: An overview. 1992.

[12] H. Li and S. Thompson. Multicore Profiling for Erlang Programs Using Percept2. In *Erlang Workshop 2013*, 2013.

[13] Tobias Lindahl and Konstantinos Sagonas. Practical type inference based on success typings. In *Proceedings of the 8th ACM SIGPLAN Symposium on Principles and Practice of Declarative Programming*, pages 167–178, New York, NY, USA, 2006. ACM Press.

[14] J. Lonnberg, M. Ben-Ari, and L. Malmi. Visualising concurrent programs with dynamic dependence graphs. In *Visualizing Software for Understanding and Analysis (VISSOFT), 2011 6th IEEE International Workshop on*, pages 1–4, 2011.

[15] Matthias S. Mller, Andreas Knpfer, Matthias Jurenz, Matthias Lieber, Holger Brunst, Hartmut Mix, and Wolfgang E. Nagel. Developing Scalable Applications with Vampir, VampirServer and VampirTrace. In *PARCO*, volume 15 of *Advances in Parallel Computing*, pages 637–644. IOS Press, 2007.

[16] P. Mutton, P. Rodgers, and J. Flower. Drawing graphs in Euler diagrams. In *Proceedings of 3rd International Conference on the Theory and Application of Diagrams*, volume 2980 of *LNAI*, pages 66–81, Cambridge, UK, March. Springer.

[17] A. Rusu, A.J. Fabian, R. Jianu, and A. Rusu. Using the gestalt principle of closure to alleviate the edge crossing problem in graph drawings. In *Information Visualisation (IV), 2011 15th International Conference on*, pages 488–493, 2011.

[18] Lucas Mello Schnorr, Arnaud Legrand, Jean-Marc Vincent, et al. Interactive analysis of large distributed systems with topology-based visualization. 2012.

[19] Gem Stapleton, Leishi Zhang, John Howse, and Peter Rodgers. Drawing euler diagrams with circles: The theory of piercings. *Visualization and Computer Graphics, IEEE Transactions on*, 17(7):1020–1032, 2011.

[20] Anne Treisman and Janet Souther. Search asymmetry: a diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114(3):285, 1985.

[21] Guido Wirtz. A visual approach for developing, understanding and analyzing parallel programs. In *Visual Languages, 1993., Proceedings 1993 IEEE Symposium on*, pages 261–266. IEEE, 1993.

[22] Kang Zhang and Gaurav Marwaha. Visputera graphical visualization tool for parallel programming. *The Computer Journal*, 38(8):658–669, 1995.

Pencil-Like Sketch Rendering of 3D Scenes Using Trajectory Planning and Dynamic Tracking

Günay Orbay
Carnegie Mellon University

Levent Burak Kara
Carnegie Mellon University

***Abstract*—We present a new non-photorealistic rendering method to render 3D scenes in the form of pencil-like sketches. This work is based on the observation that the dynamic feedback mechanism involving the human visual system and the motor control of the hand collectively generates the visual characteristics unique to hand-drawn sketches. At the heart of our approach is a trajectory planning and tracking algorithm that generates the sketch using a dynamic pen model. A set of target strokes are generated from the silhouette lines, edges, and shaded regions which serve as the target trajectory for a closed-loop dynamic pen model. The pen model then produces the rendered sketch, whose characteristics can be adjusted with a set of trajectory and tracking parameters. We demonstrate our approach with examples.**

I. Introduction

We describe a pencil-like sketch rendering method of 3D scenes using a range of styles spanning novice drawers to trained artists. The proposed method produces the visual features unique to sketching through a dynamic modeling of the drawing process. The perceived style of the resulting sketches can be manipulated using the parameters of the system.

As suggested in [1], hand-drawn sketches exhibit a variety of other artifacts such as overtracing, hooks at the stroke ends, tonal variation in stroke intensities, lifting versus non-lifting strokes, and layered hatching and cross hatching. These features typically vary based on the skill level of the sketcher, the sketcher's particular precision during sketching, the level of detail to be included in the sketch, and on the dynamics of the pen, hand and the arm. Collectively, these phenomena give rise to a rich set of stylistic variations in sketches, which have been difficult to represent and reproduce algorithmically.

As one step toward addressing this challenge, we describe a method to incorporate and control such effects using a two step process. This involves a stroke field design and dynamic stroke tracking. At the heart of our approach is the design of a 3D pen trajectory from render buffers of a 3D scene through edge detection, region clustering and stroke generation algorithms. When combined with a dynamic pen model, this approach helps produce the unique characteristics observed in pencil sketches. The first step of stroke field design identifies the silhouette, edge and hatching regions, and establishes the stroking behaviors to be applied to them. This results in a 3D target pen trajectory that serves as a reference to be tracked. The second step involves the tracking of this reference by a dynamic pen model. This model enables a rich set of artifacts including a variation in tracking accuracy, stroke skipping, overshoots, pen pressure, non-lifting strokes, and muscle jitter through a set of dynamic parameters.

II. Related Work

A. Line Rendering of 3D Objects

1) Silhouette and Edge Rendering: We discuss the prior work on non-photorealistic rendering of silhouettes and edges in three subgroups. The first group of techniques aim to identify the most representative set of silhouettes and edges that best articulate the shape. DeCarlo *et al.* [2] described suggestive contours, Kalogerakis *et al.* [3] introduced a real-time rendering method that uses learned curvatures variations, Zhang *et al.* [4] utilized the Laplacian of the surface illumination for contour detection. Inspired by these works, we use discrete Laplacian kernels on the depth and normal buffers to identify the silhouettes and sharp edges from input images.

The second group of work focus on rendering identified silhouettes and edges in prescribed styles. Markosian *et al.* [5] described real-time rendering algorithms combined with stylized stroke rendering. Yeh and Ouhyoung [6] developed stroke rendering algorithms that helps mimic Chinese inking styles. Hertzmann *et al.* [7] describe an algorithm to enable curve analogies by learning a statistical model from an input pattern and replicate it on another curve. Barla *et al.* [8] developed statistical models that can learn and produce a wider range of patterns. Brunn *et al.* [9] developed a multi-resolution framework for encoding and generating curve patterns. While these studies cover a large variety of stylistic rendering, little or no emphasis is given to pencil-like renderings. This leaves the simulation of characteristic features contained in pencil sketches an open challenge. This work aims to enable the generation of such features within a wide variety of stylistic effects.

2) Hatching Rendering: Winkenbach and Salesin [10] developed stroke textures to render both textures and tone with line drawings. Salisbury *et al.* propsed a similar approach that produces scale-dependent renders on different scales of sketches. Winkenbach and Salesin [11] later extended their work to rendering parametric surfaces with lines following parametric derivatives on the surface. Salisbury *et al.* [12] proposed an interactive design system that allows the users to quickly design directional fields to support line renderings.

Inspired by the works of Winkenbach *et al.* , a multitude of methods were developed to render lines by tracing the directional fields computed from 3D geometries. Hertzmann and Zorin [13] presented a set of algorithms to calculate hatching lines that follow directional fields defined on surfaces. Praun *et al.* [14] proposed a method for real-time rendering using tonal maps. Palacios and Zhang [15] described an interactive method for field design on surfaces using rotational tensor symmetry. Paiva *et al.* [16] proposed a physically inspired directional field design method that calculates fluid-based hatching strokes. Jodoin *et al.* [17] also used sample drawing methods that enables the reproduction of recorded hatching patterns. Kalnins *et al.* [18] incorporated similar pattern learning and reusing techniques in an interactive non-photorealistic rendering system. Similarly, Kalogerakis*et al.* [19] target learning hatching preferences of an artist from a sample sketch that is drawn over a 3D render.

Most of the prior work in this area is based on the observation that hatching strokes are commonly drawn in directions that closely align with the underlying geometry. Although this observation is valid for in contexts where precise and accurate 3D renderings are desired, it may not apply in other contexts involving casually drawn pencil sketches. In this work, we base our hatching patterns on the shape of the shaded regions.

III. RENDERING APPROACH

Our approach consists of two main steps: (1) Reference trajectory design (2) Trajectory tracking via a dynamic pen model. Within each step, a number of parameters are noted to facilitate the discussions in Section IV. These parameters are highlighted at their point of definition.

A. Reference Trajectory Design

This step takes as input the conventional render, normal and depth buffers of a 3D scene. From these buffers, it produces the curves and regions to be rendered with strokes. We use the depth and normal buffers to identify the silhouettes and edges, and the render buffer for shading region identification and tonal control.

From these buffers, our method generates a 2D stroke pattern that traces the identified curves, and shades the regions. For each stroke generated during this process, its position, shape, length, and intensity variation is computed using the render buffers. Spatial and angular perturbations are then applied to the strokes to simulate the inaccuracies with the human motor control system. Finally, this stroke pattern is transformed into a single, time-dependent 3D space curve. The x and y components of this resulting curve control the stroke's final shape in the image plane, while the z coordinate dictates whether the stroke leaves a trace on the paper, and if so, the stroke's thickness and intensity. The transformation from a 2D pattern to a 3D curve allows the subsequent pen model to follow a 3D path, thereby allowing a rich set of visual artifacts to be created on its projection to the image plane.

1) Silhouettes and Edges: To design the stroke trajectory for the silhouettes and edges of an object, we use conventional edge detection methods on the depth and normal buffers. However, other methods such as suggestive contours [2], [3], [4] could also be used without loss of generality. In our implementation, we use the Laplacian convolution kernel on the depth and normal buffers to identify the curves at which the depth and normal vectors vary abruptly. (Fig. 1). When combined, this results in the extraction of silhouettes and sharp edges of the model as a set of polylines. From this point onward, we do not distinguish between a silhouette or an edge curve, and instead simply refer to them as ***siledge*** curves.

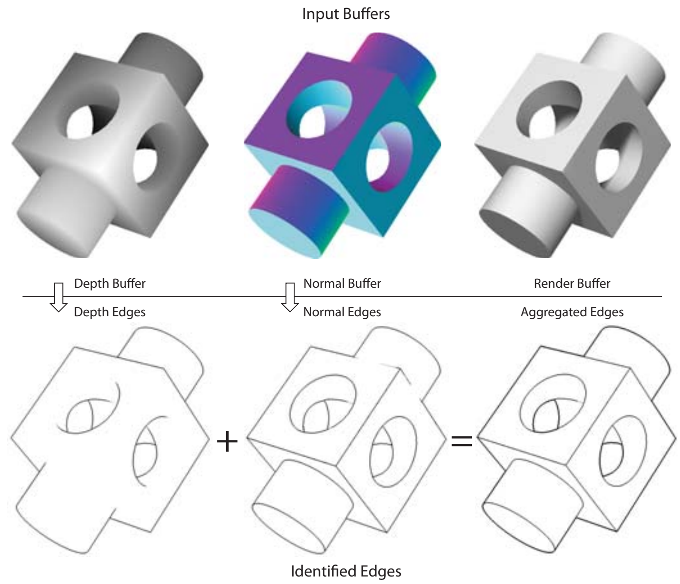


Fig. 1. The input buffers: Depth, normal, and render buffers. The silhouettes calculated from the depth buffer and the sharp edges identified from the normal buffer are combined to identify the silhouettes and edges. The render buffer is later used for hatching.

While this analysis may produce clean siledges , it may also produce disconnected islands of such curves. In either case, we assemble all siledge curves into a single chain using their spatial proximity. In addition to its spatial configuration, the siledge path also maintains information regarding the underlying image intensity from the render buffer, which is later used to generate the height (z)coordinates.

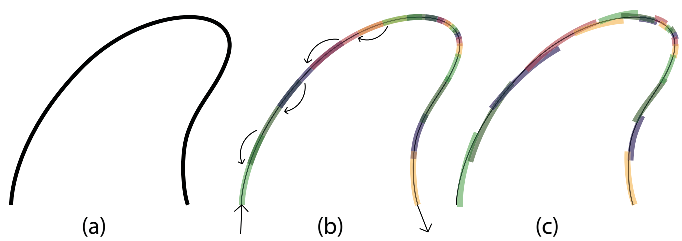


Fig. 2. Each siledge is converted into a series of strokes. The stroke lengths are determined adaptively according to the curvature along the siledge. The magnitude of overlap between consecutive strokes is determined probabilistically from a normal distribution. The strokes are altered by random perturbations in orientation, length, and scale. Strokes are rendered thicker than normal for demonstration.

The next step involves stroke design on this siledge path. The local curvature of the path has an influence on the length of the strokes to be generated. High curvature regions typically result in shorter strokes, while low curvature sections can be traced using longer strokes. Once stroke lengths are determined, we use a normal distribution to generate variations in the amount of overlap between consecutive strokes, as a function of the strokes' lengths. Figure 2b demonstrates this idea. We control the average length of the strokes and the overlap with two parameters, $\mathbf{P_{L_{mean}}}$ and $\mathbf{P_{overlap}}$, respectively. Finally, the stroke orientations, lengths and positions are randomly perturbed. We do so using a single parameter $\mathbf{P_{siledge}}$ that concurrently controls the standard deviations of these attributes, as shown in Figure 2c.

Once the size, shape and position of the strokes are determined, points are sampled along the stroke to mimic equal time intervals during traversal. Similar to stroke length determination, we simulate the effect of the pen slowing down around high curvature regions using curvature-adaptive sampling. High curvature regions are thus sampled more densely compared to low curvature regions. In the following sections, we discuss the effects of the pen speed on tracking, and thus the stroke renditions produced by our pen model.

2) Hatching: We use the render buffer to identify the regions to be hatched for shading. For this, we apply a low pass Gaussian filter to the render image and use an intensity cut-off value ($\mathbf{P_{int-cutoff}}$) to isolate the shading regions. The first row of Figure 3 illustrates the idea. This process results in a set of bounded regions, each treated as a separate area to be shaded. Within each region, a hatching pattern is designed that maintains a flow congruent with that region's geometry.

If the previously computed siledge path crosses through an identified region, that region is further subdivided into smaller regions. Since siledges correspond to discontinuities in the depth and normal maps, this subdivision helps our system produce distinct hatching patterns within each subregion. This desirably helps the hatching to preserve such discontinuities.

As shown in the second row of Figure 3, the identified regions often exhibit complex boundaries. This prevents a straightforward adoption of a hatching direction, as the resulting hatching behavior would require frequent pen-up/pen-down motions. Artists typically prefer to shade such regions by dividing it into multiple smaller groups that can be continuously hatched with a series of strokes. Inspired by this observation, we compute the medial axis of the region, which reduces our analysis to a group of branches. To identify the salient regions to be distinguished, we seek to merge branches that are connected with smooth transitions. This is done by studying the curvature of a compound curve that is formed by joining the two branches. If the discrete curvature of the compound curve at the joint is less than a threshold, the branches are merged (Figure 3f). This process is repeated until no more mergers are available. Next, short branches are identified and removed, thereby leaving a small set of skeletal curves forming the axes of the subregions to be hatched.

We next generate a hatching pattern for each identified region using each region's medial axis branch as a reference. Figure 4 illustrates this process. For a given region, the hatching strokes are created from arc segments sharing a fixed radius. This mimics the arc-like strokes caused by the motion of the fingers and the wrist of the artist about a virtual pivot

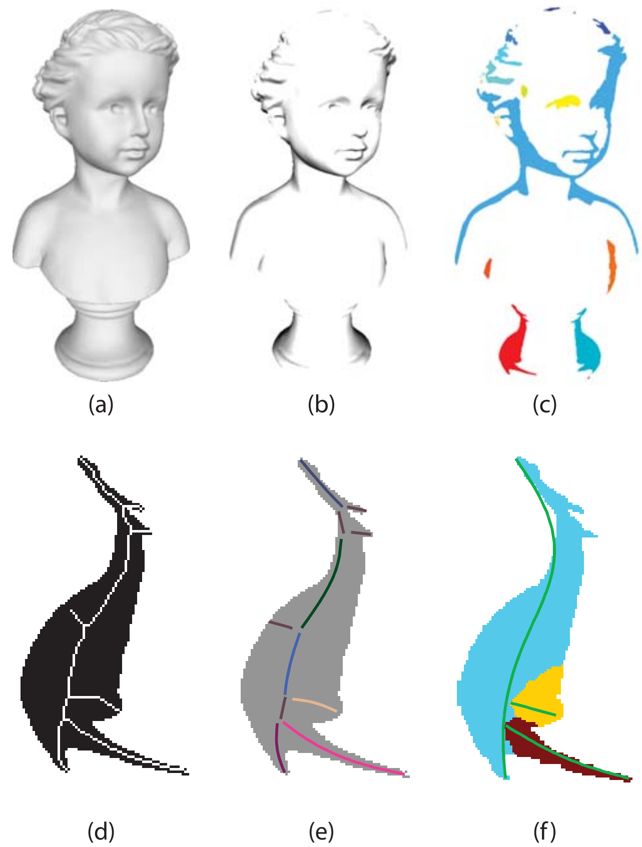


Fig. 3. Identification of hatching regions. (a) The render buffer is processed to determine (b) the hatching regions and shading tones (c) resulting in islands of regions. (d) For each region, the skeleton is calculated (e) and the branches are segmented. (f) Continuous branches are merged to produce compound branches, from which a reduced set of regions are determined.

point. The orientation of the hatching strokes are chosen such that the generated pattern is a function of the region's shape. For slender regions, the strokes are configured such that they are always offset by a prescribed angle relative to the medial axis. In such cases, the hatching pattern closely follows the shape of the medial axis. However, this requirement is relaxed for wider regions so that the long hatching strokes produced for wider regions do not form fan-like patterns in an effort to follow the tight turns of the medial axis. This effect is formulated by assigning an offset angle along the medial axis, and smoothing this angle vector in relation to the region's width along the medial axis[1]. After the hatching strokes are generated, parts of the strokes falling outside the region are masked out, leaving the hatching only in the region to be shaded.

Similar to siledges, perturbations are applied to the orientation, position and size of the hatching strokes in order to simulate the inaccuracies of the human arm. This is controlled with a single parameter $\mathbf{P_{hatching}}$. Additionally two other

[1]Note that the medial axis encodes the distance to the closest boundary points, thereby allowing the thickness of the region to be trivially computed as the medial axis is traversed.

parameters in this process control the mean frequency of the hatching strokes $\mathbf{P_{htc-freq}}$, and the deviation of the hatching frequency within the region $\mathbf{P_{htc-freq-dev}}$. Similar to the siledge strokes, hatching strokes are lastly sampled to generate points along their traversal. Since hatches are formed by constant radius arcs, the curvature-sensitive sampling produces equally sampled points along these strokes.

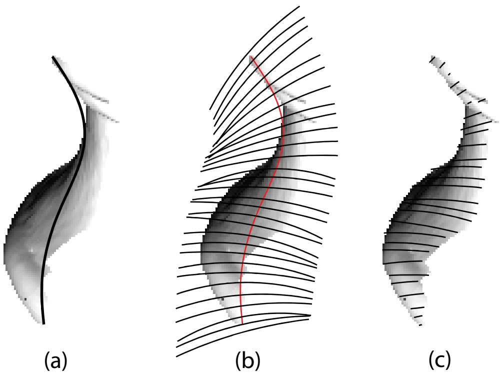


Fig. 4. Hatching stroke generation. A series of arcs are oriented along the medial axis. The arcs are trimmed with the hatching region, and perturbed in orientation and scale generating the final hatching strokes.

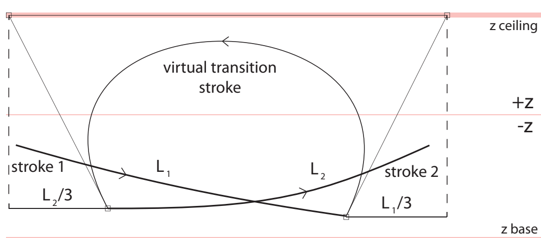


Fig. 5. The virtual transition strokes connect consecutive strokes. The strokes attain z coordinates based on the underlying intensities. The virtual transition strokes ascend into $+z$ space, though not exceeding $\mathbf{z_{ceiling}}$. The follow through of the transition strokes parallel to the virtual paper is determined by the underlying stroke's length.

3) Stroke Connections: At this point, the image intensity values along the siledge and hatching strokes (obtained from the render image) are used to create the height (z) coordinates for each stroke. This is facilitated by considering a virtual paper surface representing the $z = 0$ plane. All strokes generated thus far lie below this plane, as this indicates the pen being in contact with the paper. Darker intensities are transformed into deeper z coordinates whereas lighter ones are interpreted as shallow depths. Image intensities in the range $[0, 1]$ are linearly mapped to the z coordinates $[0, \mathbf{P_{z_{base}}}]$, where $\mathbf{P_{z_{base}}}$ is the parameter governing the maximum penetration depth as shown in Fig. 5.

After stroke heights are computed, successively generated strokes are attached to each other using virtual transition strokes. This simulates the pen lifting off the surface after one stroke ends, and reentering the paper for the next stroke. We model this effect using cubic Beziér curves as shown in Fig. 5. The design of these strokes is critical, as they govern segments of the trajectory that our dynamic pen model will be tracking. To this end, we introduce a control parameter $\mathbf{P_{z_{ceiling}}}$ that determines the maximum height the pen is allowed to ascend during transitions. Additionally, to account for longer strokes causing higher resistance for the pen to change direction at the liftoff point, we extend a vector proportional to the stroke's length. We use the right triangle formed by the ceiling, and the extension vector of the stroke to determine the tangent vectors of the Beziér curve. This allows a natural *follow through* for each of the transition strokes, based on the strokes they connect. Using this formulation, we generate a single trajectory that the dynamic pen model will next follow. Figure 6 shows the final 3D trajectory produced in this way.

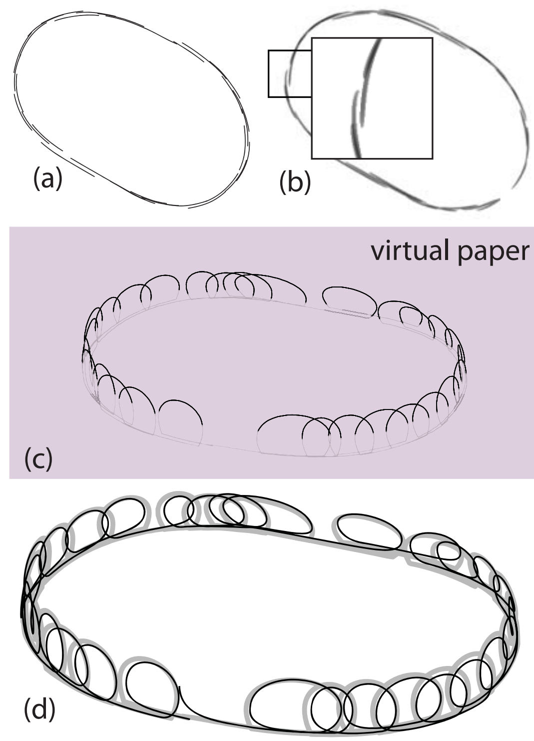


Fig. 6. (a) A series of strokes sampled in 2D (b) The resulting rendered strokes. (c) 2D strokes are transformed into 3D reference trajectory that intersects the virtual plane at multiple locations. (d) The reference trajectory (gray) is tracked using a dynamic pen (black), producing the render in (a).

B. Dynamic Pen Model

The calculated reference trajectory is next tracked by a dynamic pen model which produces the final rendered strokes (Fig. 7). The model consists of a closed loop control system representing the visual system and the motor control of the human. A free point mass represents the lumped pen, hand, and arm masses. The input to the system is the continuous 3D reference trajectories computed previously, and the output is the actual 3D trajectory traced by the point mass. A muscle model mathematically equivalent to a PID controller is used for tracking. This formulation is similar to that used in [20]. We enhance this model by adding random muscle jitter to the guidance force applied to the point mass. The jitter is modeled by a first order low pass filter that rejects frequencies above a cut-off frequency. The equivalent mass is controlled by the parameter $\mathbf{P_{mass}}$, while the amount of jitter and the filter cut-off frequency are linearly controlled by the parameter $\mathbf{P_{jitter}}$.

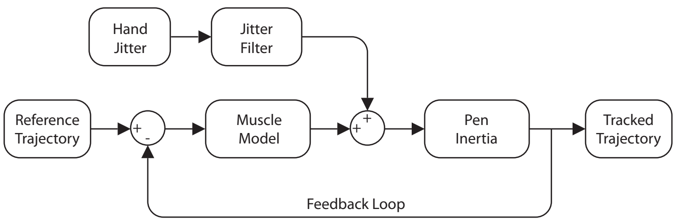


Fig. 7. Dynamic pen model is a closed loop control system that guides a pen/arm mass along a reference trajectory. The force on the pen (*i.e.* the force that the muscle model applies to the pen inertia) is perturbed to model the jitter in the motor control system.

We model the motion in three dimensions as three independent systems of equations and use a Runge Kutta solver to obtain the pen's motion in space as a function of time. The dynamic pen properties can be adjusted in the (x, y) and z directions independent from each other. Figure 6d shows the reference trajectory (gray) and the resulting pen trajectory (black) produced using this process. The final rendered strokes are computed from the intersections of the calculated pen trajectory with the virtual plane. The segments of the pen trajectory under the $z = 0$ plane form the strokes to be rendered. Furthermore, using the penetration depths of the trajectory, intensity values are assigned along the strokes similar to the way they were computed in Section III-A3.

IV. Results and Discussions

A. Performance

Our method is implemented in MATLAB. The hatching region extraction described in Section III-A2 is currently the most time consuming step. The computational cost of other calculations including stroke generation and perturbation, reference trajectory design, reference tracking using the dynamic pen model and image differencing are negligible compared to the region extraction step. The time required for the calculations depend on the resolution of the render buffers, the number of silhouette edges generated, and the total area of the regions to be hatched. The overall processing times varies between 30 seconds to 3 minutes of a Core 2 Dup computer.

B. Examples

Figure 8 presents example renderings produced by our system. The input renders are obtained from software that can export such buffers. The blow out windows show various phenomena generated from our dynamic tracking, including variations in siledge strokes, continuous versus distinct hatching strokes, cross hatching strokes, varying pen pressure and stroke intensities. In these examples, we have manually used two layers of hatching strokes where hatching directions make about 60 degrees between each other. Figure 9 shows the variations in sketch styles that can obtained by varying the described sketch parameters.

C. Limitations

Our region segmentation analysis is able to produce accurate partitions for smooth boundary regions. However, in cases where the boundaries exhibit significant waviness, the algorithm may produce fragmented regions. These regions, when hatched, can cause undesirable hatching patterns that are visually unappealing (in Fig. 8, the left leg of the second child from the right). A similar situation occurs when a short siledge is trapped within a large hatching region. This causes fragmented regions, where the hatching angles change abruptly within the region. One solution may involve a region analysis algorithm that also takes into account additional factors such as intensity variations within the region, together with the image skeleton approach we use.

V. Conclusion

We described a new non-photorealistic rendering method to render 3D objects in the form of pencil-like sketches. Our method is based on the observation that the dynamic feedback mechanism involving the human visual system and the motor control of the hand collectively generates the visual characteristics unique to hand-drawn sketches. To this end, we developed a stroke planning and dynamic tracking algorithm that produces a sketch render using silhouette, edge and hatching strokes. This approach allows visual artifacts unique to hand drawn sketches to be reproduced.

The mapping between our algorithmic parameters and the three sketch descriptors we present provides a convenient basis for studying the stylistic variations produced by our system. We believe this opens a new avenue for future work where the technical parameters can be learned and mapped to the descriptor parameters using data-driven approaches, thereby providing a link between artistic and algorithmic languages. We also believe our system may also serve as a training and visualization tool, as our approach helps produce a temporal progression of the sketch.

References

[1] K. Eissen and R. Steur, *Sketching: drawing techniques for product designers*. Bis, 2007.

[2] D. DeCarlo, A. Finkelstein, S. Rusinkiewicz, and A. Santella, "Suggestive contours for conveying shape," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 22, no. 3, pp. 848–855, Jul. 2003.

[3] E. Kalogerakis, D. Nowrouzezahrai, P. Simari, J. McCrae, A. Hertzmann, and K. Singh, "Data-driven curvature for real-time line drawing of dynamic scenes," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 1, p. 11, 2009.

[4] L. Zhang, Y. He, X. Xie, and W. Chen, "Laplacian lines for real-time shape illustration," in *Proceedings of the 2009 symposium on Interactive 3D graphics and games*. ACM, 2009, pp. 129–136.

[5] L. Markosian, M. Kowalski, D. Goldstein, S. Trychin, J. Hughes, and L. Bourdev, "Real-time nonphotorealistic rendering," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1997, pp. 415–420.

[6] J. Yeh and M. Ouhyoung, "Non-photorealistic rendering in chinese painting of animals," *Journal of System Simulation*, vol. 14, no. 6, pp. 1220–1224, 2002.

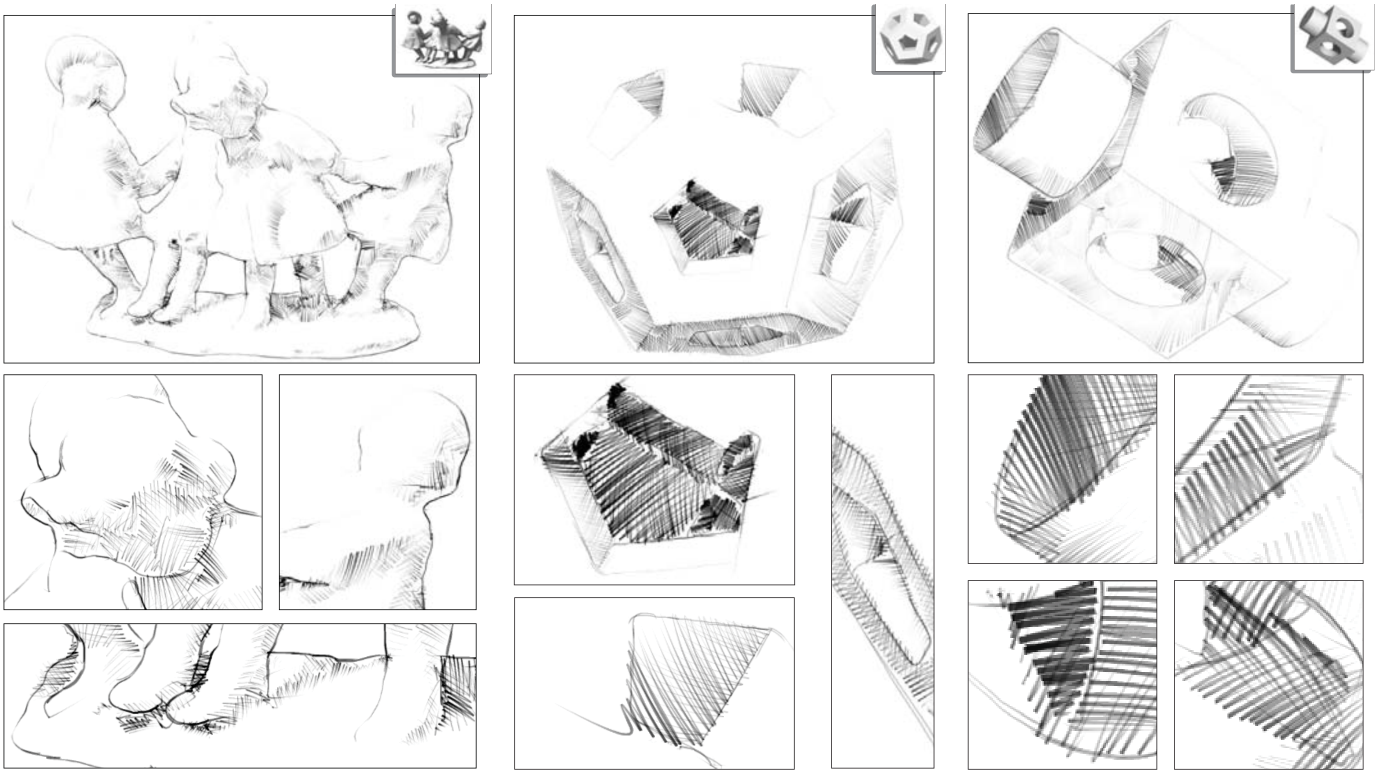


Fig. 8. Results on various models. Note the hatching behaviors and dynamic effects illustrated in the blow out views.

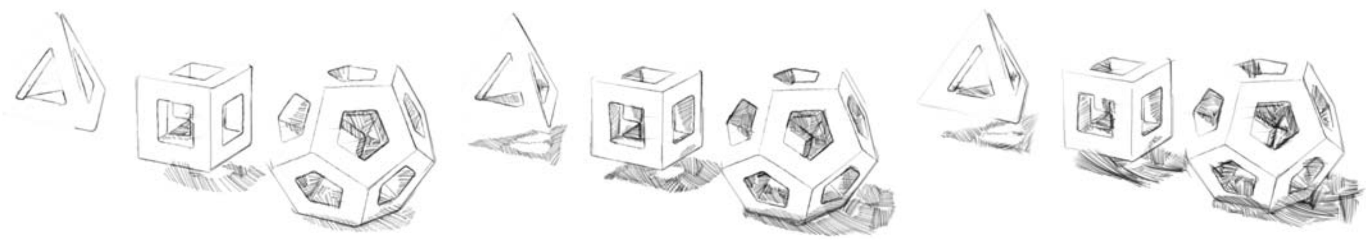


Fig. 9. The renders resulting from different values of sketch parameters. Please note the varying amounts of strokes and different hatching behaviors.

[7] A. Hertzmann, N. Oliver, B. Curless, and S. Seitz, "Curve analogies," in *Proceedings of the 13th Eurographics workshop on Rendering*. Eurographics Association, 2002, pp. 233–246.

[8] P. Barla, S. Breslav, J. Thollot, F. Sillion, and L. Markosian, "Stroke pattern analysis and synthesis," in *Computer Graphics Forum*, vol. 25, no. 3. Wiley Online Library, 2006, pp. 663–671.

[9] M. Brunn, M. Sousa, and F. Samavati, "Capturing and re-using artistic styles with reverse subdivision-based multiresolution methods," *International Journal of Image and Graphics*, vol. 7, no. 4, pp. 593–615, 2007.

[10] G. Winkenbach and D. Salesin, "Computer-generated pen-and-ink illustration," in *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. ACM, 1994, pp. 91–100.

[11] ——, "Rendering parametric surfaces in pen and ink," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 469–476.

[12] M. Salisbury, M. Wong, J. Hughes, and D. Salesin, "Orientable textures for image-based pen-and-ink illustration," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1997, pp. 401–406.

[13] A. Hertzmann and D. Zorin, "Illustrating smooth surfaces," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 517–526.

[14] E. Praun, H. Hoppe, M. Webb, and A. Finkelstein, "Real-time hatching," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 2001, p. 581.

[15] J. Palacios and E. Zhang, "Rotational symmetry field design on surfaces," *ACM Trans. Graph.*, vol. 26, Jul. 2007.

[16] A. Paiva, E. Vital Brazil, F. Petronetto, and M. Sousa, "Fluid-based hatching for tone mapping in line illustrations," *The Visual Computer*, vol. 25, no. 5, pp. 519–527, 2009.

[17] P. Jodoin, E. Epstein, M. Granger-Piché, and V. Ostromoukhov, "Hatching by example: a statistical approach," in *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*. ACM, 2002, pp. 29–36.

[18] R. Kalnins, L. Markosian, B. Meier, M. Kowalski, J. Lee, P. Davidson, M. Webb, J. Hughes, and A. Finkelstein, "Wysiwyg npr: Drawing strokes directly on 3d models," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 755–762, 2002.

[19] E. Kalogerakis, D. Nowrouzezahrai, S. Breslav, and A. Hertzmann, "Learning Hatching for Pen-and-Ink Illustration of Surfaces," *ACM Transactions on Graphics*, vol. 31, no. 1, 2011.

[20] D. House and M. Singh, "Line drawing as a dynamic process," *Pacific GraphicsŠ 07*, pp. 351–360, 2007.

Exploring Local Optima in Schematic Layout

Daniel Chivers
Department of Computer Science
The University of Kent, Canterbury
Email: dc355@kent.ac.uk

Peter Rodgers
Department of Computer Science
The University of Kent, Canterbury
Email: P.J.Rodgers@kent.ac.uk

***Abstract*—In search-based graph drawing methods there are typically a number of parameters that control the search algorithm. These parameters do not affect the fitness function, but nevertheless have an impact on the final layout. One such search method is hill climbing, and, in the context of schematic layout, we explore how varying three parameters (grid spacing, the starting distance of allowed node movement and the number of iterations) affects the resultant diagram. Although we cannot characterize schematics completely and so cannot yet automatically assign parameters for diagrams, we observe that when parameters are set to values that increase the search space, they also tend to improve the final layout. We come to the conclusion that hill-climbing methods for schematic layout are more prone to reaching local optima than had previously been expected and that a wider search, as described in this paper, can mitigate this, so resulting in a better layout.**

I. Introduction

Search-based methods for graph drawing have been successfully used for a number of years. Despite taking a relatively long time to produce layouts, they have a number of advantages, including that of targeting a fitness function that explicitly includes layout metrics, so allowing a direct measure of the quality of a graph.

Schematic layout, often called the metro map layout problem, is a variant on the graph drawing problem where a number of aesthetics are present, including the requirement that edges are restricted to a limited number of angles, typically octilinearity. Search is a common method for attempting to solve the schematic layout problem, perhaps because force directed techniques struggle to meet the aesthetic requirements [5]. As a result this problem seems suitable for further investigation in improving the effectiveness of search in graph layout.

Current search-based methods for schematic layout include hill-climbing and mixed integer programming. The results from hill-climbing methods for schematic layout have been positively compared to some published maps [8], and allow easy implementation of layout criteria. Current hill-climbing methods also provide plenty of room for performance improvements on criteria calculations. Mixed integer programming techniques are able to escape local minima, producing high-quality schematics at the cost of optimisation time [6]. However, it is hard to implement new criteria and there is less potential for optimization.

Parameter optimization in search has been widely studied [1] and relates to the problem of metaoptimization [7]. Search-based methods for the general graph drawing problem include simulated annealing [4] and genetic algorithms [2]; however, these methods make a wide search of the problem space, which requires many recalculations of fitness, so are not considered feasible with the computationally heavy fitness function required in schematic layout. It is one of the goals of the research described in this paper to see if a wider search can be conducted, whilst keeping runtime within computationally sensible bounds. This implies that both performance improvements to the fitness calculation are needed, and that a narrower search algorithm than the more general methods is required. Due to the large performance improvement potential and ease of addition of new criteria, we have chosen to use hill-climbing for our automated schematic layout method.

Any search-based algorithm relies on the setting of a number of parameters. In the case of hill climbing, the parameters we considered in this paper were: grid spacing; the starting distance of node movement; and the number of iterations that the algorithm lasted for. We were interested in seeing if we could increase the search of the problem space by varying these parameters. As they are typically set in an ad-hoc manner, we aimed to discover if current hill-climbing schematic layout had a tendency to reach local optima, indicating that current search-based methods are prone to sub-optimal results. As a consequence, we hoped to improve the layout of schematic diagrams drawn by search-based methods.

The method for calculating metrics described in this paper is considerably faster than previous techniques, hence permitting more alternatives to be examined and so widening the search of the problem space. We examined four well known metro maps and our results indicate that the variation in final layout as the parameters change has underlying trends, but the result is unpredictable for specific parameter sets. In addition, the difference in final layout can be large, often between sets of parameters that have values that are close together. We interpret this as indicating that the system reaches a relatively poor local optima frequently and that there is no general characterization of parameter sets for any of the maps. As a result, in general, increasing the search space by attempting multiple layouts with different parameters results in a better schematic than can be reached by a single search using a single parameter set.

In the remainder of this paper, Sect. II provides an overview of the algorithm used to layout schematics, along with a description of the performance improvements and the algorithm parameters. Section III provides the results from the testing process, as well as our interpretation. Finally, Sect. IV gives our conclusions and outlines possible future work.

II. Overview of the Method

Inspired by the methods developed in [8], our automated layout algorithm uses a multicriteria hill-climbing search

TABLE I. MAPS USED

Map	Junctions	Stations	Edges
Washington	9	77	53
Vienna	10	80	63
Mexico City	24	123	120
Sydney	24	151	103

TABLE II. PARAMETERS AND VALUES USED

Parameters	Values
Grid Spacing	8, 10, 12, 14
Start Distance	13, 14, 15, 16, 17
Iterations	8, 12, 16, 20

technique. This method operates by attempting to lower a set of measurable criteria by performing modifications to the schematic. There are two stages: firstly, the nodes are positioned; and secondly, the labels are positioned. In this paper we targeted performance improvements on the first, node positioning, as this is the major bottleneck.

There are three key parameters to the method. These are:

1) **Grid Spacing**: A grid is placed over the canvas, and each schematic node must be positioned on a grid point. This parameter defines the grid resolution in pixels. When altered, this parameter affects the start layout as the nodes are initially snapped to the grid. It will also alter the number of potential sites that they can be positioned in when they move, and the distance by which they can move.
2) **Start Distance**: This parameter defines the initial (and maximum) distance the nodes can be moved, in terms of grid positions. A method is applied to reduce this distance over the duration of the optimization, as explained below.
3) **Iterations**: In an iteration, every node and cluster of nodes is examined to see if its location can be improved. Once an iteration has completed, the distance the nodes can move in the following iteration is evaluated. The initial iteration always uses the start distance, and the last iteration always uses a distance of one grid space. The distance is represented by an integer, and decreases along a floored linear interpolation between the initial and final iterations, so is not always reduced between iterations.

The method uses a series of criteria to calculate the aesthetic fitness. Each of these criterion has its own metric for calculating a value on the current schematic representing how well it adheres to the criterion. Using this value we can then calculate, weight, and sum it with all other values to produce a total fitness for the current schematic. The algorithm will attempt to reduce the total fitness value by moving around nodes, or clusters of nodes, and recalculating the required values. A node can be either a junction (station with a degree of greater than 2) or a line bend point. A total fitness value of zero would indicate that all criteria have been perfectly met. The criteria include:

1) **Octilinearity.** Lines should be at multiples of $45°$.
2) **Edge Length.** Line sections between nodes should be a standard length.
3) **Line Straightness:**
   a) **Total.** The entirety of the line should be as straight as possible.
   b) **Through Nodes.** Lines sections passing through junctions should be kept as straight as possible.
4) **Edge Crossings.** Lines should not cross other lines.
5) **Occlusion.** Nodes should not occlude parts of any edges.

TABLE III. OVERALL TIME IMPROVEMENTS (IN MINUTES)

Diagram	Avg. Before	Avg. After	Speedup (times faster)
Washington	36.872	4.630	8.0
Vienna	51.929	10.581	4.9
Mexico	234.982	27.475	8.6
Sydney	518.813	61.340	8.5

TABLE IV. LAYOUT TIME IMPROVEMENTS BY CRITERIA (IN MINUTES)

Criteria	Avg. Before	Avg. After	Speedup (times faster)
Octilinearity	1.251	0.040	31.3
Edge Length	6.534	0.041	159.4
Line Straightness	13.289	3.062	4.3
Edge Crossings	14.671	1.870	7.9
Occlusions	110.223	5.257	21.0

Nodes can be moved in eight directions; *North*, *North-East*, *East*, *South-East*, *South*, *South-West* and *West*.

The initial implementation of the search method in [3] was not optimized for performance, making the layout slow and so attempting the type of experimentation here was infeasible. As a consequence, we spent some effort improving the performance of the method in order to make it run faster. The main performance increase is gained by caching all individual criteria fitness values for nodes and edges and re-using these as much as possible. We detect graph items that have moved and only recalculate the fitness values that were affected; these are then summed with the unaffected, previously calculated, values.

In some cases, in particular edge crossings and occlusion, detecting items that have been affected by a node movement is not trivial as a change in the position of a single edge can affect edges or nodes along its length. In order to avoid having to re-check the moved edge with every other edge and junction, we place a second grid over the entire schematic. At the start of layout, each edge is examined and the edges that pass through grid cells are identified. This edge location grid is updated each time an edge is moved. Using such a grid to monitor the location of edges speeds up the testing for edge crossings and occlusions when checking for each, as the method can identify a subset of all nodes and edges as potential occlusion or crossing candidates by the grid squares in which changes have been made.

Table I lists the maps on which the testing has been performed, along with the number of junctions, stations and edges. The maps were chosen as being representative of reasonably sized schematics that demonstrated different characteristics and for which we could easily access the data. The table is listed by ascending order of the total number of junctions and stations. Table III shows the overall time performance increases gained by the algorithm improvements on each map. As seen in the table, there is a large performance increase from the implemented changes, averaging a 7.5 times speed improvement across the four schematics. This improvement in run time made the testing feasible by allowing us to carry out the required experiments in a reasonable time frame. Table IV

TABLE V. VIENNA RESULTS

Rank	Map ID	Start Distance	Iterations	Grid Spacing	Fitness
1	46	15	20	10	0.972
2	78	17	20	10	0.981
3	58	16	16	10	0.986
4	62	16	20	10	0.986
5	74	17	16	10	0.986
6	42	15	16	10	1.037
7	30	14	20	10	1.135
8	70	17	12	10	1.180
9	54	16	12	10	1.195
10	10	13	16	10	1.317
...					
40	53	16	12	8	2.174
...					
80	68	17	8	14	4.233

shows a breakdown of the time improvements on a per-criterion basis. There was a substantial performance increase per criterion, averaging 44.8 times faster. This is higher than the performance increase on the overall running time due to the algorithm performing other tasks that have not been optimized, such as label placement.

The timings were performed on an ASUS Eee Pad Transformer TF101 running the Android operating system, version 3.2.1. The device runs on a 1GHz NVIDIA Tegra 2 with 1GB of RAM.

To test how changing the parameters impacts on the final diagram, we developed a test rig that would allow us to explore a variety of settings for the chosen parameters. The range of values for the parameters were decided by informal testing and are given in Sect. III. The testing rig outputs images of each schematic, and a file containing the fitness value of each. From this data, we investigated how modifying the parameters relates to the final fitness value. Of course, we can also pick the best layout (the one with best fitness) from those generated to be our output schematic.

III. RESULTS AND ANALYSIS

In this section we present the results from the testing of each of the four maps. We have chosen Vienna as an example in which to go into more detail because it provides good variation in layout between the different runs. Diagrams generated from the other examples can been seen in the Appendix.

Table II lists the values used for each parameter. Running a map from Table I with all possible parameter configurations from Table II results in a total of 80 layouts. Table V shows an abridged table of the Vienna schematic results from the testing rig along with the parameter settings for each. It shows the top ten best schematics by fitness value, the median schematic and the single worst. An immediate trend can be seen from this table with a grid spacing value of ten appearing in all the best fitness values. For other maps, data given in the Appendix, only Sydney shows a similar trend, the other two maps do show some grouping by grid spacing, but in these cases it is less conclusive as the best grid also appears much lower down the ranking. For the four maps, only two share the same grid spacing in the best map, and there is no correlation between map size and grid spacing. As a result, it appears that grid spacing is perhaps the most important parameter to explore;

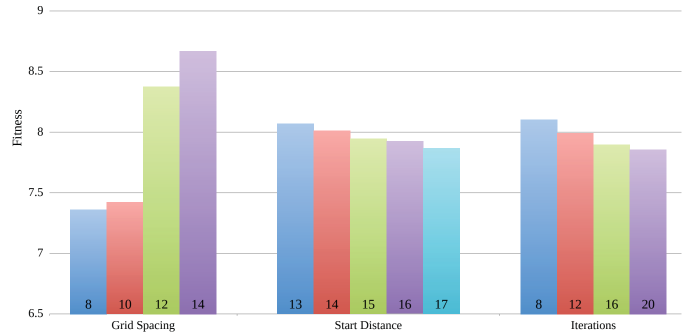


Fig. 1. Cumulative mean fitness value for all maps combined. Note: The Y-Axis starts from 6.5

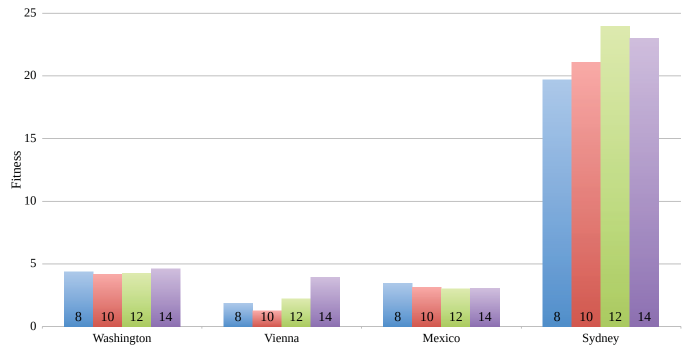


Fig. 2. Mean grid spacing against fitness value

however, there are no patterns evident that would allow the derivation of an ideal grid spacing for a particular map.

From Table V we can also see that there is a continuing improvement in fitness at the top of the list. This pattern, where the best fitness is found for only one set of parameters is also shown in two other maps (see the Appendix). We believe this is an indication that the system is not converging on the optimum solution, and so an even wider examination of the search space may be required to achieve the best final layout.

Figure 3 shows the original geographic layout of Vienna, used as the starting point by the algorithm. Figures 4, 5 and 6 show the best, median and worst final schematics for Vienna respectively. We have included the median, as this is the expected layout when parameters are arbitrarily chosen from the ranges used in this paper. In this case, the fitness of the best diagram is 44.71% of the median. Various cases of local optima are visible in the median Vienna diagram, Fig. 5. For example, the median diagram has generally worse line straightness than the best diagram, Fig. 4, which can be seen in nearly all lines. Many more cases of local optima can be seen in the worst diagram, Fig. 6, which along with much poorer line straightness, has multiple edges which break the octilinearity criterion.

Besides Vienna, we have also included the geographic and best layout of Sydney from the tests, shown in Figures 7 and 8 respectively. An abridged table of results for Sydney can be found in the Appendix. Geographic and best images of Mexico and Washington, along with results, can also be found in the Appendix.

When the parameters are set in an ad-hoc fashion, and

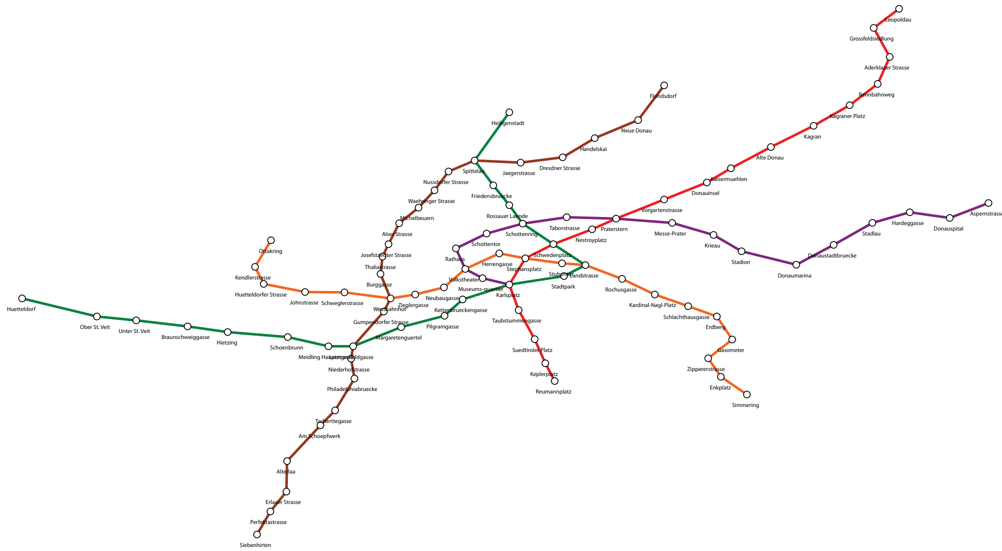


Fig. 3. Vienna - Geographic Map

Fig. 4. Vienna - Rank 1 (Fitness = 0.972)

the wider search done here is not performed (as in the case in most current layout methods), the expected output layout is the median fitness value, and we can investigate how much our best diagram improves on this. In the case of Washington the best is 95.68% of the median, for Mexico the best is 86.79% of the median, and for Sydney the best is 91.28% of the median. As the percentage for Vienna was 44.71%, this implies that in some cases, something close to the best diagram can be found with a wide range of parameters, so allowing a more constrained (and faster) search to be applied. In other cases, a more restricted search will produce a much worse diagram.

It has been mentioned that a small change in the algorithm parameters can make a very large difference to the resultant layout. An example of this is the top ranked Mexico layout which has a fitness of 2.667, with parameters: start distance 13, iterations 20, grid spacing 14. A change from 20 to 16 iterations (one step in our testing) results in a drop to rank 50 and a fitness of 3.217.

Figure 1 shows the cumulative mean fitness for each parameter value. The graph indicates that, for number of iterations and starting distance, a larger parameter value is likely to lead to a layout with a better fitness; these trends also hold true on the map-specific graphs for start distance and iterations shown in the Appendix. Increasing both of these parameters increases the search space, so indicating that an even wider search will tend to produce a better diagram. Grid spacing, which, as discussed previously, has the greatest effect upon the resulting layout's fitness, tends to produce better results when a smaller value is used.

However, Fig. 2, which shows the mean fitness value of each layout produced by varying the grid spacing, does not show this trend on a by-map basis. This graph has been included because, unlike search distance and iterations, a clear trend cannot be seen in the data. Although grid spacing has a large effect on the fitness value of the produced layouts, there is no single optimum grid spacing value for all layouts, and each

Fig. 5. Vienna - Rank 40 (Fitness = 2.174)

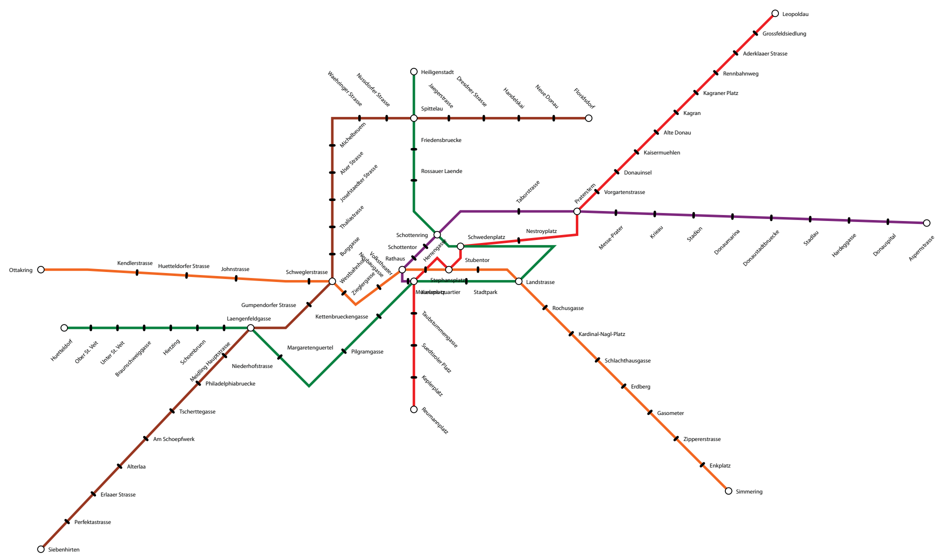


Fig. 6. Vienna - Rank 80 (Fitness = 4.233)

has a specific value at which the best layouts are produced. Interestingly, Mexico and Sydney, which are comparatively large maps with similar sizes, display opposite trends in terms of grid spacing, with Mexico favouring the larger spacing values and Sydney the smallest. Further investigation is required by examining a larger variation of grid spacing in order to try to identify any global trends.

IV. Conclusion

When implementing any multicriteria search method, many parameters are used to configure the algorithm, including those studied in this paper. With the algorithm we present here, the fitness of the best schematic layout is typically considerably better than a layout that might be found by an ad-hoc method for assigning parameter values. As a result, our algorithm generates better layouts than previous schematic search-based methods.

We have shown that for all four maps there is a slight trend towards better layouts being produced by more iterations and a greater start distance. This behaviour is expected, as increasing these parameters allows the system to evaluate more station positions, but at the potential cost of greater run time. Unlike number of iterations and start distance, grid spacing did not display a clear trend for improving fitness across all schematics, but results obtained did indicate that each schematic has a 'best' grid spacing value at which a lower fitness is more commonly produced. Although trends have been identified in all parameters, the best layouts are not when all the parameters are at their best settings as indicated by the trends; there is considerable variation.

We have begun work on examining parameters for improving layout and characterizing maps. However, much future work needs to be done. Firstly, other parameters may be investigated, for example, the number of bend points on diagrams and those that control clustering. Secondly, the current

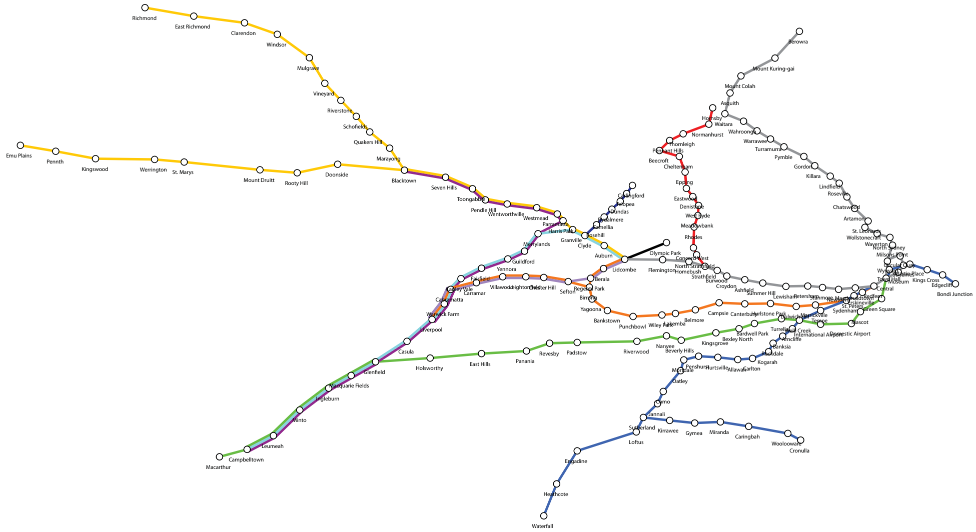


Fig. 7. Sydney - Geographic Map

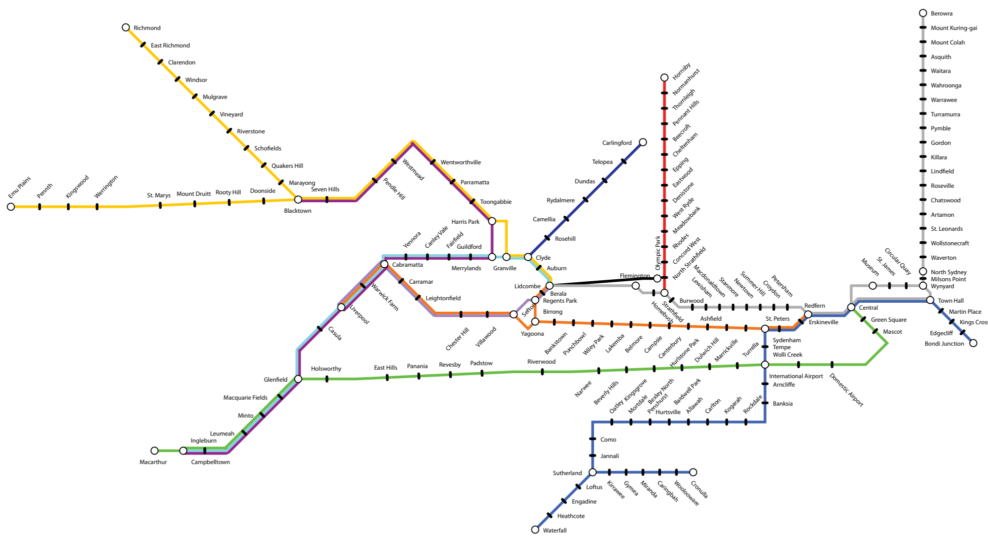


Fig. 8. Sydney - Rank 1 (Fitness = 19.387)

parameters have not been fully explored, including looking at the limits of the trends currently identified in this paper. We could also develop more sophisticated characterizations of diagrams such as geographic density variance, or examine if radial and orbital schematic characteristics have an effect and so can aid the choice of parameter value. Finally, we have evidence that the most optimal layout has not yet been reached with our current method, and so further investigation with a wider search is an important next step towards generating the best possible schematic layouts.

References

[1] Thomas Bäck and Hans-Paul Schwefel. An overview of evolutionary algorithms for parameter optimization. *Evol. Comput.*, 1(1):1–23, March 1993.

[2] Jürgen Branke, Frank Bucher, and Hartmut Schmeck. Using genetic algorithms for drawing undirected graphs. In *The Third Nordic Workshop on Genetic Algorithms and their Applications*, pages 193–206, 1996.

[3] Daniel Chivers and Peter Rodgers. Gesture-based input for drawing schematics on a mobile device. In *Proceedings of Information Visualization*, pages 127–134, 2011.

[4] Ron Davidson and David Harel. Drawing graphs nicely using simulated annealing. *ACM Trans. Graph.*, 15(4):301–331, October 1996.

[5] Seok-Hee Hong, Damian Merrick, and Hugo A. D. do Nascimento. The metro map layout problem. In *Proceedings of the 2004 Australasian symposium on Information Visualisation - Volume 35*, APVis '04, pages 91–100, Darlinghurst, Australia, 2004. Australian Computer Society, Inc.

[6] Martin Nöllenburg and Alexander Wolff. A mixed-integer program for drawing high-quality metro maps. In *Proceedings of the 13th international conference on Graph Drawing*, GD'05, pages 321–333, Berlin, Heidelberg, 2006. Springer-Verlag.

[7] Ibrahim Osman and Gilbert Laporte. Metaheuristics: A bibliography. *Annals of Operations Research*, 63(5):511–623, October 1996.

[8] Jonathan Stott, Peter Rodgers, Juan Carlos Martínez-Ovando, and Stephen G. Walker. Automatic metro map layout using multicriteria optimization. *IEEE Transactions on Visualization and Computer Graphics*, 17(1):101–114, 2011.

Appendix

TABLE VI. FITNESS STATISTICS BY MAP

Schematic	Mean	Median	Mode	SD	IQR
Washington	4.358	4.258	4.634	0.222	0.489
Vienna	2.327	2.197	2.221	1.043	0.601
Mexico	3.179	3.139	3.073	0.306	0.304
Sydney	21.924	21.926	23.232	1.681	2.675

TABLE VII. STATISTICS BY PARAMETER VALUE

Parameters	Mean	Median	Mode	Standard Deviation	IQR
Grid Spacing					
8	7.354	4.116	19.730	7.223	5.471
10	7.421	3.879	4.144	8.004	6.143
12	8.361	3.624	2.221	9.090	6.461
14	8.653	4.434	4.634	8.353	5.907
Start Distance					
13	0.043	4.178	19.663	8.264	5.318
14	7.992	4.205	19.660	8.268	5.466
15	7.928	4.133	19.730	8.215	5.468
16	7.911	4.133	19.730	8.193	5.486
17	7.861	4.105	2.221	8.200	5.339
Iterations					
8	8.081	4.205	2.221	8.197	5.425
12	7.976	4.145	2.221	8.213	5.323
16	7.884	4.144	2.221	8.221	5.542
20	7.848	4.144	2.221	8.227	5.488

TABLE VIII. WASHINGTON RESULTS

Rank	Map ID	Start Distance	Iterations	Grid Spacing	Fitness
1	226	17	8	10	4.074
2	202	15	16	10	4.116
3	206	15	20	10	4.116
4	214	16	12	10	4.116
5	201	15	16	8	4.122
6	205	15	20	8	4.122
7	213	16	12	8	4.122
8	225	17	8	8	4.135
9	170	13	16	10	4.144
10	174	13	20	10	4.144
…					
40	211	16	8	12	4.258
…					
80	209	16	8	8	4.777

TABLE IX. VIENNA RESULTS

Rank	Map ID	Start Distance	Iterations	Grid Spacing	Fitness
1	46	15	20	10	0.972
2	78	17	20	10	0.981
3	58	16	16	10	0.986
4	62	16	20	10	0.986
5	74	17	16	10	0.986
6	42	15	16	10	1.037
7	30	14	20	10	1.135
8	70	17	12	10	1.180
9	54	16	12	10	1.195
10	10	13	16	10	1.317
…					
40	53	16	12	8	2.174
…					
80	68	17	8	14	4.233

TABLE X. MEXICO RESULTS

Rank	Map ID	Start Distance	Iterations	Grid Spacing	Fitness
1	96	13	20	14	2.667
2	108	14	16	14	2.667
3	112	14	20	14	2.667
4	124	15	16	14	2.667
5	128	15	20	14	2.667
6	139	16	16	12	2.802
7	143	16	20	12	2.802
8	155	17	16	12	2.802
9	159	17	20	12	2.845
10	94	13	20	10	2.847
…					
40	151	17	12	12	3.073
…					
80	81	13	8	8	4.111

TABLE XI. SYDNEY RESULTS

Rank	Map ID	Start Distance	Iterations	Grid Spacing	Fitness
1	317	17	20	8	19.387
2	257	14	8	8	19.660
3	261	14	12	8	19.660
4	265	14	16	8	19.660
5	269	14	20	8	19.660
6	241	13	8	8	19.663
7	245	13	12	8	19.663
8	249	13	16	8	19.663
9	253	13	20	8	19.663
10	305	17	8	8	19.719
…					
40	262	14	12	10	21.237
…					
80	263	14	12	12	24.332

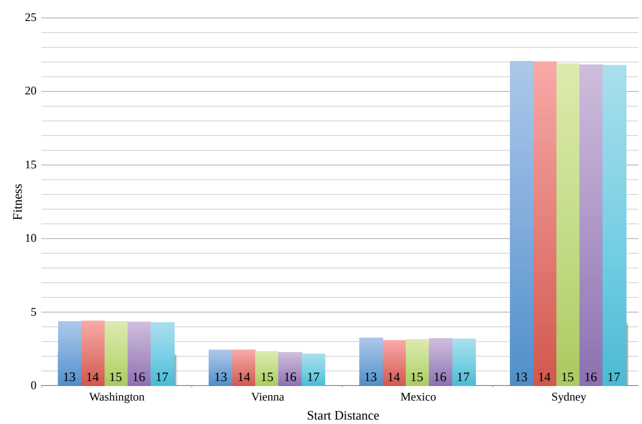


Fig. 9. Mean start distance against fitness value

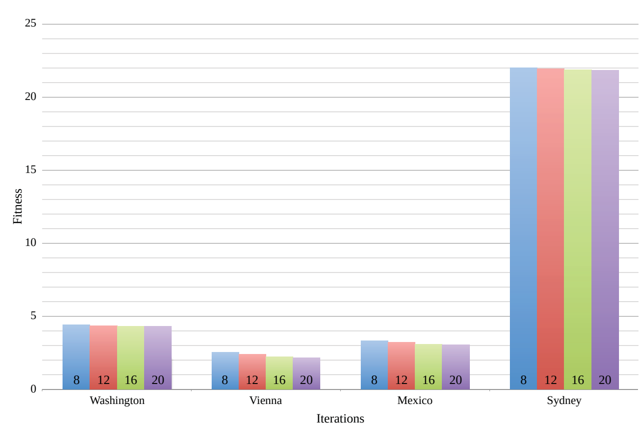


Fig. 10. Mean number of iterations against fitness value

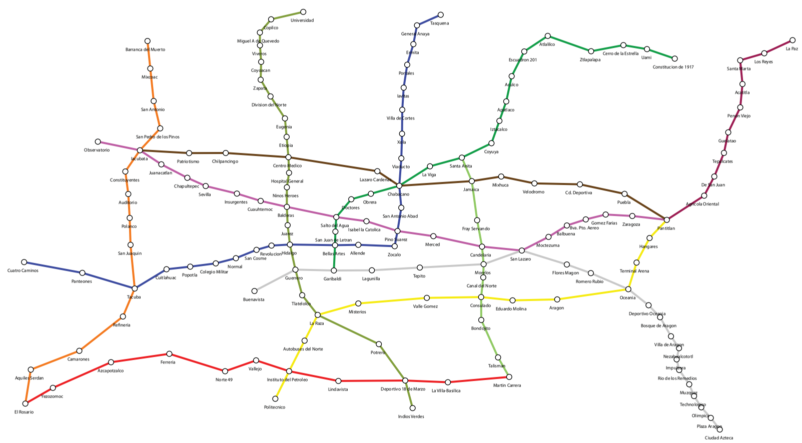


Fig. 12. Mexico - Geographic Map

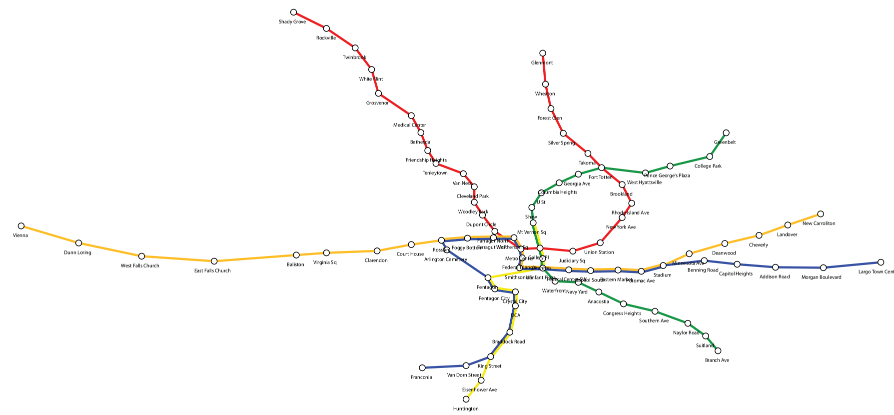


Fig. 11. Washington - Geographic Map

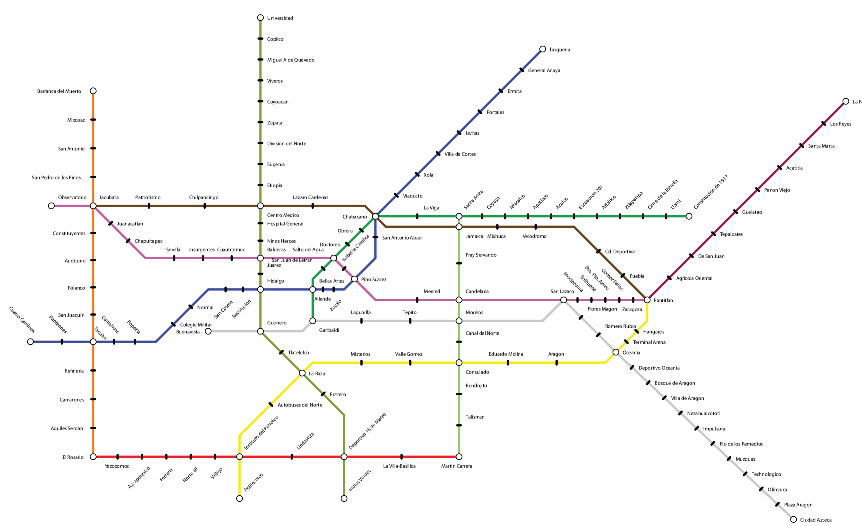


Fig. 14. Mexico - Rank 1 (Fitness = 2.667)

Fig. 13. Washington - Rank 1 (Fitness = 4.074)

From use cases to task trees through resources

Paolo Bottoni
Department of Computer Science, Sapienza University of Rome
Viale Regina Elena 295, 00161, Rome, Italy
bottoni@di.uniroma1.it

***Abstract*—We propose a systematic approach to the integrated construction of high-level specifications of use case, task, and data models, with reference to the resources involved in a successful interaction, distinguishing different types of resources and their origins. Relations between tasks are analysed by inspecting the resources in their pre- and post-conditions.**

***Index Terms*—Resources, Essential Use Cases, Task Models**

I. Introduction

We address the problem of maintaining consistency between Software Engineering (SE) and Human-Computer Interaction (HCI) outcomes of requirement analysis for interactive software, while keeping their respective concerns separated, via an approach based on a logic of resources.

As to SE, Use Cases, introduced in [11], and now part of the Unified Modeling Language (UML) standard, are at the core of the Unified Process [10]: as the central artifact of the *requirements* workflow, they support dialogue between stakeholders and structure the development process. However, there is a lack of prescriptions on how to describe use case behaviours, and how much detail on actual user interaction to incorporate in a description. With Essential Use Cases one abstracts from concrete aspects of the User Interface (UI), thus allowing flexible and adaptable interaction models and avoiding premature commitment [7], while use case content can be derived from a systematic decomposition of user goals.

In HCI, task analysis decomposes high-level tasks into simpler subtasks, down to an atomic level. UI aspects vary from mouse movements in the Goal, Operations, Methods, Strategies (GOMS) approach to user, system or interactive activities, in ConcurTask Trees (CTT) [13]. The Cameleon Reference Framework (CRF) [6] proposes task and domain models as a starting point for deriving the Final UI, via a sequence of transformations down to specific widget selection.

Despite their different origins and goal, both methods produce a user-centred view of the system, and specify which parts of a process are under whose responsibility. Also, they both require some aspects of process control, e.g. sequentialisation, iteration or optionality, to be specified. It is therefore important that these common aspects be kept consistent through the requirement process, while avoiding duplication of efforts, or giving prominence to one approach over the other.

We propose a systematic approach to the integrated derivation of use case, task and data models from a high-level specification of interactive process in terms of resources. For each process, type, nature and origin of resources are defined, providing the basis for the domain model. Then, we analyse which resources are used to achieve the process functional goals to model individual steps as rewriting rules on sets of resources, deriving control aspects from the relations between premises and conclusions of different rules. Refinements also derive from non-functional requirements, e.g. interaction or security policies. The structured collections of rules provide descriptions of use case behaviours and task models.

Section II presents two case studies. Section III introduces our approach, applications being discussed in Section IV. Section V refines the functional modeling and Section VI relates resources to task and domain models. We discuss related and future work and conclusions in Section VII.

II. Case studies

We introduce two case studies, whose operations follow a typical pattern: 1) *activate* the behaviour, 2) *set up* the execution context, 3) *perform* the execution. The activation phase often simply coincides with the trigger event, but might require confirmation or authentication, the setup phase can progressively acquire information, and performance can follow some specific protocol. In the first case study a banking system enables users to: 1) gain access to operations, 2) check the balance, 3) withdraw money, 4) make a transfer to a different account, and 5) exit a session. Access must be gained prior to selecting any operation, but concrete interfaces may pose specific requirements. For example, an ATM-policy may require that access be regained after withdrawal or transfer. The case study highlights: 1) relations between use cases; 2) consideration for physical resources; 3) injection of non-functional requirements; 4) reuse of specifications. In the second case study, a ticket is bought after indicating origin, destination and date of a trip. As the price of a ticket cannot be established before all data is available, setup is sequentialised; buying a one-way or a return ticket involves different interactions; usability may require that data be obtained via specific interactors. We highlight: 1) derivation of variants from a basic definition of the service; 2) modeling of dependencies among resources; 3) management of transactions.

III. Modeling of resources

We categorise resources via the metamodel of Figure 1, at whose core is the metaclass `Resource`, stereotyped as `Type`, together with additional, not shown, constraints. If a type stands for a collection of resources then `isCollection()` returns `true`. Resources have *attributes*, at most one of which is designated as an *identifier* (`isID()` returns `true`).

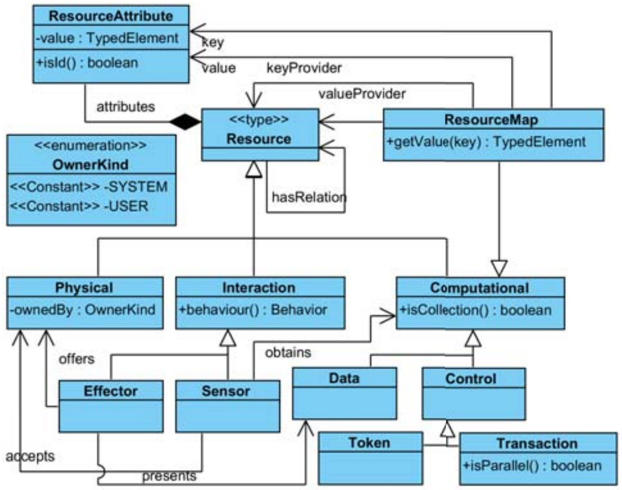


Figure 1. The metamodel for resource description.

`Physical` resources are managed via specific input and output peripherals. While they exist, they are in the availability of some interacting agent, as indicated by `ownedBy`.

`Computational` resources are of type `Data`, `Control` or `ResourceMap`. The first models data – owned, produced, or obtained by the system – relevant to an interaction. Control resources model mechanisms for constraining system behaviour. The `Token` subtype indicates resources forcing synchronization or sequentialization, while `Transaction` resources are consumed when a transaction is completed or has failed. If `isParallel` is `true`, a number of independent instances of the transaction can be executed, failing or succeeding independently, and the resource is consumed when all the launched transactions are terminated. The type `ResourceMap` represents relations between resources, through attribute values: a map pairs a value for an attribute deemed as `key` in a `keyProvider` with a value for an attribute deemed as `value` in a `valueProvider`.

`Interaction` resources model the interface with the environment. The system `obtains` computational resources and `accepts` physical ones via `Sensors`, using `Effectors` to `offer` physical resources to the environment, or `present` views of computational resources. The attribute `visible` for `Sensors` indicates runtime sensor accessibility.

The meta-association `hasRelation` gets instantiated by modelers to define arbitrary relations between resource types. The latter can be refined into collections of types.

We derive type systems from the metamodel, and represent resources as terms of the form $\langle resName \rangle (list)$, where $resName$ is the type of the resource (with lowercase initial) and $list$ is a list of attribute-value pairs. When all attributes in a resource type are relevant to a specification, we simply list their values (or associated variables) in the same order as in the type specification, omitting identifiers, otherwise we list only the relevant attributes. We give directly the value of an attribute when it is the only one besides the identifier.

IV. Modeling of behaviours

The modeling of an interactive process starts with the construction of a resource type model as an instance of the metamodel in Figure 1 and the definition of pre- and post-conditions for some *Behavior* B, in terms of statements about the presence of resources and constraints on their attributes. We identify the first action in B as the *trigger*, represented as an action by which a sensor accepts a physical resource or obtains information from the user. We classify conditions on resources associated with a process B into six categories.

1) *Enabling* pre-conditions (En) define resources, e.g. sensors, which must be present to enable triggering of B.
2) *Required* conditions (Rq) specify resources which must be present, before B is triggered, to ensure the achievement of its overall goal, but whose presence can only be assessed during the execution of B.
3) *Provided* conditions (Pv) represent information obtained via some sensor concurring to ensure success of B.
4) *Constraints* (Cx) compare attribute values, either implicitly by equality of variable names or direct use of expressions, or explicitly via OCL syntax.
5) *Holding* post-conditions (Hd) specify the configuration of computational resources guaranteed to hold after B.
6) *Interface* post-conditions (Ui) concern availability of interaction and physical resources to further activities.

A resource term in En, Rq, or Pv but not in Hd or Ui is *consumed*, while one only in Hd or Ui is *produced*. Otherwise, the term is *preserved*, possibly with updated attribute values. We also use the notation $En(B)$, $Rq(B)$, etc.

The internal articulation of B often depends on non-functional requirements, e.g. interface or security policies, rather than functional aspects. For example, the *perform* phase can require some confirmation of success, while the *activate* phase might involve some authorisation dialogue.

Relations between types induce partial orders on the flow of events needed to bring post-conditions to hold. Events are associated with rules, formed by a set of constraints and several components, each being a collection of (possibly empty) sets of resource specifications. **Context** (abbreviated $\mathcal{X}$) defines a set of resources which need to be present for the rule to be applied, but which are left unaltered. **Requested** ($\mathcal{R}$) are resources consumed or modified by the rule. **ReqOptional** ($\mathcal{R}_o$) is a list of non-empty sets of resources which, if present when the rule is fired, can alter its effect. **Neg** ($\mathcal{N}$) is a collection of sets of resources, each describing a configuration which must not be present when the rule is fired. **Ensured** ($\mathcal{E}$) are resources guaranteed to be present after rule application. **EnsOptional** ($\mathcal{E}_o$) is a list of sets of resources describing complementary effects if the resources in the corresponding $\mathcal{R}_o$ set were used. If a set is empty in $\mathcal{E}_o$, then the resources in the corresponding $\mathcal{R}_o$ are simply consumed. Finally, **Cond** (Γ) is a list of sets of constraints on values of attributes in the Γ, $\mathcal{R}$, $\mathcal{R}_o$ and $\mathcal{X}$ components, and for assigning values in the $\mathcal{E}$ and $\mathcal{E}_o$ components. The syntax $\mathbf{Rule} := \mathbf{Name} : [\mathcal{X}\ '?']\ [\mathcal{R}]['!'\ \mathcal{N}][:\ \mathcal{R}_o]\ '=>'\ \mathcal{E}[';'\ \mathcal{E}_o][':'\ \Gamma]$ gives a compact concrete representation. All compartments are optional, but at least one of $\mathcal{X}$, $\mathcal{R}$ or $\mathcal{R}_o$ should be non-empty.

Once a process B is given in terms of pre- and post-conditions, a set of rules is derived by identifying dependencies between resources, based on occurrences of the *obtains* relation and constraints in $Cx(B)$. Moreover, one creates rules s.t. the union of their $\mathcal{X}$ and $\mathcal{R}$ parts covers $Hd(B) \cup Ui(B)$, and iterates the process until all $\mathcal{X}$ and $\mathcal{R}$ are in $En(B) \cup Rq(B)$. The generated rules realise B, as can be reconstructed via a set of proof trees [1] satisfying the obligation to obtain $Hd(B) \cup If(B)$ (or $Us(B)$) starting from $En(B) \cup Rq(B)$.

A. Banking

We present a simplified banking system using several computational types and one sensor type. A resource $authMap(cstmrId, authValue)$ associates authentication and identification information, e.g. a PIN and a user name; a collection $actMap(cstmrId, accIdSet)$ relates identification information to accounts: $act(accId, balance)$ represents an account. Gaining access corresponds to producing a token $axs(accId)$ when $axsInfoSnsr()$ obtains access information matching $axsInfo(cstmrId, authValue, acctId)$. Finally, $selModeInfo(mode)$ records the current operation mode and $amt(value)$ stores the amount to be withdrawn or transferred. For transfer, $tgtAct(accId)$ is the target account. Maps and accounts are persistent resources while the others have limited lifespans. Table I shows the conditions involved with functional and interactive aspects of fund transfer.

Table I
CONDITIONS FOR *Transfer*.

Enabling	$\{selSnsr(\text{“}chs''$, true$)$, $selMd(\text{“}chs'')$, $selIdInfo(A)\}$
Required	$\{act(accId = A, balance = ABal)$, $act(accId = T, balance = TBal)\}$
Provided	$\{amt(V)$, $tgtAct(T)\}$
Constraints	$\{ABal \geq V\}$
Holding	$\{act(accId = A, balance = ABal - V)$, $act(accId = T, balance = TBal + V)$, $selIdInfo(A)\}$

As $Ui(\texttt{Transfer}) = \emptyset$, $selSnr()$ is consumed. A quick analysis also tells us that no dependency exists between amt and $tgtAct$. The *Transfer* behaviour is realised with four rules, presented in Table II, using abbreviations for their names. In order for the final rule, `PerformTransfer` to cause the post-conditions to hold, it needs to obtain amount and target in the setup phase. As these data are in Pr, they are obtained via two sensors with the corresponding rules. The rules in Table II are not specific to *Transfer*; e.g. `ObtainAmount` can also be employed for *Withdraw*. Sensors are made available in the *activate* phase, which requires a single rule.

If we now consider the obligations for the conditions in Table I, we obtain two derivation trees, of which one is represented (in a concise form) in Table III, while the other executes `OA` before `OT`, as the acquisition of target or amount can take place in any order. The top premise is satisfied when the pre-conditions hold, so that we can define

Table II
THE RULES FOR TRANSFERING.

`ST:`	$\{selMdSnsr(\text{"}C\text{"}, \texttt{true})$, $selModeInfo(\text{"}C\text{"})\}$ $=> \{selModeInfo(\text{"}T\text{"})$, $tgtActSnsr(\texttt{true})$, $amtSnsr(\texttt{true}\}$
`OT:`	$\{tgtActSnsr(\texttt{true})\} => \{tgt(T)\}$
`OA:`	$\{amtSnsr(\texttt{true})\} => \{amt(V)\}$
`PT:`	$\{selIdInfo(A)\}$? $\{selModeInfo(\text{"}T\text{"})$, $amt(V)$, $tgt(T)$, $act(A, BA)$, $act(T, BT)\} =>$ $\{act(A, BA1)$, $act(T, BT1)\}$: $\{BA >= V$, $BA1 = BA - V$, $BT1 = BT + V\}$

event flows according to the expression: `SelectTransfer` (`ObtainTarget`|`ObtainAmount`) `PerformTransfer`.

Table III
SEQUENTS IN THE PROOF OF THE POST-CONDITIONS OF *Transfer*.

$$\cfrac{\cfrac{\cfrac{\cfrac{\Gamma \vdash selSnsr(\text{"}C\text{"}), selMode(\text{"}C\text{"}), acts, selIdInfo(A), \Delta}{\Gamma, \texttt{ST} \vdash amtSnsr(), tgtSnsr(), selMode(\text{"}T\text{"}), acts, selIdInfo(A), \Delta}}{\Gamma, \texttt{OA} \vdash amt(V), tgtSnsr(), selMode(\text{"}T\text{"}), acts, selIdInfo(A), \Delta}}{\Gamma, \texttt{OT} \vdash amt(V), tgtAct(T), acts, selIdInfo(A), selMode(\text{"}T\text{"}), \Delta}}{\Gamma, \Gamma', \texttt{PT} \vdash act(A, BA - V), act(T, BT + V), selIdInfo(A), \Delta}$$

B. Ticket reservation

In the ticketing case study, the setup and perform phases are more complex. We discuss the case for a single traveller, one-way, ticket, modeled as a physical resource $tkt(tktId$, $orig$, $dest$, $date$, $price)$, based on the information about the date, origin and destination of the required trip. This information is maintained in $dateInfo(date)$ and two instances of $plcInfo(plcId, role)$, with `ORG` and `DST` the possible values for $role$. A resource $pmtInfo(pmtId)$ abstracts from payment details. A successful use case produces a physical ticket presenting the information in $tktInfo(tktId$, $trpId$, $itnry$, $date$, $price$, $state)$ and a persistent record of the transaction in $tnscn(transId)$. Resource maps $itnrMap(transId$, $tktId)$ relate transactions to emitted tickets, where $tktID$ maintains the identity of the ticket associated with the transaction identified by $transId$. The function $newTktId()$ manages the generation of new ticket identifiers. The set of resource types is complemented by sensors for obtaining user inputs and effectors for presenting the final price and for emitting tickets. The same process can be used to buy ticket for any type of trip (e.g. by train or by plane) and different organisations can refine these resources to manage specific types of trip.

The use case conditions state that if: (1) there is a place available for a trip on the itinerary and for the dates entered by the user, and (2) the user pays the correct amount, then: (a) a ticket is produced, (b) the transaction is recorded, and (c) the availability of the trip is decreased accordingly.

We assume that the use case is activated by accessing some environment, enabling sensors mentioned in the En compartment, for obtaining *date*, *orig* and *dest*, and that in the *setup* phase the user *provides* the corresponding resources in Pr. These are required to evaluate the due payment, which is then obtained from the user via a suitable sensor. The chain of dependencies $\textbf{Trp} \triangleright \textbf{PmntInfo} \triangleright \textbf{TktInfo}$ results. Hence, for the *perform* phase to end with a rule `AcceptPayment`, the system has first to retrieve a trip, and set the price for a passenger through a rule `FindTrips`, then to obtain the payment information through a $pmtInfoSnsr$ sensor, which is made available only if the trip has a seat available.

This buying process is adapted to the case of return trips by adding optional parts to manage the choice of a return ticket, and to several passengers, by obtaining a resource $psgNmbInfo(numb)$, so as to check availability with respect to the value in $numb$, also used to multiply the price of a trip. Collections must be used for journeys formed by several legs.

V. Refining behaviours

While functional models should not be concerned with the concrete UI or unrequired control structures, security or usability policies, modelled via resource dependency or refinement, might require specific forms of control, as a result of the consequent rule refinement.

Indeed, if some relation is imposed from $\mathbf{S_k}$ to $\mathbf{S_l}$, then either the terms s_k and s_l have to appear together in any rule involving either of the two, or a rule creating an instance of s_l has to use an instance of s_k in $\mathcal{X}$, $\mathcal{R}$ or $\mathcal{N}$. For example, a single sensor type **CS** may be initially defined to represent a generic interaction to *obtain* information of type **CD**. If these types are then refined into **CS1** and **CS2**, respectively **CD1** and **CD2**, the rule $\{cs()\} => \{cd(X)\}$ is replaced by $\{cs1()\} => \{cd1(X1)\}$ and $\{cs2()\} => \{cd2(X2)\}$. With reference to *GainAccess* from Section IV-A and conditions in Table IV, the enabling pre-condition is the availability of the access sensor. Users provide matching pairs of identifying and authentication elements, and select one out of the accounts they own. The ensured effect is a token to access the account and the enabling of the selection sensor.

Table IV
Conditions for *GainAccess*.

Enabling	$\{axsSnsr()\}$
Requested	$\{authMap(X,V),\ actMap(X,AS)\}$
Provided	$\{axsInfo(X,V), selIdInfo(A)\}$
Constraints	$\{AS.contains(A))\}$
Holding	$\{axs(A),\ selMode(\text{“}C''), selIdInfo(A)\}$
Interface	$\{selMdSnsr(\text{“}C'', \mathtt{true})\}$

We can derive three rules from Table IV. The first, `GetAxsInfo`, directly derives from the *obtains* relation between sensor and data for access information, while `GetSelInfo` derives from the existence of the constraint on being the identified account an authorised one. Since $selIdInfo(acctId)$ has to be provided, the associated sensor resource must be activated. Finally, `CheckAxs`, exploits this information in the context of the other resources.

`GetAxsInfo`: $\{axsSnsr()\} => \{axsInfo(X,V),$
$selSnr(visible = \mathtt{true})\}$
`GetSelInfo`: $\{actMap(X,AS), axsInfo(X,V)\}$?
$\{selIdSnr(visible = \mathtt{true})\} => \{selIdInfo(A)\}$:
$\{AS.contains(A)\}$
`CheckAxs`: $\{selIdInfo(A)\}$? $=>$
$\{axs(A),\ selMdSnsr(\text{“}C'', \mathtt{true}),\ selMode(\text{“}C'')\}$

If **AxsSnsr** is refined into **IdSnsr** and **AuthSnsr**, obtaining customer'sidentity and credentials for authorisation (and **AxsInfo** is acordingly refined) then rule `GetAxsInfo` is replaced by two rules obtaining the separate pieces of information. If a relation (i.e. an instance of the `hasRelation` meta-association) exists between resource types, say $\mathbf{R_1} \triangleright \mathbf{R_2}$, so that producing the second resource requires the existence of the first, then we add a synchronisation token to the $\mathcal{E}$ component of a rule producing a term $r_1()$, to be consumed in, or in the context of, the production of $r_2()$. Then, $\{cs_i()\} => \{cd_i(X_i), csTkn(type = cs_{i+1})\}$ and $\{cd_i(X_i)\}?\{cs_{i+1}(), csTkn(type = cs_{i+1}\} => \{cd_{i+1}(X_{i+1})\}$ obtain one at a time the refined resources.

Similarly, when a relation $\mathbf{R_1} \triangleright \mathbf{R_2}$ derives from conditioning the production of $\mathbf{R_2}$ on a check on attributes of $\mathbf{R_1}$, the check can be reified into a control token enabling the rule which produces $\mathbf{R_2}$. For example, $BA>=V$ in `PerformTransfer` can be checked immediately after obtaining the amount and reified into a resource $suffBal(A)$, so that information about the target is obtained only if the transfer can eventually succeed: `CheckBalance`: $\{selMode(\text{“}T''),$ $amt(V), axs(A),\ act(A,BA)\}$? ! $\{suffBal()\} => \{suffBal()\}$, while `ObtainTarget` would be modified to: $\{amtSnsr(\mathtt{true}), suffBal()\} => \{amt(V)\}$.

Refining a sensor/info pair can result into a collection of independent pairs to be managed together so that progress is allowed only when all the data resource terms have been obtained. This can be achieved by providing a **SubmitSnsr** sensor type, from which a **SubmitTkn** token type, reifying information on global data availability, and an associated rule `ObtainSubmit`: $\{submitSnsr(visible = \mathtt{true})\} => \{submitTkn()\}$ are derived. In this case, a rule would be required to make the sensor visible only when all needed information has been obtained, following a template where the meta-variable $REQUIRED\_INFO$ is to be instantiated to some collection of resource terms forming the $\mathcal{X}$ component.

`EnableSubmitTempl`: $REQUIRED\_INFO$?
$\{submitSnsr(visible = \mathtt{false})\} =>$
$\{submitSnsr(visible = \mathtt{true})\}$

The separation of functional aspects from policy-dependent ones allows forms of reuse through rule refinement. For example, *GainAccess* can be realised differently via an ATM or an Internet banking system. With $axsSnsr$ refined into **idSnsr/authSnsr**, in the ATM case the $idSnsr()$ will accept a badge, i.e. a physical $idElem()$, whose ownership is transferred from the customer to the system, and a relation $\mathbf{idInfo} \triangleright \mathbf{authSnsr}$ is imposed. The badge-accepting rule makes the $authSnsr()$ resource available to rule `ObtainAuthInfo`. In the Internet version, both parts of the information have to be submitted simultaneously. Hence, we require the pair $sbmtSnsr()$ / $sbmtTkn()$, with $\mathbf{IdInfo} \triangleright \mathbf{SbmtTkn}$ and $\mathbf{AuthInfo} \triangleright \mathbf{SbmtTkn}$, and $REQUIRED\_INFO$ instantiated to $\{idInfo(), authInfo\}$. Rules `ObtainSubmitAxs` and `EnableSubmitAxs` are formed, while $sbmtTkn()$ is introduced in the $\mathcal{R}$ component of `CheckAxs`. Analogous considerations hold for the ticketing example, if one wants to obtain separately information about one-way or return trips.

Security concerns lead to further refinements. As an example, `PerformTransfer` only required identification of the source account, but not that access to this account was authorised. If so needed, a security resource $axs(actId)$ can be included in the $\mathcal{E}$ compartment, to ensure that the value of $actId$ be the same as in $selIdInfo(actId)$. Interaction and security policies can also interact: `PerformTransfer` is not concerned with access after performing an operation involving money, but ATM and Internet require specific solutions. For the ATM, $authSnsr()$ must be present after its application, if the $idElem()$ is still owned by the system. In the Internet case, $axs(A)$ is maintained through the whole session, so it appears in the $\mathcal{X}$ component of the rules for each operation.

VI. Derivation of models

As shown in Section IV-A, the partial orderings induced by derivation trees, and extended to rule refinements, define acceptable event flows in behaviour realisation. This draws a bridge with typical task analysis languages, e.g. ConcurTaskTrees, based on LOTOS temporal relations [13]. The *activate-setup-perform* pattern identifies three corresponding high-level abstract tasks. The specific dependencies among resources for these tasks, and progressive refinement of resources and rules, give rise to both a hierarchical decomposition of tasks into subtasks, where a subtask is identified with a rule, until rules are expressed in terms of atomic resources. As an example, Figure 2 represents the two possible developments for *GainAccess* for ATM and Internet.

Adopting UsiXML [15] as a UI specification language, the presence in the behaviour model of specific types of interaction resources constrains the Abstract UI model (AUI) to contain abstract interaction objects presenting specific facets, namely a *control* facet for a sensor obtaining resources of type `Control`, *input* facet for a sensor obtaining `Data` resources and the *output* facet for effectors presenting `Data` resources.

The identification of collections of resources and transactional sub-behaviours can also drive the transformation process from the AUI to the CUI model, based on interaction patterns [16]. For example, at the AUI level, one requires that a collections of sensors for control tokens be made available through components in a single abstract container. When a graphical UI is produced, these will be organised as *CommandArea* or *ButtonGroup* panels, to express the common aspect of the behaviour to which they give access.

The identified types of resources provide a natural basis for constructing domain models for the interactive system, represented as class diagrams in UML or as collections of tags from the domain language of UsiXML. Resource maps would be translated to association classes or to relationships. These can also provide information on objects involved in a task, complementing the CTT description. Moreover, the distinction between the `Data`, `Interaction`, and `Control` resource meta-types can be used to associate stereotypes such as *Entity*,

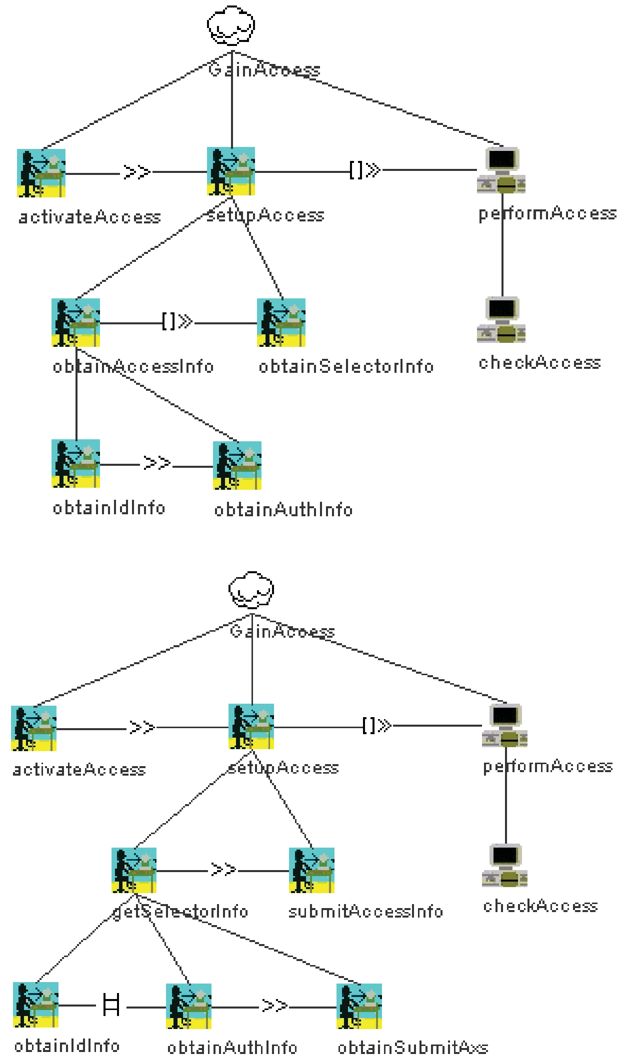


Figure 2. Two CTTs for *GainAccess* in the ATM and Internet cases.

Boundary or *Control* to the resulting domain types, thus facilitating the construction of simplified analysis models.

By analysing relations between use cases one can prepare test scenarios as in [5]. Use cases are independent of one another, except for `extend` or `include` dependencies. However, the execution of a behaviour B_1 can depend on some previous B_2 where $(Hd(B_1) \cup Ui(B_1)) \cap (En(B_2) \cup Rq(B_2)) \neq \emptyset$. If $(En(B_2) \cup Rq(B_2)) \cap (Hd(B_1) \cup Ui(B_1)) = \emptyset$, we lift temporal relations to behaviours and say that B_1 *enables* B_2 ($B_1 >> B_2$). Conversely, if B_1 consumes some resource needed by B_2 we say that B_1 *disables* B_2 ($B_2 [> B_1$). For example, in the ATM version of the banking case, $axs(actId)$ is in $\mathcal{R}$ but not in $\mathcal{E}$ in *Transfer* and *Withdraw*, so it is consumed, but is in $\mathcal{X}$, hence maintained, in *Check*. The following relations result (using initials of use case names).

(**enable**) $>>= \{(\mathrm{G,C}), (\mathrm{G,W}), (\mathrm{G,T}), (\mathrm{G,E})\}$

(**disable**) $[>= \{(\mathrm{T,W}), (\mathrm{W,T}), (\mathrm{C,W}), (\mathrm{C,T}), (\mathrm{W,E}), (\mathrm{T,E}), (\mathrm{E,W}), (\mathrm{E,T})\}$

While W and T behave in the same way with respect to the other cases, they also disable session exiting (of course exiting disables all other cases). To avoid this we can re-introduce selection resources in their Ui components, but with a new modality only allowing one to gain access or exit the session. Suitable rules can then be defined. Admissible sequences are described by the high-level CTT in Figure 3.

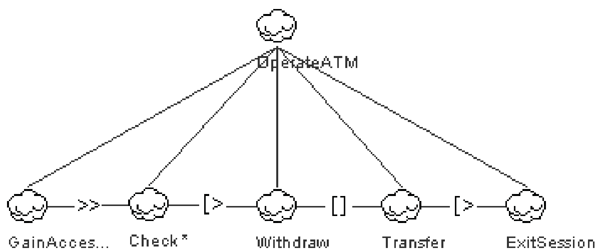


Figure 3. Temporal operators between use cases.

VII. Related Work and Discussion

In [14], two semantics are proposed for use cases and CTT. Both are given as sets of (partially ordered) sets of elementary actions, defining admissible sequences of concurrent steps. Finite state machines model the progress of use cases and task models and verify their consistence [4]. State Transition Networks are derived from CTTs and used to derive dialogue control in [12]. In all these cases, the event flow associated with the use case or the task model is given as an a priori. Our approach allows the derivation of concurrent flows from an analysis of resource dependency.

Giese and Heldal propose a systematic derivation of Statecharts, reasoning backwards from post-conditions for all possible flows of a use case [8]. We support the incremental integration of alternate flows and model refinement and generation of specific interactions from a same functional core.

The resource meta-model could supplement the integration of CTT and Use Case metamodels in [2]. In [9] a M2M transformation is proposed from use case flows to activity diagrams, with decision nodes generated for exceptional flows.

In CRF, abstract and concrete user interfaces are derived from a task model via suitable transformations [15]. The technique is applicable to models derived from our approach and resource refinement can suggest specific patterns.

We specify interactive systems in terms of resources to be produced or consumed in use cases, systematically deriving domain models from the identified types of resources and behavioural specifications in terms of sequences of rule applications. Resources and rules are refined as the analysis progresses, or based on interaction or security policies. Abstract transformations are decomposed into atomic subtasks, represented by resource rewriting rules. The approach provides an intuitive, yet formally founded, specification of use cases, avoiding both over-formalisation and text ambiguities, and separates functional from interaction concerns, as well as from special cases, avoiding premature commitment. Failures are modeled in terms of missing or inadequate resources.

We have focused on behaviours progressing in discrete steps, according to the *activate-setup-perform* pattern. Future work will consider other types of patterns (e.g. from [3]). For example monitoring processes are not to be completed, but can be observed to check the current situation. Finally, we plan to integrate existing tools for use case or task modeling.

References

[1] J.-M. Andreoli. Logic programming with focusing proofs in linear logic. *J. Log. Comput.*, 2(3):297–347, 1992.

[2] R. Bastide. An integration of task and use-case metamodels. In *Proc. HCI (1)*, volume 5610 of *LNCS*, pages 579–586. Springer, 2009.

[3] R. Biddle, J. Noble, and E. D. Tempero. Essential use cases and responsibility in object-oriented development. In *Proc. 25th ACSC*, pages 7–16, 2002.

[4] B. Bomsdorf and D. Sinnig. Model-based specification and validation of user interface requirements. In *Proc. HCI International (1)*, volume 5610 of *LNCS*, pages 587–596. Springer, 2009.

[5] L. Briand and Y. Labiche. A UML-based approach to system testing. *SoSyM*, 1:10–42, 2002.

[6] G. Calvary, J. Coutaz, D. Thevenin, Q. Limbourg, L. Bouillon, and J. Vanderdonckt. A unifying reference framework for multi-target user interfaces. *Interacting with Computers*, 15(3):289–308, 2003.

[7] L. L. Constantine and L. A. D. Lockwood. *Software for use: a practical guide to the models and methods of usage-centered design*. Addison-Wesley, 1999.

[8] M. Giese and R. Heldal. From informal to formal specifications in UML. In *Proc. UML 2004*, volume 3273 of *LNCS*, pages 197–211. Springer, 2004.

[9] J. Gutiérrez, C. Nebut, M. Escalona, M. Mejas, and I. Ramos. Visualization of use cases through automatically generated activity diagrams. In *Proc. Models 2008*, volume 5301 of *LNCS*, pages 83–96. Springer, 2008.

[10] I. Jacobson, G. Booch, and J. Rumbaugh. *The Unified Software Development Process*. Addison-Wesley, 1999.

[11] I. Jacobson, M. Christerson, P. Jonsson, and G. Övergaard. *Object-oriented software engineering - a use case driven approach*. Addison-Wesley, 1992.

[12] K. Luyten, T. Clerckx, K. Coninx, and J. Vanderdonckt. Derivation of a dialog model from a task model by activity chain extraction. In *Proc. DSV-IS 2003*, volume 2844 of *LNCS*, pages 203–217. Springer, 2003.

[13] G. Mori, F. Paternò, and C. Santoro. CTTE: Support for developing and analyzing task models for interactive system design. *IEEE TSE*, 28(8):797–813, 2002.

[14] D. Sinnig, P. Chalin, and F. Khendek. Common semantics for use cases and task models. In *Proc. IFM 2007*, volume 4591 of *LNCS*, pages 579–598. Springer, 2007.

[15] A. Stanciulescu, Q. Limbourg, J. Vanderdonckt, B. Michotte, and F. Montero. A transformational approach for multimodal Web user interfaces based on UsiXML. In *Proc. ICMI 2005*, pages 259–266. ACM, 2005.

[16] J. Tidwell. *Designing Interfaces*. O'Reilly, 2006.

Formalizing Concept Diagrams

Gem Stapleton, John Howse, Peter Chapman, Aidan Delaney, Jim Burton
University of Brighton
Email: {g.e.stapleton,john.howse,p.b.chapman,a.j.delaney,j.burton}@brighton.ac.uk

Ian Oliver
Navteq Europe
Email: ian.oliver@nokia.com

***Abstract*—A range of diagrammatic logics have been developed, such as Venn-II and existential graphs, and much is now known about their expressive limitations. For instance, Shin's Venn-II has the expressive power of monadic first-order logic. There are a family of logics that extend the syntax of Euler diagrams and, excluding concept diagrams, the most expressive of these is believed to be generalized constraint diagrams. These diagrams are second-order in that, under some circumstances, they allow existential quantification over subsets of the universe. In this paper, we formalize the syntax and semantics of concept diagrams. In addition, we present a series of expressiveness results, comparing concept diagrams with predicate logic. In particular, we establish that the concept diagram logic is equivalent to second-order predicate logic where we can quantify over sets and binary relations, so concept diagrams are highly expressive compared with other diagrammatic logics. We further demonstrate that some fragments of concept diagrams are at least as expressive as the corresponding fragment of predicate logic.**

I. Introduction

Typically, the diagrammatic logics that have been developed to date are somewhat limited in expressiveness. Many of them are restricted to a fragment of monadic first-order logic [1], [2], [3], [4], [5], where only one place predicate symbols are permitted. Some diagrammatic logics include two-place predicates symbols, such as constraint diagrams [6] which are equivalent to some unknown fragment of dyadic first-order logic [7]. Generalized constraint diagrams were proposed which include very limited quantification over subsets of the universal set, so they are second-order but their exact expressiveness is unknown [8].

Here, we focus on concept diagrams [9] which have been designed to formally specify ontologies in a visual manner. By contrast to constraint diagrams, concept diagrams are very expressive and do not have any reading order issues concerning quanitifiers, as we establish in this paper. Moreover, as demonstrated [10], they overcome various usability problems associated with constraint diagrams. Additionally, they increase expressiveness over generalized constraint diagrams by allowing quantification over subsets (in an unlimited way) and binary relations. In terms of practical application, concept diagrams have been designed for visually modelling ontologies; see [11] for examples of their use.

This paper has two main contributions: (a) it defines a formal syntax and semantics of concept diagrams, and (b) it establishes that concept diagrams are as expressive as second-order predicate logic where we can quantify over sets and binary relations and we discuss how various fragments coincide (or not), extending [12]. Our main expressiveness result demonstrates that concept diagrams significantly advance the state-of-the-art in terms of what can be expressed using diagrammatic logics. From an ontology specification perspective, the main contributions of the paper demonstrate that concept diagrams can be used as a rigorous visual alternative to symbolic notations, such as description logics [13], and are capable of specifying reasonably complex ontologies, beyond what was previously possible using diagrams. Section II provides an informal introduction to concept diagrams. In section III we define a vocabulary over which concept diagrams and second-order predicate logic formulae will be constructed. Section IV presents our formalization of concept diagrams. Our main expressiveness result, that concept diagrams are equivalent to dyadic second-order predicate logic, is given in section V. In addition, we identify when natural fragments of each notation coincide with regard to expressiveness. Finally, we conclude in section VI.

II. An Overview of Concept Diagrams

Concept diagrams are based on Euler diagrams which visualize relationships between sets. We refer the reader to [10] for a discussion on the design of concept diagrams. Since they were introduced for specifying ontologies, we use common description logic terminology for the vocabulary over which their syntax is defined: sets will be referred to as concepts and binary relations as roles. Thus, concepts label curves and, as shown below, roles label arrows.

As an example, suppose that the individual Erica is a Person who is marriedTo only the Person Peter and that Erica owns exactly two pets (identified by the role ownsPet), both of which are Dogs, including a Beagle called Minnie. A concept diagram asserting this information uses three closed curves to represent the concepts Person, Dog, and Beagle; see figure 1. The concepts Person and Dog are disjoint and Beagle is subsumed by Dog. The named individuals are represented by labelled dots, with their location telling us of which concepts they are instances; for example, Minnie is located inside the curve labelled Beagle. In general, individuals will be represented by trees to allow uncertain information to be represented, as we shall see in a moment, and distinct trees represent distinct individuals.

The fact that Erica owns a set of Pets is visualized by the use of the arrow, labelled ownsPet, hitting an unlabelled curve. This unlabelled curve is drawn inside Dog, to assert that the image of the relation ownsPet, when its domain is restricted to Erica, is subsumed by Dog. The fact that this unlabelled curve contains two trees tells us that Erica owns two Dogs. Erica's dog that is not called Minnie could be either a Beagle or not a Beagle. This uncertainty is captured by the use of the (unlabelled) tree with two nodes, one

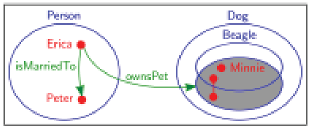


Fig. 1. Visualizing role information.

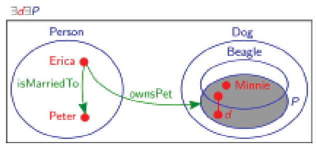


Fig. 2. Formal version of figure 1.

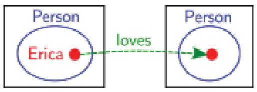


Fig. 3. Arrows between boxes.

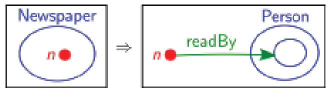


Fig. 4. Free variables.

inside both the Dog and Beagle curves (for the case when this dog is a Beagle) and the other inside the Dog curve but outside the Beagle curve (for when this dog is not a Beagle). Finally, the shading is used to assert that the only dogs owned by Erica are those represented by the trees. That is the shading provides the *exactly* information: in a shaded region, all elements must be represented by trees. Strictly speaking, the unlabelled curve and the unlabelled tree in this diagram are acting as existentially quantified variables; this will be reflected in our formal concept diagram syntax. Here, the omitted existential quantifiers and variable names did not introduce ambiguity, which allowed the diagram to be less cluttered. The formal version with the quantifiers and variables can be seen in figure 2.

The arrows used in figure 1 convey information about the image of a role under a domain restriction, such as Erica isMarriedTo Peter and only Peter, that is $im(marriedTo|_{\{Erica\}}) = \{Peter\}$, where $im(r)$ denotes the image of the relation r and we use $r|_S$ to mean r under the domain restriction to the set S. We use dashed arrows to represent partial information, such as Erica loves some Person; Erica may love many things, but all we know that she loves at least one Person. Further, we do not know whether the Person Erica loves is distinct from herself. A concept diagram expressing this information can be seen in figure 3. Of note is that the arrow connects diagrammatic syntax placed in different boxes. This is done to ensure that we have not made any assertion about whether the Person Erica loves is distinct from Erica. If the arrow had instead hit the curve labelled person, then this would assert that Erica loved all people, of which there is at least one.

A diagram asserting that every Newspaper is readBy only a subset of Person can be seen in figure 4. Here, the labelled dot inside the curve labelled Newspaper is a free variable, and so is equivalent to a universally quantified variable. The diagram expresses that if *n* is a Newspaper then the set of things *n* is read by is subsumed by Person.

III. A Common Vocabulary

As is typical, formulae (concept diagrams or predicate logic statements) are formed over some specified vocabulary:

Definition 1: We have the following sets:

1) $\mathcal{I}$ is a set of **individual names**.
2) $\mathcal{VI}$ is a countably infinite set of **individual variables**.
3) $\mathcal{C}$ is a set of **concepts**.
4) $\mathcal{VC}$ is a countably infinite set of **concept variables**.
5) $\mathcal{R}$ is a set of **roles**.
6) $\mathcal{VR}$ is a countably infinite set of **role variables**.

Following the standard definitions of predicate logic formulae, individual names, concepts and roles correspond to constants, predicates of arity 1, and predicates of arity 2, respectively. The variables of each type similarly correspond to predicate logic syntax.

The set, $\mathcal{R}$, of roles gives rise to a set of *inverse* roles, denoted $\mathcal{R}^-$. In particular, we define $\mathcal{R}^- = \{\mathrm{r}^- : \mathrm{r} \in \mathcal{R}\}$. The set $\mathcal{VR}^-$ is similarly defined. These derived inverse roles thus extend the vocabulary and they will be used as labels on arrows in concept diagrams. We assume that all of the above sets are pairwise disjoint (have no elements in common) and adopt the standard approach of interpreting the elements of these sets over some universe:

Definition 2: An **interpretation** is a pair, $\mathcal{INT} = (U, \cdot^{\mathcal{INT}})$, where

1) U is a non-empty set, called the **universal set**,
2) for each element, I, in $\mathcal{I}$, $\mathrm{I}^{\mathcal{INT}}$ is an element of U,
3) for each element, C, in $\mathcal{C}$, $\mathrm{C}^{\mathcal{INT}}$ is a subset of U,
4) for each role, r, in $\mathcal{R}$, $\mathrm{r}^{\mathcal{INT}}$ is a binary relation on U, and
5) for each inverse role, r^-, in r^-, $(\mathrm{r}^-)^{\mathcal{INT}}$ is the inverse of $\mathrm{r}^{\mathcal{INT}}$.

An **extension** of $\cdot^{\mathcal{INT}}$ to variables ensures

1) for each element, i, in $\mathcal{VI}$, $i^{\mathcal{INT}}$ is an element of U,
2) for each element, C, in $\mathcal{VC}$, $C^{\mathcal{INT}}$ is a subset of U,
3) for each element, r, in $\mathcal{VR}$, $r^{\mathcal{INT}}$ is a binary relation on U, and
4) for each inverse role, r^-, in $\mathcal{VR}^-$, $(r^-)^{\mathcal{INT}}$ is the inverse relation of $r^{\mathcal{INT}}$.

We adopt the notation $\mathcal{INT}[x \mapsto e]$ to mean a function (in particular, an interpretation or an extended interpretation) equal to $\mathcal{INT}$ except x is mapped to e.

IV. Formal Syntax and Semantics

First, we present the abstract syntax for *unitary* diagrams. These diagrams form the building blocks of concept diagrams. In our definition, the trees are called **spiders** [2] and are drawn from a countably infinite set, $\mathcal{S} = \{\sigma_1, \sigma_2, ..., \sigma_n, ...\}$, and the **curves**[1] are drawn from a countably infinite set, $\mathcal{K} = \{\kappa_1, \kappa_2, ..., \kappa_n, ...\}$, where $\mathcal{S}$ and $\mathcal{K}$ are disjoint. The regions that can be described as being inside some (or no) curves and outside the rest of the curves are called *zones*. Arrows can be one of two types, dashed or solid, and can be sourced and targeted on either spiders or curves. Formally, an **arrow** is of the form $(s, l, t, \circ)$ where s and t are spiders or curves, l is in $\mathcal{R} \cup \mathcal{R}^- \cup \mathcal{VR} \cup \mathcal{VR}^-$ and $\circ \in \{\dashrightarrow, \rightarrow\}$. Shading can be placed in any region in the diagram. Whilst not illustrated in our previous examples, we also use $=$ to assert that two spiders represent the same individual. We adopt the convention that drawn diagrams will be given names like, for example, d which will be denoted by δ at the abstract level.

[1]Note that these are not curves in a mathematical sense, but just abstract names that correspond to actual curves in a drawn diagram.

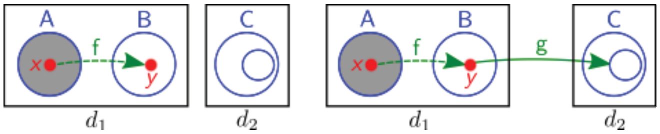


Fig. 5. Unitary and basic concept diagrams.

Definition 3: A **unitary diagram** is a tuple

$$\delta = (\Sigma, K, \lambda, Z, Z^*, \eta, \tau, A)$$

whose components are defined as follows.

1) $\Sigma = \Sigma(\delta) \subset \mathcal{S}$ is a finite set of *spiders*.
2) $K = K(\delta) \subset \mathcal{K}$ is a finite set of *curves*.
3) $\lambda = \lambda_\delta \colon \Sigma \cup K \to \mathcal{I} \cup \mathcal{VI} \cup \mathcal{C} \cup \mathcal{VC}$ is an injective function such that
   a) λ labels each spider, σ, with an element of $\mathcal{I} \cup \mathcal{VI}$, and
   b) λ labels each curve, κ, with an element of $\mathcal{C} \cup \mathcal{VC}$.
4) $Z = Z(\delta)$ is a set of **zones** such that
$$Z \subseteq \{(in, K - in) : in \subseteq K\}.$$
5) $Z^* = Z^*(\delta) \subseteq Z$ is a set of **shaded zones**.
6) $\eta = \eta_\delta \colon \Sigma \to \mathbb{P}Z - \{\emptyset\}$ is a function that returns the **location** of each spider.
7) $\tau = \tau_\delta$ is a reflexive, symmetric relation on Σ that identifies whether two spiders are joined by an equals sign; $(\sigma_1, \sigma_2) \in \tau$ means σ_1 is joined to σ_2.
8) $A = A(\delta)$ is a finite set of arrows such that for all $(s, l, t, \circ)$ in A, s and t are in $\Sigma \cup K$.

In figure 5, d_1 has abstract syntax δ_1 as follows: $\Sigma(\delta_1) = \{\sigma_1, \sigma_2\}$ (there are two spiders); $K(\delta_1) = \{\kappa_1, \kappa_2\}$ (there are two curves); $\lambda_{\delta_1}(\sigma_1) = x$, $\lambda_{\delta_1}(\sigma_2) = y$, $\lambda_{\delta_1}(\kappa_1) = A$ and $\lambda_{\delta_1}(\kappa_2) = B$; $Z(\delta_1) = \{(\emptyset, \{\kappa_1, \kappa_2\}), (\{\kappa_1\}, \{\kappa_2\}), (\{\kappa_2\}, \{\kappa_1\})\}$ (there is no zone inside both κ_1 and κ_2); $Z^*(\delta_1) = \{(\{\kappa_1\}, \{\kappa_2\})\}$ (the zone inside κ_1 but outside κ_2 is shaded); $\eta_{\delta_1}(\sigma_1) = \{(\{\kappa_1\}, \{\kappa_2\})\}$, $\eta_{\delta_1}(\sigma_2) = \{(\{\kappa_2\}, \{\kappa_1\})\}$ (the spider σ_1 is in the zone inside κ_1 but outside κ_2; the spider σ_2 is in the zone inside κ_2 but outside κ_1); $\tau_{\delta_1} = \{(\sigma_1, \sigma_1), (\sigma_2, \sigma_2)\}$ (note that these occurrences of $=$ are not represented in the diagram since their presence would lead to unnecessary clutter: they would just assert, for example, $x = x$); $A(\delta_1) = \{(\sigma_1, f, \sigma_2, \dashrightarrow)\}$ (the arrow is sourced on σ_1, labelled f, targeted on σ_2, and is dashed).

The diagram d_2 has abstract syntax δ_2, where: $\Sigma(\delta_2) = \emptyset$ (there are no spiders); $K(\delta_2) = \{\kappa_3, \kappa_4\}$ (there are two curves); $\lambda_{\delta_2}(\kappa_3) = C$, $\lambda_{\delta_2}(\kappa_4) = X$, where X is the 'hidden' label used on the unlabelled curve; $Z(\delta_2) = \{(\emptyset, \{\kappa_3, \kappa_4\}), (\{\kappa_3\}, \{\kappa_4\}), (\{\kappa_3, \kappa_4\}, \emptyset)\}$ (there is no zone inside κ_4 but outside κ_3); $Z^*(\delta_2) = \emptyset$ (there are no shaded zones); $\eta_{\delta_2} = \emptyset$ (because there are no spiders); $\tau_{\delta_2} = \emptyset$ (because there are no spiders); $A(\delta_2) = \emptyset$ (there are no arrows).

Unitary diagrams form the building blocks of concept diagrams. As in figure 5, we can place arrows between their syntactic components, giving basic concept diagrams:

Definition 4: A **basic concept diagram** is a tuple, $\Delta = (\mathcal{D}, \mathcal{A})$, where

1) $\mathcal{D}$ is a finite set of unitary diagrams such that for any pair of distinct unitary diagrams, δ_1 and δ_2, in $\mathcal{D}$
   a) $\Sigma(\delta_1) \cap \Sigma(\delta_2) = \emptyset$, and
   b) $K(\delta_1) \cap K(\delta_2) = \emptyset$.
2) $\mathcal{A} = \mathcal{A}(\Delta)$ is a finite set of arrows such that for all $(s, l, t, \circ)$ in $\mathcal{A}$, $s, t \in \Sigma(\Delta) \cup K(\Delta)$ where
   a) $\Sigma(\Delta) = \bigcup_{\delta \in \mathcal{D}} \Sigma(\delta)$, and
   b) $K(\Delta) = \bigcup_{\delta \in \mathcal{D}} K(\delta)$, and

   for all unitary diagrams, δ, in $\mathcal{D}$ it is not the case that $s \in \Sigma(\delta) \cup K(\delta)$ and $t \in \Sigma(\delta) \cup K(\delta)$.

The basic concept diagram in figure 5 has abstract syntax $(\mathcal{D}, \mathcal{A})$ where $\mathcal{D} = \{\delta_1, \delta_2\}$, assuming that δ_1 and δ_2 are the abstract syntaxes of the unitary diagram d_1 and d_2, respectively, and $\mathcal{A} = \{(\sigma_2, g, \kappa_4, \rightarrow)\}$. In addition, each of d_1 and d_2 are basic concept diagrams. For instance, d_1 is $(\{\delta_1\}, \emptyset)$.

We saw, in figure 4, $\Rightarrow$ can be used as a logical connective in concept diagrams. We incorporate all standard logical connectives, as well as quantifiers, as seen in figure 2.

Definition 5: A **concept diagram** is defined as follows. Any basic concept diagram is a concept diagram. Given a concept diagram, δ, $\neg\delta$ is a concept diagram. Given concept diagrams δ_1 and δ_2, we also have $(\delta_1 \wedge \delta_2)$, $(\delta_1 \vee \delta_2)$, $(\delta_1 \Rightarrow \delta_2)$ and $(\delta_1 \Leftrightarrow \delta_2)$ as concept diagrams. Given a concept diagram, δ, and a variable, v, chosen from $\mathcal{VI} \cup \mathcal{VC} \cup \mathcal{VR}$, $\exists v\, \delta$, and $\forall v\, \delta$ are concept diagrams.

Turning our attention to the semantics, the meaning of a unitary concept diagram is determined by how its individual pieces of syntax are related to each other. These relationships give rise to a set of constraints that describe the class of interpretations that are its models. Returning to figure 5, d_1 informally expresses the following:

1) The union of the sets represented by its zones is U:
$$\big((U{-}A){\cap}(U{-}B)\big){\cup}\big(A{\cap}(U{-}B)\big){\cup}\big(B{\cap}(U{-}A)\big) = U.$$
2) The shaded zone contains only elements represented by spiders: $A \cap (U - B) \subseteq \{x, y\}$.
3) The spiders, σ_1 and σ_2, (strictly, spider labels x and y) represent elements in the set represented by their locations: $x \in A \cap (U - B) \wedge y \in B \cap (U - A)$.
4) The two spiders represent distinct elements, since they are not joined by $=$: $x \neq y$.
5) The only arrow, $(\sigma_1, f, \sigma_2, \dashrightarrow)$, tells us that the element represented by σ_1 is related to at least the element represented by σ_2 under f: $im(f|_{\{x\}}) \supseteq \{y\}$.

To define the semantics, we start by translating a unitary diagram into a set of *semantic conditions*, as just illustrated informally.

Definition 6: Let δ be a unitary diagram and let $\mathcal{INT} = (U, \cdot^{\mathcal{INT}})$ be an extended interpretation. The **semantic condition** for δ given $\mathcal{INT}$, denoted $sc(\delta, \mathcal{INT})$ is the conjunction of the following conditions:

1) **The Curves Condition** The union of the sets represented by the zones is equal to U: $\bigcup_{(in,out)\in Z(\delta)} (in, out)^{\mathcal{INT}} = U$ where

$$(in, out)^{\mathcal{INT}} = \bigcap_{\kappa\in in} \lambda(\kappa)^{\mathcal{INT}} \cap \bigcap_{\kappa\in out} (U-\lambda(\kappa)^{\mathcal{INT}}).$$

2) **The Shading Condition** Every shaded zone contains only elements represented by spiders:

$$\bigwedge_{(in,out)\in Z^*(\delta)} (in, out)^{\mathcal{INT}} \subseteq \{\lambda(\sigma)^{\mathcal{INT}} : \sigma \in \Sigma\}.$$

3) **The Spiders' Location Condition** Each spider, σ, represents an element that lies in one of sets represented by the zones in its location:

$$\bigwedge_{\sigma\in\Sigma(\delta)} \lambda(\sigma)^{\mathcal{INT}} \in \bigcup_{(in,out)\in\eta_\delta(\sigma)} (in, out)^{\mathcal{INT}}.$$

4) **The Spiders' Distinctness Condition** Any two distinct spiders, σ_1 and σ_2, represent the same element if and only if they are joined by $=$, that is:

$$\bigwedge_{(\sigma_1,\sigma_2)\in\tau_\delta} \lambda(\sigma_1)^{\mathcal{INT}} = \lambda(\sigma_2)^{\mathcal{INT}} \quad \text{and}$$

$$\bigwedge_{(\sigma_1,\sigma_2)\in(\Sigma(\delta)\times\Sigma(\delta))-\tau_\delta} \lambda(\sigma_1)^{\mathcal{INT}} \neq \lambda(\sigma_2)^{\mathcal{INT}}.$$

5) **The Arrows Condition** For each arrow with source s, label l and target t, the image of $l^{\mathcal{INT}}$ under the domain restriction to $\lambda(s)^{\mathcal{INT}}$ equals $\lambda(t)^{\mathcal{INT}}$, if it is of type $\rightarrow$, otherwise it is a superset of $\lambda(t)^{\mathcal{INT}}$:

$$\bigwedge_{(l,s,t,\rightarrow)\in A(\delta)} im(l^{\mathcal{INT}}|_{\lambda(s)^{\mathcal{INT}}}) = \lambda(t)^{\mathcal{INT}} \quad \text{and}$$

$$\bigwedge_{(l,s,t,\dashrightarrow)\in A(\delta)} im(l^{\mathcal{INT}}|_{\lambda(s)^{\mathcal{INT}}}) \supseteq \lambda(t)^{\mathcal{INT}}.$$

The curves condition for d_2 in figure 5 is

$$\big((U-\lambda(\kappa_3)^{\mathcal{INT}})\cap(U-\lambda(\kappa_4)^{\mathcal{INT}})\big)\cup\big(\lambda(\kappa_3)^{\mathcal{INT}}\cap(U-\lambda(\kappa_4)^{\mathcal{INT}})\big) \cup \big(\lambda(\kappa_4)^{\mathcal{INT}} \cap (U - \lambda(\kappa_3)^{\mathcal{INT}})\big) = U.$$

Now, given an interpretation $\mathcal{INT} = (U, \cdot^{\mathcal{INT}})$, this curves condition will be true under some extensions of $\mathcal{INT}$ but not under others. Since d_2 does not contain any other syntax, the remaining semantic conditions are trivially true under any extension of $\mathcal{INT}$. The semantic conditions for unitary diagrams form the basis of semantic conditions for basic concept diagrams:

Definition 7: Let $\Delta = (\mathcal{D}, \mathcal{A})$ be a basic concept diagram. The **semantic condition** for Δ given $\mathcal{INT}$, denoted $sc(\Delta, \mathcal{INT})$ is the conjunction of the semantic conditions for the unitary diagrams, δ, in $\mathcal{D}$ together with the arrows condition for $\mathcal{A}$ (i.e. the same as the arrows condition for a unitary diagram, but over $\mathcal{A}$).

In figure 5, the basic concept diagram $(\{\delta_1, \delta_2\}, \{(\sigma_2, g, \kappa_4, \rightarrow)\})$ has semantic condition:

$$sc(\delta_1) \wedge sc(\delta_2) \wedge im(g^{\mathcal{INT}}|_{\lambda(\sigma_2)^{\mathcal{INT}}}) = \lambda(\kappa_4)^{\mathcal{INT}}.$$

Given an extended interpretation, we can identify whether the semantic condition for a basic concept diagram Δ is true. However, in the more general case we have non-basic diagrams which can involve quantification. In these cases, to determine truth within an extension we need to be able to adjust the extension appropriately. For example, if we are to show that a diagram of the form $\exists x\Delta$ is satisfied by $\mathcal{INT}$ then we need to be able to change $x^{\mathcal{INT}}$, so that x maps to something that makes $sc(\Delta, \mathcal{INT})$ true. We now proceed to define a notion of truth in an extended interpretation, which forms the basis of a definition of satisfaction in an interpretation.

Definition 8: Let $\mathcal{INT} = (U, \cdot^{\mathcal{INT}})$ be an extended interpretation. Let Δ be a concept diagram. We say that Δ is **true** under $\mathcal{INT}$ in the following circumstances:

1) If Δ is a basic concept diagram then if $sc(\Delta, \mathcal{INT})$ is true under $\mathcal{INT}$ then Δ is true under $\mathcal{INT}$.
2) If $\Delta = \neg\Delta_1$ then Δ is true under $\mathcal{INT}$ whenever Δ_1 is not true under $\mathcal{INT}$.
3) If $\Delta = \Delta_1 \wedge \Delta_2$ then Δ is true under $\mathcal{INT}$ whenever Δ_1 and Δ_2 are both true under $\mathcal{INT}$.
4) If $\Delta = \Delta_1 \vee \Delta_2$ then Δ is true under $\mathcal{INT}$ whenever Δ_1 or Δ_2 is true under $\mathcal{INT}$.
5) If $\Delta = \Delta_1 \Rightarrow \Delta_2$ then Δ is true under $\mathcal{INT}$ whenever Δ_1 is not true under $\mathcal{INT}$ or Δ_2 is true under $\mathcal{INT}$.
6) If $\Delta = \Delta_1 \Leftrightarrow \Delta_2$ then Δ is true under $\mathcal{INT}$ whenever Δ_1 and Δ_2 are both true under $\mathcal{INT}$ or both not true under $\mathcal{INT}$.
7) If $\Delta = \exists x\Delta_1$ then Δ is true under $\mathcal{INT}$ whenever there exists e such that Δ_1 is true under $\mathcal{INT}[x \mapsto e]$.
8) If $\Delta = \forall x\Delta_1$ then Δ is true under $\mathcal{INT}$ whenever for all appropriate e (e.g. if x is in $\mathcal{VI}$ then e is an element of U), Δ_1 is true under $\mathcal{INT}[x \mapsto e]$.

Definition 9: Let $\mathcal{INT} = (U, \cdot^{\mathcal{INT}})$ be a interpretation. Let Δ be a concept diagram. The diagram Δ is **satisfied** by $\mathcal{INT}$ whenever Δ is true under all extensions of $\mathcal{INT}$. If $\mathcal{INT}$ satisfies Δ then $\mathcal{INT}$ is a **model** for Δ. In addition, Δ **axiomatizes** the set of interpretations that are its models.

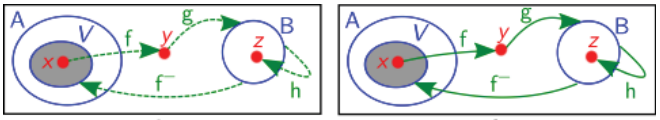
V. Expressiveness

The contribution of this section is to compare the expressiveness of concept diagrams with predicate logic. Predicate logic formulae are defined in the standard way, as is the notion of satisfaction and a model, over the vocabulary defined in section III. The specific logic that we consider is as follows:

Definition 10: We define **terms** and **formulae** as follows:

1) Terms: if x and y are elements of $\mathcal{I} \cup \mathcal{VI}$, $x = y$ is a term.
2) Formulae:
   a) every term is a formula,
   b) if x is an element of $\mathcal{I} \cup \mathcal{VI}$ and c is an element of $\mathcal{C} \cup \mathcal{VC}$ then $c(x)$ is a formula,
   c) if x and y are elements of $\mathcal{I}\cup\mathcal{VI}$ and r is an element of $\mathcal{R}\cup\mathcal{VR}$ then $r(x, y)$ is a formula, and
   d) if x is an element of $\mathcal{VI} \cup \mathcal{VC} \cup \mathcal{VR}$ and f and g are formulae then so too are $\neg f$, $f \wedge g$, $f \vee g$, $f \Rightarrow g$, $f \Leftrightarrow g$, $\forall xf$, and $\exists xf$.

Note that the above definition does not allow the construction of formulae of the form $c(s)$ and $r(s, t)$ where s and



d_1 d_2

Fig. 6. Translating concept diagrams to second-order predicate logic.

t are second-order variables. The standard definition of the semantics of predicate logic formulae is essentially given by a re-write of definition 9, replacing diagrams with formulae, with case 1 being replaced by the obvious definition of truth in an interpretation for the atomic formulae.

In order to allow us to compare succinctly the expressiveness of fragments of each logic, we will write $\mathcal{CD}[X]$ to mean the fragment of concept diagrams defined over the vocabulary X; $\mathcal{PL}[X]$ is similarly defined for predicate logic fragments. The full fragment of concept diagrams is denoted $\mathcal{CD}[\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{VC}, \mathcal{R}, \mathcal{VR}]$ and, for example, the dyadic first-order fragment of predicate logic will be denoted by $\mathcal{PL}[\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{R}]$; since $\mathcal{VC}$ and $\mathcal{VR}$ are not specified in $[\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{R}]$, this means that formulae are constructed without using any of their elements thus making it first-order. Excluding $\mathcal{VC}$ and $\mathcal{VR}$ reduces expressiveness since we can no longer define properties like being even. We will show that $\mathcal{CD}[\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{VC}, \mathcal{R}, \mathcal{VR}]$ is equivalent in expressiveness to $\mathcal{PL}[\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{VC}, \mathcal{R}, \mathcal{VR}]$ and start by giving a strict definition of equivalence in expressive power:

Definition 11: Given two fragments, F_1 and F_2, we say that F_1 and F_2 have **the same expressive power**, denoted $F_1 \equiv F_2$, whenever they can axiomatize the same classes of interpretations. If F_1 can axiomatize all of the classes of interpretations that F_2 can axiomatize then F_1 has **at least the expressive power** of F_2, denoted $F_1 \geq F_2$.

We will establish, for all $X \subseteq \{\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{VC}, \mathcal{R}, \mathcal{VR}\}$, that $CD[X] \equiv PL[X]$ or $CD[X] \geq PL[X]$ and, in particular, that the two 'full' fragments have the same expressiveness (theorem 4). To illustrate the idea, we will use the two concept diagrams in figure 6. Each of these diagrams represents the various combinations of sources and targets for arrows (i.e. the four combinations arising from spiders and curves being used as sources and targets). The translation is directly informed by the semantic condition for the diagram. We will translate each of the conditions that give rise to the semantic condition in a systematic way. First, we translate d_1:

1) The translated curves condition is:

$$\forall i \quad \big((\neg A(i) \wedge \neg B(i) \wedge \neg V(i)) \vee (A(i) \wedge \neg B(i) \wedge \neg V(i)) \vee (A(i) \wedge \neg B(i) \wedge V(i)) \vee (\neg A(i) \wedge B(i) \wedge \neg V(i))\big);$$

each of the disjuncts arises from a zone (e.g. $(A(i) \wedge \neg B(i) \wedge \neg V(i))$ comes from the zone – blurring the distinction between curves and their labels – $(\{A\}, \{B, V\})$. Similarly to the curves condition telling us that all elements must lie in the set represented by a zone, the translated condition tells us that all elements must 'lie in' one of the disjuncts.

2) The translated shading condition is:

$$\forall i((A(i) \wedge \neg B(i) \wedge V(i)) \Rightarrow (i = x \vee i = y \vee i = z));$$

the single shaded zone gives rise to the antecedent and, since we know that elements in the sets represented by shaded zones must be represented by the spiders, the consequent tells us that i must equal one of these elements, too.

3) The translated spiders' location condition is:

$$A(x) \wedge \neg B(x) \wedge V(x) \wedge \neg A(y) \wedge \neg B(y) \wedge \neg V(y) \wedge \neg A(z) \wedge B(z) \wedge \neg V(z);$$

each spider represents an element in the set represented by the region in which it is placed, in the case of x we know that x is, informally, in $A \cap (U - B) \cap V$, corresponding to the formula $A(x) \wedge \neg B(x) \wedge V(x)$.

4) The translated spiders' distinctness condition is:

$$x \neq y \wedge x \neq z \wedge y \neq z$$

since no two spiders represent the same element.

5) The translated arrows condition is the conjunction of the translations of the four arrows (blurring the distinction between labels and the items they are labelling in d_1):
   a) $(x, f, y, \dashrightarrow)$ translates to $f(x, y)$ since this dashed arrow only tells us that x is related to (at least) y under f.
   b) $(y, g, B, \dashrightarrow)$ translates to
   $$\forall i\big(B(i) \Rightarrow g(y, i)\big)$$
   since this dashed arrow tells us that y is related to at least the elements in B (equivalently, everything in B is related to by y) under g.
   c) $(B, h, y, \dashrightarrow)$ translates to
   $$\exists i\big(B(i) \wedge h(i, z)\big)$$
   since this dashed arrow tells us that at least one element in B is related to z
   d) $(B, f^-, V, \dashrightarrow)$ translates to
   $$\forall i\big(V(i) \Rightarrow \exists j(B(j) \wedge f(i, j))\big)$$
   since this dashed arrow tells us that every element in V is related to by at least one element in B under f^-.

Diagram d_2 has a similar translation except for the arrows: they are dashed rather than solid:

(a) $(x, f, y, \rightarrow)$ translates to:
$$f(x, y) \wedge \forall i\big(f(x, i) \Rightarrow i = y\big)$$
since this solid arrow tells us that x is related to exactly y under f.

(b) $(y, g, B, \rightarrow)$ translates to:
$$\forall i\big(B(i) \Rightarrow g(y, i)\big) \wedge \forall i\big(g(y, i) \Rightarrow B(i)\big)$$
since this solid arrow tells us that y is related to all and only the elements in B under g.

(c) $(B, h, y, \rightarrow)$ translates to:
$$\exists i\big(B(i) \wedge h(i, z)\big) \wedge \forall i \forall j\big((B(i) \wedge h(i, j)) \Rightarrow j = z\big)$$
since this solid arrow tells us that at least one element in B is related to z and elements in B are not related to any other elements.

(d) $(B, f^-, V, \rightarrow)$ translates to

$$\forall i\Big(V(i) \Rightarrow \exists j\big(B(j) \wedge f(i,j)\big)\Big) \wedge \forall i \forall j\big((B(i) \wedge f(j,i)) \Rightarrow V(j)\big)$$

since this solid arrow tells us that every element in V is related to by at least one element in B under f^- and, in addition, elements in B are only related to elements in V under f^-.

Here, d_2 is inconsistent: x is related to only one element, y, under f which is not in B but the arrow labelled f^- implies that x is related to at least one element in B.

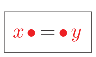
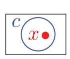
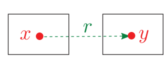
General translations of unitary diagrams to second-order predicate logic can be defined along the lines of the above example and they easily extend to basic concept diagrams. Therefore, all concept diagrams can be translated, extending the translation in the obvious inductive way. Importantly, it can be shown that the translation of any given concept diagram preserves semantics, thus establishing the following result:

Theorem 1: Let $\Delta = (\mathcal{D}, \mathcal{A})$ be a basic concept diagram. If Δ axiomatizes class of interpretations $\mathcal{M}$ then there exists a formula F that also axiomatizes $\mathcal{M}$.

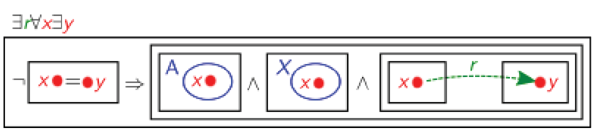
Therefore, concept diagrams are at most as expressive as dyadic second-order logic, justified by a simple induction argument with the base case provided by theorem 1:

Theorem 2: $\mathcal{CD}[\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{VC}, \mathcal{R}, \mathcal{VR}] \leq \mathcal{PL}[\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{VC}, \mathcal{R}, \mathcal{VR}]$.

Our attention now turns to providing translations from formulae of predicate logic to concept diagrams. The translation for atomic formulae is specified by[2]

$(x = y) \mapsto$  $c(x) \mapsto$  $r(x, y) \mapsto$ 

The translation just defined extends in the obvious inductive way to non-atomic formulae. We illustrate the idea by considering: $\exists r \forall x \exists y\Big(\neg(x = y) \Rightarrow \big(A(x) \wedge X(x) \wedge r(x, y)\big)\Big)$ which translates to:



Theorem 3: $\mathcal{PL}[\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{VC}, \mathcal{R}, \mathcal{VR}] \leq \mathcal{CD}[\mathcal{I}, \mathcal{VI}, \mathcal{C}, \mathcal{VC}, \mathcal{R}, \mathcal{VR}]$.

Theorems 2 and 3 thus combine to give us theorem 4:

Theorem 4: The two full fragments are equivalent in expressive power.

We observe that the translation from a basic concept diagram, Δ, to predicate logic involves the introduction of variables, from $\mathcal{VI}$, that were not present in Δ. However, this is the only time an item from the vocabulary gets introduced in the translation process. Following a similar style to the arguments given above, we therefore have the following theorems:

Theorem 5: $\mathcal{CD}[X] \equiv \mathcal{PL}[X]$ provided X includes $\mathcal{VI}$.

Theorem 6: $\mathcal{CD}[X] \geq \mathcal{PL}[X]$ provided X does not include $\mathcal{VI}$.

[2] Strictly, the translation returns the abstract syntaxes of the diagrams.

VI. Conclusion

We have provided a formalization of concept diagrams and established that they have the same expressive power as the defined fragment of second-order logic. This result demonstrates that, in so far as diagrammatic logics are concerned, concept diagrams are highly expressive. In particular, concept diagrams are more expressive than any previously formalized diagrammatic logic that uses Euler diagrams as a basis. The approach we have taken to establish our expressiveness result is constructive, providing translations between the two logics that preserve semantics. Translations between description logics and predicate logic have already been established [15]. Thus, using these previously existing translations, we are able to translate between description logics and concept diagrams. In terms of applications, description logics are used for ontology specification. Thus, our work now allows concept diagrams to be used as a formal, visual alternative to symbolic logics, such as description logics, for ontology specification. A key part of our future work is to establish which fragments of concept diagrams correspond to description logics that have efficiently decidable inference problems, which is important in the area of ontology specification.

References

[1] E. Hammer, *Logic and Visual Information*. CSLI Publications, 1995.

[2] J. Howse, G. Stapleton, and J. Taylor., "Spider diagrams," *LMS J. of Computation and Mathematics*, 8:145–194, 2005.

[3] K. Mineshima, M. Okada, Y. Sato, and R. Taakemura, "Diagrammatic reasoning system with Euler circles: Theory and experiment design," in *Diagrams*, Springer, 2008, pp. 188–205.

[4] S.-J. Shin, *The Logical Status of Diagrams*. CUP, 1994.

[5] N. Swoboda and G. Allwein, "Using DAG transformations to verify Euler/Venn homogeneous and Euler/Venn FOL heterogeneous rules of inference," *J. on Software and System Modeling*, 3(2):136–149, 2004.

[6] S. Kent, "Constraint diagrams: Visualizing invariants in object oriented modelling," in *Proceedings of OOPSLA97*. ACM Press, 1997, pp. 327–341.

[7] A. Fish, J. Flower, and J. Howse, "The semantics of augmented constraint diagrams," *J. of Visual Languages and Computing*, 16:541–573, 2005.

[8] G. Stapleton and A. Delaney, "Evaluating and generalizing constraint diagrams," *J. of Visual Languages and Computing*, 19(4):499–521, 2008.

[9] I. Oliver, J. Howse, G. Stapleton, E. Nuutila, and S. Törma, "A proposed diagrammatic logic for ontology specification and visualization," in *International Semantic Web Conference*, 2009.

[10] P. Chapman, G. Stapleton, J. Howse, and I. Oliver, "Deriving sound inference rules for concept diagrams," in *IEEE Symposium on Visual Languages and Human-Centric Computing*. IEEE, 2011.

[11] J. Howse, G. Stapleton, K. Taylor, and P. Chapman, "Visualizing ontologies: A case study," in *International Semantic Web Conference 2011*. Springer, 2011, pp. 257–272.

[12] G. Stapleton, J. Howse, P. Chapman, I. Oliver, and A. Delaney, "What can concept diagrams say?" in *Diagrams 2012*. Springer, 2012, pp. 291–293.

[13] F. Baader, D. Calvanese, D. McGuinness, D. Nadi, and P. Patel-Schneider. (eds), *The Description Logic Handbook*. CUP, 2003.

[14] G. Stapleton, J. Taylor, J. Howse, and S. Thompson, "The expressiveness of spider diagrams augmented with constants," *J. of Visual Languages and Computing*, 20:30–49, 2009.

[15] A. Borgida, "On the relative expressiveness of description logics and predicate logics," *Artificial Intelligence*, 82(1-2):353–376, 1996.

A Locally Nameless Visual λ-calculus

Peter Chapman
University of Brighton
E-mail: p.b.chapman@brighton.ac.uk

Abstract

λ-calculus can be seen as both a prototypical programming language, and as a basis for computation itself. Its syntax, whilst minimal, is not necessarily the most accessible. In order to address this perceived short-coming, we present a visualisation of the untyped λ-calculus, together with some examples of its computational rules. In particular, we present a locally nameless calculus, à la de Bruijn, removing the need for reasoning rules about capture avoiding substitutions. Instead, our notation focuses on the computation itself. We evaluate the syntax using known visual design principles.

1. Introduction

Alonzo Church introduced the λ-calculus in a series of papers in the 1930s, demonstrating its use as a model for computation. In particular, it solved one of Hilbert's problems (independently solved by Turing, using what would become known as Turing Machines, and Gödel) by demonstrating there were unsolvable problems in a relatively simple system. Similar to Turing machines, the untyped λ-calculus is relatively sparse, consisting of three term constructors and two reduction rules.

In the intervening decades, the untyped λ-calculus has been augmented with simple types, and more complex constructions such as dependent types [13] (the interested reader is directed to [2] for more details). All varieties generally share the same compact notation, albeit with extra constructors and reduction rules. To provide a visual notation for the untyped λ-calculus, then, could provide a basis for visual notations of much more descriptively rich programming languages.

Visual notations for reasoning have been employed to good effect in first-order and propositional logic. Examples include the Euler-Venn family of visual languages (for a thorough survey, see [14]). Some of these notations have been implemented in automated theorem provers and proof assistants [16], [3]. Their effectiveness for human perception and reasoning has been investigated [11], and the results suggest the diagrammatic notation can be easier to understand than traditional sentential approaches. Propositional and first-order logic, however, contain more inference rules than the λ-calculus, and thus we can be reasonably confident that a visual system for the latter could be beneficial, especially for new users of the λ-calculus.

1.1 Structure of the paper

In section 2, we give an overview of the untyped λ-calculus, and the locally nameless variant owing to de Bruijn. This section is intended as a brief introduction of the necessary material to make the paper self-contained. Those familiar with λ-calculus can safely skip this section. Following that, in section 3, we look at alternative approaches to visualising the λ-calculus. In section 4, we present our visual notation, illustrating the constructions and rules with many examples. The notation is evaluated against frameworks for effective visual designs in 5, and we conclude in section 6.

2. λ-calculus

We give a brief overview of the necessary material needed for the paper. The interested reader is directed to [2] for a much more thorough treatment.

The terms of the untyped λ-calculus are defined inductively as follows:

$$M \equiv x \mid \text{app}(M_1\ M_2) \mid \lambda x.M$$

where x is drawn from some set of variables. Other than atomic terms, then, there are only two constructions, called *application* and (λ-)*abstraction*. From now on, we will write application as $M_1 M_2$ instead of $\text{app}(M_1\ M_2)$. In an application $M_1 M_2$, we say that M_1 is *applied to* M_2. Application is left-associative, so for example, the term $(M_1 M_2) M_3$ would be written as $M_1 M_2 M_3$.

Abstraction can be seen as a generalisation operator: given a term M and a variable x, the term $\lambda x.M$ is M

where the (free) occurrences of x are liable to be replaced. Of course, there may be no (free) occurrences of x within the term M: this situation is known as *vacuous abstraction*. For example, the term $\lambda x.y$ is a vacuous abstraction: there are no occurrences of x in the term y. The term M in $\lambda x.M$ is referred to as the *body* of the abstraction.

In order to make sense of abstraction and its subtleties, we need to characterise variable occurrences as either free or bound. The *free variables* of a term M, denoted $FV(M)$, are defined on the structure of M as follows:

$$\begin{array}{rcl} FV(x) & = & \{x\} \\ FV(M_1 M_2) & = & FV(M_1) \cup FV(M_2) \\ FV(\lambda x.M_1) & = & FV(M_1) - \{x\} \end{array}$$

and, similarly, the *bound variables* of a term M, denoted $BV(M)$, are defined on the structure of M as follows:

$$\begin{array}{rcl} BV(x) & = & \emptyset \\ BV(M_1 M_2) & = & BV(M_1) \cup BV(M_2) \\ BV(\lambda x.M_1) & = & BV(M_1) \cup \{x\} \end{array}$$

For example, the variable x appears both free and bound in $x(\lambda x.xy)$: the first occurrence is free, and the second is bound. The variable y occurs free.

The untyped λ-calculus has two rules[1], known as α-conversion and β-reduction. The former will be less interesting, for the simple reason that we will consider a nameless calculus where α-conversion has no meaning. Intuitively, however, α-conversion allows us to rename bound variables, in such a way that no free variables become bound. In other words, we can identify the terms $\lambda x.x$ and $\lambda y.y$ using α-conversion: we replace every occurrence of the bound x by a y, and no free variables become bound as a result. However, we do not identify the terms $\lambda x.xy$ and $\lambda y.yy$: upon renaming the bound variable x to y, we have bound the free variable y which occurs in the former term. α-conversion is used to rename all bound variables so they are distinct from free variables. This conversion greatly simplifies implementation of the other rule, β-reduction.

As observed earlier, abstraction is a method for turning a term into a function. The abstraction $\lambda x.M$ creates a function which, when applied to some term N, will replace all free occurrences of x in M by N. In order to describe this, we need the notion of substitution. We use the notation $[N/x]M$ (read "N for x in M"), and define it on the structure of M as follows:

$$\begin{array}{rcl} [N/x]\,x & = & N \\ [N/x]\,y & = & y \\ [N/x]\,(M_1 M_2) & = & ([N/x]\,M_1)([N/x]\,M_2) \\ [N/x]\,(\lambda x.M_1) & = & \lambda x.M_1 \\ [N/x]\,(\lambda y.M_1) & = & \lambda y.\,[N/x]\,M_1,\ y \notin FV(N) \end{array}$$

[1] At this time, we are not considering η-conversion.

In the final clause, the side condition $y \notin FV(N)$ will remove the case where the variable y will become bound in $\lambda y.[N/x]M_1$. This substitution scheme is hence sometimes known as *capture-avoiding substitution*. In order to ensure that naïve substitution is capture avoiding (i.e. if we removed the side condition on the final clause), we use the Barendregt variable convention [2]: all bound variables should be named differently from free variables.

We are now in a position to define the computational rule β-reduction, denoted $\rightarrow_\beta$:

$$(\lambda x.M)N \rightarrow_\beta [N/x]M$$

As an example, consider the term $(\lambda x.(\lambda y.x))z_1 z_2$. Since application is left-associative, we get the following chain of reductions:

$$(\lambda x.(\lambda y.x))z_1 z_2 \rightarrow_\beta ([z_1/x](\lambda y.x))z_2 = (\lambda y.z_1)z_2 \rightarrow_\beta z_1$$

We use the notation $\rightarrow_{\beta *}$ to denote a sequence of 0 or more instances of $\rightarrow_\beta$. So, the above example would be written as $(\lambda x.(\lambda y.x))z_1 z_2 \rightarrow_{\beta *} z_1$. Any term of the form $(\lambda x.M)N$ is known as a *reducible expression*, or *redex*.

Some λ-calculus terms have special names, which we will use later in the paper:

$$\begin{array}{rclrcl} \lambda x.x & = & \mathbf{I} & \lambda x.(\lambda y.x) & = & \mathbf{T} \\ \lambda x.(\lambda y.y) & = & \mathbf{F} & \lambda x.(\lambda y.(\lambda z.xz(yz))) & = & \mathbf{S} \\ \lambda x.xx & = & \omega & \omega\omega & = & \Omega \end{array}$$

Note that $\mathbf{T}$ is often denoted $\mathbf{K}$. In implementations of the λ-calculus, most of the effort is devoted to ensuring that substitution really is capture avoiding. A number of approaches have been investigated to alleviate this burden on the user, including the nominal approach [7] and implemented in Nominal Isabelle [15]. However, we will instead provide a visualisation of the locally nameless syntax of de Bruijn, known as de Bruijn indices [6].

The idea is simple: in order to ensure that all bound variables are named differently from free variables, we simply remove the names of the bound variables and instead use their relative positions. Because it is the position of the variable in the body of an abstraction which is important, not the name, we replace the name by a positional index number. For example, the term $\mathbf{K}$ becomes $\lambda\lambda 1$. This statement can be read as follows: there are two abstractions, and the body of the abstraction consists solely of the first abstracted variable. Similarly, the term $\mathbf{S}$ can be written as $\lambda\lambda\lambda\ 13(23)$. Now, when we apply this term to other terms, it is clear there will be no capturing of free variables.

3. Related Work

There are a number of visual programming languages, however we will look here at other visual approaches to the λ-calculus. The VEX system of Citrin, Hall and Zorn [5] is most similar in spirit to what we are presenting here: it is a method for visualising the untyped λ-calculus. VEX can offer both visualisations of the terms themselves, and visualisations of the reduction rules. However, their syntax contains explicit names for bound variables: it is necessary to have a number of rules for α-conversion. However, the motivation for explicit names is that VEX was designed as a model for identifying the free and bound variables of an expression, and for allowing users to investigate their interplay. Since our motivation is somewhat different: providing a visualisation of the computations, rather than getting bogged down in details about substitution, it is expected that the two systems will be visually distinct.

Moreover, it would appear that nested abstractions may not be possible: "abstractions may only have a single parameter" suggests that expressions such as $\lambda xy.x$ (which is just notational short-hand for $\lambda x.(\lambda y.x)$) are not representable. Further, no examples are presented in [5] which include nested abstraction.

In [10], a Visual Lambda notation and interactive environment is developed. It uses a bubble notation, similar to VEX, but nested abstractions are shown as bubbles within bubbles. Application is displayed by the overlapping of one bubble over another. In a case such as $(M_1 M_2) M_3$ the bubbles for M_1 and M_2 both overlap the bubble for M_3, whereas in the term $M_1(M_2 M_3)$ the bubble for M_1 overlaps the bubble for M_2, but not the bubble for M_3. This design choice removes the need for brackets, or other delimiters, in the syntax of the language. However, it is not always clear, when parsing the visual expressions, which are acceptable reductions, and which are not.

In [4], Chien and Buehre present a notation for the untyped λ-calculus based on Mind Maps. It is implemented within the Mind Map project, and is computable. It uses a tree structure, with nodes for variables, applications, and abstractions. Variable names are explicit, and thus there will be some overhead in terms of ensuring capture avoiding substitution. Various lines are used to link parts of the term to other parts, and the system performs the α-conversion. It is visually somewhat dense, but offsets this by extensive use of colours.

Addressing a similar motivation, but using only minimal visualisation, is the Penn lambda calculator[2]. Similar to [4], the calculator provides a tree structure for λ-calculus terms, however the nodes are simply text.

As an introduction to λ-calculus reasoning for beginners, there is the *Alligator Eggs!* game [1]. The system there is representative of untyped λ-calculus, and is nameless: colour is used to associate variables with the appropriate λ-calculus term.

[2] Available at http://www.ling.upenn.edu/lambda/

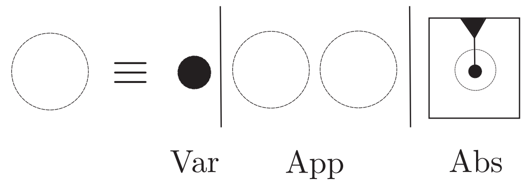


Figure 1. Visual λ terms

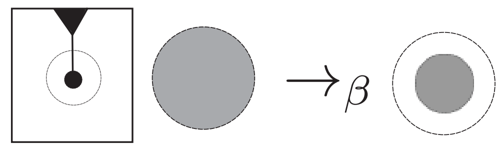


Figure 2. β-reduction

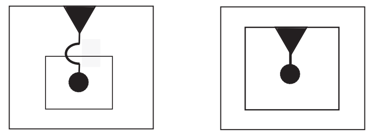
4. The Calculus

The visual analogy we use is based on the language used for substitution as "dropping a term into a hole". In other words, an abstraction (certainly a locally nameless abstraction) can be viewed as a term possibly containing some holes. In the case of vacuous abstraction, it contains no holes. β-reduction is then the process of dropping a term into the appropriate hole, to create a new term. However, before we get to reduction, we present the syntax of the visual terms in figure 1.

We represent a variable as a solid circle (which may be named, although this is not shown in figure 1). An arbitrary term is an empty circle, with a dotted border. Application (App) is represented as the placing of two terms next to each other. Abstraction (Abs) is represented as a box, containing a funnel on its upper surface, connected to a variable-like hole in an arbitrary term.

Of course, this representation of the definition is overly specific: it suggests that there is exactly one instance of the abstracted variable occurring free in the arbitrary term. The abstraction constructor should be taken more as a pattern: an abstraction is a box, with zero or more funnels on its upper edge, each connected to a single variable-like hole in an arbitrary term. Note that the presence of two funnels does not imply that the abstraction is of two variables at once. Rather, it suggests that there are two occurrences of the abstracted variable in the body of the abstraction.

Before coming to examples, we give the representation of the β-reduction rule. In figure 2, we see the process of substitution: we have picked up the grey term, dropped it into the funnel on the abstraction term, and it falls into the appropriate place in the body of the abstraction. Once the term has been dropped in, the outer box is removed. Again, this visual definition is overly specific: in order to present a β-reduction visually, we have to specify how many times the abstracted variable occurs in the body of the abstrac-



$\lambda xy.x = \lambda x.\lambda y.x = \mathbf{T}$ $\lambda xy.y = \lambda x.\lambda y.y = \mathbf{F}$

Figure 3. Vacuous substitution: $\mathbf{T}$ and $\mathbf{F}$

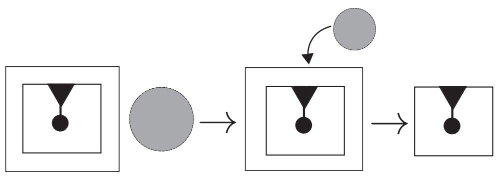


Figure 4. Vacuous substitution: $\mathbf{F}$

tion. What we do not have to specify, however, is whether or not there are any captures of free variables: the abstracted variable (the hole) has no name, and so it is not possible to create any bound variables by substitution.

With the problem of overspecificity in mind, we argue that the best way to demonstrate the types of term structures in the calculus is via concrete examples, rather than attempting to give visual definitions for every case. Through the examples, we will explain some of the design choices made with the notation.

4.1 Examples

Our first example is two instances of vacuous abstraction: the first being a case where the second abstraction is vacuous, and the second a case when the first abstraction is vacuous. In figure 3, we see the terms $\mathbf{T}$ and $\mathbf{F}$. In the former, we see that the funnel from the outer box passes into the inner box, and moreover the inner box has no funnel. The lack of funnel is what makes the abstraction vacuous. If we were to apply this term to another, say M, then, we would drop M into the inner box via the funnel, and the outer box would be removed.

Figure 4 shows the visual β-reduction rule in a dynamic way, in the case of vacuous substitution. We see a visualisation of the term $\mathbf{F}M$, where M is the grey circle. Firstly, we look at the abstraction: the term $\mathbf{F}$. Why does it represent $\mathbf{F}$? The outer box, containing no funnel, means that the variable abstracted from the body does not occur in the body. So, when we drop a term onto the abstraction, the body does not change. However, the outer box is still removed in this process. Thus, we are left with the body of the abstraction (in this case, the term $\mathbf{I}$) after a β-reduction. Figure 4 then represents the reduction $(\lambda y.\mathbf{I})M \rightarrow_\beta \mathbf{I}$. For some users, it may be easier to see why the terms reduce in the way they do from the visual representation, than through

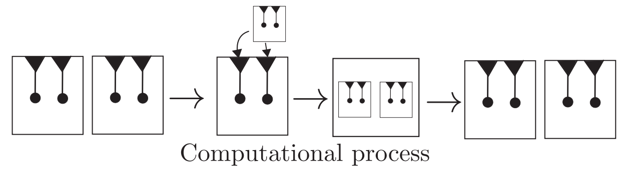


Figure 5. Ω reduction steps

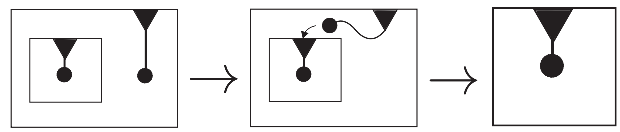


Figure 6. Reduction inside an abstraction

examination of the term $(\lambda xy.y)M$.

Our third example is one where there is more than one occurrence of an abstracted variable in the body of an abstraction. In particular, we look at ω and Ω. In figure 5, each of the left-hand terms represents ω. We see a box with two funnels, each connected to exactly one hole. Since the holes are the only parts of the body, it means we have represented $\lambda x.xx$. Then, ω applied to itself proceeds as follows: we pick up the second term in order to drop it into the funnels of the abstraction. Since there are two funnels, we produce two copies, and drop one down each funnel. Then, since we have performed the substitution, we remove the outer box. In other words, we have demonstrated visually that $\Omega \rightarrow_\beta \Omega$.

Figure 5 demonstrates a design choice in the notation. It would be possible to have a single funnel, and then a split to each of the holes (space restrictions mean there is no figure for this). However, using that visual notation could create drawability and perception issues. The former could arise when we have an arrangement of terms in the body of an abstraction which could not be drawn without crossing the funnels. The latter could arise when the different angle of some connections from funnels to holes cause it to have prominence in a diagram, which would be unwarranted[3].

We can also reduce inside an abstraction. Figure 6 shows such a reduction. When inside an abstraction, we can drop the hole of one abstraction M_1 (now treated like a term) into the funnel of another term M_2 it were possible to do so had the outer box of M_1 not existed. By this, we mean that, had the outer abstraction instead simply been a term, we could have performed the substitution. We still remove the inner box, since the reduction is otherwise normal. Thus, we have represented the reduction $\lambda x.(\lambda y.y)x \rightarrow_\beta \lambda x.x$.

In figure 7, we see the process for building complex λ-terms. In particular, we see how to build the term $\mathbf{S}$. We start with the inner most abstraction: $\lambda z.xz(yz)$. The ab-

[3] c.f. the *common fate* principle of [17].

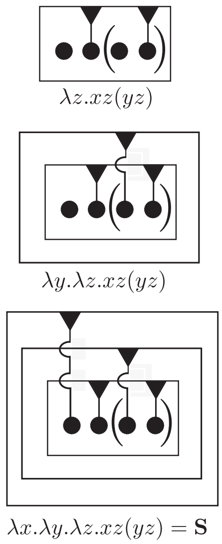


Figure 7. Building complex terms

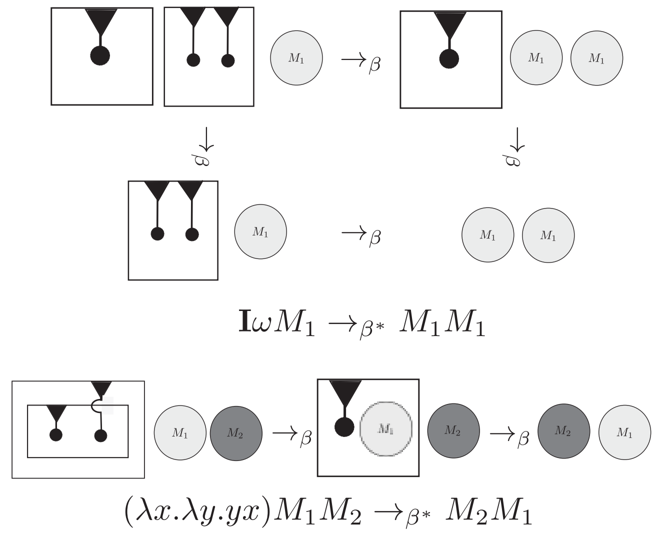


Figure 8. Redexes in different orders

stracted variable (z) occurs twice, so we need to build a box with two funnels connected to each occurrence of the hole for z. We fit the holes into the pattern of the body. If we were to stop at this stage, we would have to label the other two (free) variables. However, we continue by looking at the next layer of abstraction: the abstraction of y. There is only one occurrence of y in the term, so we connect a single funnel from the new box to the appropriate variable. At this stage, the z and y occurrences have become nameless. The final abstraction is of a single variable, and the funnel must reach the body of the first abstraction, hence it crosses the two inner boxes. Note that this term does not contain any redexes, and this is clear from the visualisation: we cannot pick any of the holes up and drop them into other funnels without crossing out of a box.

We now consider a more complicated example, where there is a choice of redex to inspect. Consider the upper part of figure 8. There are two possible redexes: treating ωM_1 as a term, and applying it to $\mathbf{I}$, and applying M_1 to ω. The former is represented in the vertical direction, and the latter in the horizontal, from the initial term. Whichever redex we chose, we are still left with a single redex, which we can β-reduce to get to the bottom right diagram.

The lower portion of figure 8 shows how seemingly minor changes in a term structure produce large changes in output. The only difference between the abstraction in the lower panel, and that in figure 6, is that the first funnel creates a hole in the second abstraction. However, when we apply two terms to the abstraction, we switch the order of the resulting application of terms. Had we applied M_1M_2 to the term in figure 6, since that term reduced to $\mathbf{I}$ we would have obtained M_1M_2 back again. Again, this is a fairly small scale example: it is quite clear from $\lambda xy.yx$ that the term does indeed switch the order of the terms upon application. In more complicated terms however, the diagram may illuminate structure which is otherwise hidden in the sentential form of the expression.

In the case that expressions are made up of many simple abstractions applied to terms, then the notation will be reasonably scaleable. However, for terms with multiple nested abstractions, or for abstractions with complex bodies, the notation could become quite visually cluttered.

5. Evaluation against cognitive frameworks

In the previous section, mainly through the use of particular examples, we demonstrated the use of our notation for locally nameless λ-calculus terms. In this section, we briefly consider whether or not the notation is adequate when judged against design principles. The principles we examine are taken from Wertheimer's Gestalt principles [17], Moody's Physics of Notations [12], and the principles of Larkin and Simon [9].

- *Proximity*. In other words, things close to each other tend to be grouped. In our case, this is partially satisfied: we only want an abstraction, and the term immediately following it, to be grouped together. It is possible that introducing brackets as barriers, to make explicit the associativity properties, could help. However, this would be at the expense of visual simplicity.
- *Similarity*. In other words, similar symbols tend to be grouped together. Similarity could include size or shape. In order to exploit this principle further, the redex under consideration could be increased in size relative to the rest of the diagram. That is an issue for

implementation in a computer system, however (see section 6).
- *Closure*. In other words, lines that form a closed region tend to be viewed as a distinct object. It is for this reason that abstractions are given as boxes.
- *Perceptual Discriminability*. In other words, distinct classes of object should be readily identifiable. A general term is a circle (filled or otherwise), a variable is a solid, smaller circle, and an abstraction is a box. These are all easily distinguishable items.
- *Semantic Transparency*. In other words, diagrams are well-matched to meaning [8]. The analogy of substitution as placing terms in holes is what we were intending to visualise. With this meaning, and the dynamic method for performing reductions, we argue that there is semantic transparency with our notation.
- *Cognitive Integration*. In other words, there should be a mechanism for combining information from separate diagrams. In our case, there is such a mechanism, it is simply concatenation. In order to apply one term to another, we place them next to each other.
- *Visual Expressiveness*. In other words, the full range of visual variables should be used. Since we have a small number of visual variables, it is easy to satisfy this principle.
- *Graphical Economy*. In other words, do not use more visual variables than is cognitively manageable. By using locally nameless abstraction, we have minimised the number of visual variables (i.e. fewer labels), and also the number of rules for the syntax.
- *Dual Encoding*. In other words, complement visualisations with textual representations. We mentioned that, when variables occur free, they should be labelled. There is nothing stopping the visualisation from being augmented with other pieces of textual information.

Of course, the preceding arguments are in no way to be taken as proof that the notation is effective for its stated task of modelling computation without wasting cognitive overhead on variable naming issues. However, the principles guided the design of the notation, which could be further refined upon empirical testing.

6. Conclusions and Further Work

We gave a brief overview of one of the main models of computation, the λ-calculus. Upon finding that the visual notations which exist are primarily concerned with identifying bound variables and α-conversion, we proposed a syntax which instead removes these issues, allowing the focus to be entirely on the computational aspects of a problem, given by β-reductions. The proposed notation was designed and evaluated against known design principles, and found to a potential candidate for further investigation.

The work described here represents an initial approach to increasing the accessibility of reasoning about computation. Empirical studies will be conducted, to refine the notation. The participants would be students new to the λ-calculus, in order to assess whether the notation correctly guides the user to make allowable reductions. Above all, the notation is designed to be dynamic: it is envisaged that users of the notation will use a computer system to drag-and-drop terms into funnels. Such a system will need to be developed. Finally, we will conduct empirical studies to discover whether the stated goal, of illuminating computation for beginners, has been achieved.

References

[1] Alligator Eggs! http://worrydream.com/AlligatorEggs/.

[2] H. Barendregt. *The Lambda Calculus: Its Syntax and Semantics*. North Holland, 1984.

[3] D. Barker-Plummer, J. Etchemendy, A. Liu, M. Murray, and N. Swoboda. Openproof - A Flexible Framework for Heterogeneous Reasoning. In *Diagrams*, pages 347–349, 2008.

[4] L. Chien and D. Buehrer. Using a Typed Mind Map as a Data Model in a TDD DICE System. *CIT*, 16(4):345–354, 2008.

[5] W. Citrin, R. Hall, and B.Zorn. Programming with Visual Expressions. In *VL*, pages 294–301, 1995.

[6] N. G. de Bruijn. Lambda Calculus Notation with Nameless Dummies, a Tool for Automatic Formula Manipulation, with Application to the Church-Rosser Theorem. *Indagationes Mathematicae*, 34:381–392, 1984.

[7] M. Gabbay and A. Pitts. A New Approach to Abstract Syntax with Variable Binding. *Formal Asp. Comput.*, 13(3-5):341–363, 2002.

[8] C. Gurr. Effective Diagrammatic Communication: Syntactic, Semantic and Pragmatic Issues. *J. Vis. Lang. Comput.*, 10(4):317–342, 1999.

[9] J. Larkin and H. Simon. Why a diagram is (sometimes) worth ten thousand words. *Journal of Cognitive Science*, 11:65–99, 1987.

[10] V. Massalõgin. Visual Lambda Calculus. Master's thesis, University of Tartu, 2008.

[11] K. Mineshima, M. Okada, Y. Sato, and R. Takemura. Diagrammatic Reasoning System with Euler Circles: Theory and Experiment Design. In *Diagrams*, pages 188–205, 2008.

[12] D. L. Moody. The Physics of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. *IEEE Trans. Software Eng.*, 35(6):756–779, 2009.

[13] B. Nordström, K. Petersson, and J. Smith. Martin-Löf's Type Theory. *Handbook of Logic in Comp. Sci.*, 5, 2005.

[14] G. Stapleton. A Survey of Reasoning Systems Based on Euler Diagrams. *ENTCS*, 134:127–151, 2005.

[15] C. Urban and C. Tasson. Nominal Techniques in Isabelle/HOL. In *CADE*, pages 38–53, 2005.

[16] M. Urbas, M. Jamnik, G. Stapleton, and J. Flower. Speedith: A Diagrammatic Reasoner for Spider Diagrams. In *Diagrams*, pages 163–177, 2012.

[17] M. Wertheimer. Gestalt Theory. *Social Research*, 11(1):78–99, 1944.

GPGPU Implementation of a Generative Modelling Language

Martin Griffin
University of Brighton
mg200@uni.brighton.ac.uk

Aidan Delaney
University of Brighton
a.j.delaney@brighton.ac.uk

Karina Rodriguez Echavarria
University of Brighton
kre@brighton.ac.uk

Abstract

We record and digitally preserve public monuments by describing them using a generative language. Such generative models allow visualisations of the monuments to be computed from descriptions that are suitable for transmission to personal mobile devices. The intent of using a personal mobile device is to make the cultural heritage experience more personal to the user. We present an implementation of a generative modelling language used primarily to describe tangible cultural heritage. Our implementation is the first to make explicit use of GPGPU computation. As such, it allows generative models to be computed on mobile devices where, currently, the GPU processing power significantly outweighs the available CPU power.

1 Introduction

The amount of 3D content available is rapidly increasing due to the popularity of 3D acquisition technologies such as photogrammetry and affordable 3D scanners in addition to traditional point-and-click modelling tools. In particular, the area of cultural heritage is benefiting from this trend as the professional and the general public alike are recording objects in the towns and cities in which they live or visit. These objects vary from smaller objects displayed in museums to larger sculptures, monuments and historical buildings. Moreover, 3D content can support the documentation of this type of tangible cultural heritage, which in turn can support its preservation and conservation. This is because 3D content serves as a more accurate representation of the objects than other multimedia content types, such as 2D images or video. Nevertheless, this content needs to be organised in digital repositories [4, 8, 11] and suitable technologies for visualisation need to be developed if the content is to be accessed when and where it is needed.

It was proposed [2] that capture, storage/management and visualisation are research challenges for the use of intelligent information and communications technologies for cultural heritage. In particular, the use of popular 3D scanning technologies at the capture stage leads to using an approach to the storage and management of data that lacks any kind of associated semantics of the shape being stored. In this case, shape is described using a collection of elementary geometric units or primitives (e.g. points, triangles, NURBS, cubes, spheres). As an example, in the model of a historical building there is little, or no information to tell whether a given point or triangle is part of the model floor, wall or roof. This approach is also seen in model produced by manipulation (e.g. extrusion, intersection) of primitives in modelling tools.

Alternatively, visual languages such as procedural shape representation place emphasis on the sequence of operations or rules performed to create a 3D shape rather than on the resulting primitives. Hence, they can address the challenge of formalizing the description of shape for any given shape or shape class. Examples of these visual languages include shape grammars, which are widely used in architecture [10], shape programming languages [5] and scripting facilities provided by some modelling tools. Hence, bridging the gap between acquired geometry (by scanning mechanisms or photogrammetry) and a procedural shape representation is on-going research and an unsolved challenge.

In this paper, we are particularly interested in the efficient representation and visualisation of 3D content, as a wide range of platforms are available to users for accessing 3D content. Particularly relevant is the current and next-generation of mobile computing devices, which do not possess such large local storage and processing capabilities. As we will explain, the CPU processing power of a mobile device is severely restricted. However, mobile devices, being personal to the owner, are seen as an ideal platform for the dissemination of cultural heritage [9] particularly for preservation, tourism and educational purposes. As such, our interest is in visualising cultural heritage on mobile devices with restricted computing capacity.

The restrictions of mobile computation can be observed by measuring the number of floating point operations per second (Flops) of which an architecture is capable. A current generation desktop CPU achieves over 70GFlops of single-precision computation, and a high-end consumer-grade desktop graphics card can achieve up to 160TFlops of single-precision computation. On the other hand a current generation tablet CPU achieves 1.2GFlops[1]. However, though the floating-point computing power of a desktop CPU and a mobile CPU are vastly different, we observe that the floating-point power of a desktop CPU is roughly comparable with the floating-point power of a mobile GPU.

This paper proposes an approach based on generative modelling technologies for the efficient storage and visualisation of 3D content. In particular, we explore the use of the Generative Modelling Language (GML) [5] for modelling architectural heritage which is stored in a state of the art 3D digital repository [4]. The main contribution of this paper is an implementation of an interpreter for this language in OpenCL on the GPGPU [7] in order to enable a more efficient representation and storage.

The paper is organised as follows. In section 2 we provide an overview of generative modelling as well as the modelling of architectural heritage using the GML language. Thereafter, in section 3 we describe our implementation of the generative modelling language interpreter using OpenCL.

2 Generative modelling and GML Overview

Rather than considering 3D models to be composed of vertices, edges and faces; generative modelling conceives of manifolds that are manipulated using a small number of algebraic operations. These operations, often referred to as Euler operations, ensure that the application of an Euler operation always produces a valid surface. These surfaces can then be converted into and meshes of polygons. The polygonal mesh can itself be interpreted explicitly as a geometric model or as a control mesh for a NURBS surface. In this way generative modelling is a formal visual language that allows us to generate complex 3D meshes as a sequence of atomic operations.

Commonly, in implementations of generative modelling, a program is written that explains the connectivity of edges and vertices. Higher-order operations such as extrusion or deletion can then be performed on the connected edges thus generating a complex model. An example can be seen in figure 1. Figure 1a is a photograph of an historical building, known as the Brighton Pavilion. This building, as many other historical structures, has a very complex shape but exhibits a great deal of symmetry and common stylistic elements. Traditional modelling of the detailed shapes would involve many millions of tiny triangles which would still show up as individual facets if the viewer gets too close to the model. A procedural model can, in principle, exploit the regularity and the stylistic elements of the Pavilion to produce a model that can be viewed at any scale and still show smooth detail over a really large object.

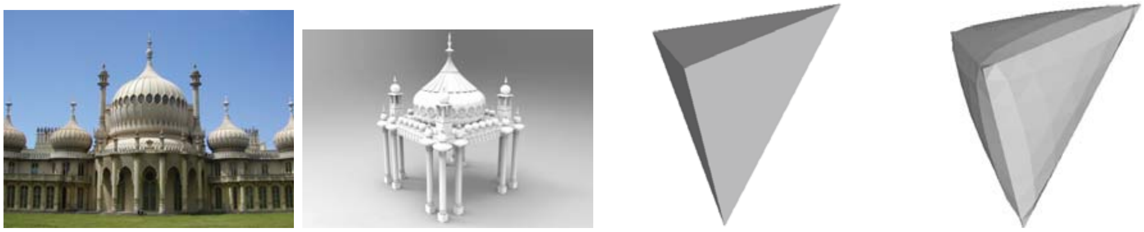
Figure 1b illustrates the generative model produced for a section of this building. This model is represented by a script and stored in a file of size 95Kb. The generated mesh, which is comparable in size to a Wavefront OBJ or COLLADA file storing the mesh, is of tens of megabytes in size. Hence, one significant advantage of generative modelling is that the interchange files are suitable for download onto a mobile device over a restricted bandwidth connection.

GML [5] is one implementation of a generative modelling language. GML is a catenative stack based language which extends PostScript [1]. An example of the GML shape description language is listed in figure 2a. Though the language itself is textual, the core operations are entirely focused on specifying visual properties of subject being modelled. A visualisation of the surface generated by lines 1 through 3 is provided in figure 2b. The leftmost pair of vertices connected by an edge is the result of the application of the Euler operation **makeVEFS** to make a new surface around the given edge on line 1 of the program. The centre graph depicts the application of the Euler operation **makeEVone** on line 2 which extends the existing structure with a new edge. In a similar manner, the rightmost graph in figure 2a demonstrates the creation of the first face of the tetrahedron using the Euler operation **makeEF** in line 3 of the program.

Figure 1c provides a visualisation of the internal GML boundary representation structure of the generated tetrahedron. This is the structure generated by the entire program in figure 2a. Finally, as the tetrahedron is the primary focus of this scene, we may wish to render it in more detail by subdividing the surface, figure 1d. The subdivided structure maintains the curvature given by the original tetrahedron.

The formal visual language, called the shape description language is the highest level, level 4, of GML's 4-tier architecture. The Euler operations on surfaces constitute level 3 of the architecture. Level 2 consists

[1] Measured using Linpack on an Intel i5, AMD R700 and an ARM Cortex A9 respectively. The actual numbers of lesser importance than the difference in scale between them.



(a) (b) (c) (d)

Figure 1: Generative modelling example. Brighton Pavillion photograph by user Xgkkp on `http://commons.wikimedia.org`.

of a boundary representation [6] structure which is subdivided using the Catmull-Clark algorithm [3] at level 1. Our ongoing implementation of a generative language interpreter focuses on the GML language and concentrates on implementing level 4 and level 3 of this architecture.

3 GML in OpenCL

As GML is a terse extension to PostScript, we have implemented a PostScript-style interpreter on a GPU for a subset of GML, called `cl-gml`. Our choice of GML subset was directed by the task at hand, namely the construction of a generative model of the Brighton Pavilion. Using a desktop implementation of GML [2] we defined the floor plan of the Pavilion as a polyhedron. Thereafter, we extruded the floor plan to produce the internal and external walls. The architecturally splendid roof and cornice decorations were also generatively described. Our OpenCL implementation replicates this production pipeline by first concentrating on implementing the generative modelling of polyhedra.

Given the 4-tier architecture of GML, our first task was to implement a PostScript interpreter for the GML shape description language. We observed that in modelling polyhedra we used a subset of PostScript and GML that makes extensive used of both the PostScript operator stack and storage of composite objects. Therefore, we focused on implementing the core GML operators that manipulated the PostScript stack. We further support the basic built-in integer, floating point number, array and 3d vector types over which these operators operate.

Level 3 of the GML architecture defines five Euler operations and their duals. Of these operations we implement the three previously introduced, namely **makeVEFS**, **makeEVone** and **makeEF**. Our choice of opeartors allow us to generate any surface of genus 0. As a consequence of this implementation decision we collapse level 2 and level 1 of the GML architecture into a more straightforward mesh representation. Removing the need for an OpenCL implementation of a the boundary representation and subdivided boundary representation structures allowed us to demonstrate the validity of our approach within limited development time. However, it is clear that supporting the generation of higher genus surfaces will require a boundary representation implementation.

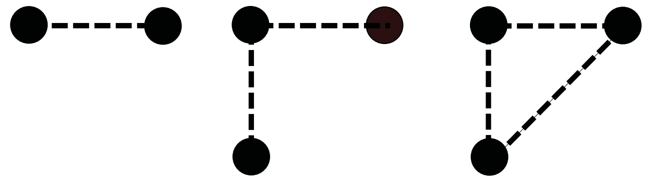
The `cl-gml` architecture is, due to the above simplifications, more straightforward to implement in OpenCL. Implementations of the PostScript stack and a heap for the composite object store were developed as an OpenCL kernel. As OpenCL provides no higher level abstractions of memory, our heap implementation features an implementation of the traditional `C` system calls **malloc** and **free** to allow dynamic memory allocation. Furthermore, we implemented a call stack with local storage, i.e. stack frames, to allow the GML operator **exec** to call functions defined in GML.

Having created a traditional heap and stack, our problem of interpreting the GML shape description language became an issue of providing portable implementations of each core GML operator. Each operator is implemented as a macro which is included in our OpenCL kernel and then compiled at runtime to produce object code that will execute on the target GPU. The implementation of the **add** operator can be seen in figure 3. The code listing demonstrates the three different specialisations of **add** supported by our implementation of GML. These are the addition of integers, floats and vectors in R^3. Furthermore, the internal stack operations **push_INT**, **push_FLOAT** and **push_VECTOR3** demonstrate that the result of the **add** operation is pushed to the internal PostScript stack.

[2] `http://www.generative-modeling.org/GenerativeModeling/software`

```
(0, −1, 0) (1, 1, 0) makeVEFS
  dup (−1, 1, 0) makeEVone
  1 index 1 index makeEF
  dup (0, 0, 1) makeEVone
  3 index 1 index makeEF
  3 index 1 index makeEF
  3 index 1 index makeEF
pop pop pop pop pop pop
```

(a)



(b)

Figure 2: Level 4 and level 3 of generating a tetrahedron in GML.

```
OPERATOR(ADD, ``add'', (
        (((INT, int a), (INT, int b)), (
        push_INT(stack, a + b);
        )), (((FLOAT, float a), (FLOAT, float b)), (
        push_FLOAT(stack, a + b);
        )), (((VECTOR3, struct Vector3 a), (VECTOR3, struct Vector3 b)), (
                push_VECTOR3(stack, (struct Vector3) { a.x + b.x, a.y + b.y, a.z + b.z });
                ))
))
```

Figure 3: A C99 macro implemenataion of the GML **add** operator.

4 Conclusion

We have implemented a subset of the operations of a visual language, called the generative modelling language, in OpenCL. The implemented operations allow the modelling of any 3D surface that does not contain "holes" i.e. is of genus 0. The restriction to genus 0 surfaces is a restriction of our proof-of-concept implementation rather than the underlying formal visual language. Our implementation allows the use of generative modelling on hardware architectures on which CPU based generative modelling is prohibitively expensive. Our target platforms include smartphones and tablets. The trend in these target architectures points to a significant increase in their graphics processing capability in excess of the increases planned for their CPU processing capabilities. However, having chosen OpenCL as an implementation platform there is no technical reason that we cannot prefer purely CPU based computation should this trend reverse itself. Our GML interpreter is has been designed and implemented so that it may provide cultural heritage visualisations well into the future. Finally, our implementation is available as an open-source project from `https://gitorious.org/cl-gml`.

Acknowledgements: Aidan Delaney is grateful to the European, France (Channel)-England INTERREG IVA European Cross-border Co-operation Programme (Project CEREEV, Ref: 4224) for the financial support towards this project. Martin Griffin and Karina Rodriguez-Echavarria are funded under the 3D-COFORM Integrating Project (FP7-ICT-231809).

References

[1] ADOBE Systems. *PostScript Language Reference Manual*. Addison-Wesley, 3 edition, 1999.

[2] D. Arnold and G. Geser. *The EPOCH research agenda for the applications of ICTs to cultural heritage Excellence in Processing Open Cultural Heritage*. Archaeolingua, Budapest, 2008.

[3] E. Catmull and J. Clark. Recursively generated b-spline surfaces on arbitrary topological meshes. *Computer-Aided Design*, 10(6), 1978.

[4] M. Doerr, K. Tzompanaki, M. Theodoridou, C. Georgis, A. Axaridou, and S. Havemann. A Repository for 3D Model Production and Interpretation in Culture and Beyond. In *Proceedings of the International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, pages 97–104, 2010.

[5] S. Havemann. *Generative Mesh Modeling*. PhD thesis, TU Braunschweig, 2005.

[6] M. Mäntylä. *An Introduction to Solid Modeling*. Computer Science Press, 1988.

[7] A. Munshi, editor. *The OpenCL Specification Version: 1.2 Document Revision: 19*. Khronos Group, 2012.

[8] Princeton University. Princeton Shape Benchmark. http://shape.cs.princeton.edu/search.html.

[9] K. Rodriguez-Echavarria, J. Kaminski, and D. B. Arnold. 3D heritage on mobile devices: Scenarios and opportunities. In *EuroMed*, pages 149–158, 2012.

[10] G. Stiny and J. Gips. Shape grammars and the generative specification of painting and sculpture. *Information Processing*, 71:1460–1465, 1972.

[11] VISIONAIR Consortium. The Shape Repository of the Visualisation Virtual Services (VVS). http://visionair.ge.imati.cnr.it/ontologies/shapes/.

Comparing Purely Visual with Hybrid Visual/Textual Manipulation of Complex Formula on Smartphones

Annemarie Harzl, Vesna Krnjic, Franz Schreiner, and Wolfgang Slany
Graz University of Technology
aharzl@ist.tugraz.at, vesna.krnjic@iaik.tugraz.at, franz.schreiner@student.tugraz.at, wolfgang.slany@tugraz.at

***Abstract*—Only very few end users have the skills to develop mobile apps such as games or animations. Visual programming languages can be very supportive for casual and first-time users, allowing the users to concentrate on the programming task rather than learning complex syntax. This is why visual programming languages are often used where children are concerned. Nevertheless, studies have shown that the advantage of visual languages tends to decrease on larger tasks or mathematical formulas. The paper distinguishes different approaches for creating formulas with end user programming languages, namely purely textual, purely visual, and hybrid approaches. In our paper we introduce Pocket Code, a new approach with visual programming and hybrid formula editing, which combines the easiness of visual programming with the effectiveness and clarity of textual formula displaying. Additionally, we present a proposal of an evaluation of the different approaches to formula manipulation in visual programming languages for smartphones.**

I. Introduction

Visual programming languages (VPL) are important for end user programming. It empowers end users with little, no, or only casual programming experience to develop programs for their individual use. The ability to write applications, to create one's own games, or to automate small tasks, can be easily learned. Due to the features of VPL they are often used when it comes to children. Especially for younger children it seems to be easier to drag & drop bricks together like in Scratch[1] than to learn a textual programming language. They can learn the basic principles of programming without bothering with the sometimes restrictive syntax of a textual programming language. There are other important benefits of visual languages as well, e.g., being able to see what command blocks are available and thus might be employed, thereby suggesting their use without the user having to know or even remember them. Regarding programming and the writing of formulas there are two clearly distinct approaches: the textual and the visual approach. With traditional textual programming languages like C or Java, developers enter statements on a standard keyboard, though modern IDEs to some degree simplify the entering of statements through context sensitive statement completion. Visual languages like Scratch use the visual approach, where even formulas are created with pre-defined graphical blocks. When complex formulas are involved, this latter approach can become unwieldy and even confusing. Some programming languages like TouchDevelop[2], which is mainly textual, pursue an approach where code and formulas are entered via an interface that reminds of pocket calculators: Statements and operators are chosen from a set of visually differentiated alternatives, but the actual visualization of the resulting statements and formulas is done in a purely textual way. In our paper we introduce Pocket Code[3], a new approach with visual programming, but textual formula representation, which combines the easiness of visual programming with the effectiveness and clarity of textual formula displaying. One of the reasons for introducing this new combination is that the presented visual programming language is optimized for the use on smartphones with their touch screens and small display sizes, where Scratch-like blocks cannot easily be accommodated due to the narrowness of the screens and the difficulty to drag and drop blocks closely nested together with one's fingers compared to when using a mouse pointer. Furthermore we will discuss different programming language approaches for editing and manipulating formulas and present a proposal how these approaches could be compared and evaluated regarding their efficiency, effectiveness, and user satisfaction.

II. Related Work

User studies have shown mixed results on the superiority of visual languages over text languages. However most work has focused on the desktop so far. For instance the empirical study [1], which has been done comparing constructibility of programs in textual versus visual languages, concludes that matrix manipulation programs can be more easily constructed in a visual language (e.g., Formd3) than in a textual language (e.g., OSU-APL and Pascal). Another empirical study [2] concludes that visual languages provide a better user experience, reduce perceived workload, and increase perceived success. On the other hand, Green et al. [3] show that dataflow visual programming languages are not consistently superior to text languages. Their study shows that some visual notations, for example the gate notation are, in fact, worse than equivalent textual notations. Neither textual languages nor visual languages are perfect and are both perceived as not supportive or confusing by end users without (much) programming experience. However the comparison of the two approaches needs to be considered for smart phones and small displays as well.

When creating formulas with end user programming languages there are three main approaches:

1) The purely textual approach like in Microsoft's Excel, where formulas are created and displayed textually. Excel is a spreadsheet application, designed mainly for adults and the use on traditional computers, where

[1] http://scratch.mit.edu/

[2] https://www.touchdevelop.com/

[3] http://www.pocketcode.org/, previously known as Catroid

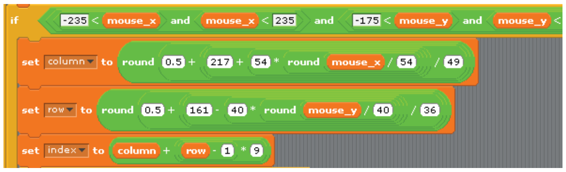


Fig. 1. Visual formula editing in Scratch.

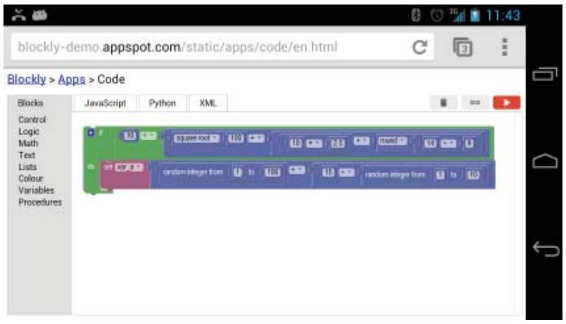


Fig. 2. Blockly in landscape mode on a mobile browser.

users can calculate values with formulas entered in Excel.

2) The purely visual approach, like in Scratch [4], Snap![4] [5], and Blockly[5] where pre-defined visual segments are used to create and to display a formula.
3) The "hybrid" approach like in Microsoft's TouchDevelop and Pocket Code's formula editor, where formulas are created using visual elements similar to a pocket calculator, but are displayed textually.

In the course of our paper we will focus on the latter two approaches.

Scratch was designed for children and the use on traditional computers. Snap!, an extension to Scratch, was designed for children and adults and the use on traditional computers (large screen, keyboard, mouse), but works on smartphones as well. In addition to Scratch's approach (see Figure 1) Snap! highlights different nesting levels of formulas in a so called "zebra"-mode: Parts of the formula are alternatingly lighter and darker colored (see Figure 3).

Blockly[6] was also mainly designed for adults and the use on traditional computers (large screen, keyboard, mouse), but works on smartphones as well. It allows to switch seamlessly between a purely visual approach similar to Scratch to purely textual ones (alternatively JavaScript, Python, and XML) and back (see Figure 2).

TouchDevelop is an application creation environment intended particularly for students or adult hobbyist programmers. It is intended to be used primarily on smartphones. The programming language is text-based but uses a few non-ASCII graphical characters for representation of the syntax, for example arrows, a recycling symbol, etc. It resembles a traditional text-based programming language, though with a specialized editor (see Figure 4) and use of annotation of program text

[4]http://snap.berkeley.edu/
[5]https://code.google.com/p/blockly/
[6]http://blockly-demo.appspot.com/static/apps/code/en.html

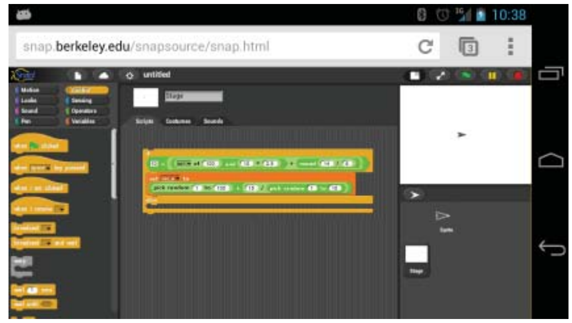


Fig. 3. Visual formula editing in Snap! in landscape mode on a mobile browser.

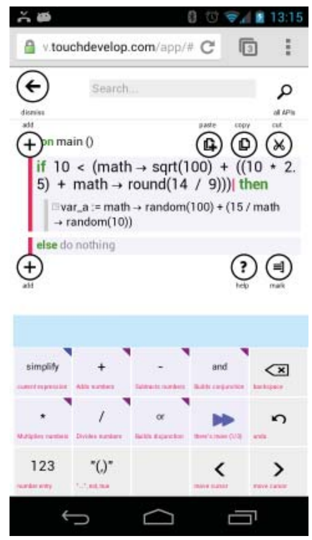


Fig. 4. Textual formula editing in TouchDevelop.

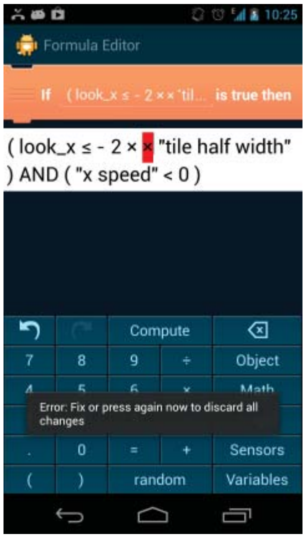
and automated re-formating. Formulas are entered visually but displayed textually, similar to Pocket Code's formula editor (see Figure 5(a)).

III. Pocket Code

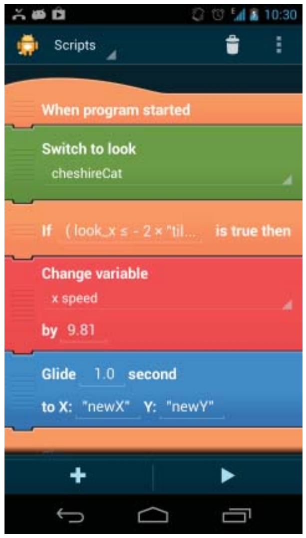
Pocket Code is a free and open source mobile visual programming system for the Catrobat language[7]. It allows users, starting from the age of eight, to develop games and animations with their smartphones. To program, the children use their Android phone, iPhone, Windows Phone, or other smartphone with an HTML5 browser. No notebook or desktop computer is needed [6]. Pocket Code is inspired by, but distinct from, the Scratch programming system developed by the Lifelong Kindergarten Group at the MIT Media Lab [4]. Similar to Scratch, our aim is to enable children and teenagers to creatively develop and share their own software. The main differences of Pocket Code in contrast to Scratch are:

1) Support and integration of multi-touch mobile devices
2) Use of mobile device's special hardware (e.g., acceleration, compass, inclination)
3) No need for a traditional PC

[7]http://catrobat.org/



(a) In this particular example the user entered a syntactically incorrect formula (two multiplication signs one after the other). After pressing the BACK-button an error message appears and the syntax error gets highlighted in red.



(b) Visual script of a program written using Pocket Code that contains the If-brick with the formula that was edited in Figure 5(a).

Fig. 5. Textual formula editing in Pocket Code.

Additionally, there are more than 30 ongoing sub-projects mostly aiming at extending Pocket Code's functionality, e.g., a 2D physics engine that will make the programming of games similar to the popular Angry Birds type of games very easy, or an extension allowing to very easily record the screen as well as sound during execution of a program and to upload it to YouTube, the high definition video being created on our server and uploaded from there to avoid high costs and lengthy file transmissions for the kids.

In the following we give a succinct overview of the design intentions of Pocket Code's formula editor:

We wanted to show the user which statements operators, variables, messages are possible, thus simplifying discoverability for the user. It should be easy to edit on a small touch screen ("use only one's thumb to enter a whole program or formula"). Text based formulas are easier to display on narrow screen (text wrapping) and at the same time well known from typical pocket calculators, calculator apps, but also from spreadsheets and even from math classes at school. It should be easy to get a preview of the current value of a formula through a "compute" button. It should provide users with an overview over current variable values. The formula editor should make it easy to position the cursor at any place in the formula, and to select larger parts of a formula. It should visualize "matching" parentheses in complicated nested expressions. The formula editor should also eliminate some syntax errors preemptively. Other syntax errors when users try to use an "unfinished" formula (see Figure 5(a)) should be highlighted. Scrolling of long formulas should be possible. Copy/cut/paste of parts of formulas should be possible. Easy undo/redo should make it easier to develop a formula.

IV. Formula Manipulation Approaches Under Study

In this section we discuss two main ways of editing and displaying formulas (the third main approach, textual creation and textual visualization of formulas, was mentioned in Section II, but will not be discussed here):

1) Visual creation and visual representation of formulas like in Scratch and Blockly. Scratch (see Figure 1) and other visual programming languages use purely visual formula editing. Different segments have to be nested within each other to compose a formula.
2) Visual creation and textual representation of formulas, like in TouchDevelop and Pocket Code. In TouchDevelop's and Pocket Code's formula editor, formulas are created via a pocket calculator-like interface (see Figures 4 and 5(a)). Statements and operators are selected visually, but the actual formula representation is purely textual. This should help save screen space and provide users with a better overview over the formula.

Visual composition of formulas can become a tedious task, because numerous visual components have to be nested within each other for more complex formulas. This is especially true for the small screens of smartphones. The screen limitations of mobile phones and the common knowledge of how to use a calculator led to our decision to display formulas textually in Pocket Code's formula editor. Most teenage and adult users know how to operate a pocket calculator and should therefore experience no problems with Pocket Code's formula editor. For smaller children, who did not use a calculator before, future usability studies will determine whether the textual or visual approach works better for them. Displaying formulas textually may be faster and easier to understand for users who are familiar with pocket calculators. In order to evaluate the two different approaches we will conduct a formal experiment that is described in the following section.

V. Evaluation of Pocket Code's Formula Editor

To assess the usability of Pocket Code's formula editor we followed the main objectives of User Centered Design (UCD) methods defined by ISO 9241-210:2010[8] including user research, interface design, and usability testing during the implementation cycles. According to the agile principles used by the software development team, the applied UCD methods followed the agile methods as well, such as inspection, heuristic evaluation, paper mockups, and thinking aloud tests [7]. Additionally to previously done formative testing we are planning to conduct a formal experiment in order to gain a summative assessment of the formula editor. This section describes the methodology that will be applied to evaluate the usability of Pocket Code's formula editor and compare it with three different programming language approaches.

A. Methodology

The purpose of the planned experiment is to provide scientific evidence to support or revoke the assertions described below. The following hypotheses are stated:

[8]http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075

1) Null hypothesis: For the manipulation of complex formulas, the calculator metaphor (the hybrid textual/visual approach) is more effective and efficient than the pure visual programming language approach.
2) Alternative hypothesis: The contrary of above null hypothesis: For the manipulation of complex formulas, the calculator metaphor (the hybrid textual/visual approach) is as good or less effective and efficient than the pure visual programming language approach.

A complex formula in this context will be clearly defined, for example something like a logical formula composed of nested expressions at least 4 levels deep, with 12 parentheses, 3 variables, 1 sensor value, 6 constants, 4 logical operators, 6 numerical constants, and 8 operators. Users will be allowed to use the phones in portrait and landscape modus. We will also experiment with different screen sizes and resolutions. Before running the experiment a pilot test will be conducted to discover errors and to obtain extra practice for the research team [8]. We will evaluate several aspects during the pilot test, like the reactions of participants, discovery of errors in the test setup, and the procedure for data processing and analysis. After the pilot test and resulting adaptations to the test procedure, the real test with the null hypothesis presented before will be conducted. The following subsection describes the experimental design and specifies the test metrics.

B. Research Questions

The aim of the study is to compare four different programming language environments described in Section II. We want to find out what type of formula creation and visualization works faster, is better understandable, and preferred by the participants, while using and manipulating complex formulas: a purely visual programming language like Snap! and Blockly, or a hybrid programming language like TouchDevelop and Pocket Code's formula editor. To answer these questions, we will conduct a comprehensive formal experiment. Details of the planned test method and the experimental setup are provided in the following section.

C. Experimental Design

To evaluate the hybrid programming environment approach of formula editing, we are going to conduct a counterbalanced formal experiment with repeated measures. The experiment will take place in a laboratory at Graz University of Technology. We will use the same test setup for all of the tests, only the smartphone for TouchDevelop will be a Windows Phone. All other programming environments will run on Android devices. For the evaluation, we will randomly select 32 participants at the age of 16, because Pocket Code is specifically targeted at teenagers. For other age groups, the programming system will we be adapted to their needs. The participants will be recruited from schools in and around Graz. None of them will have any previous programming experience. The participants will be randomly distributed in four groups. In either case, the participants will spend two hours in the experiment, first learning the basics of the system from a tutorial, and then trying to accomplish a series of tasks. First, they will be asked to create a very simple program in order to get familiar with the programming environment. The order of the tasks will be counterbalanced (each group works with the different programming environments in a different order) between the groups to avoid learning bias [9]. After each task the participants will be asked to fill out a feedback questionnaire for the purpose of collecting subjective qualitative data. The dependent variables that will be measured are

1) time spent for solving each task, while using different programming environments
2) successfully finished tasks
3) tasks finished with help and
4) the number of errors occurred
5) quality of the programs created by the participants, rated by the test team.

Different applications, tasks, and time are going to be the dependent variables. In addition to the quantitative data, we will collect qualitative data as well. After each task, we are going to initiate a discussion with the participants and try to get as much information as possible about their subjective experiences, information about what they liked and disliked, and what would have made the programming language more compelling, more useful, or easier to use.

The data will be collected from three different sources. The most relevant data will be compiled during the task execution. After each task, users will fill out a feedback questionnaire to get some subjective feedback from the participants.

ACKNOWLEDGMENT

Many thanks to the Catrobat team members[9].

REFERENCES

[1] R. K. Pandey and M. M. Burnett, "Is it easier to write matrix manipulation programs visually or textually? An empirical study," in *IEEE SYMP. VISUAL LANGUAGES*, 1993, pp. 24–27.

[2] T. Booth and S. Stumpf, "End-user experiences of visual and textual programming environments for Arduino," in *End-User Development*, ser. Lecture Notes in Computer Science, Y. Dittrich, M. Burnett, A. Mrch, and D. Redmiles, Eds. Springer Berlin Heidelberg, 2013, vol. 7897, pp. 25–39. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-38706-7_4

[3] T. Green and M. Petre, "When visual programs are harder to read than textual programs," in *Proceedings of the Sixth European Conference on Cognitive Ergonomics (ECCE 6)*, 1992, pp. 167–180.

[4] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. S. Silver, B. Silverman, and Y. B. Kafai, "Scratch: programming for all," *Commun. ACM*, vol. 52, no. 11, pp. 60–67, 2009.

[5] D. Garcia, L. Segars, and J. Paley, "Snap! (build your own blocks): tutorial presentation," *J. Comput. Sci. Coll.*, vol. 27, no. 4, pp. 120–121, Apr. 2012.

[6] W. Slany, "A mobile visual programming system for Android smartphones and tablets." in *VL/HCC*, M. Erwig, G. Stapleton, and G. Costagliola, Eds. IEEE, 2012, pp. 265–266.

[7] D. Brown, *Agile User Experience Design: A Practitioner's Guide to Making It Work*. Elsevier Science, 2012.

[8] E. van Teijlingen and V. Hundley, "The importance of pilot studies," *Nurs Stand*, vol. 16, no. 40, pp. 33–6, 2002.

[9] B. A. Kitchenham, S. L. Pfleeger, L. M. Pickard, P. W. Jones, D. C. Hoaglin, and K. E. Emam, "Preliminary guidelines for empirical research in software engineering," *IEEE Transactions on Software Engineering*, vol. 28, pp. 721–734, 2002.

[9] http://catrob.at/credits

VLC Demo/Poster

KeyScretch on Android Tablets and Smartphones

Gennaro Costagliola, Mattia De Rosa, Vittorio Fuccella
University of Salerno
Via Ponte Don Melillo
84084 Fisciano (SA), Italy
{gencos, matderosa, vfuccella}@unisa.it

Abstract

KeyScretch is a text entry method for mobile devices equipped with touch-screens, based on a menu-augmented soft keyboard. It improves the previously studied menu-based methods by enabling the interpretation of compound strokes, corresponding to the input of particularly frequent character sequences. Here we describe the design of an application we developed for the Android system, runnable on tablets and smartphones.

1 Introduction

Text entry is one of the main tasks for mobile device users. Mostly, text is entered to compose SMSs, even though the growing computing capacity of the devices suggests that they will be increasingly used for richer text-based applications. Nevertheless, text entry on mobile devices is inefficient mainly due to their small size.

KeyScretch is a text entry method for mobile devices equipped with touch-screens, based on a menu-augmented soft keyboard. A menu-augmented soft keyboard is a keyboard whose keys have a radial menu surrounding them. Each menu item contains a (particularly frequent) character. The menu is shown as soon as the key is pressed and the selection of a menu item through a radial swipe is a shortcut to enter the character associated to it. *KeyScretch* improves the previously studied menu-based methods by enabling the interpretation of compound strokes, corresponding to the input of particularly frequent character sequences.

The *KeyScretch* text entry method has been presented in previous papers [2, 3]. The results of an analysis of the method's performance on a stylus-based device showed that *KeyScretch* significantly outperforms the baseline tapping-based method, with a crossover occurring after about two hours and a half. Here we describe the design of an application we developed for the Android system, runnable on tablets and smartphones. Special attention was required in the design of the interface, in particular to cope with the problem of occlusion.

The rest of the paper is organized as follows: the next section briefly describes the *KeyScretch* method; Section 3 presents the design of the *KeyScretch* Android application; finally, Section 4 offers our conclusions and outlines the future work.

2 The *KeyScretch* Text Entry Method

KeyScretch has been recently introduced by Costagliola et al [2, 3]. The method allows the user to enter text on a soft keyboard through both *taps* and more articulated gestures. Each gesture is initiated on a key and is driven by the key surrounding menu. A space character is inserted by ending the gesture inside the character key area or through a direct use of the space bar.

KeyScretch can be instantiated differently for different languages. Recently tested instances include an optimized one for the English language with the characters 'o', 'a', 'e' and 's' associated to each of the sides of the character key, and an another one optimized for the Italian language with the vowels 'a', 'e', 'i' and 'o' associated to each of the sides of the character keys, as shown in Figure 1. With this menu layout, the interaction sequence necessary to enter the Italian text *ciao gente* (*hello folks*, in English) is shown in Figure 2. The string is ten characters long but it can be entered with a sequence of four strokes (*taps* or *gestures*).

KeyScretch has been extensively evaluated. In particular, the performance of the above introduced Italian instance has been assessed in a study with eight users, showing that *KeyScretch* significantly outperforms the traditional method based on tapping on a soft keyboard. In particular, a speed crossover was obtained after about two hours and a half of practice, with a final (after about six hours of practice) improvement of about 17%.

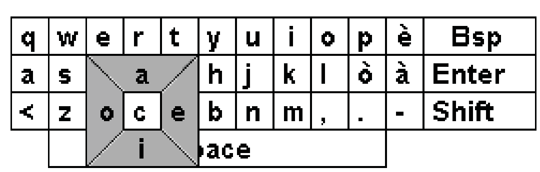


Figure 1: The QWERTY keyboard layout augmented with a menu.

Another interesting result regarding the method is the estimation of the expert performance. We used a predictive model, which revealed that specific instances of the method enable improvements within the range of 30%-49% against the traditional method on the Qwerty layout, according to the language used. Furthermore, the method always outperforms the simple menu-based method.

3 The Android Application

A screenshot of the *KeyScretch Keyboard* Android application is shown in Figure 3. As in most designs, when the user presses a key, a magnifying of the pressed key is shown above the pressed key. To cope with occlusion we also preferred to show the menu in the same frame, rather than around the key.

While the user performs a gesture, its trace is shown in red and when the gesture is completed the recognized template is shown in green for a fraction of a second as a feedback. We also preferred to show the menu permanently in an angle of the keyboard, as in previous experiences we have had some clues that this could help the user in finding a menu item more quickly.

The keyboard prototype was implemented as a customization of the Android *SoftKeyboard* sample project and provided as an Android input service in order to allow the use of the keyboard within any Android application that requires text input, such as email clients, notepads, browsers, and other applications, including a typing video game specifically developed for *KeyScretch* and presented in [1].

The original view was slightly modified to enable the visualization of the menu on the peripheral keys on the right side of the layout. The very low frequency of use of the *scretches* with the peripheral keys on the left side of the layout, instead, allowed us to leave unchanged their distance from the margin.

The application was tested on two different devices: a *Samsung Galaxy Tab* tablet equipped with a 7" screen, at a resolution of 1024x600, with the Android 2.2 operating system installed and an *LG P760 Optimus L9* mobile phone, equipped with a 4.7" display with 540×960 resolution and running the Android 4.0.4 operating system.

The keyboard was used for different experiments aimed at measuring the user performance with *KeyScretch*. During these experiments, besides obtaining quantitative data, we carefully observed the participants and gathered their impressions, both through questionnaires and as freeform feedback. They generally expressed positive comments and their satisfaction level was good. Some issues reported by the participants were related to the size of the keyboard and to the feedback of the gesture recognizer. As regards the former, it is worth noting that, due to the presence of the menu, the keys must be reduced of some extent. As a consequence, the use of the method on very tiny devices might be uncomfortable or error prone and the use of a dictionary or of a language model could greatly improve the user experience. As regards the latter, the participants did not show particular difficulties in using the menu in the learning stages, although the most experienced of them complained that the menu sometimes hid part of the keyboard. This was due to our choice of leaving it active even after the key was released, in order to give the user a feedback on the recognition of the gesture. Special care should be used in choosing the right timing.

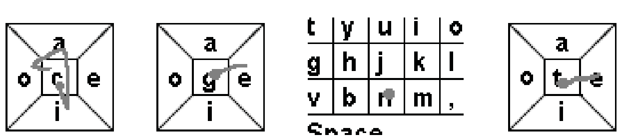


Figure 2: The 4 strokes needed to enter the text ‘ciao gente’.

Figure 3: A screenshot of the *KeyScretch Keyboard* Android application (Italian instance).

4 Conclusions and Future Research

We have presented the design of a keyboard for Android implementing the *KeyScretch* text entry method. The design was driven by our experience, obtained in past user studies. As future work, we are currently working on different techniques to improve the user experience, most of them being related to the accuracy of the method: a chunk-level correction mechanism to reduce the correction activity, an ad hoc designed spell corrector and an effective feedback mechanism telling the user the recognized template *scretch*, through which the user could be immediately conscious of possible typing errors.

References

[1] G. Costagliola, M. De Rosa, V. Fuccella, and F. Torre. Typejump: A typing game for keyscretch. In *Visual Languages and Human-Centric Computing (VL/HCC), 2012 IEEE Symposium on*, pages 249 –250, sept. 2012.

[2] G. Costagliola, V. Fuccella, and M. Di Capua. Text entry with keyscretch. In *Proceedings of the 16th international conference on Intelligent user interfaces*, IUI ’11, pages 277–286, New York, NY, USA, 2011. ACM.

[3] G. Costagliola, V. Fuccella, and M. Di Capua. Interpretation of strokes in radial menus: The case of the keyscretch text entry method. *Journal of Visual Languages and Computing*, 2013. http://dx.doi.org/10.1016/j.jvlc.2013.02.003.

Reviewer's Index

A
Junia Coutinho Anacleto
Tim Arndt
Danilo Avola

B
Arvind K. Bansal
William A. Beasley
Andrew Blake
Paolo Bottoni
Jeff Browne

C
Alfonso F. Cardenas
Augusto Celentano
Ing-Ray Chen
Shu-Ching Chen
Wei Chen
William Cheng-Chung Chu
Yuan-Sun Chu
Gennaro Costagliola
Philip Cox
Alfredo Cuzzocrea

D
Sergiu Dascalu
Andrea De Lucia
Aidan Delaney
Vincenzo Deufemia
Pierpaolo Di Bitonto
Tiansi Dong
David H. C. Du

E
Larbi Esmahi

F
Filomena Ferrucci
Daniela Fogli
Manuel J. Fonseca
Rita Francese
Vittorio Fuccella
Kaori Fujinami
David Fuschi

G
Ombretta Gaggi
Nikolaos Gkalelis
Angela Guercio

H
Niklas Hallberg
Robert Heller

I
Carlos A. Iglesias
Pedro Isaias

J
Qun Jin
Joaquim A Jorge
Erland Jungert

K
Levent Burak Kara
Yau-Hwang Kuo

L
Robert Laurini
Lian Li
Fuhua Lin
Hong Lin
Alan Liu
Jonathan Liu

M
Paolo Maresca
Andrea Molinari

O
Mitsuhiro Okada

P
Sethuraman Panchanathan

Ignazio Passero
Joseph J. Pfeiffer, Jr.
Antonio Piccinno
Giuseppe Polese

R
Michele Risi
Peter Rodgers
Teresa Roselli
Veronica Rossano

S
Giuseppe Scanniello
Tevfik Metin Sezgin
Buntarou Shizuki
Peter Stanchev
Gem Stapleton
Mahbub Syed
Mahbubur Syed

T
Christian Timmerer
Genny Tortora

V
Giuliana Vitiello

W
Danli Wang

Y
Kazuo Yana
Chengjiu Yin
Atsuo Yoshitaka

Z
Ing Tomas Zeman
Kang Zhang
Roberto Zicar

Author's Index

A
Elizabeth Acosta-Gonzaga, 112, 131
Mohamed A. Ahmed, 38
Carmelo Ardito, 79
Saulius Astromskis, 55
Ryan R. Azevedo, 71

B
Robert Baker, 156
Pierfrancesco Bellini, 20
Matt Benatan, 9
Alex N. Borges, 71
Paolo Bottoni, 176
Ivan Bruno, 20
Paolo Buono, 106
Jim Burton, 182

C
Augusto Celentano, 3
Daniele Cenni, 32
Shi-Kuo Chang, 43
Peter Chapman, 182, 188
Ji Cheng, 66
Daniel Chivers, 168
Ashraf ul-Islam Chowdhury, 90
Bee Bee Chua, 61
Luigi Colazzo, 137
Catarina Costa, 71
Peiling Cui, 51

D
Vincenzo Del Fatto, 90
Aidan Delaney, 182, 194
Giuseppe Desolda, 79, 106
Gabriella Dodero, 90
Tiansi Dong, 51

E
Karina Rodriguez Echavarria, 194
Nathan Eloe, 14

F
Boyang Fu, 85

G
Bin Gao, 43
Luca Geraci, 90
Martin Griffin, 194
Angela Guercio, 122

H
Roman Hak, 75
Annemarie Harzl, 198
John Howse, 182

J
Andrea Janes, 55

K
Toshiyuki KAMADA, 95
Eun-Young Kang, 100
Levent Burak Kara, 162
Yasushi KODAMA, 95
Vesna Krnjic, 198
Liqun Kuang, 43

L
Rosa Lanzilotti, 79, 106
Petra le Roux, 128
Jennifer L. Leopold, 14, 145
Greg Lessard, 85
Michael Levison, 85
Huiqing Li, 156
Chia-Yung Lin, 66

M
Paolo Maresca, 122
Silvio Meira, 71
Sergio Mendonça, 71
Andrea Molinari, 137

N
Paolo Nesi, 20, 32
Gary Ng, 100
Kia Ng, 9

O
Ian Oliver, 182
Günay Orbay, 162

P
Jae H. Paik, 100
Michela Paolucci, 20
Kang-Lin Peng, 66
Nicholas J. Pioch, 150
Arno Puder, 100

R
Mohamed Abdur Rahman, 26
Aldo Ramirez-Arellano, 131
Jonathan Reid, 85
Rodrigo G. C. Rocha, 71
Peter Rodgers, 156, 168
Hunvil Rodrigues, 100
Luigi Romano, 118

S
Chaman L. Sabharwal, 14, 145
Franz Schreiner, 198
Taketo SHIMOHIGOSHI, 95
Alberto Sillitti, 55
Wolfgang Slany, 198
Lidia Stanganelli, 122
Gem Stapleton, 182
Giancarlo Succi, 55

T
Yuki TERAWAKI, 95
Manori Thakur, 100
Simon Thompson, 156
Jui-Yuan Tsai, 66
Yu-Chee Tseng, 66

W
Niels R. Walet, 112
Yingze Wang, 43
Colin Ware, 150
Makoto WATANABE, 95
William Wright, 150

Y
Kazuo YANA, 95
Ilmi Yoon, 100

Z
Tomas Zeman, 75

Poster/Demo Presenter's Index

C

Gennaro Costagliola, A-3

D

Mattia De Rosa, A-3

F

Vittorio Fuccella, A-3

DMS

2013

Proceedings of the Nineteenth International Conference on Distributed Multimedia Systems

Printed by
Knowledge Systems Institute
3420 Main Street
Skokie, Illinois 60076
(847) 679-3135
office@ksi.edu | www.ksi.edu
Printed in USA, 2013
ISBN 1-891706-34-9 (Paper)
ISSN 2326-3261 (Print)
ISSN 2326-3318 (Online)