ROMANIAN CONFERENCE ON HUMAN-COMPUTER INTERACTION

Proceedings of the 12th Romanian Conference on Human-Computer Interaction RoCHI 2015, 24-25 September
Bucharest, Romania

Edited By:

Marian Dârdală
Bucharest University of Economics

and

Traian Eugen Rebedea
Politehnica University of Bucharest

MATRIX ROM
Bucharest, 2015

Published by

MATRIX ROM

C.P. 16-162
062510 – Bucharest, Romania
Phone: +4021 4113617, Fax: +4021 4114280
E-mail: office@matrixrom.ro
www.matrixrom.ro

MatrixRom publishing house is a certified publisher by the National Council for Scientific Research in Higher Education (Consiliul Naţional al Cercetării Ştiinţifice din Învăţământul Superior)

Cover graphic design: *Sabin-Corneliu Buraga*, A.I.Cuza University, Iasi
Graphic identity: *Adrian Mironescu*, Idegrafo

Conferinta Natională de Interactiune Om-Calculator (2015)

ISSN 2344 – 1690

http://rochi.utcluj.ro/proceedings/en/index.php

TABLE OF CONTENTS

ORGANIZERS

Conference Chair
Marian Dardala, Bucharest University of Economic Studies, Romania

Program Committee Chair
Traian Eugen Rebedea, University Politehnica of Bucharest, Romania

Organizing Committee Chair
Catalin Emilian Boja, Bucharest University of Economic Studies, Romania

RoCHI Conference Website Administrator
Teodor Stefanut, Technical University Cluj-Napoca, Romania

PROGRAM COMMITTEE

Alexandru Balog, ICI Bucharest, Romania

Sabin-Corneliu Buraga, A.I.Cuza University, Iasi, Romania

Grigoreta Sofia Cojocar, Babes-Bolyai Univeristy, Cluj-Napoca, Romania

Karin Conninx, Hasselt University, Belgium

Marian Dardala, Bucharest University of Economic Studies, Romania

Mihai Dascalu, University Politehnica of Bucharest, Romania

Philippe Dessus, Universite Grenoble Alpes, France

Paul Fogarassy-Neszly, BAUM Engineering, Romania

Peter Forbrig, Rostock University, Germany

Dorian Gorgan, Technical University Cluj-Napoca, Romania

Adriana-Mihaela Guran, Babes-Bolyai University, Cluj-Napoca, Romania

Adrian Iftene, A.I.Cuza University, Iasi, Romania

Dragos Daniel Iordache, ICI Bucharest, Romania

David Lamas, Tallinn University, Estonia

Ana-Maria Marhan, Institute of Philosophy and Psychology of the Romanian Academy, Bucharest, Romania

Cristian Marian Mihaescu, Universitatea din Craiova, Romania

Alin Moldoveanu, University Politehnica of Bucharest, Bucharest

Philippe Palanque, Universite Toulouse, France

Oscar Pastor, Valencia University of Technology, Spain

Lyn Pemberton, Brighton University, UK

Dorin Mircea Popovici, Universitatea Ovidius din Constanta, Romania

Costin Pribeanu, ICI Bucharest, Romania

Traian Eugen Rebedea, University Politehnica of Bucharest, Romania

Adriana-Elena Reveiu, Bucharest University of Economic Studies, Romania

Cristian Rusu, Pontificia Universidad Católica de Valparaíso, Chile

Carmen Santoro, ISTI-CNR, Italy

Dominique Scapin, INRIA, France

Marcin Sikorski, Gdansk University, Poland

Christian Stary, University of Linz, Austria

Teodor Stefanut, Technical University Cluj-Napoca, Romania

Stefan Trausan-Matu, University Politehnica of Bucharest, Romania

Dan-Ioan Tufis, Research Institute in Artificial Intelligence of Romanian Academy, Bucharest, Romania

Jean Vanderdonckt, Universite Catholique de Louvain, Belgium

Gabriel Zamfir, Bucharest University of Economic Studies, Romania

Panayiotis Zaphiris, Cyprus University of Technology, Limasol, Cyprus

ORGANIZING COMMITTEE

Adriana Elena Reveiu, Bucharest University of Economic Studies, Romania
Gabriel Zamfir, Bucharest University of Economic Studies, Romania
Felix Titus Furtuna, Bucharest University of Economic Studies, Romania
Cristian Ionita, Bucharest University of Economic Studies, Romania
Alexandru Barbulescu, Bucharest University of Economic Studies, Romania

ADDITIONAL REVIEWERS

Lenuta Alboaie, A.I.Cuza University, Iasi, Romania
Felix Albu, Valahia University of Targoviste, Romania
Adriana Alexandru, ICI Bucharest, Romania
Victor Bacu, Technical University of Cluj-Napoca, Romania
Oana Balan, University Politehnica of Bucharest, Romania
Dragos Catalin Barbu, ICI Bucharest, Romania
Verginica Barbu Mititelu, Research Institute in Artificial Intelligence of Romanian Academy, Bucharest, Romania
Tiberiu Boros, Research Institute in Artificial Intelligence of Romanian Academy, Bucharest, Romania
Dumitru Clementin Cercel, University Politehnica of Bucharest, Romania
Mihaela Colhon, University of Craiova, Romania
Denisa Copandean, Technical University of Cluj-Napoca, Romania
Liviu Cotfas, Bucharest University of Economic Studies, Romania
Felix Titus Furtuna, Bucharest University of Economic Studies, Romania
Florin Girbacia, Transilvania University of Brasov, Romania
Valentina Marinescu, University of Bucharest, Romania
Cornelia Melenti, Technical University of Cluj-Napoca, Romania
Vasile Danut Mihon, Technical University of Cluj-Napoca, Romania
Maria Gheorghe-Moisii, ICI Bucharest, Romania
Florica Moldoveanu, University Politehnica of Bucharest, Romania
Gabriel Neagu, ICI Bucharest, Romania
Stefan Negru, A.I.Cuza University, Iasi, Romania
Mihaela Ordean, iQuest Technologies, Cluj-Napoca, Romania
Marian Padure, Babes-Bolyai University, Cluj-Napoca, Romania
Camelia Mihaela Popa, Institute of Philosophy and Psychology of the Romanian Academy, Romania
Marinel Popescu, ICMET Craiova, Romania
Dan Rotar, Vasile Alecsandri University of Bacau, Romania
Emil Stanescu, ICI Bucharest, Romania
Dan Stefanescu, University of Memphis, USA
Serban Ungureanu, A.I.Cuza University, Iasi, Romania
Adrian Zafiu, University of Pitesti, Romania

Conference organized with the support of:

Bucharest University of Economic Studies

National Institute for Research & Development in Informatics – ICI Bucharest

RoCHI – ACM SIGCHI Romania

Foreword

In 2015, the Romanian Human-Computer Interaction Conference – RoCHI has reached its 12th edition. It has already become a well-known tradition for RoCHI – the Romanian group of ACM SIGCHI association to organize the national HCI conference each year. Like all the previous editions, RoCHI 2015 provides an interdisciplinary forum for the exchange of ideas, expertise and research results in the field of human-computer interaction. However, this edition of RoCHI has an element of novelty compared to the previous ones, as it is the first edition of the conference that has adopted English as the main language for the conference. We hope that this switch to English will open up the conference to researchers from abroad and will transform RoCHI in the following years from an important national scientific event to a conference that will become interesting at least for the European HCI community.

Previous editions of RoCHI were organized in Bucharest (2004 and 2012 - University Politehnica of Bucharest; 2006 – The Bucharest University of Economic Studies, 2010 - National Institute for Research and Development in Informatics, 2011 - Institute for Artificial Intelligence „Mihai Draganescu"), Cluj-Napoca (2005 - "Babes-Bolyai" University, 2009 and 2013 - Technical University of Cluj-Napoca), Constanta (2007 and 2014 - Ovidius University) and Iasi (2008 - "Alexandru Ioan Cuza" University). This year's edition, RoCHI 2015, is organized for the second time by The Bucharest University of Economic Studies, on Thursday and Friday, September 24-25, 2015.

The conference is publishing within this proceedings two invited papers and 26 regular papers organized in six thematic sessions. These papers were selected from a total of 36 submissions that were sent to RoCHI 2015. After a careful review process, each submission being assigned to a minimum of three and maximum of five reviewers, a total number of 26 short and long papers were accepted.

The invited paper "Human Computer Interaction and societal impact – Can HCI influence public policy making and IT politics?", is signed by a proficient researcher from the HCI community: Jan Gulliksen, Dean of the School of Computer Science and Communication from the KTH Royal Institute of Technology, Stockholm. The topic of the paper aims to bring new insights on how HCI research can impact society at its most and the author advocates for research that has practical relevance and that can contribute to changing and improving society.

In the second invited paper "Social Computing, HCI, Social Change and Peace", authored by Panayiotis Zaphiris, the focus is on another inter-disciplinary research, which starts from the concept of Social Computing and shows how digital technologies can help with peace building. Dr. Panayiotis Zaphiris is another renowned researcher in HCI and Multimedia and he is the Dean of The School of Fine and Applied Arts at Cyprus University of Technology

The first session in the proceedings and also in the conference program is *Usability and evaluation*. The session contains papers on topics that are at the core of HCI research, such as usability studies, assessment of functionality, and models for software design. The four papers are presenting comparisons of different software products software based on their usability and functionality (Constanta Zoie Radulescu, Alexandru Balog, Lidia Bajenaru, Delia Mihaela Radulescu - *Multi-criteria Decision Making software products - a comparison and ranking in terms of usability and functionality*), the results of using a design thinking approach for weight management programs (Paul Doney, Marc Fabri – *Seeking New Insights: A design thinking approach to the development of persuasive technology aimed at supporting clients of a weight management program*), a model for students motivation of using social networks (Dragos Daniel Iordache, Alexandru Balog, Costin Pribeanu, Vincentas Lamanauskas - *A motivational model for Facebook acceptance by university students*), and a study of the accessibility of official websites for local authorities in Romania (Costin Pribeanu, Maria Gheorghe-Moisii, Paul Fogarassy-Neszly - *Accessibility of Romanian municipal websites – conformance with WCAG2*).

The *Gaming and simulation* section contains four papers which are addressing various issues such as: simulation of natural phenomena (Sandra Al-Assaf, Dorian Gorgan - *Realistic Simulation of Environmental Phenomena – Snow Fall and Accumulation*), providing a better interaction for blind or visually-impaired people (Vlad Trifanica, Alexandru Butean, Alin Moldoveanu, Diana Butean - *Gamepad vibration methods to help blind people perceive colors* and Oana Balan, Alin Moldoveanu, Florica Moldoveanu, Alexandru Butean -

Developing a navigational 3D audio game with hierarchical levels of difficulty for the visually impaired players), and automatic assessment of strategies in computer games (Mircea Catalin Catana, Dorian Gorgan – *Analyzing Computer Game Strategies through Visual Techniques*).

In the *Natural language interfaces* section, we have collected five papers. They are addressing several important topics at the border between HCI and natural language processing: automatic text generation (Dragos Alexandru Cojocaru, Ștefan Trăușan-Matu - • *Text Generation Starting from an Ontology*), the design of foreign language learning using syntactic motivated patterns (Mihaela Colhon - *A Design Framework for Foreign Language Learning Applications*), opinion mining and summarization from large datasets (Bogdan Marchiș, Alexandru Țifrea, Mihai Volmer, Traian Rebedea – *Opinion Summarization for Hotel Reviews*), automatic paper annotation using semantic analysis (Ionut Cristian Paraschiv, Mihai Dascalu, Stefan Trausan-Matu, Philippe Dessus - *Automated Paper Annotation with ReaderBench*), and conversational agents for a better interaction with knowledge bases (Dorin Rotarescu, Adrian Bogatu, Traian Rebedea, Stefan Ruseti - *Conversational Agent that Models a Historical Personality*).

The *Mobile human computer interfaces* contains several papers on a novel and popular topic in HCI. The papers are presenting several applications and technologies related to mobile and wearable devices: ultraviolet monitor application for a smart watch (Andrei-Bogdan Baran, Adrian Iftene – *Disruptive Technologies – UV Protect – Smart Watch Application*), learning basic geometric shapes using a mobile device (Bogdan Troanca, Alexandru Butean, Alin Moldoveanu, Oana Balan – *Introducing basic geometric shapes to visually impaired people using a mobile app*), mobile indoor localization and navigation (Szabolcs Orban, Teodor Stefanut – *Indoor Localization and Navigation Using Phone Sensors and a 3D Model of the Building*), and smartphone usage for detecting election frauds (Paul Strîmbeanu, Alexandru Butean, Florica Moldoveanu – *An approach for detecting ID frauds in a traditional voting system using a smartphone stand*).

The four papers from the *Interaction in virtual spaces* session discuss topics such as: user socialization within 3D virtual spaces (George Marian Hotca, Teodor Stefanut, Dorian Gorgan – *Socialization Techniques in Virtual 3D Space*), gesture recognition in a CAVE environment (Saleh Salous, Julien Newton, Laure Leroy, Safwan Chendeb – *Gestural Recognition by a Four-Kinect Module in a CAVE "Le SAS"*), developing a low-cost immersive VR system (Mihai Chifor, Teodor Stefanut – *Immersive Virtual Reality application using Google Cardboard and Leap Motion technologies*), and real-time simulation of crowd dynamics using GPUs (Dan Răzvan Ilieș, Adrian Sabou and Dorian Gorgan – *Real Time Visualization of Crowd Dynamics Scenarios*).

In the *Interactive web applications* section, there are five papers covering a wide range of themes: real time video processing for the web (Cristian Ionita, Alexandru Barbulescu – *Real-time Video Processing in Web Applications*), enhanced searching using semantic information (Cristian Neamtu, Adrian Iftene – *Diversifying Search Results Using Semantic Resources*), motivation for usage of Facebook by Romanian students (Alexandru Balog, Costin Pribeanu, Ion Ivan – *Motives and characteristics of Facebook use by students from a Romanian university*), video search and classification (Daniel Gavril, Adrian Iftene – *Movies Classification*), and polyglot text-to-speech (Paul Fogarassy-Neszly, Zlatomir Zinveliu, Costin Pribeanu – *A software component for polyglot text-to-speech synthesis: beta testing results*).

We cannot end this preface without expressing our appreciation to the members of the scientific committee and to the volunteer reviewers from the RoCHI group who helped for selecting the best papers to be presented at the conference. Moreover, we acknowledge the efforts of all the persons involved in the organization of the RoCHI 2015 Conference and thank them for their efforts!

Bucharest, August 2015

Editors,

Marian Dardala

Traian Eugen Rebedea

Human Computer Interaction and societal impact – Can HCI influence public policymaking and IT politics?

Jan Gulliksen
KTH Royal Institute of technology
Lindstedtsvägen 3,
10044 Stockholm,Sweden
Gulliksen@kth.se

ABSTRACT
Digitalization is a concept to encompass the fastest growing trend for societal development of modern times. Digitalization happens everywhere, in schools, in business, in health care and in our private life. The digitalization gives also opportunities for political decisions and policy making to support turning our society into a modern and efficient society.

Research and research funding organizations are becoming more and more aware of the need to conduct research that proves some form of utility to the society and has some form of practical impact. There are several different ways of making research that has practical relevance and that can contribute to changing and improving society. This keynote paper aims at discussing ways to plan, conduct research with the aim of improving the society and also show how we should make use of our research knowledge and positions to influence politics and public policy making.

Author Keywords
Digitalization; policymaking, usability, accessibility

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Computers and Society; Public Policy Issues.

INTRODUCTION
Information and communications technology (ICT) is an area that substantially will shape the economy of the future; it already provides half of our productivity growth. In the European union (EU) there is currently a quarter of the EU population that has never used the Internet. Yet, in the near future, 90% of jobs will require some level of digital literacy; we risk a "digital divide ". Internet usage is lower among groups already at heightened risk of socio-economic exclusion.

DIGITALIZATION
Digitalization has now become the leading term for the situation in which digital technologies contributes to advancing developments in society and transforming most areas and activities. Consequently there is a need to define the concepts more thoroughly. We can distinguish two major types of digitalization

- Information digitalization – a process in which information is transformed from analogue to digital, making the information structurable, searchable and accessible through the digital channels.
- Societal digitalization – the social and human-revolving process that gradually becomes increasingly difficult to distinguish at all from any part of life. This means that individuals and organizations can communicate and exchange information with other people, organizations and their environment in new ways. Digitalization and the use of ICT-based solutions can help to increase the accessibility and efficiency of both the business and public administration.

DIGITAL AGENDA OF EUROPE
The Digital Agenda presented by the European Commission forms one of the seven pillars of the Europe 2020 strategy that sets objectives for the growth of the European Union (EU) by 2020. The Digital Agenda proposes to better exploit the potential of Information and Communication Technologies (ICTs) in order to foster innovation, economic growth and progress.

- **Digital single market** – on May 6, 2015 the European Commission issued a Strategy for the Digital Single Market. The purpose is to facilitate internet use and services between the European countries without geo-blocking, excessive roaming fees or other factors preventing internet based trade and services, just as it is free to travel, work and trade anywhere in EU.
- **Interoperability and standards** –To be able to enhance the interoperability of devices, applications, data repositories, services and networks, EU must review its standard-setting policy and promote appropriate rules for intellectual property rights.
- **Online trust and security** – The European commission has presented measures on network and information security and the fight against cyber attacks to combat cybercrime, child pornography and breaches of privacy and personal data security.
- **Fast and ultra-fast Internet access** – EU works to establish next generation access networks (NGAs) to achieve competitively priced fast and ultra fast Internet access for all.

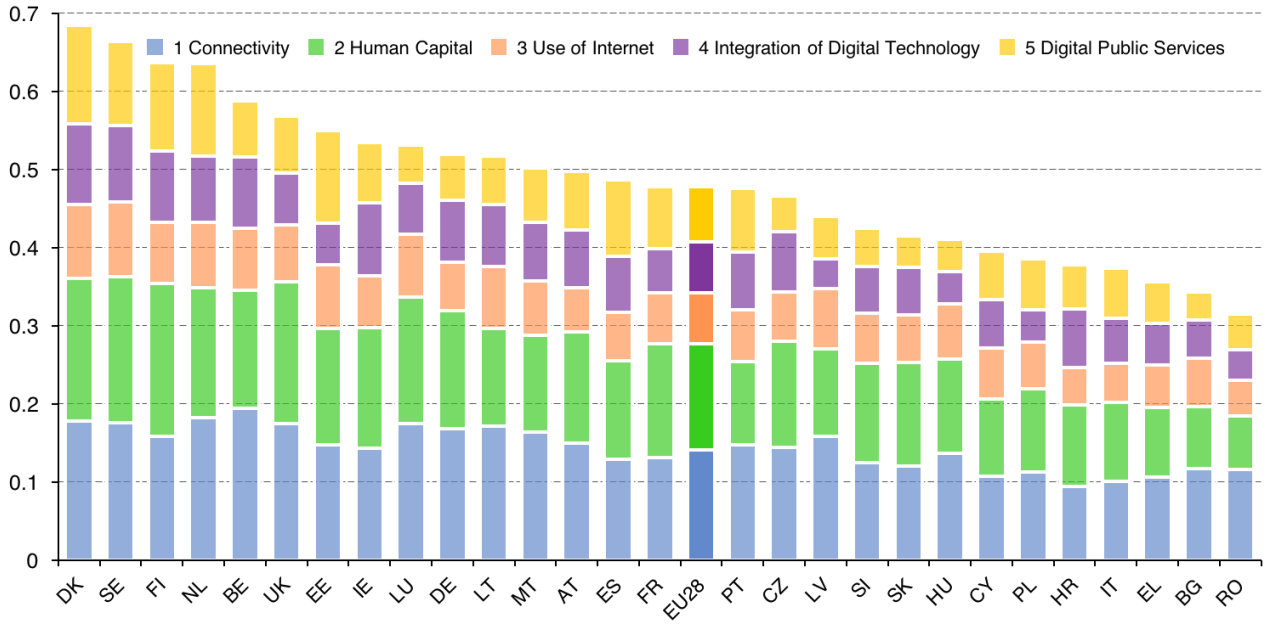

Figure 1: The Digital Economy and Society Index ranks countries based on 5 parameters; Connectivity, Human Capital, Internet Usage, Integration of Digital Technology and Digital Public Services. (Retrieved from the Digital Agenda Scoreboard 2015-09-01)

- **Research and innovation** – with the purpose of catching up with the main competitors in ICT research and digital innovation EU invests in world-class ICT research and innovation through the Horizon 2020 research funding programs to boost growth and jobs via innovative Public-Private Partnerships
- **Enhancing digital literacy, skills and inclusion** – Many parts of the EU population are still excluded from media literacy in the digital environment. EU is facing a crisis of a shortage of employees with digital skills across the EU. In 2013 the EC established the Grand Coalition for Digital Jobs and Skills to promote employment of jobs requiring digital skills across EU. It is a multi-stakeholder partnership to facilitate collaboration among business and education providers, public and private actors to take action attracting young people into ICT education, and to retrain unemployed people.
- **ICT-enabled benefits for EU society** – EU should particularly focus on the following areas: climate change, managing the ageing population, digitalization of content and intelligent transport systems.

Hence the digital agendas clearly outline the goals and ambitions of the IT politics of Europe at a Union level. To be able to fully benefit of these goals, the activities must be manifested on a national level through supporting decisions, activities and similar goals. Therefore many of the European countries have developed their own digital agendas, endorsing the goals of the European union that makes sense to the respective country and additionally to set their own ambitions and goals.

MEASURING AND BENCHMARKING DIGITALIZATION

One way of supporting the development of the digitization of the country is by providing means to assess and benchmark to what extent the goals of the digital agenda are fulfilled and how the development relates to the development of the neighbouring countries. The European Commission has developed a tool for this purpose. The Digital Agenda Scoreboard 1 measures progress of the European digital economy. It provides a mechanism through which every user can adapt the statistics to suit their own needs. As an example (see figure 1) the Digital Economy and Society Index shows a ranking of the European countries in relation to a few central parameters. Such indexes and visualizations helps countries show and set more ambitious goals and to ramp up their investments in the digital agenda

DIGITAL CHAMPIONS FOR SOCIETAL IMPACT

To fully get the power out into the countries there was a need for ambassadors that could help promote and localize the goals to the national needs and expectations. This was the starting point for the initiative to appoint digital champions.

In 2012 the EU commission launched the appointment of digital champions to vitalize the political discussion, to localize EU's missions within the digital agenda and to provide insightful input to the development of politics and policy making within the region. The digital champions have focused their activities on the development of digital

[1] http://ec.europa.eu/digital-agenda/en/digital-agenda-scoreboard

skills, fostering digital inclusion, fostering an open and secure Internet as well as driving different activities to impact the development of the society through digitalization.

The initiative to support the increasing digitization of European union countries by appointing Digital Champions was launched by the EU President Jose Manuel Barroso and the EU Vice-President Neelie Kroes in February 2012. It was inspired by the United Kingdom, which appointed Martha Lane Fox as UK Digital Champion and asked her to advise and challenge the government on how to make greater and faster progress to get people online. Since then 26 European Union Member States have appointed their own national Digital Champions.

According to the initial initiative a Digital Champion would have the following goals:

- be able to network with, understand and energize the industry, education and voluntary sectors;
- be dynamic and high-profile, and therefore able to focus media attention on the initiative; and
- be appointed by central government, based on expertise and merit, ideally reporting to leaders.

The Digital Champions are ambassadors for the Digital Agenda for Europe who are appointed by their Member States to help every European become digital. Each member state has defined their own Champion's role, but in essence it is to promote the role and use of information and communication technologies by connecting people, policies and sectors.

The Digital Champions have different profiles. Some work for government authorities or public bodies such as national libraries, others are entrepreneurs, academics or representatives of civil society. What is uniting them is their passion to help people take advantage of the digital Opportunities by actively working on and promoting digital inclusion, e-government, ICT education, digital skills and entrepreneurship, etc.

Each Digital Champion develops initiatives in their own countries to get people more digital. Digital Champions encounter similar challenges across the European Union. By ensuring regular contact between the Champions, the European Commission provides a platform to discuss and compare action at grass-roots level, and mobilizes citizens and businesses to take advantage of the digital economy. The Champions also advise the Commission on its policies. They meet at least twice a year, and more often virtually.

DIGITAL CHAMPIONS JOINT MISSION STATEMENT

In September 2014 the Digital champions joined forces and expressed a joint mission statement to help every European become digital and benefit from an inclusive Digital Society [2]. The background to the statement was the observation that throughout Europe, politicians and citizens are facing similar changes: our population is aging, we can no longer automatically rely on governments and businesses to generate jobs and we are progressing from an information society to a networked society. These changes bring along significant challenges; digitalization causes concerns about employment and privacy, as well as about social isolation and alienation. However, the growing importance of Internet-driven economic growth is also creating enormous potential for our joint economy and society if we act in a timely and appropriate manner.

In terms of jobs, up to 900,000 digital jobs risk going unfilled by 2020 without pan-European coordination, while 1.2 million jobs could be created through ICT infrastructure construction. But action is also required for reasons beyond the economical; going digital can improve European citizens' lives on many levels. In order to optimally exploit the growth potential of the digital economy and maximize societal benefits of digitalization, European citizens and businesses needed to be mobilized.

As Digital Champions, we express what we stand for and what we see as our current concerns and priorities in the joint mission statement. We describe the main challenges that we see on the road towards achieving an inclusive digital society for all Europeans. These challenges simultaneously represent our main working areas, how we will strive to overcome these challenges.

- **Towards a Digitally Skilled Society** – Our societies need more ICT skills, at work and at home. We must help everybody enhance their ICT skills;
- **Towards an Open Internet –** Everybody has the right to have access to content, apps and services of their choice, everywhere and at any time. We must work to ensure this;
- **Towards a Connected Continent –** Our digital society benefits from faster Internet connectivity. It makes business easier, life more comfortable and prepares us for the future. We welcome the single market and net neutrality and the end of roaming premiums and red tape. We must make Europe more connected;
- **Towards a Secure Digital Environment –** A flourishing digital society is a digitally secure society. Without security there is no trust, without trust we cannot optimally benefit from our digital possibilities. We must work to ensure safety online;
- **Translating Tech Trends to Societal benefit –** Though technology trends can, and do, make a huge difference to society as a whole, the more 'niche' trends still too often remain within the confines of the so-called 'tech-bubble'

[2] https://ec.europa.eu/digital-agenda/en/news/digital-champions-joint-mission-statement

in which they are developed. We must ensure European society at large can optimally benefit from more cutting edge tech-trends at an earlier stage;

- **Endorsing current initiatives and developing future activities –** As Digital Champions we are also active in endorsing, supporting and promoting other parties' activities that support our mission to help every European become digital and benefit from an inclusive digital society, both on a European and on a national level.

HCI RESEARCH AND DIGITALIZATION

Human Computer Interaction (HCI) is a truly multi-disciplinary research field covering such diverse areas as technology, engineering and computer science, on one side, economics, behaviour and social sciences on the other side, but also using design other creative sciences. Although a lot of basic research can and should happen within the field of HCI it is mostly considered as an applied field of research. This means that the research methodologies, tools and techniques for data gathering, analysis and synthesis involves a rich toolbox of data gathering methods such as ethnographic field studies with observations and interviews, experimentation according to the traditions from psychology, constructivist methods and various research methods through design, just to mention a few. With the practical applications right at hand research methods that can be contextualised and applied in practice has a certain place.

Action research methodologies (Reason & Bradbury, 2001) are a family of methods for practice-oriented research that recognize and cherish the obvious fact that we have goals and ambitions with our research, ambitions to improve, develop and change our life. In action research the actual change created through the research project is equally important as the scientific knowledge gathered (ibid).

DISCUSSION

Europe is a heterogeneous region with some of the worlds most highly developed countries and some that are struggling with a weak infrastructure, financial problems, low literacy levels and a high level of unemployment. The challenges and opportunities in developing such a region is very different to the greatest countries in the world, with different languages, lacking collaboration between neighbouring countries and obstacles to competition. The digital agenda has been formulated to set high goals for the entire region and to provide means of benchmarking. The role of digital champions has been defined to drive the development and constitute a channel between the European commission and the local development. The role means recognition and high visibility on a policy level, but at the same time no resources have been allotted to drive the work. The activities are based on volunteer work and other means of resources.

Given the independence of the digital champions, many have chosen the prioritize issues of importance for the development of the society as a whole. Many are developing issues relating to accessibility and social responsibility. Others are focusing on digital skills and the overall needs for more people that are able to work with ICT development. Other again are working to support the local startup business or promoting other activities to decrease unemployment. The needs and priorities are different from region to region.

One of the core values that control the development is the need to emphasize ICTs role for deriving equitable use, that every person, regardless of capabilities should have equal opportunities to be a member of the society in every possible aspect. Regardless of the obstacles to access is based on function, skills, and language or for economical reasons. Equal access is a human right, guiding all the work.

CONCLUSIONS

Digitalization of our society is the single most important process of change that we currently are living through in our society today. Digitalization creates new businesses, new jobs and completely changes the business models and work of traditional companies. To succeed and prosper in the future you need to acknowledge and perhaps also strive to lead the process of digitalization. Research and development need to embrace the process of digitalization and contribute to the development and analyse its consequences to make the development sustainable for the society.

HCI as a field has the knowledge and breadth to cover all necessary aspects of digitalization. Its multidisciplinary nature as well its practice-oriented traditions help to research the complex problems of the society of today. HCI researchers should engage more in the development of the society, more in the effects that the development potentially could have on the society and select research problems that have a bigger potential to move the development of the society in the right direction. Understanding the societal impact and planning for impact becomes increasingly more important

REFERENCES

1. Reason, P., & Bradbury, H. (Eds.). (2001). *Handbook of action research: Participative inquiry and practice*. Sage.
2. Gulliksen, J., Borälv, E., Elvelid, J., Hadley-Kamptz, I., Krusell, J., & Liss-Larsson, N. (2014) En digital agenda i människans tjänst – en ljusnande framtid kan bli vår. Statens Offentliga utredningar SOU 2014:13, Regeringskansliet. Short version in English online: https://digitaliseringskommissionen.se/wp-content/uploads/2014/06/Summary-report-ICT-schools-March-2014.pdf
3. Gulliksen, J., Carlsson, L., Borälv, E., Bergman, S., Hadley-Kamptz, I., Krusell, J., Persson-Stenborg, A., Richter, A. (2015) Gör Sverige i framtiden – digital kompetens. Statens Offentliga utredningar SOU 2015:28, Regeringskansliet. Short version in English online: https://digitaliseringskommissionen.se/wp-content/uploads/2014/06/Summary-Digital-Skills-report-March-2015.pdf

Social Computing, HCI, Social Change and Peace

Panayiotis Zaphiris,
Cyprus University of Technology
Limasol, Cyprus
panayiotis.zaphiris@cut.ac.cy

ABSTRACT

This paper reports work regarding the design, development and evaluation of a surface computing application to support collaborative decision making.. The domain-independent application, so called *Ideas Mapping*, builds on the principle of Affinity Diagramming to allow participants to analyze a problem and brainstorm around possible solutions while they actively construct a consensus artifact -- a taxonomy of their ideas. During idea generation, *Ideas Mapping* replicates physical post-it notes on a multi-touch tabletop. Additional functionality supports student collaboration and interaction around the organization of ideas into thematic categories associated with the problem at hand. We report on the functionality and user experience while interacting with the application which was designed and developed using a user-centered approach. We also report initial findings regarding the affordances of surface computing for collaborative decision making.

Author Keywords
surface computing, user experience, collaborative decision making, CSCL

ACM Classification Keywords
K.3.1 Computer Uses in Education: Collaborative learning

General Terms
Design, Human Factors

INTRODUCTION
A multi-touch interactive tabletop can support collaboration, allowing different patterns of turn taking, negotiation and interaction [5, 2]. In this paper we report the design, development and evaluation of a surface computing application that supports idea generation, collaborative decision making and group artifact construction. The paper starts by covering related research literature and continues with the description of the design and development of *Ideas Mapping* and its use in two studies aiming to understand the affordances of surface computing for collaborative decision making. The paper concludes with a discussion of the key findings and makes suggestions to researchers and practitioners.

EXISTING KNOWLEDGE
The work reported in this paper draws from literature in the areas of Human Computer Interaction (HCI) methods (Affinity Diagramming) and Multi-touch interactive surfaces. The current state of the art in these areas is briefly summarized in this section.

Affinity Diagramming
HCI techniques exist to facilitate discussion in groups and to extract ideas from users' initial conceptual models. For example, the Kawakita Jiro diagrammatic method [8], also known as Affinity Diagramming, is a team-based knowledge elicitation technique. It is used for grouping information into categorical domains [10] and bears similarities to open card sorting. Users write down items of knowledge or descriptions on sticky notes and then organize the notes into groups before creating group headings. These methods are useful to HCI specialists as techniques for creating and analyzing categorizations of knowledge and are considered among the foremost usability methods for investigating a user's (and groups of users') mental model of an information space [9]. In affinity diagramming, the method is enforced in teams usually working on a shared whiteboard or large piece of paper. They are encouraged to communicate their reasoning verbally; thus, collaborative team decisions upon consensus lead to category cluster formation [1].

Multi-touch Interactive Tabletops

Multi-touch interactive tabletops have recently attracted the attention of the HCI and Computer Supported Collaborative Learning (CSCL) communities. Based on preliminary evidence from the education and computer-science literature, Higgins et al. [7] provide a review of the technological characteristics of multi-touch interactive tabletops and their pedagogical affordances. Overall, as pointed out by Higgins et al. [7], most of what we know in this area concerns technical issues related to interaction of users with the technology, but we know little about the use and value of multi-touch tabletops on collaborative learning situations within formal educational settings. Below we summarize some recent empirical evidence related to multi-touch tabletops and learning.

Multi-touch tabletops have been used with disabled user groups to promote development of social skills. SIDES, for example, is a four-player cooperative computer game designed to support adolescents with Asperger's syndrome to practice social skills and effective group work during their group therapy sessions [11]. SIDES provided an engaging experience for this audience who remained engaged in the activity the entire time and learned from the activity (unlike typical behavior of this population) [11]. Similarly, StoryTable has been used to facilitate collaboration and social interaction for children with autistic spectrum disorder with positive effects [4].

StoryTable was initially designed to support children's storytelling activity in groups [3]; Evaluation of StoryTable showed that it enforced cooperation between children during the storytelling activity, by allowing simultaneous work on different tasks, while forcing them to perform crucial operations together in order to progress [3]. In some other work, multi-touch tabletops have been studied for their added benefits compared to single-touch tabletops. Harris et al. [5] contrasted groups of children in multi-touch and single-touch conditions and found that children talked more about the task in the multi-touch condition while in the single-touch condition; they talked more about turn taking. Furthermore, the technology is considered engaging For example, the overall (perceived) usefulness and benefit of using interactive tabletops in collaboration contexts was assessed in a recent experiment by [2] with 80 participants. That study showed that groups in the tabletop condition had improved subjective experience and increased motivation to engage in the task.

With regards to using tabletops in formal learning settings, a series of studies are currently being conducted as part of the SynergyNet project [7]. SynergyNet goes beyond using single tables to studying a network of tabletops that can communicate with each other. SynergyNet focuses on how this technology can best support collaboration within small groups, while undertaking the development of curricula and tabletop applications for classroom integration [7]. A recent SynergyNet study contrasted groups of children in multi-touch and paper-based conditions to examine the differences in their collaborative learning strategies [7]. The authors found that student groups in the multi-touch condition maintained better joint attention on the task than groups in the paper-based condition. Another recent SynergyNet study examined NumberNet, a tool designed to promote within and between group collaboration in a mathematic classroom using a network of tabletops [6]. In this study, pilot results from 32 students showed significant knowledge gains from pre to post testing.

DESIGN METHODOLOGY

We adopted a strongly user-centered approach, emphasizing the engagement of students and instructors in all phases of the design process. Four university students and three instructors were involved, contributing to design elements of the application.

First, through low-fidelity paper-based prototypes, we simulated a collaborative activity with four students around a (turned-off) tabletop using paper and pencil. The scenario involved "the creation of a computer games industry in Cyprus and the factors involved." First, students generated ideas individually for 10 minutes. They wrote a (physical) post-it note for each new idea. Next, the ideas appeared one-by-one on the table and became subject to discussion, after a brief explanation from their originator, in an effort to categorize them in thematic units. Students revisited and changed ideas, rejected less promising ones, and generated new ideas during a collaborative decision making process leading to their thematic categorization. Finally, the activity concluded with a consensus of the main factors (i.e., resulting thematic categories) involved in the creation of a computer games industry in Cyprus. After the completion of the activity, instructors (who observed and kept records of all interactions during the activity) and students discussed the potential surface computing application and contributed to elements of the design from their own viewpoints.

Following the low-fidelity design discussions and analysis of user needs, a prototype Beta version application was developed in Action Script 3.0, for a multi-touch tabletop, the MagixTable. The application, so called *Ideas Mapping*, was designed to be domain-independent with a mild learnability curve. Our participants were called back to collaborate on different scenarios using *Ideas Mapping* and provide feedback on its user experience and further suggestions for improvement. Evaluation sessions took place in a fully equipped usability lab and all sessions were video recorded and analyzed. *Ideas Mapping* was optimized and finalized in three major iterative cycles of design, development and evaluation.

OVERVIEW OF THE APPLICATION

Overall, *Ideas Mapping* is designed to support idea generation, collaborative decision making and group artifact construction. The application builds on the principle of Affinity Diagramming to allow participants to analyze a problem and brainstorm around possible solutions while they actively construct a consensus artifact; namely, a taxonomy of their ideas. This is done in three stages:

Stage 1: With a scenario at hand, each collaborator generates new ideas. Ideas are typed into a web application (producing an XML file associated with *Ideas Mapping*) through the use of a mobile device (laptop, tablet, smartphone connected to the Internet). The need for the integration of mobile devices and a web application emerged from a constraint imposed by the MagixTable (also true for other platforms such as the MS Surface) -- that text entry can be done from one pre-existing keyboard at a time. For the kind of activity we sought, this constraint would be significant. To resolve this problem, we developed four virtual keyboards on the tabletop (one for each user). However, users experienced difficulties typing extended ideas on the virtual keyboard during stage 1; the keyboard interaction suffered from input latency and mistyping issues. Thus, the use of mobile devices for input via a web application was considered as a practical solution to this problem for stage 1. This problem demonstrates both the still existing technical limitations of tabletops but also the importance of user input in developing applications for such technologies.

Stage 2: Next, the ideas are presented one-by-one, as digital post-it notes in the middle of the tabletop surface and

become subject to discussion amongst the collaborators. For each idea, collaborators make an effort to categorize it in a thematic unit. Functionalities include:

- Each post-it note must be categorized before the next one appears. If controversy exists, an idea can be placed in the "Decide Later" depository to be revisited upon the categorization of other ideas. Post-it notes are automatically oriented to face their contributor, which encourages them to elaborate on the idea. This functionality was implemented as a result of users' feedback and is consistent with previous work by [12] showing that orientation can play an important role in collaborative interactions around tabletops by signifying ownership and directing attention.
- Thematic units can be created by any participant using the virtual keyboard. Once a participant begins the categorization of an idea (e.g., either begins to type a thematic unit or simply touches the post-it note), others must wait as only one keyboard is presented at any given time. Thematic units can be renamed if needed.
- Participants can drag and drop a post-it note over a thematic unit to categorize it. Post-it notes can be manipulated in order to move them across the surface, rotate and resize them.
- In this stage participants cannot edit ideas, or generate new ideas notes, and thematic units cannot be deleted. These design decisions aimed to scaffold the collaborative activity by allowing time for learners to consider all contributed ideas before making significant decisions.

Stage 3: In this last stage, more flexibility is given to the participants to finalize their taxonomy. In addition to the above, users can now edit ideas or generate new ones, delete ideas or thematic units that are less promising, and reallocate ideas into thematic units for a better fit. Overall, students engage in a collaborative decision making process, leading to the construction of a group artifact -- a taxonomy of their ideas.

STUDIES WITH IDEAS MAPPING

To examine the affordances of surface computing for collaborative decision making two studies were conducted with groups of university students: a small pilot study and a larger scale investigation.

THE PILOT STUDY

Participants and Setting:

Four university students, aged between 22-27 years old, were recruited to participate in a short activity around the tabletop. The scenario involved the "creation of an action plan that can improve university students' experiences at the Cyprus University of Technology, including social and educational aspects." The session was video recorded and analyzed.

Video Analysis and Preliminary Findings

An exploratory approach was used to trace the kinds of interactions amongst the collaborators and the technology and to better understand the role of tabletops in supporting learning. General research questions guided our video analysis such as: what kinds of interactions take place around the tabletop? and what evidence is present regarding the value of multitouch interactive tabletops for collaborative decision making?

One of the researchers considered the video corpus in its entirety – a total of 57 minutes. Most interaction occurred during the 2nd and 3rd stages of Ideas Mapping, which became the focus of the analysis. The researcher repeatedly watched the video, marked segments of interest, and created transcripts, in an effort to categorize the types of discourse and gestures used by the group members around the tabletop. A preliminary coding scheme is presented in Table 1. This coding scheme will be further refined as more studies are conducted in this context. Understanding collaborative decision making around tabletops is currently limited. It is thus important to establish a coding scheme of the interactions evident around this technology (particularly, the synergetic dialog and physical gestures) to be able to examine the phenomenon further. Ultimately, the coding scheme should help us examine interesting patterns of collaborative decision making around multi-touch interactive tabletops.

Spoken Contributions
• Information Sharing – Defining/describing/identifying the problem
• Proposing – Proposing a thematic unit/new idea
• Elaborating – Building on previous statements, Clarifying
• Negotiating meaning – Evaluation of proposal, Questioning/ answering, Expressing agreement/disagreement, Providing arguments for/against
• Stating consensus – Summarizing ideas, Metacognitive reflections
• Other talk – Tool-related talk, Social talk, Laughter
Gesture Contributions
• Communicative Gestures – Show on the table without touching, Dominating/blocking gestures
• Touch Gestures – Resize, Rotate, Type, Move something across, Random touching or touching to explore

Table 1: Preliminary Coding Scheme

Overall, the pilot study provided initial evidence that the CSCL setting encouraged and stimulated discussion and physical interaction around shared artifacts.

LARGER SCALE INVESTIGATION

Participants

To further examine the value of multitouch interactive tables for collaborative decision making, we recruited

postgraduate students in Cyprus to discuss a scenario related to peace. The sample was composed of 17 postgraduate students enrolled in a CSCL/CSCW course at a public university in Cyprus, aged between 22-45 years old (M=30).

The participants were divided into five groups: 3 groups of 3 students and 2 groups of 4 students, suitable for the four-sided tabletop. Group members were familiar with working together through other course learning activities. All, but one student, had no prior experience with using a multi-touch tabletop.

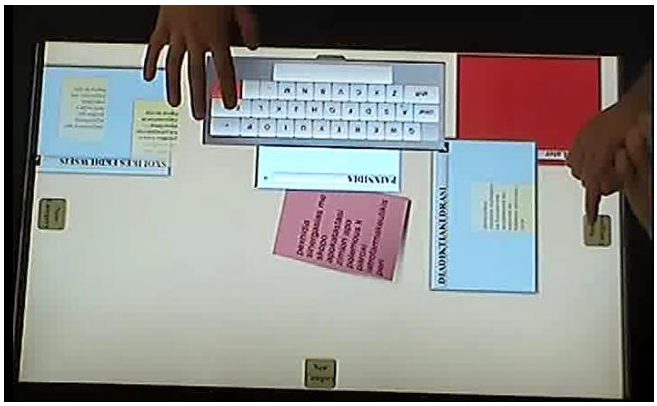

Figure 2: Categorization of ideas it in thematic units

Procedures

In this study there was a preparatory phase before students engaged in group work around the tabletop. That is, Stage 1 of *Ideas Mapping* was completed in distance, during the week before the tabletop investigation. The preparatory week aimed to allow students to research the scenario and think at their own pace. During the preparation week, students were tasked to investigate the topic, think creatively and record at least 10 ideas into the *Ideas Mapping* web application.

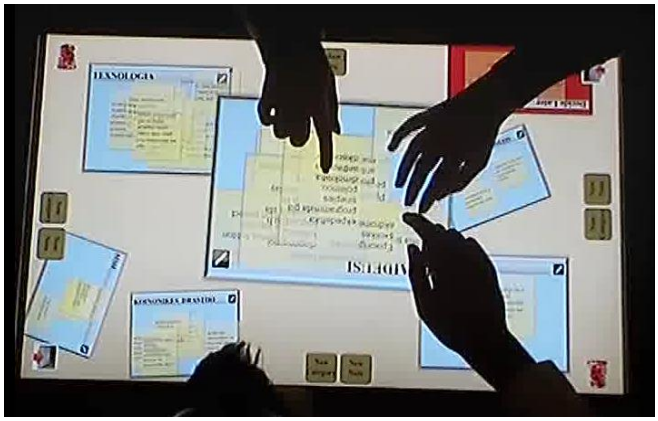

Figure 3: Consensus on a group artifact

The following scenario was presented to the students: "Your team works at a non-governmental organization dealing with global peace. Your project is to create a roadmap of actions to promote global peace using technology." The specific scenario was chosen for it to be thought-provoking and without obvious answers to it. The goal was to stimulate critical thinking, dialog, and creative problem solving. Cyprus is a country in a long lasting political conflict. Thus, the topic was both personally important for the student participants, but also required their emotional and mental engagement.

The next phase involved collocated collaboration around the tabletop. Following the prep week, each group met face-to-face and engaged in collaborative work as described in Stages 2 and 3 of *Ideas Mapping*. Briefly, the ideas of each group were presented on the tabletop one-by-one. Students engaged in discussion and physical interaction with the tabletop in an effort to categorize the ideas in thematic units (i.e., taxonomy of ideas).

Data collection

The sessions of all five groups were video recorded for subsequent utterance coding and analysis. To complement the video data, a questionnaire was administered to all participants soon after the completion of the activity. The questionnaire aimed to assess students' perceptions of the collaborative learning experience and the usability of the surface computing application.

Video Analysis

An extensive video analysis of the data was carried out but due to the scope and space limitations of this paper the results will be presented elsewhere.

Quantitative Data Analysis

The questionnaire included 30 Likert-type items with a 7-point agreement response scale (from 1: completely disagree to 7: completely agree). These items measured three constructs of interest: (1) *Collaboration Support*, assessing the extent to which students thought the technology supported their collaboration such as, "The technology helped me work

effectively in my group", "The technology met my needs as a collaborator"; (2) *Learning Experience*, assessing the extent to which students were satisfied with their learning experience overall, such as "Overall, my collaborative learning experience was positive", "I am satisfied with my experience through this activity", and (3) *Usability Satisfaction* (adapted from Lewis, 1995), assessing the extent to which students were satisfied with the usability of the system such as, "It was simple to use this system", "I can effectively complete my work using this system", "I like using the interface of this system".

A total of 17 students completed the questionnaire. First, the internal consistency for each subscale was assessed using Cronbach's alpha; all 3 subscales had acceptable internal consistency (Cronbach's alphas > .80). Then, subscale mean scores were calculated for every participant (i.e., an un-weighted composite score for each participant on each subscale) followed by computation of descriptive statistics. As shown in Table 2, means were well above the midpoint of the 7-point response scale for all three measures, suggesting that the technology was positively endorsed by the participants overall. Specifically, the participants thought the technology supported their collaboration (*M*=5.53, *SD*= .22), and were satisfied with their learning experience (*M*=5.77, *SD*= .51). With regards to the third measure, participants found the system usable overall (*M*=4.93, SD= .77), but individual item means pointed to some aspects which may need improvement. The rating average was lower for three particular items in this scale, suggesting that we should improve the way participants recover from mistakes ("The system gives error messages that clearly tell me how to fix problems" *M*=3:00 and "Whenever I make a mistake using the system, I recover easily and quickly" *M*=3.36), as well as extend the application to include more functionality ("This system has all the functions and capabilities I expect it to have" *M*=3.88).

The questionnaire also included an open-ended question concerning the pros and cons of using tabletops for collaborative learning activities. We reviewed students' responses to identify themes. Several students commented on how the tabletop promoted collaboration, helped them maintain attention to the task and was enjoyable to use. For example, one of the participants commented: *"The tabletop helped us collaborate and the resulting product was a group effort. It helps you pay attention. I also found it very enjoyable".* Often, students pointed out the capabilities of the system that enabled effective collaboration, such as "*It was nice all of us could use the tools at the same time, to rotate a note, to make it larger to read, or to put it in the box to revisit later."* On the negative side, a few participants found the virtual keyboard difficult to use and that the system needed improvement in handling mistakes, which was consistent with the findings from the quantitative data. These results confirmed our views regarding the affordances of multi-touch tabletops to support collaboration activities and also contributed to further refinement of *Ideas Mapping*.

Subscale	# Items	Cronbach's Alpha	M (SD)
1. Collaboration Support	6	.94	5.53 (.22)
2. Learning Experience	5	.96	5.77 (.51)
3. Usability Satisfaction	19	.97	4.93 (.77)

Table 2: Subscales statistics and descriptive statistics (N=17)

DISCUSSION - CONCLUSION

This study reports on the functionality and user experience while interacting with a multitouch application which was designed and developed using a user-centered approach. We also report initial findings regarding the affordances of surface computing for collaborative decision making.

Ideas Mapping builds on the principle of Affinity Diagramming to allow participants to analyze a problem and brainstorm around possible solutions while they actively construct a consensus artifact -- a taxonomy of their ideas. We feel *Ideas Mapping* makes the Affinity Diagramming technique more collaborative. By allowing for an extension sorting activity, it provides a way for participants to negotiate around an emerging group artifact and make sense of challenging problems, such as how to promote world peace using technology.

We further have evidence that the CSCL setting of the study, and surface computing more generally, encouraged and stimulated dialog and collaborative work around an authentic problem . Following the individual generation of ideas, *Ideas Mapping* supported a 2-stage collaborative activity that promoted ideas sharing, negotiating, sorting and constructing a group artifact while coming to a consensus.

Moreover, we believe that traditional user experience evaluation methods (e.g. questionnaires) were useful for evaluating Ideas Mapping. However qualitative evaluation (e.g. video analysis and the establishment of a coding scheme) is also important; such methods can reveal interesting patterns of interactions amongst the participants and with the technology beyond what is self-reported.

Below, we identify some implications of this work for future research and practice in the fields of HCI and CSCL.

Suggestions to Practitioners:

1. Designers should focus on engaging students and instructors in the design process of educational surfaces computing applications.
2. Current interactive tabletop technologies come with a lot of user interface limitations. These should be taken into account when designing applications for such surfaces.

3. The CSCL setting of the study encouraged and stimulated active dialogue with a problem at hand and a multitouch interactive tabletop application to support them.
4. Self-reported measures showed that students positively endorsed the use of multitouch interactive tabletops for small group work.

Suggestions to Researchers:

1. The proposed coding scheme can be applied and extended to more studies in the area.
2. New qualitative analysis methodologies for evaluating user experience are needed.
3. The role of surface computing in promoting dialogue around sensitive topics (like peace) is an interesting area for further research.
4. A framework for using surface computing for collaborative decision making in general (especially related to sensitive issues) can be developed and tested.

ACKNOWLEDGEMENT

This project is funded under the Cyprus Research Promotion Foundation's Framework Programme for Research, Technological Development and Innovation 2009-2010 (DESMI 2009-2010), co-funded by the Republic of Cyprus and the European Regional Development Fund, and specifically under "Bilateral Cooperation" (ΔΙΑΚΡΑΤΙΚΕΣ/ΚΥ-ΣΛΟ/0411)

REFERENCES

1. Beyer, H., & Holtzblatt, K. (1998). Contextual design: Defining customer-centered systems. California: Morgan Kaufmann Publishers CA, USA.
2. Buisine, S., Besacier, G., Aoussat, A., Vernier, F. (in press). How do interactive tabletop systems influence collaboration? *Computers in Human Behavior*.
3. Cappelletti, A., Gelmini, G., Pianesi, F., Rossi, F., & Zancanaro, M. (2004). Enforcing cooperative storytelling: First studies. In *Proceedings of the 4th IEEE Intentional Conference in Advanced Learning Technologies*, Joensuu, Finland.
4. Gal, E., ,N. Goren-Bar, D., Pianesi, F., Stock, O., Zancanaro, M. & Weiss, P. (2009). Enhancing social communication of children with high-functioning autism through a co-located interface. AI & SOCIETY, 24 (1), 75-84, DOI: 10.1007/s00146-009-0199-0
5. Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., et al. (2009). Around the table: Are multipletouch surfaces better than single-touch for children's collaborative interactions? In *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning* (pp. 335–344). ISLS.
6. Hatch, A., Higgins, S., Joyce-Gibbons, A. & Mercier, E. (2011) NumberNet: Using Multi-touch Technology to Support Within and Between Group Mathematics Learning. In Proceedings of the 9th International Conference on Computer Supported Collaborative Learning (pp. 176-183). International Society of the Learning Sciences.
7. Higgins, S.E., Mercier, E.M., Burd, E., & Hatch, A., (2011) Multi-touch Tables and the Relationship with Collaborative Classroom Pedagogies: A Synthetic Review. International Journal of Computer-Supported Collaborative Learning. 6 (4), 515-538. DOI: 10.1007/s11412-011-9131-y
8. Kawakita, J. (1975). KJ method: A scientific approach to problem solving. Tokyo, Japan: Kawakita Research Institute.
9. Martin, S. (1999). Cluster analysis for Web site organization. Internetworking 2, 3. Retrieved February 29, 2012, from http://www.internettg.org/newsletter/dec99/cluster_analysis.html
10. Nielsen, J. (1994). Usability engineering. CA, USA: Morgan Kaufman.
11. Piper, A. M., O'Brien, E., Morris, M. R., & Winograd, T. (2006), SIDES: A cooperative tabletop computer game for social skills development. In *Proceedings of the 6th International Conference on Computer Supported Cooperative Work*, Alberta, Canada.
12. Wigdor, D., & Balakrishnan, R. (2005). Empirical investigation into the effect of orientation on text readability in tabletop displays. In Proceedings of Computer-Supported Cooperative Work (Vol. 18, p. 22). Springer.Retrieved 15.06.2010 from http://www.springerlink.com/index/p62165406wv725r4.pdf.

Multi-Criteria Decision Making Software Products - A Comparison and Ranking in Terms of Usability and Functionality

Constanta Zoie Radulescu, Alexandru Balog, Lidia Bajenaru, Delia Mihaela Radulescu
National Institute for R&D in Informatics
Bucharest, Romania
radulescucz@yahoo.com,alexb@ici.ro,lidia.bajenaru@yahoo.com,
delia.mihaela2010@gmail.com

ABSTRACT
A large number of multi criteria decision making (MCDM) software products have been developed in academia and business. The selection of the appropriate MCDM software product to solve a given decision problem is a difficult task. Users need adequate MCDM software that meets quality standards regarding usability and functionality, utility, reliability and computer efficiency. In this paper we make a comparison of six MCDM software products regarding usability and functional suitability. Then we rank the MCDM software products included in comparison, first for criteria of usability and functionality and second only for criteria of usability. Ranking the six MCDM software products is a multi-criteria problem itself.

Author Keywords
Multi criteria decision making; Usability; Software Quality; Functional Suitability; Alternatives; Criteria.

ACM Classification Keywords
H.4. Information Systems Applications: Types of Systems: Decision support.

INTRODUCTION
Choosing an appropriate decision problem structure, an appropriate decision method and a decision software product in order to solve a given decision problem are important steps in achieving good decisions.

A MCDM problem consists in selecting an alternative from a set of several available alternatives. The alternatives are characterized by several criteria which are usually in conflict. MCDM offer various advantages: they allow the consideration of conflicting criteria, provide a structure and an organization that guide a transparent analysis process and can handle both qualitative and quantitative criteria [4].

With the advance of modern computing technology, a great number of software products that support multi-criteria decision were realized. Various software products have been developed based on MCDM methods. In OR/MS Today' October 2014 for Decision Analysis Software Survey (http://www.orms-today.org/surveys/das/das.html) 38 software packages were listed. Selection of an appropriate MCDM software product for solving a given decision problem is a difficult task. Users often seek adequate MCDM software that satisfies some minimal quality standards in terms of usability, utility, reliability and computer efficiency.

In recent years, the research on standards for software quality has begun to gain great importance. The quality is determined not only by the software product, but also by the context in which it is used: the particular users, tasks and environments. The usability attributes which contribute to software quality will include the style and properties of the user interface, the dialogue structure, and the nature of the functionality.

The paper is organized as follows. In the next section we present short considerations to ISO Standards for Software Quality. Then, we present the common input data involved in general MCDM decision problems and criteria for selection a MCDM method.

We make a comparison of six MCDM software products regarding usability and functional suitability. Finally we rank MCDM software products included in comparison, first for criteria of usability and functionality and second only for criteria of usability. Ranking the six MCDM software products is a multi-criteria problem itself. Our aim is to help (a) potential users to select a suitable MCDM software product that is more compatible with their needs and (b) software developers to improve the MCDM interactive software products.

THE ISO STANDARDS ON SOFTWARE QUALITY
According to series ISO/IEC 25000 [8], a quality model is a "defined set of characteristics and of relationships between them, which provides a framework for specifying quality requirements and evaluating quality". The software quality models have been analyzed for many years. The ISO/IEC 9126 is the best known reference in this area [2]. This standard provides a very general quality model for software products, based on a set of 6 quality characteristics (Functionality, Reliability, Usability, Efficiency, Maintainability, Portability) and 27 sub-characteristics. This standard has been replaced by ISO/IEC 25010 [9], which updates the previous quality model in various ways.

An important characteristic for selection of a software product is usability. There are many researches that aim to measure usability. Usability is the effectiveness, efficiency, and satisfaction with which specified users achieve

specified goals in particular environments. There is no consensus agreement on this definition; it might refer to the user interface, ease of use or user friendliness [1, 3, 5].

The ISO standards provide a general conceptual framework for defining the quality model for complex systems with a substantial software component. To be of practical use, these standards must be tailored to the specific class of software systems under consideration. This may not be a simple task, especially when these software systems do not fit well with the systems considered in the classical software products, such as ERP, command & control, embedded systems. This is the case of MCDM software products, which possess a number of characteristics that greatly differentiate them from the above systems.

MULTI-CRITERIA DECISION PROBLEMS

Multi-criteria decision problems share the following common input data:

- A set of decision makers (individual or team decision makers / analysts / experts) assessing and finally selecting the most appropriate solution in relation to requirements.

- A set of alternatives to be evaluated and from which the most suitable alternative will be selected.

- A set of criteria. Each alternative is evaluated with the help of criteria taking into account the preferences of the decision maker and a rating scale (quantitative or qualitative). To each criterion can be associated a coefficient of importance (weight).

After setting the input data, the MCDM method to solve the decision problem is chosen. Literature is rich with different types of MCDM methods [6, 7, 10]. There is no single MCDM method which can be a superior method for solving all decision-making problems. Different researchers have different point of views on this issue. The selection of a MCDM method must take into account: (a) the type of the decision problem, (b) the number of the alternatives considered, (c) the criteria features, (d) the easiness of use and (e) the decision maker skills.

COMPARATIVE ANALYSIS OF MCDM SOFTWARE PRODUCTS

Various MCDM software products or decision support systems (DSS) have been developed to support the use of MCDM in practice. Besides computational support for implementing the methods and the calculation of the results, the software usually provide various ways to also support other phases of the process, such as construction of the model and analysis of the results. Especially, the graphical user interfaces can provide various possibilities to view the process and the results, and consequently make the understanding of the results more transparent.

From the set of MCDM decision software products we selected six MCDM software products in order to perform our analysis, mainly based on the availability of some demo or trial versions of the software product.

The MCDM software products considered in this paper are following:

1. **1000Minds** (Free for academic purposes) – The software product supports decision-making, prioritization and the discovery of decision makers' preferences. 1000Minds is based on PAPRIKA (Potentially All Pairwise Rankings of All Possible Alternatives) method and is Web-based software with a tab-based interface.

2. **Analytica** (Lumina Decision Systems, Inc.). It helps in building business models or policy analysis. Has intuitive influence diagrams for creating models and allows communicating clearly with colleagues and clients. Analytica has Object-oriented visual interface, with which one can implement practically any method. Analytica has various graph-building and pre-defined modules available, for example, for MAUT, optimization, and risk analysis.

3. **Criterium Decision Plus 3.0** (InfoHarvest). It can be used for managing the entire decision process. Criterium Decision Plus includes Direct Tradeoffs, basic MAVT software with AHP functionality, larger models, powerful graphics and extensive options for supporting decision making.

4. **V.I.S.A. Decisions** (SIMUL8 Corporation Ltd) - It allows weighing up all the factors using a considered and sound process and documents how decision was made and why it was the right outcome for future reference. V.I.S.A. Decisions is based on MAVT method.

5. **Multicrit** (ICI Bucharest) - Multicrit assists decision-makers in structuring and analyzing complex problems, following which decisions can be made as close to the purpose. Multicrit is based on a set of multi-criteria decision-making methods: TOPSIS, ONICESCU and WSM held in a methods base.

6. **Logical Decisions** (Logical Decisions). It allows evaluating choices by considering many variables at once, separating facts from value judgments and explaining choices to others. Provides a variety of methods for assessing attribute weights and has many results displays. Logical Decisions is based on the MAVT software with the AHP functionality.

We make a comparison between these six MCDM software products in term of the interfaces and functions which they have in user interaction (usability and functional suitability). Comparative analysis of the MCDM software products is presented in Table 1.

Ranking the six MCDM software products is a multi-criteria problem itself. We rank the six MCDM software products considered, based on the nine criteria defined, using the PAPRIKA method.

Table 1. Comparative analysis of the MCDM software products

Criteria	Alternatives					
	1000Minds	Analytica	Criterium Decision Plus	Logical Decisions	Multicrit	V.I.S.A
Usability						
1. Language of the user interface	English	English	English	English	Romanian	English
2. Decision Process interface support	Yes	No	No	NO	No	No
3. Visual scoring	No	Yes	Yes	Yes	No	Yes
4. Visual weighting	No	Yes	Yes	Yes	Yes	Yes
5. Level of user expertise	2	3	3	3	2	3
Functional suitability						
1. Supported MCDM methods	PAPRIKA, MAUT/MAVT	MAUT/MAVT	MAVT, AHP, SMART	AHP, MAUT	TOPSIS, ONICESCU,SAW	MAVT
2. Hierarchical method	No	Yes	Yes	Yes	No	Yes
3. AHP/Pair wise comparison	Yes	No	Yes	Yes	No	No
4. Modeling by uncertainties	No	Yes	No	No	No	No

The preference weights (determined by pairwise comparisons in the PAPRIKA method) are:
1. Language of user interface: 0.019,
2. Decision Process interface support: 0.278,
3. Visual scoring: 0.093,
4. Visual weighting: 0.130,
5. Level of user expertise: 0.074,
6. Supported MCDM methods: 0.259,
7. Hierarchical method: 0.037,
8. AHP/Pair wise comparison: 0.037,
9. Modeling by uncertainties: 0.074.

The Radar Chart of criteria weights is presented in Figure 1.

In Table 2 the evaluation matrix and the alternatives ranking are presented. The optimal MCDM software product for this criteria (usability and functionality) and defined preference weights is Criterium Decision Plus.

For criteria considered only for Usability the evaluation matrix and the alternatives ranking are presented in Table 3. The optimal MCDM software product for the usability criteria and the defined preference weights is 1000Minds. Analytica, Criterium Decision Plus, Logical Decisions and V.I.S.A have equal ranks.

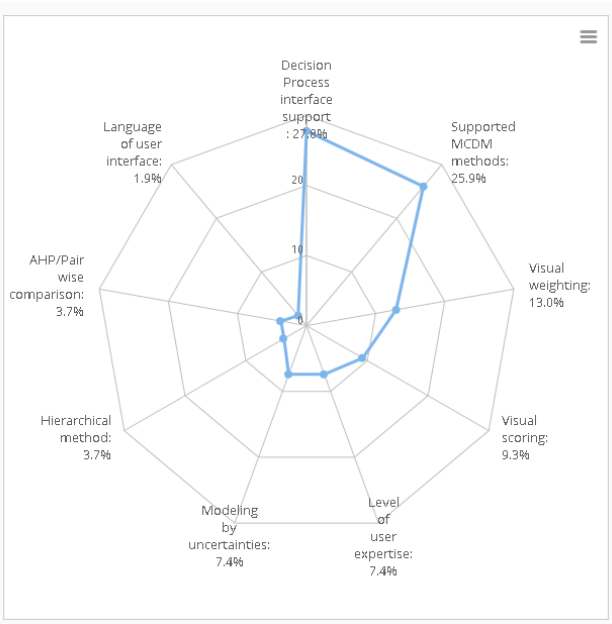

Figure 1. Radar Chart of criteria weights

Table 2. The ranked alternatives for 9 criteria (usability and functionality)

Alternatives	Criteria										
	Language of user interface	Decision Process interface support	Visual scoring	Visual weighting	Level of user expertise	Nr. of supported MCDM methods	Hierar-chical method	AHP/ Pair wise comparison	Modeling by uncert-ainties	Rank	Total score
Criterium Decision Plus	English	No	Yes	Yes	3	3 methods	Yes	Yes	No	1st	0.57
1000Minds	English	Yes	No	No	2	2 methods	No	Yes	No	2nd	0.54
Logical Decisions	English	No	Yes	Yes	3	2 methods	Yes	Yes	No	3rd	0.46
Multicrit	Romanian	No	No	Yes	2	3 methods	No	No	No	4th	0.44
Analytica	English	No	Yes	Yes	3	1 method	Yes	No	Yes	5th	0.35
V.I.S.A	English	No	Yes	Yes	3	1 method	Yes	No	No	6th	0.28

Table 3. The ranked alternatives for 5 criteria (usability)

Alternatives	Criteria						
	Language of user interface	Decision Process interface support	Visual scoring	Visual weighting	Level of user expertise	Rank	Total score
1000Minds	English	Yes	No	No	2	1st	0.55
Analytica	English	No	Yes	Yes	3	2nd=	0.45
Criterium Decision Plus	English	No	Yes	Yes	3	2nd=	0.45
Logical Decisions	English	No	Yes	Yes	3	2nd=	0.45
V.I.S.A	English	No	Yes	Yes	3	2nd=	0.45
Multicrit	Romanian	No	No	Yes	2	6th	0.18

CONCLUSION

In this paper, we have compared six MCDM software products on the terms of the usability (interfaces) they provide and functionality. For these six MCDM software products we considered two ranking problems. The first ranking problem is to rank the six MCDM software products for criteria of usability and functionality. For this type of software product the interface for definition of the decision problem is very important. The second ranking problem is to rank the six MCDM software products only for criteria of usability.

The conclusion from analysis is that all of these MCDM software products provide excellent support for the decision process beginning with problem formulation and continuing through to evaluation, results and sensitivity analysis. Nonetheless, there are distinctions in particular methods used, the interfaces and in results presentation.

The analysis shows that characteristics of the selected MCDM software products are similar to each other. This can be explained by standardized multiple-criteria process implemented in such software. Common trend in the analyzed MCDM software products seems to be multi-purpose software providing several methods for various decision problems. This flexibility requires certain expertise from the user to use such software.

Analysis and ranking obtained may be considered by software developers such as recommendations for developing this type of interactive software products. This research addresses also to decision makers who wish to purchase a MCDM software product.

Future research directions will consider a greater number of usability criteria as well as reliability and computer efficiency criteria.

ACKNOWLEDGMENTS

This research was supported by a PN 09 23 02 06 project of the National Authority for Scientific Research.

REFERENCES

1. Atoum, I., Bong, C.H. (2015). Measuring Software Quality in Use: State-of-the-Art and Research Challenges. *ASQ. Software Quality Professional*, 17, 2: 4–15.
2. Balog, A.(ed.). (2004). *Calitatea sistemelor interactive. Studii şi experimente*. Ed. Matrix Rom, Bucuresti.
3. Bačíková, M., Porubän, J. (2014). *Domain usability, user's perception.* In Human-Computer Systems Interaction: Backgrounds and Applications, Zdzisław S. Hippe, Juliusz L. Kulikowski, Teresa Mroczek, Jerzy Wtorek (eds), Springer, New York, 15–26
4. Belton, V., Stewart, T. (2002). *Multiple Criteria Decision Analysis. An Integrated Approach*. Kluwer Academic Publishers, Massachusetts.
5. Carvajal, L., Moreno, A.M., Sanchez-Segura, M.I., Seffah, A. (2013). Usability through software design. *IEEE Transactions on Software Engineering*, 39, 11:1582–1596.
6. Doumpos, M., Zopounidis, C. (2014). *Multicriteria Analysis in Finance*, Series: Springer Briefs in Operations Research, Springer, New York.
7. Figueira, J., Greco, S., Ehrgott, M. (2005). *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer, New York.
8. ISO/IEC 25000:2005: *Software Engineering – Software Product Quality Requirements and Evaluation (SQuaRE) – Guide to SQuaRE*, 2005.
9. ISO/IEC 25010:2011: *System and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE)* – System and Software Quality Models , 2011.
10. Rădulescu, C.Z., (2013), *Decizie Multicriterială, metode şi aplicaţii în dezvoltare durabilă***,** Ed. Printech, Bucureşti

Seeking New Insights: A Design Thinking Approach to the Development of Persuasive Technology Aimed at Supporting Clients on a Weight Management Program

Paul Doney
Leeds Beckett University
Leeds, UK
p.doney@leedsbeckett.ac.uk

Marc Fabri
Leeds Beckett University
Leeds, UK
m.fabri@leedsbeckett.ac.uk

ABSTRACT
The application of persuasive technology has been shown to be effective in a weight management context. However, it has been observed that the impact is not as significant as predicted. The aim of this project was to investigate whether a Design Thinking approach could generate new insights that could be used to drive the development of an innovative application to help people on their weight management journey. Findings show that although no radically new user needs were identified, the needs that users did express most pertinently are not effectively met by currently available technology. Further, we examined the Design Thinking approach itself and sought to identify criteria for the design of successful insight gathering activities, through end-user engagement. We summarize these at the end of this paper.

Author Keywords
Persuasive technology; design thinking, weight management, service design, behavior change.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces: User-Centered Design.

INTRODUCTION
There is a significant body of literature and numerous commercial applications that offer support to people who have a personal goal of weight loss and/or weight management. The conceptual models resulting from the work of Fogg [2] and Oinas-Kukkonen [10] have provided frameworks to aid the development and evaluation of such systems. These models in themselves are underpinned by theories of behavior change, notably Theory of Planned Behavior, Transtheoretical Model, Social Cognitive Theory [14]. Such applications have been shown to have a positive impact [14, 3], yet there is concern that there is little certainty as to which feature/s are associated with a positive impact on weight change, or improvements in attrition levels on a weight management program [8]. Further to this it has been suggested that whilst the impact of Internet based support can be shown to be positive, this impact is small, variable and not sustainable [5] .

The Persuasive System Design model offered by Oinas-Kukkonen [10] is very comprehensive, offering four categories of persuasive system support (Primary Task, Dialogue, System Credibility and Social) each having 7 principles and illustrating each principle with numerous examples. In a study of health and wellbeing apps with a stated aim of positive behavior change [6] the application of these models has been found to be limited. It was found that self-monitoring was universally used for primary task support along with pre-defined suggestions to make goal setting straightforward. Dialogue support was also utilized, most commonly in the form of reminders. However, targeted dialogue was identified as underutilized. Social Support was also identified as minimal and typically limited to sharing widely through mechanisms such as Facebook links. Crucially the study concluded that none of the applications under review exhibited any characteristics that reflected the user group or context, with no opportunity for tailoring to address this.

The project reported on in this paper set out to seek new insights into how technology could be applied to the weight management context. The longer term aim being to develop new digital tools that are fully underpinned by recognized theories and models, but without losing sight of the need to reflect the requirements of end users; and build tools that not only provide useful functionality, but specifically to build tools people want to use.

To respond to the challenge of increasing relevance to end users, a Design Thinking approach was adopted and relationships were established with end users, facilitated by organizations that offer weight management services to the health service and local authorities.

DESIGN THINKING APPROACH
Design Thinking or Service Design Thinking is a human-centered design methodology that uses co-design and intuitive problem-solving techniques to match people's needs with what is technologically feasible and organizationally viable [1]. End-users are involved at every stage. It is typically applied to deal with difficult, multi-dimensional problems that lack recognizable requirements and solutions – so-called "wicked problems" [11].

The first step in the process is to build empathy and the designers are directed to begin by setting aside pre-conceptions and knowledge of existing solutions. Damien Newman's squiggle illustration [9] captures this mindset and in particular the almost chaotic early stages of a Design Thinking approach:

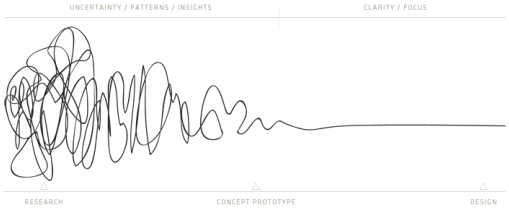

Figure 1: The Process Design Squiggle by Damien Newman

The squiggle celebrates the fact that at the beginning of a problem solving process, we take a seemingly directionless route from not knowing how to address the problem, gradually settling on a clear understanding as we gain insights, knowledge and a strong empathy for the target end users of the service. As well as the pursuit of empathy with end users, at the core of Design Thinking is the promotion of experimentation and, counter to the term Design Thinking, the bias towards action and doing.

Design Thinking as a methodology emerged as an approach separate from existing approaches, such as Human Centered Design, in the 1990's, promoted particularly by IDEO with their Deep Dive approach to problem solving [7]. IDEO identifies Design Thinking as a 'deeply human' process, one that trusts in our ability to be intuitive, to recognize patterns, to construct ideas that are emotionally meaningful as well as functional [1]. Subsequently Design Thinking has been further developed by others, including The Hasso Plattner Design Institute. The Institute have offered a framework that identifies five overlapping and inter-related activities which they have labeled as follows [12]:

Empathize

1. Define
2. Ideate
3. Prototype
4. Test.

It is this framework that has been adopted to drive the initial stages of this project; however, there are no hard edges between these stages, the stages themselves just providing a roadmap rather than a rigid framework.

RESEARCH AND INVESTIGATION

The project was initiated by a scoping activity. In keeping with the Design Thinking philosophy this initial stage was approached without pre-conceptions. To gain empathy for target end users and derive a clearer definition of the problem, some insight gathering activities involving clients of weight management programs were undertaken. The insights emerging from those activities were then used to drive an ideation process, designed to come up with new ideas for how the end users can be supported in their weight management journey. The project has reached this point.

Input into these processes was derived from a number of sources, including people engaged on structured weight loss programs run by two UK based organizations specializing in face-to-face support (MoreLife UK Ltd.[1] and ABL Health Ltd.[2]) weight loss program organizers and facilitators employed by collaborating organizations, Design Thinking practitioners (Uscreates Ltd.[3]) and academics from Leeds Beckett University.

Co-Scoping

The process was initiated by undertaking a co-scoping activity involving Leeds Beckett and weight management professionals - led by Design Thinking professionals. After agreeing the key goals of the project overall, a "What, Who, When, How" exercise was undertaken to reveal and share initial unknowns regarding the problem and the context. The questions that were captured provided a touchstone for assessing progress throughout the insight gathering phase of the project. Once the point was reached where new questions were difficult to formulate, the questions were added to post-it notes so they could be visualized as a whole and themes and groupings identified. The questions divided into two fairly clear groupings:

- Those that related to the individuals and their experience of seeking to lose weight e.g. "what triggers positive or negative eating behaviors"; "how do you quantify success" etc.
- Those that related to the broader context of weight loss and weight loss programs e.g. "do those that drop out of programs have any common characteristics"; "what are the success criteria for the program" etc.

Examining the blocks of questions and drawing on prior experience of working with clients using a Design Thinking approach, techniques to be applied to the engagement with stakeholders were identified. Structured interviewing was identified as a suitable approach to work with employees of weight management programs, to explore the wider context of weight loss and weight loss programs. Whereas a more engaging insight gathering workshop, using playful activities, was devised for working with individuals taking part in weight management programs. The particular activities used were designed by the Design Thinking practitioners, drawing on their prior experience of similar workshops and client groups

[1] http://www.more-life.co.uk/
[2] http://www.ablhealth.co.uk/
[3] http://uscreates.com/

Insight Gathering

An insight gathering workshop was run with one group of clients as a pilot, after which some small adjustments were made before running the workshop again with four further groups. There was no attempt to select participants; the groups worked with were those that were part of existing groups supported by the two program providers. In total 29 individuals contributed to the workshops, Table 1 shows the age and gender profile of participants.

	Male	Female
Over 50	13	5
31-50	3	6
18-30	0	2

Table 1: Age and Gender Profile of Participants

The workshop room was deliberately set with the use of flowers, music and dressed tables to foster a sense of occasion and generate positive energy within the room.

Prior to engaging participants with planned activities the potential of the project was explained with a strong focus on the positive impact their contribution could have. Thereby generating a sense of being an integral part of the process and project.

The initial workshop consisted of four separate activities, each designed to provide different insights into the participants' life experience. Activities were:

Activity 1: Life Pies

The Life Pie activity asked participants to consider their life priorities and then consider how they spent their time in their normal daily lives; with the latter task being captured on a circle divided in 'pie slices'. A short period of reflection followed to explore the differences between the outputs of the two tasks.

Figure 2: Life Pie Activity

Typical responses:

- Priorities: family, health, work
- Activity: work, TV, chores

Activity 2: Circles of Influence

Participants considered how their experience of weight loss was influenced both positively and negatively, and by whom. Firstly focusing on people that were very close to them such as family and friends; then widening their thinking to their environment such as neighbors, work colleagues and their local physical environment. Finally considering influences at the wider societal level, such as media and politics. All thoughts were captured on either pink (negative) or green (positive) post-it notes and organized in concentric circles to build a picture of these influences. Again there was a short pause to reflect on the output, particularly the mix of colors that developed.

Figure 3: Circles of Influence Activity

Typical responses:

- Inner
  - 'Husband takes care of children so I can exercise'
  - 'Husband wants cakes and desserts'
- Middle
  - 'Healthy options at canteen'
  - 'Biscuits at work'
- Outer
  - 'Weight loss blogs'
  - 'Air brushed models'

Activity 3: Modeling Measurement

Participants were given Play-Doh, a modeling compound usually played with by children, and invited to model the means by which they tracked their weight. Whilst the output was limited, the aim of this approach was to re-ignite the energy within the room and drive engagement by adding an increased sense of fun.

Figure 4: Modeling Measurement Activity

Activity 4: Hurdles and Drivers

The final task was Hurdles and Drivers, where participants were asked to consider hurdles they had experienced or

they perceived others had experienced when seeking to lose weight. Suggestion were captured on strips of masking tape that were placed on the floor to simulate hurdles; a workshop facilitator was then positioned by the hurdle and the participants were asked to offer ways in which the hurdle could be confronted or overcome. With the facilitator needing 5 suggestions to move on before another hurdle was requested.

Figure 5: Hurdles and Drivers Activity

Typical responses:

- Socialising
    - Reduce calories in advance
    - Check menu beforehand
- Setbacks
    - Ask for help
    - Recognize past success

The pilot workshop demonstrated that the overall approach was effective, but took too long to be easily repeated given the availability of participants. The Modeling Measurement was recognized as being overly time consuming for the insights gathered. This activity was therefore dropped and the insights associated with it collected through a brief discussion at the end of subsequent workshops. At the trial workshop the activity did have a noticeable impact on the mood of participants by increasing energy levels. However, at subsequent workshops it was seen that the increasing playfulness of the remaining activities kept participants focused and engaged when the Modeling Measurement was dropped.

All the insights were gathered as part of the activities, on post-its, masking tape and diagrams.

The data gathered was analyzed briefly on an on-going basis and after the 5th session it was recognized that very few, if any, new insights were emerging.

A mapping exercise was carried out on the insights gathered against the initial question set generated, confirming that the majority had been addressed and triangulated through repeated workshops.

Ideation

To initiate the ideation phase of the project the expertise of Design Thinking professionals was again utilized to facilitate a workshop which was attended by representatives from Leeds Beckett and from the design agency that will undertake any future application development. All the material that emerged from the insight gathering workshops was gathered and jointly examined in a structured way. The responses to each of the key workshop activities: Life Pies, Circles of Influence and Hurdles and Drivers were explored to seek out patterns and themes. The input from the weight management companies and the group facilitators was drawn on to provide a common understanding of the context in which the workshops had been run.

Through the Life Pie exercise it was possible to derive common traits and issues regarding how people spent their time when contrasted with their life goals. The three key findings that were identified were:

- People spend less time on health and wellbeing than their priorities suggest
- They spend even less time with Family and Friends than their priorities suggest
- Both work and watching TV are environments without physical activity and where unhealthy eating may occur frequently

The Circles of Influence activity drew responses that reflected the breadth of the differing life experiences of the participants. However, there were still some key insights to be derived:

- Inner Circle
    - Family and friends make an equally positive and negative contribution.
    - Own bad habits and those of others are difficult to overcome
    - A close supportive social circle provides a strong motivator
- Middle Circle
    - Temptations that can be controlled at home but are difficult to manage away from home
    - Sense that people react to events and encounters, rather than seek to actively steer a path toward positive actions and choices
- Outer Circle
    - Level of self-efficacy key to determining the analysis and response to societal level influences.

The Hurdles and Drivers exercise produced an output that addressed concerns similar to those emerging from prior exercises, but in a much more focused way. Consequently

the analysis at this point became a distillation of all the data collected to identify a number of key challenges that participants are commonly faced with, and perceived ones that present a threat to their continued progress towards meeting their weight loss goals.

These key challenges were:

- Managing Low Mood
- Facing up to the Long Haul
- Overcoming Temptation
- Securing Support

DISCUSSION

Considering the first two challenges: Managing Low Mood and Overcoming Temptation. These are dynamic issues, they occur without warning, those on a weight management journey can seek to plan their lives to manage them – with or without supporting technology – but it is the unexpected and unforeseeable that will generate these problems in a manner that can result in loss of commitment or focus.

Facing up to the Long Haul and Securing Support are planning issues. However, these could be easily overlooked at the start of a weight management journey, which can be started based on initial enthusiasm. Considering these issues as part of initial planning would be an investment that is more likely to have an impact later in a journey.

It would be difficult to argue that these four challenges cannot be addressed by techniques suggested by existing models. The Persuasive System Design model is comprehensive in nature and offers numerous principles that could be applied. For example Managing Low Mood could be addressed by support that provides praise or rewards, or through social facilitation or expert advice. All of which are principles in the PSD model [10].

It is notable that the issues that emerged are not expressed in a way that reflects any technology based support system. Although initially participants were aware of the longer term aim of developing a technology based support system, talk of technology was minimized during the workshop and there was no attempt to steer any discussion or output.

It is also notable that the issues do not address goal setting or weight tracking, which have been identified as the key components of many currently available health applications [6].

However, at first glance the four key challenges seem perfectly obvious. There were however some subtle nuances to the participant responses that were not anticipated. For example the desire to be part of a support network could have been predicted as all participants were part of a group that met regularly. However, the nuance that emerged was the primacy of support from close families and friends and the difficulty in securing that in a way that participants found positive. Likewise managing temptation could have been predicted, but the particular nuance that emerged was the impact of temptation in environments that could not be easily controlled.

A number of key lessons were learnt from the workshops. By limiting the introduction of perceived existing knowledge and understanding, the discussion was liberated from the desire to look for solutions – in this environment a strong empathy can develop and new insights can emerge. By gathering input through activities that had elements of playfulness, physical movement and visualization the engagement of participants remained high and the workshops flowed easily.

Capturing input as part of the exercises, particularly by visual means, allowed for a "quick and dirty" analysis to be carried out by all workshop participants. It is notable that the Circles of Influence and Hurdles and Drivers tasks generated much more positive energy than the life pies activities. The latter being the only task where participants captured their input separately, making it hard to share and reflect on.

CONCLUSION

This paper presented findings from an insight gathering process, working closely with the clients of face-to-face weight management programs. The key aim was to explore the use of Design Thinking as a means to generate new insights, which could then be further developed to plot a course towards a deeper understanding of how technology can be deployed to support those on a weight loss journey.

The Design thinking approach described by the Hasso Plattner Design Institute [12] was utilized on the project, and the co-analysis workshops run with clients as part of this process are described in some detail.

Five insight gathering workshops were completed, followed by an ideation workshop to explore the output and identify key themes and issues. The four key issues that emerged were:

- Managing Low Mood
- Facing up to the Long Haul
- Overcoming Temptation
- Securing Support

Although it cannot be argued that the key focus of these issues could not have been predicted, there are some nuances that are worthy of note. The desire for effective support was primarily focused on close family and friends, rather than others on a weight management journey. Temptation was seen as a particular challenge when it was perceived that the exposure to temptation was outside the individual's control. The need to secure support and consider a realistic time scale were identified as key initial planning issues that are often over looked. Reflecting on this, the Design Thinking approach did enable us to "see the familiar with new eyes", and gather new insights from the

obvious and sometimes the mundane. This was one of the key aims of our approach.

Reflecting on the experience of running the workshops a number of key characteristics that drive success have been identified:

1. The atmosphere at the start of workshops can be lightened by the use of staging to communicate a sense of occasion.
2. The focus of participants can be clearly intensified by the presentation of their role as one that was integral to the project and of greater importance than the facilitators.
3. The use of playful and physical activity can keep the energy in a room and commitment to a workshop, at a high level.
4. The gathering of data in a visual form allows for a 'quick and dirty' analysis that fully involves participants as well as facilitators.
5. The generation of output that could be easily shared and viewed collectively can be a very effective way to maintain momentum and engagement.

Having identified key issues of concern to existing clients of weight management services, the next step is to work on a co-design activity that will contribute further to the ideation stage of the project. Moving towards the prototyping of an application that can then be developed in an Agile manner and reflects both user group and context.

ACKNOWLEDGMENTS

Thanks to all clients of MoreLife and ABL Health that contributed to the workshops as well as employees of MoreLife, ABL Health, Uscreates and Carbon Imagineering. Part of this project was funded through a research cluster development grant by Leeds Beckett University.

REFERENCES

1. Brown, T. & Wyatt, J., 2010. Design Thinking for Social Innovation. *Stanford Sociaol Innovation Review,* pp. 8(1):29-35.
2. Fogg, B. J., 2009. *A behavior model for persuasive design.* s.l., s.n.
3. Foster, C., Thorogood, M., Richards, J. & Hillsdon, M., 2013. Remote and web 2.0 interventions for promoting physical activity. *Cochrane Database of Systematic Reviews,* Issue DOI: 10.1002/14651858.CD010395.pub2.
4. IDEO, n.d. *About IDEO.* [Online] Available at: http://www.ideo.com/about/ [Accessed 26 May 2015].
5. Kohl, L. F., Crutzen, R. & deVries, N. K., 2013. Online Prevention Aimed at Lifestyle Behaviors: A Systematic Review of Reviews.. *Journal of Medical Interent Reserach,* p. 15(7):e146.
6. Langrial, S. et al., 2012. *Native Mobile Applications For Personal Well-Being: A Persuasive Systems Design Evaluation.* Hochiminh City, PACIS.
7. Myerson, J., 2001. *IDEO: masters of Innovation.* Miicihigan: University of Michigan.
8. Neve, M., Morgan, P. J., Jones, P. R. & Collins, C. E., 2010. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. *Obesity Reviews,* pp. 11(4):306-21.
9. Newman, D., 2011. *Calling for Entries:Have you used the Squiggle.* [Online] Available at: http://centralstory.com/think/tag/the-squiggle/ [Accessed 26 May 2015].
10. Oinas-Kukkonen, H. & Harjumaa, M., 2009. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems.*
11. Rittel, H. W. & Webber, M. M., 1973. Dilemmas in a general theory of planning. *Policy sciences,* 4(2), pp. 155-169.
12. Stanford D.School, 2010. *Bootcamp Bootleg.* [Online] Available at: http://dschool.stanford.edu/blog/2010/12/17/2010-bootcamp-bootleg-is-here/ [Accessed 26 May 2015].
13. Vandelanotte, C., Spathonis, K. M., Eakin, E. G. & Owen, N., 2007. Website-delivered physical activity interventions a review of the literature. *American Journal of Preventative Medicine,* pp. 33:54-64.
14. Webb, T. L., Joseph, J., Yardley, L. & Michie, S., 2010. Using the Internet to Promote Health Behavior Change: A Systematic Review and Meta-analysis of the Impact of Theoretical Basis, Use of Behavior Change Techniques, and Mode of Delivery on Efficacy. *Journal of Medical internet Research,* 12(1).

A Motivational Model for Facebook Acceptance by University Students

Dragos Daniel Iordache, Alexandru Balog, Costin Pribeanu

ICI Bucharest, Romania

Bd. Maresal Averescu nr.8-10, Bucuresti

{iordache,alexb,pribeanu}@ici.ro

Vincentas Lamanauskas

Siauliai University, Lithuania

P.Vishinskio Str. 25, Siauliai

v.lamanauskas@ef.su.lt

ABSTRACT

As social networking websites become more and more widespread there is an increasing interest to analyze and explain their usage. Facebook is a web-based service that is very popular among university students. In this paper a motivational model is presented that explains Facebook acceptance by university students with two key factors: perceived usefulness and perceived enjoyment. The results show that perceived enjoyment is a stronger determinant than perceived usefulness.

Author Keywords

Facebook, motivational model, TAM, structural equation modeling.

ACM Classification Keywords

D.2.2: Design tools and techniques. H5.2 User interfaces.

INTRODUCTION

Social networking websites in general and Facebook in special are interactive environments that are supporting communication, collaboration, exchange of information / resources, and self-advertising. Facebook gained a lot of popularity in the last decade and is widely used by university students [10, 23].

Several studies showed that social networking websites are both useful [2, 16, 17, 20] and enjoyable [18, 19]. Since for many users Facebook became part of their everyday life there is an increasing interest to explain its acceptance.

Motivation plays an important role in the acceptance of information technology by influencing both the actual use and the intention to continue using the system in the future [7]. The motivational model explains the technology acceptance with two key drivers: extrinsic motivation and intrinsic motivation. While the former is instrumental, being goal oriented, the latter is hedonic, being related to pleasure and inherent satisfaction created by a specific activity [8, 24].

Although many authors mention that social networking websites are enjoyable interactive environments, relatively few approaches exist to measure the perceived enjoyment by university students interacting on Facebook.

In this paper a motivational model for the acceptance of Facebook by university students is presented that explains the intention to use it in the future with two key factors: perceived usefulness (extrinsic motivation) and perceived enjoyment (intrinsic motivation).

The rest of this paper is organized as follows. In the next section we present the theoretical background and model conceptualization. Then the empirical validation of the model and the estimation results are discussed. The paper ends with conclusion and intention of future work.

THORETICAL BACKGROUND AND HYPOTHESES

Motivation and technology acceptance

The motivational model takes its roots from self-determination theory that highlighted two main forms of motivation: extrinsic and intrinsic. Deci et al. [9] show that intrinsically motivated behaviors are the prototype of self-determination: people engage in activities for their own pleasure and satisfaction.

A well-known model aiming to explain and predict technology acceptance on a large variety of technologies is TAM (Technology Acceptance Model), developed by Davis [6], and Davis et al. [7]. TAM focused on two main drivers, perceived ease of use and perceived usefulness that determine the intention to use a technology. In further studies the motivational model was adapted to technology acceptance by adding the perceived enjoyment as an intrinsic motivation to use a technology [8].

In the context of technology acceptance, the extrinsic motivation was conceptualized as perceived usefulness, defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" [6]. Intrinsic motivation was conceptualized as perceived enjoyment, defined as "the extent to which the activity of using a specific system is perceived to be enjoyable in its own rights, aside from any performance consequences resulting from system use" [8].

Perceived usefulness relates to the extent to which the users perceive that external characteristics of a system will aid them in performing a task or a set of tasks. Perceived usefulness is considered to have a positive direct effect on the behavioral intention and on the attitude towards using a system. Potential adopters develop a positive attitude and are willing to further use if they observe that the system delivers positive outcomes with regard to their job performance [7].

Perceived enjoyment is also able to significantly influence the intention to use since users who experience pleasure or enjoyment are more likely to form a positive attitude and intention to use it than others [8, 14, 18].

Usage of Facebook in educational contexts

Brown & Adler [4] pointed out that Internet is able to support and expand various aspects of social learning where knowledge is created socially through discussion with others. Social learning means meeting, active participation, critical thinking, information and content sharing [4, 23].

Understanding the usage of social learning technologies is needed for a modern school that seeks to adapt the teaching strategies to student's lifestyle [2]. Mazman & Usluel [20] investigated Facebook adoption for educational use. Their findings show that the perceived usefulness has a significant influence on the adoption of Facebook.

The use of Facebook in university contexts creates social capital which in turn can bring many benefits for users [10, 12, 17, 21]. In a previous work, a multidimensional model was proposed that is capturing more accurately the bridging social capital embedded in the Facebook network of a university student [22].

Lampe et al. [17] investigated the usefulness of Facebook as information source. Their study shows how FB users convert the social capital in information (another form of capital).

As Deci et al. [9] pointed out, motivation and educational outcomes are closely related. They argued that promoting self-determination in the forms of intrinsic motivation and autonomous internalization is beneficial for both the individual and the society [9].

Lin & Lu [19] employed a motivational model to analyze why people use social networking websites. They found that perceived enjoyment is the main driver of the continuation intention. Lee et al. [18] explored the perceived community value of Facebook and found that the experiential value was the most important outcome of information sharing in a social network.

Research model and hypotheses

The research model is presented in Figure 1. We hypothesized that the intention to continue using Facebook (INT) is influenced by two factors: perceived usefulness (PU) and perceived enjoyment (PE). Perceived usefulness is influenced by the perceived enjoyment [7].

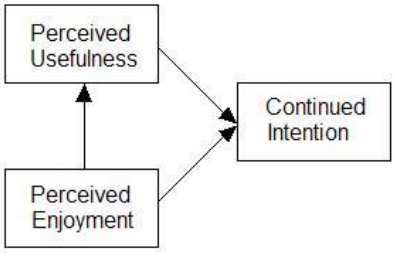

Figure 1. The research model

The following hypotheses are proposed in this study:

H1. Perceived enjoyment has a positive effect on the perceived usefulness (PE → PU).

H2. Perceived usefulness has a positive effect on the intention to continue using Facebook (PU →CI).

H3 Perceived enjoyment has a positive effect on the intention to continue using Facebook (PE →CI).

METHODS

Data collection and sample

In order to empirically test the model a questionnaire was developed in the context of a larger study focusing on Facebook use in an educational context. The variables of interest for this paper are presented in Table 1.

A total of 152 students (110 female, 42 male) from two universities in Lithuania were asked to answer questions related to demographics (age, gender), enrollment (university, faculty, year of study), FB usage (size of their FB network, frequency of use, minutes per day), and to evaluate items on a 7 points Likert-type scale.

Most participants are undergraduate (144) aged from 18 to 45 (M=23.47, SD=5.62), the majority (79%) being between 18 and 25 years old. The size of their social network varies from 10 to 1000 FB friends (M=280.38, SD=190.28).

Table 1. Variables

Item	Description	M	SD
PU1	Using Facebook I can better present myself	3,53	1,71
PU2	Using Facebook I can better present my university work to other people	4,13	1,67
PU3	Using Facebook I am better informed about events of interest in my university	5,06	1,62
PU4	Using Facebook I get useful information from university people	4,99	1,41
PU5	Using Facebook I have better access to university related materials	4,74	1,59
PU6	Using Facebook I stay in touch with my colleagues from home	4,96	1,76
PE1	I have fun using Facebook	4,45	1,66
PE2	Using Facebook is enjoyable	4,63	1,53
PE3	Using Facebook is entertaining	4,45	1,64
PE4	Using Facebook is pleasant	4,39	1,50
CI1	I intend to continue using Facebook in the future	5,46	1,36
CI2	It is likely that I will continue using Facebook in the future.	5,55	1,28
CI3	I will regularly use Facebook in the future.	4,95	1,54

The constructs were operationalized by using and / or adapting existing scales in the literature [5, 8].

Analytical procedures

Data analysis was carried out using the SPSS 16.0 for Windows. Structural Equation Modelling (SEM) with AMOS 7.0 software was applied to test the model. Testing was carried out in accordance with a two-step approach [1, 3] including measurement and structural models.

ANALYSIS AND RESULTS

Descriptive statistics

All mean scores except the mean of PU1 are greater than 4.0 (neutral), showing a positive perception regarding the motivation and the intention to continue using Facebook in the future.

The standard deviations ranged from 1.28 to 1.76, indicating a fairly narrow spread of scores around the

mean. Both univariate and multivariate outliers were searched in the data set and since none of the cases appeared to be extreme, all the data were kept for analysis.

Data normality was investigated in terms of skewness and kurtosis. The values were all within the recommended level [13], supporting the moderate departure from normality for all variables.

Measurement model

Based on an analysis of relevant research [13], the following goodness-of-fit measures were used in this study: normed chi-square (χ^2/df), Tucker-Lewis index (TLI), comparative fit index (CFI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA). In summary, to support model fit it is desirable to exhibit: the normed chi-square should be less than 3, TLI and CFI should exceed 0.95, SRMSR should be less than 0.05, and RMSEA should be less than 0.05 and its 90% confidence interval to be not excessively wide.

The results for initial measurement model show a range of fit results that are indicative for a poor model fit. We examined the standardized residuals and the modification indexes [3]. Two items, PU1 and PU6, were eliminated. As it could be observed, the rest of four items are related to the university life and the students' university work.

The modified measurement model fits acceptable with the data. Although the χ^2 test is significant (χ^2=89.58, *df*=41, *p*=0.000), the other fit indices indicated a good fit with the data: χ^2/df=2.185, TLI=0.850, CFI=0.963, SRMR=0.057, RMSEA=0.089 and its 90% confidence interval is fairly narrow (i.e., 0.064-0.114).

We examined the convergent and discriminant validity of the model using the procedure outlined by Fornell and Larcker [11].

All standardized item loadings were statistically significant (*t*-values > 1.96). The item reliability (R^2) values are above the suggested standard of 0.50 [13], with exception of PU2 (0.40). Cronbach's alpha values are acceptable for all three constructs (see Table 2). The composite reliability (CR) values ranged from 0.862 to 0.939. These values are above the minimum level of 0.70 [13], indicating an adequate reliability. The values of average variance extracted (AVE) are all above the minimum level of 0.50 [13], ranging from 0.613 to 0.804, confirming convergent validity.

The discriminant validity of constructs was examined through the squared correlations test [11]. The results in Table 2 show that the square root of the AVE for each construct is greater than the correlations involving the construct thus provides evidence of adequate discriminant validity.

Table 2. Results of discriminant validity

	Alpha	CR	AVE	PU	PE	CI
PU	0.855	0.862	0.613	**0.783**		
PE	0.935	0.939	0.795	0.330	**0.891**	
CI	0.917	0.924	0.804	0.363	0.585	**0.897**

Notes: The bold diagonal numbers are the square root of AVE

Structural model

A structural equation modeling (SEM) was carried on to test the fit between the research model and the data. The structural model presented in Figure 2 shows the standardized path coefficients, the item loadings, and the explained variance (R^2) for the dependent variables.

The analytical results showed that PE has a significant positive influence on PU (β=0.33, p<0.001) so H1 is supported. Both PU and PE have a significant effect on CI (β=0.19, p<0.001, respectively β=0.52, p=0.018), providing support for hypotheses H2 and H3. The path coefficients show that the perceived enjoyment has the most important contribution to the continued use.

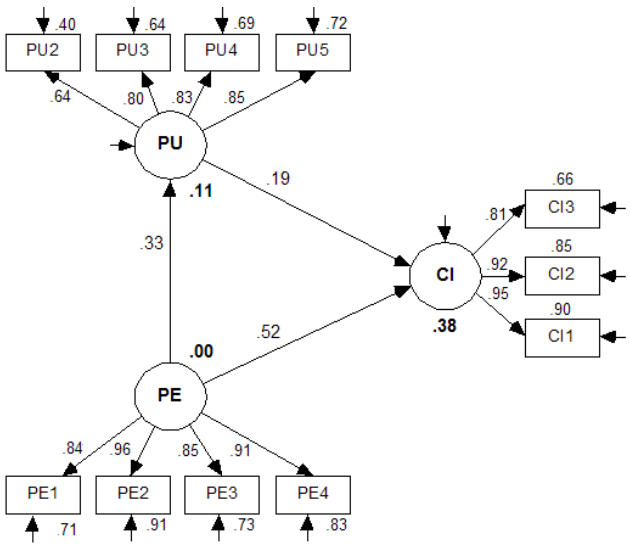

Figure 2. The structural model estimation results

The model explained 38% of the variance in the continued intention to use Facebook (R^2=0.38).

DISCUSSION

This study found that for university students in Lithuania the perceived enjoyment is a stronger determinant of Facebook adoption than the perceived usefulness. The findings confirm the results of other studies that found that Facebook is perceived as an enjoyable environment [18, 19].

As many authors pointed out, websites have a dual nature being both utilitarian and hedonic. For systems perceived as mainly hedonic, the perceived enjoyment is a stronger predictor of the behavioral intention to use than perceived usefulness [14]. The findings of this study confirm the nature of Facebook as mainly hedonic.

At the same time it is obvious that Facebook is very popular among university students. Having in mind that university students are strong users of social networking websites is important to find some effective ways of using FB for academic purposes. For, example, the research carried out in Turkey showed that FB can be effectively applied in distance learning [21].

However, the main problematic and unclear issue remains how Facebook can enhance student learning outcomes [15]. For this purpose it is necessary to reconsider the existing approaches of work with students. In this regard,

the motivation of students for using Facebook should be exploited for stimulating the acquisition of new knowledge, improvements of individual learning abilities, and active participation.

CONCLUSION AND FUTURE WORK

The main contribution of this study is a theoretically grounded and empirically validated motivational model, measuring the contribution of extrinsic and intrinsic motivation to the continued use of Facebook by university students. Model estimation results are revealing that the perceived enjoyment is the main driver of Facebook usage for Lithuanian university students.

There are inherent limitations of this work since the study is exploratory. The sample used in this study is small, at limit for SEM requirements. The students came from only two universities and few faculties.

Future work will focus on refining and extending the scale, in order to analyze external variables that are influencing the Facebook adoption. Then the evaluation instrument will be administrated in at least five universities from two countries in order to analyze the influence of cultural factors.

ACKNOWLEDGEMENT

This work was supported in part from a Romanian grant financed by ANCS under TEHSIN 0923 0207.

REFERENCES

1. Anderson, J.C., Gerbing, D.W. (1988). Structural Equation Modelling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin 103 (3)*, 411-423.
2. Arteaga Sánchez, R., Cortijo, V., & Javed, U. (2014). Students' perceptions of Facebook for academic purposes. *Computers & Education, 70*, 138-149.
3. Byrne, B. (2001). *Structural Equation Modeling with AMOS. Basic Concepts, Applications, and Programming*. Lawrence Erlbaum Association, Publishers.
4. Brown, J. S., & Adler, R. P. (2008). Open education, the long tail, and learning 2.0. *Educause review*, *43*(1), 16-20.
5. Chiu, C. M., Hsu, M. H., Sun, S. Y., Lin, T. C., & Sun, P. C. (2005). Usability, quality, value and e-learning continuance decisions. *Computers & Education*, *45*(4), 399-416.
6. Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13 (3),* 319-340.
7. Davis, F.D., Bagozzi, R.P., Warshaw, P.R. (1989). User acceptance of computer technology: A comparison of two theoretical models, *Management Science, 35 (8),* 982-1003.
8. Davis, F.D., Bagozzi, R.P., Warshaw, P.R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology, 22 (14),* 1111-1132.
9. Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational psychologist*, *26*(3-4), 325-346.
10. Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12(4),* 1143-1168.
11. Fornell, C., Larcker, D. F. (1981). Evaluating structural equations models with unobservable variables and measurement error. *Journal of Marketing Research, 18 (1),* 39-50.
12. Frank, K. A., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. *Sociology of Education, 77(2),* 148-171.
13. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. (2006). *Multivariate Data Analysis*. 6th Ed., Prentice Hall.
14. Heijden, H. van der (2004). User acceptance of hedonic information systems. *MIS Quarterly 28 (4),* 695-704.
15. Irwin, Ch., Ball, L., Desbrow, B., Leveritt, M. (2012). Students' perceptions of using *Facebook* as an interactive learning resource at university. *Australasian Journal of Educational Technology, 28* (7), 1221-1232
16. Lamanauskas V, Slekiene V, Balog A, Pribeanu C (2013) Exploring the usefulness of social networking websites: a multidimensional model. *Studies in Informatics and Control 22(2),* 175-184.
17. Lampe C., Vitak, J., Gray, R., Ellison N.B. (2012) Perceptions of Facebook's value as information source. *Proceedings of CHI 2012*, 3195-3204, ACM
18. Lee, M.K.O., Cheung, C.M.K., Chen, Z. (2005). Acceptance of Internet-based learning medium: the role of extrinsic and intrinsic motivation. *Information and Management, 42 (3),* 1095-1104.
19. Lin, K. Y., & Lu, H. P. (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior, 27(3),* 1152-1161.
20. Mazman, S. G., & Usluel, Y. K. (2010). Modeling educational usage of Facebook. *Computers & Education, 55(2),* 444-453.
21. Ozmen, B., Atici, B. (2014). The effects of social networking sites in distance learning on learners' academic achievements. *European Journal of Open, Distance and e-Learning*, 17 (2), 60-74.
22. Pribeanu, C., Balog, A., Lamanauskas, V., Slekiene, V. (2015) Social capital in Facebook social networks as perceived by Lithuanian university students. *Journal of Baltic Science Education 14(1),* 132-141.
23. Selwyn, N. (2009). Faceworking: exploring students' education- related use of Facebook. Learning, Media and Technology, 34(2), 157-174.
24. Venkatesh, V., & Speier, C. (2000). Creating an effective training environment for enhancing telework. *International Journal of Human-Computer Studies, 52(6),* 991-1005.

Accessibility of Romanian Municipal Websites – Conformance with WCAG2

Costin Pribeanu, Maria Gheorghe-Moisii
ICI Bucharest, Romania
Bd. Maresal Averescu nr.8-10, Bucuresti
{pribeanu, moise}@ici.ro

Paul Fogarassy-Neszly
BAUM Engineering
Str. Traian Mosoiu nr. 8, 310175 Arad
pf@baum.ro

ABSTRACT

Several initiatives at European level exist that promote e-inclusion and web accessibility. Despite all concerns, the accessibility of municipal websites is still low. This paper presents a study regarding the conformance of 60 municipal websites in 2014 with WCAG2 (Web Content Accessibility Guidelines). A comparison of results with similar data in 2011 shows a decrease in web accessibility that could be explained by the lack of regulations at national level and bad practices in web development. Moreover, the evaluation revealed instability of results in time which makes it difficult assessing the progress.

Author Keywords

Web accessibility, WCAG2, accessibility checking tools, conformance, municipal websites.

ACM Classification Keywords

D.2.2: Design tools and techniques. H5.2 User interfaces.

INTRODUCTION

Ensuring equal access to the information technologies for all citizens requires identification and removal of barriers affecting their use of web by people with disabilities.

In 2006, the Riga Ministerial Declaration [15] expressed a political will of EU member countries to develop an inclusive IT to overcome exclusion and improve economic performance, employment opportunities, and quality of life. Two years later, the European Commission (EC) issued a document proposing measures to achieve an accessible information society [3]. A web accessibility study in EU countries revealed a low level of conformance with accessibility guidelines in 2009 [5]

Several initiatives at European level exist that promote e-inclusion in general and web accessibility in particular: the European Disability Strategy 2010-2020, the Action Plan for eGovernment 2011-2015, and the Digital Agenda for Europe, to mention just few. The European Commission (EC) proposed to ensure fully-accessible public sector websites by 2015 [4].

Starting with 2012, the basis of accessibility requirements is the Web Content Accessibility Guidelines (WCAG2) that was issued in 2008 by W3C (World Wide Web Consortium). WCAG2 specifies three levels of conformance (A - lowest, AA, and AAA - highest) [20]. For EU public websites the AA level of conformance is required.

Despite all concerns, the accessibility of public websites is still low. Several studies show that there is little progress in time. According to Hanson & Richards [7] some improvements seem to be related more to changes in web technology and coding practices than to adherence to accessibility guidelines.

In this paper a study regarding the conformance of municipal websites with the Web Content Accessibility Guidelines – WCAG2 is presented. The evaluation was carried on in 2014, on a sample of 60 municipal websites.

The rest of this paper is organized as follows. In the next section we present the main concerns and initiatives at international and European level as well as some recent results regarding the accessibility of Romanian public websites. Then the evaluation results are presented and discussed. The paper ends with conclusion and intention of future work.

WEB ACCESSIBILITY

Web accessibility initiatives

W3C launched the Web Accessibility Initiative (WAI) to develop strategies, guidelines and resources to support web accessibility [19]. Web accessibility means that people with disabilities can perceive, understand, navigate, and interact with the web.

An important step of this initiative was to develop web content accessibility guidelines. The first version (WCAG 1.0) was published in 1999 [20]. The second version was published in 2008 (WCAG 2.0) and this is the reference that is recommended for accessibility policies. There are four key principles that underlie WCAG 2.0: perceivable, operable, understandable and robust [21].

Web accessibility in Europe

Many differences exist between European countries as regarding the regulations and measures for accessibility. According to the MeAC report [5] most countries have a strong policy on web accessibility for public websites. The average strength is between moderate and strong.

Nevertheless, the accessibility status was not good: only 12.5% of governmental websites passed the accessibility level A (automatic checkpoints) of WCAG1 in 2007 [5]. The evaluation was repeated in 2008 and 2009. In general, government websites are more accessible than other websites of public interest. As Cullen et al. [5] mentioned, although the results are showing progress, many websites fail to maintain compliance in time.

In 2012, the European Commission proposed a new directive to better support Member States to achieve their national commitment on web accessibility [4].

Monitoring web accessibility is the responsibility of each country. However, there are few accessibility initiatives that focus on municipal websites. Nietzio et al. [9] reported an accessibility initiative to improve accessibility of municipal websites in Norway. The eGovMon project aimed at integrating benchmarking in a collaborative framework including all stakeholders.

There are also few accessibility studies targeting the local e-government in Europe [8, 9, 10, 12, 13]. The work of Kopackova et al [8] on a sample of 39 Czech websites revealed that results in 2008 are worse than in 2006. Ruano [16] analyzed the development of information and communication technologies in Spanish municipalities. The results suggest a relationship between the population size and e-government capacity that is reflected in a higher compliance to accessibility rules for the websites of bigger municipalities.

Web accessibility in Romania

Statistical data provided by the Romanian Authority for Disabled People (ANPH), for September 2014 mention 727,187 people with various disabilities, out of which 107,821 are visually impaired people (14.83%.). Over 50% of them have severe visual impairments [1].

Although a concern for web accessibility exist, there is a lack of clear policies and action plans to improve it. There is no current action of monitoring the accessibility of municipal web sites. In the she study of Olsen from 2008, Romania was ranked the 22nd country in website accessibility [11].

Few studies are available that assess the conformance with WCAG2 of Romanian public websites. Studies focusing on municipal websites are even fewer and their results show that little progress has been made. The study of Colesca [2] shows that in 2007 most websites (88%) do not have alternative text for images.

Two accessibility evaluations checked the conformance with WCAG2 on a sample of 30 municipal websites in 2010 [12] and 2011 [13]. A comparison of results showed that accessibility is not preserved in time and several specific errors are varying in time and across the web site [13].

A recent accessibility study targeting the district websites in Bucharest (administrative divisions organized as municipalities) also revealed a low conformance with WCAG2 as well as low usability for the visually impaired user [14].

EVALUATION RESULTS

Method and instrument

The sample used in this study includes the first 60 Romanian towns ranked upon population, according to the 2011 census. The total population of these towns is 7862.1 thousands inhabitants which represents 39% from the total population of Romania.

The evaluation was carried on in November – December 2015. In order to test the degree to which accessibility is maintained each website was revisited three times in March 2015.

For each web site the home page was validated. We decided to evaluate only the homepage for two reasons. First, almost all pages have a header, a horizontal menu and one or two vertical menus on the left respectively right side. This organization is replicated on all pages of the website. Since the difference regarding the content between different web pages is relatively small, validating two pages would conflate the number of errors. Second, municipal websites have different information architectures. Apart from this lack of consistency which is perceived by a non-resident user, the differences in content diminish the relevance of comparison between websites for the second web page.

The conformance with WCAG2 (level AA) was assessed by using the Total Validator tool, v8.7.0. Total Validator (TV) is an accessibility checking tool for HTML code, broken links, WCAG1, and WCAG2 (any level). The tool is available on the web [16]. According to Vigo et al. [18] TV performs well across various types of website.

The accessibility score for WCAG2 conformance level A was computed as total number of accessibility errors level A. Accessibility errors level AA, HTML, and link errors were also collected.

Summary of results

A summary of evaluation results is presented in Table 1 that includes total number of errors, number of websites with errors (N), maximum, average and standard deviation.

Table 1. Summary of results

Categories	Errors	N	max	M	SD
WCAG2 A	4876	60	495	81.27	107.99
WCAG2 AA	1334	36	361	37.06	64.37
HTML	11424	57	2338	200.42	342.18
Parsing	1033	45	128	22.96	38.65
Link	2010	52	366	29.49	74.82

A number of 57 homepages out of 60 have HTML errors. Five of them have more than 500 errors. There are also many homepages with parsing (N=45) and link errors (N=52). Another accessibility issue is the large number of links on the homepage that is varying from 41 to 627 with an average of 191.13 (SD=132.01). A number of 42 websites have more than 100 links on the home page which makes it difficult to use by people using a screen reader.

An analysis of results using Person correlation coefficient shows that the websites with many WCAG2A errors also have many HTML errors ($r=.67$, $p<0.01$), parsing errors ($r=.34$, $p<0.05$), and many links on the homepage ($r=.29$, $p>0.05$).

Overall, 4876 WCAG2 errors were detected. The average number of error per web page is 81.27 (SD=107.99) with a minimum of 1 and a maximum of 495 errors. Only 36 towns had WGAG2 AA errors with an average of 37.06 (SD=64.37).

A grouping of towns according to the total number of WCAG2A errors (accessibility score) is presented in Table 2. None of the home pages passed the lowest level of conformance. Only 8 (13.33%) websites had 10 errors or less. Almost all websites having up to 10 errors have no WGAG2 AA errors so the main conformance problems are related to level A errors.

Table 2. Websites on total number of WCAG2A errors

Accessibility score	Number	Percent
1-10 errors	8	13.33
11-20 errors	5	8.33
20-50 errors	19	31.67
50-100 errors	17	28.33
Over 100 errors	11	18.33
Total	**60**	**100.00**

The 11 websites in the last category have together 2832 WGAG2A errors accounting for 58% from the total.

Main error types

A more detailed analysis of results reveals several aspects regarding the conformance to WCAG2 level A.

Most errors are related to the first WCAG2 principle (perceivable). The total number of errors is 3729 (M=62.15, SD=82.35) with a maximum of 464. From these, two error types are more frequent: the lack of text alternative for non-text content (20.16% from total) and the lack of text description for link (17.29% from total).

Other frequent accessibility errors are: tags instead of CSS (15.34%), improper ordering of heading elements (13.41%), and different links having the same link text (5.39%). The main WCAG2 error types are presented in Table 3.

Table 3. Main types of WCAG2A errors

Principle / Guideline	No	%
1. Perceivable	**3729**	**76.48**
Alternative text	983	20.16
Link description	843	17.29
Tags instead CSS	748	15.34
Headings ordering	654	13.41
Labels for controls	163	3.34
Table description	135	2.77
Title for controls	68	1.39
Other	135	77
2. Operable	**524**	**14.85**
Confusing links	263	5.39
Same link text	219	4.49
Stuttering effect	175	3.59
Other	23	0.47
3. Understandable	**73**	**1.50**
4. Robust	**613**	**12.57**
Total	4876	100.0

Most frequent error related to the third principle (understandable) is the lack of a mechanism that allows users to explicitly request changes of context. The error related to the last principle (robust) is the lack of unique IDs in a document, which accounts for 12.57% in the total number of errors.

Comparison with 2011 data and discussion

A comparison with the results from 2011[13] is presented in Table 4. Overall, the total number of WCAG2 A errors in 2014 is higher than in 2011. The difference of 730 errors (17.6%) is pretty high and shows that municipal websites failed to maintain the accessibility level over time.

Table 4. Number of WCAG2A errors - comparison with 2011

Principle / Guideline	2014		2011	
	No	%	No	%
Alternative text	983	20.16	1242	29.96
Link description	843	17.29	395	9.53
Tags instead CSS	748	15.34	1016	24.51
Heading ordering	654	13.41	137	3.30
Unique IDs	613	12.57	118	2.85
Confusing links	263	5.39	661	15.94
Same link text	219	4.49	136	3.28
Stuttering effect	175	3.59	144	3.47
Labels for controls	163	3.34	103	2.48
Table description	135	2.77	75	1.81
Other	80	1.64	1403	2.87
Total	**4876**	**100.0**	**4146**	**100.0**

A comparison on the number of websites in a given error range is presented in Table 5. From the 12 municipalities with 10 errors or less in 2011, we found only 5 that maintain this accessibility level in 2014.

Table 5. Websites grouping – comparison with 2011

Accessibility score	2011	2014
1-10 errors	12	8
11-20 errors	5	5
20-50 errors	18	19
50-100 errors	11	17
Over 100 errors	14	11
Total	**60**	**60**

In order to check how accessibility is varying in a shorter period of time, the webpages were revisited weekly in March 2015. The total number of WCAG2A errors was with 123 higher at the end of March 2015 than in December 2014.

There are several factors that contribute to the low level of web accessibility.

First, there are no regulations at national level as regarding the conformance with WCAG2 guidelines. The existing accessibility guide (mandatory for public administration websites) dates from 2008 (issued by the Ministry of Communication and Information Technology).

Second, there is no accessibility statement on the websites. A good practice requires web developers to mention the level of conformance or at least an accessibility reference (e.g. WCAG1 or WCAG2).

Third, the developer is not always mentioned on the website. Only 25 out of 60 websites were developed by software companies that mentioned their name. A number of 7 websites were developed locally by the IT department of the municipality. For almost a half of websites it was not possible to find out the name of the developer.

Last but not least, it is apparent that there is no quality procedure that enforces the accessibility checking before a

new release or update. Moreover, the large number of link, parsing and HTML errors suggest that in many cases there is no testing at all. The instability of the accessibility level makes it difficult for the disabled user to find out what is new on the website.

There are some inherent limitations of this work. First, there are only 60 towns included in the sample and the selection was not random. Second, although the decision of validating only the home page was previously justified, relying on only one web page is an inherent limitation of the study [6]. Third, automated accessibility checking has its own limitations as highlighted by Vigo et al. [18]. Nevertheless, it is a cost effective method to monitor a large number of websites in a relatively short period of time.

CONCLUSION AND FUTURE WORK

The main contribution of this paper is a wider picture of municipal websites accessibility. Overall, the accessibility of the municipal web is still low, with many errors that are violating the first principle of WCAG 2.0. In order to be used by people with disabilities, the web sites content has to be perceivable.

The results of this study should stimulate stakeholders to take measures in order to ensure the conformance level required at European level. The first priority is to establish clear accessibility requirements at national level. Previous evaluation results suggest that without regulations on web services procurement it is unlikely that a progress will be made in the next years. Second priority is to implement the COM 721 recommendation regarding the monitoring of web accessibility at national level [4].

The fact that accessibility is not preserved in time and that several specific errors are varying from month to month shows that user centered approach is not a practice in web development. A systematic software engineering approach requires testing for conformance before each new release.

ACKNOWLEDGEMENT

This work is supported by the IT2V research project (29DPST/2013), financed by UEFISCDI under the PNCDI II Innovation Program.

REFERENCES

1. ANPH (2014). *Statistical Bulletin*. National Authority for Disabled Persons.
2. Colesca, S.E. (2007). An assessment of the quality of the Romanian websites. *Informatica Economica 2(47)*, 26-33.
3. COM (2008) 804. *Towards an accessible information society*. European Commission, Brussels,
4. COM (2012) 721. *Directive of the European Parliament and of the Council on the accessibility of public sector bodies' websites*. European Commission, Brussels, 3.12.2012.
5. Cullen, K., Kubitschke, L., Boussios, T., Dolphin, C., Meyer, I. (2009). *Study on Web accessibility in European countries: level of compliance with latest international accessibility specifications, notably WCAG 2.0, and approaches or plans to implement those specifications*, European Commission.
6. Hackett, S., Parmanto, B. (2008). Homepage not enough when evaluating web site accessibility. *Internet Research, 19(1)*, Emerald. 78-87.
7. Hanson, V., Richards, J. (2013). Progress in web accessibility? ACM Transactions on the Web 7(1), 2:1-2:30.
8. Kopackova, V., Michalek, K., Cejma, K. (2010). Accessibility and findability of local e-government websites in the Czech Republik. *Universal Access in the Information Society 9*, 51-61.
9. Kuzma, J. (2010). Accessibility design issues with UK e-government sites. *Government Information Quaterly 27*, 141-146.
10. Nietzio, A., Olsen, M.G., Eibegger, M., Snaprud, M. (2010). Accessibility of eGovernment web sites: towards a collaborative retrofitting approach. *Computer Helping People with Special Needs, LNCS 6179*, 468-475, Springer.
11. Olsen, M.G. (2008). *How Accessible is the Public European Web*. http://www.mortengoodwin.net/publicationfiles/how_accessible_is_the_european_web.pdf
12. Pribeanu C, Fogarassy-Neszly P (2011). A review of municipal web sites for accessibility: A computer-aided evaluation approach. *Studies in Informatics and Control*, 20(3), 265-272.
13. Pribeanu C., Marinescu R.D., Fogarassy-Neszly P., Gheorghe-Moisii M. (2012). Web accessibility in Romania: the conformance of municipal websites to web content accessibility guidelines. *Informatica Economica Journal 16(1)*, 28-36.
14. Pribeanu C, Fogarassy-Neszly P, Pătru A (2014). Municipal websites accessibility and usability for blind users: preliminary results from a pilot study. *Universal Access in the Information Society*, 13(3), 339-349.
15. *Riga Ministerial Declaration* (2006). Riga, Latvia. Available at: http://ec.europa.eu/information_society/events/ict_riga_2006/doc/declaration_riga.pdf
16. Ruano, J.M. (2014) E-Government strategies in Spanish local governments. *Local Government Studies 40(4)*, 600-620.
17. Total Validator. Available at: http://www.totalvalidator.com/
18. Vigo, M., Brown, J., Conway, V. (2013). Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests. *Proccedings W4A 2013*, Article 1.
19. WAI (1997). Web Accessibility Initiative, W3C. Available at: http://www.w3.org/WAI/
20. WCAG1 (1999). Web Content *Accessibility Guidelines 1.0*, W3C, 1999. Available at: http://www.w3.org/TR/WCAG10/
21. WCAG2 (2008) *Web Content Accessibility Guidelines 2.0*, W3C, 2008. Available at: http://www.w3.org/TR/WCAG20/

Realistic Simulation of Environmental Phenomena – Snow Fall and Accumulation

Sandra Al-Assaf, Dorian Gorgan
Computer Science Department, Technical University of Cluj-Napoca
Str. G. Baritiu 28, 400027, Cluj-Napoca, Romania
sandra.alassaf@gmail.com, dorian.gorgan@cs.utcluj.ro

ABSTRACT
One particular case of the intersection between the photorealistic presentation and the dynamics of 3D scene of objects is snow in nature, which presents behavioral aspects of the environment such as accumulation, particle movement, and redistribution, with the added time element that allows change and progress analysis. Due to the complexity of the phenomenon, little to no attention has been given to the realism of the interaction, the attention being channeled only on the visual aspects of it. The focus of this paper is to analyze different approaches of representing snow fall and accumulation effects in real time in such a digital ensemble in order to add a layer of realism and accurately reflect natural processes. The emphasis is put on the actual manifestations that take place in nature and the micro physics that pertain to the processes. Comparisons are made between different existing approaches to snow accumulation and a new particle hierarchy is imposed to suit both the technological requirements and to maintain the realism of scenes.

Author Keywords
Photorealistic Presentation; Environmental Phenomena; Snowing Phenomenon; Snow Fall; Snow Accumulation.

ACM Classification Keywords
H.5.2 User Interfaces, I.3.7 Three-Dimensional Graphics and Realism, I.6.8 Types of Simulation.

INTRODUCTION

Virtual environments are digitally generated mediums that represent settings which contain 3D visualization support. Their use is wide spread in today's applications, bringing diversity in domains such as games, physics simulation software, tourism applications. Their intended purpose is to depict the world as we perceive it, and, as such contains 3 critical dimensions: visual, temporal and behavioral.

The visual dimension is generally achieved from detailed models representing entities, complex textures and application positions, and, of course, the most crucial aspects for realism in visualization, lights and shadows. While the visual aspect has been thoroughly dissected throughout the years, minimal consideration has been given to the plane created by the relation between behavior and time.

One particular case of the intersection between the last two dimensions is snow in nature: it presents behavioral aspects of the environment (accumulation, particle movement, redistribution) with the added time element that allows change and progress analysis. Due to the complexity of the phenomenon, little to no attention has been given to the realism of the interaction, the attention being channeled only on the visual aspects of it.

The focus of this paper is to analyze different approaches of representing snow fall and accumulation effects in real time in such a digital ensemble in order to accurately reflect natural processes. The emphasis is put on the actual manifestations that take place in nature and the micro physics that pertain to the processes. Comparisons are made between different existing approaches to snow accumulation and a new particle hierarchy is imposed to suit both the technological requirements and to maintain the realism of scenes. The micro physics element is completely new to the scene and provides a needed layer of realism to better emulate the real world.

The paper is structured as follows. The Related Works section presents an overview of the sources of the research foundations for this study and a comparison to the current approaches. The next section presents the theoretical analysis, complete explanations of choices, algorithms and mathematical models used to develop the solution and the main contributions. The testing and evaluation section presents metrics and the final experimental results. Finally, the paper concludes on the main contributions, and future research directions.

RELATED WORKS

This section is a quick overview of the research in the related domains such as computer graphics and interactive systems, 3D environmental practices, project management and elaboration, microphysics, snow fall in real time and accumulation techniques. It is meant as a discussion covering the major inspirations on which the premise of the project has been centered.

In nature, snow accumulation on a surface implies a great deal of factors. Snow usually reaches surfaces that are directly visible from an aerial view. This is not universally true however, given the disturbances in motion created by wind currents of different intensities. That is why, snow can actually reach surfaces that are hidden or partially covered

by other objects. The difficulty arises, not only in the motion, but also in the way the particles are laid out and the manner in which they interact once they reach the ground/a snow patch. Having relatively small masses, ice crystals are not as stable as ice would be, but not as volatile as any liquid. Such a conglomeration of micro particles presents varying behaviors depending on the current state of the surroundings.

Snow collection on surfaces is the predominant theme of discussion in natural phenomenon modelling for virtual worlds. It is complex in the sense that it depends on a great number of factors and that is has a great impact on the speed of any application.

Willmore and Fermeglia present in [1] all the relevant aspects of wind influence and snowflake motion, accumulation of snow corresponding to the amount fallen as indicated by the snowflake weight, and finally, redistribution of snow.

Wind influence in [1] is modelled by a force or impulse that is applied to the snowflake which gains a velocity according to its weight in the direction specified by the wind. The main observation here is that there is no concern for altitude in this approach causing a diminished behavioral realism. Wind can be present at one altitude and be completely changed in direction or intensity or even absent at another. This not only gives liveliness to the scene but also allows modelling of actual current behaviors and snow distribution accordingly.

Snowflake motion depends on the wind and mass [1]. The approach was the fact that the mass of a snowflake was a predefined value, whereas in nature, mass is directly influenced by pressure, supersaturation with respect to water of the air and creation time from the vapor stage. Modelling these factors has the effect of making an environment that can be easily customized and that can support variation of any kind.

Visually, [1] states that a 3D representation is the best method to portray individual snowflakes. This is indeed true, but due to the fact that having 3D objects involves visibility calculations for each polygon and that typical particle systems contain thousands of such models, another technique was chosen. Partly inspired by the simplicity of particle system [2], the approach is to have a plane consisting of four vertices with a texture representing a cross section of a snowflake. This plane is always oriented according to the camera view direction, and, as such, the GPU is less strained than in the case of full 3D objects rendering. The disadvantages described in [2] apply here as well: unfortunately there is no influence of the medium reflected on the snowflake: the lighting, no shadows portrayed and the colors are not altered under any circumstances. Such a compromise was found small in comparison to the great benefit of speed.

The paper also proposes individual collisions per particle. This means that in a bigger system each snowflake will check for collisions with each other object. While this presents many disadvantages in regards to speed and can produce a bottleneck in large test scenes with millions of particles, the approach is not fundamentally flawed. An adaptation of it has been made in order to reduce collision numbers but maintain an even higher number of particles. As such, quanta entities are introduced that contain exactly what their names hint at: a quanta or measure of snowflakes. It is a point in space having tied to a particle system that acts as a single individual when it comes to colliding with other objects. While the actual collision with the fallen snow surface may not be as smooth as in Willmore and Fermeglia's case, the particle numbers increase dramatically.

The height at each point in [1] is determined by the density of the snow. Again, the problem of mass arises, the approach developed considering, other than density, the volume of snow by being directly influenced by the size and weight of ice crystals.

Redistribution of snow has been modelled as closely possible to the approach presented in [1]. Although more intuitive than practical, the model proposes a threshold that has to be respected by the Laplacian of each point. While the modelled snow surfaces are not the same in representation (height maps as images versus adjacency list of mesh vertices in 3D models), the same principle can be easily applied. The drawback of this is that that threshold has absolutely no connection to the actual parametric representation of the climate.

The aspects presented encapsulate the main features analyzed and adapted from [1], while the next big influence in the accumulation approach is taken from Festenberg's famous research paper [3], detailing a diffusive manner to snow cover creation. Concepts introduced by Festenberg have a high level of detail: the snow is presented in the most realistic manner, treating cases such as snow bridges, different approaches to texturing per location in patch, accumulation on surfaces that are not visible from the top view approach, snowflake dusting and flutter. While these techniques present a great naturalness to the digital medium, they utilize slow algorithms that use a lot of memory and cannot be used in vast stages. The scenes are all static, and presented as snapshots rather than interactive media. That is why, for the purpose of real time simulations, the proposed representation algorithms could not be used.

This experiment does however focus mostly on the behavioral realism, so a technique from [3] was adapted in our research, and that is snow distribution according to the diffusive coefficient. Festenberg explains the fact that there was a need to replace snow recursive redistribution for stability measures and depth buffer displacement

techniques. This statement is what drove the project in the search for the optimal solution.

Paper [3] proposes that the snow profile change its steepness according to the quality of the snow: whether snowflakes accumulate water between ice crystals or have a higher ice to water ratio. As such, this aspect ties in very well with the proposed climate modelling scheme, where the entire purpose is to actually see the impact the temperature and water levels have on the resulting scenery. The explanations regarding the efficiency of this method in contrast to the recursive accumulation have not been expanded upon.

Snow falls in form of snowflakes: a collection of ice crystals formed through aggregation or condensation The difficult part is not modelling how snowflakes fall or positioning them, but actually simulating exactly how they come to be as they are, how they gain their mass and shape. The studies presented in [4], entitled „Cloud Microphysics", not only explain the exact implications of cloud types, atmospheric density, and altitude, but also present different growth patterns in snowflakes and shape modifications according to each formation process.

The most important benefit of [4] is the understanding of how an ice crystal is affected by the temperature and water saturation. It also presents calculation methods in order to determine mass and radius for ice crystals, general snowflake densities and a broad overview of the actual ice crystal fusion into a large – snowflake – structure.

SNOW NATURAL PHENOMENON

Snowing Phases

Snow behavior as a natural phenomenon contains three major phases: (1) creation or snowflake composition process according to the atmospherically conditions/ parameters; (2) fall or particle motion; and (3) accumulation or distribution of snow on surfaces in relation to external forces. These are the constituents of the research and the main results that should take part in completing all milestones for the entire system. There are the following requirements regarding each phase of the snowing phenomenon.

Snow Creation

The system has to create and render snowflakes in each particle system of size and weight corresponding to the atmospheric conditions at the time of its creation.

Snow Fall and Motion

The system has to begin the simulation and correspondingly the fluid motion of the snowflakes in the 3D environment without interruptions or visible stuttering. Moreover the motion of the snowflake is influenced by the wind currents defined by the user, the gravitational force, and the mass of the snowflakes.

Snow Accumulation

The system has to animate the entire process of snow surface growth in real time at a speed perceivable by the human eye, the volume of fallen snow on a surface has to be equal to the volume of deposited snow, and the accumulation shape is in accordance to the atmospherically conditions.

Parameters of Snow Phenomenon

Particle System Model

The snowflakes particles are defined by the following parameters, as in [5]:

- *Temperature* expressed in degrees Celsius at the location and moment of snowflake creation;
- *Density* or number of units of snowflakes per particle system size;
- *Formation Time* or the time taken in seconds to develop ice crystals from the vapor stage;
- *Position* or initial location of the entire particle system expressed in centimeters and relative to the scene origin point.

Wind Model

The following parameters describe the wind model:

- *Direction* expressed as a unit vector that aims to portray the general movement of the wind current;
- *Speed* expressed in meters per second that defines the intensity of the wind current;
- *Minimum Altitude* expressed in meters, portrays the minimal height relative to the ground at which the wind current stops affecting objects;
- *Maximum Altitude* expressed in meters, portrays the maximal height relative to the ground at which the wind current stops affecting objects.

Accumulation Phenomenon

The snow redistribution over the upper surface of 3D objects is described by the concept of *Laplacian Threshold*.

External forces

There are two main driving forces for such a system:

- *Gravity* in physics can be defined as the attraction between all material objects. The most visible manifestation of it can be viewed between items in the scene and the earth [6].
- *Wind* is an air transition described by speed and direction. Wind is an impulse that is applied on a force in a specific direction with a degree of intensity to it [7]. The idea when using wind, given the fact that the system already has an attached mass to it, was to apply a force on an object for every frame that it is in the wind altitude range.

Critical Situations

Different environmental states produce dissimilar particles, collisions have an impact on the future locations of snowflakes and wind can alter the expected location for any entity. These are just a few of the circumstances that have to be considered while approaching such a theme. They are critical not because they represent an exception, but because they encompass the most common occurrences that are also the most difficult to portray.

In order to depict the behavioral aspects of this specific environment, a number of situations have been foreseen and treated. These are the main cases to be evaluated in order to convey a complete image of a dynamic winter landscape:

- Particle differentiation according to parameters;
- Collision detection between appropriate object types;
- Particle system decomposition and regrouping at collision;
- Modification of snow surfaces at collision;
- Appropriate redistribution of snow surfaces after collision and certain conditions are met;
- Trajectory change of particle systems when external forces are applied;
- Accumulation beneath obstacles;
- Atmospheric changes according to wind influences.

These will be further expanded later on, but it is important to take note of them for now.

SYSTEM COMPONENTS AND CONCEPTUAL MODELS

Objects in 3D Virtual Environment

In order to populate the scene of objects to portray a realistic medium for snow to fall in, modelling and use of polygonal meshes has been employed. Meshes are structures that represent 3 dimensional objects containing the information necessary to accurately depict real world entities: vertices, faces and edges, [**Error! Reference source not found.**]. Each object in the created scene has been built from primitive shapes (cones, cubes and spheres) welded together, while others (the particles for example) are simply planes.

Snowflake Considerations

A snowflake begins as a collection of water molecules that freeze into ice crystals if a particle of matter is present in the cloud atmosphere and later aggregate together to form a larger entity. That is why, when modelling the process of their creation the following phases have been considered: (a) Vapor phase; (b) Individual crystal phase; and (c) Aggregated phase.

In the vapor phase, the most relevant aspects in order to determine the supersaturation with respect to water are the temperature and pressure of the cloud location. These factors later influence how „wet" the snow is, and as a result, the final section shape of the accumulation on surfaces. Supersaturation influences mass, shape and radius as well. The growth in mass over time influence the way snowflakes interact with the medium: how they are affected by the gravitational force, the speed of their fall, the way they respond to external influences such as wind.

Snowflakes Quanta

The entire collection of snowflakes is contained inside a volume, which in these particular research cases is a cube. The center point of the „quanta" represents the ‚parent' node for each and every instance in the particle system: once the parent node translates, or rotates, so do each of the snowflakes. Although not as dynamical, it permits the system to manipulate a greater number of particles because of the way collisions are handled.

Decomposition of Quanta

Snowy scenes typically contain obstacles under which snow cannot be accumulated. This is caused by the fact that the objects block the path of the snowflake, which can no longer reach its corresponding ground location. The experiments consider four cases of relationships between quanta and the obstacle (Figure 1).

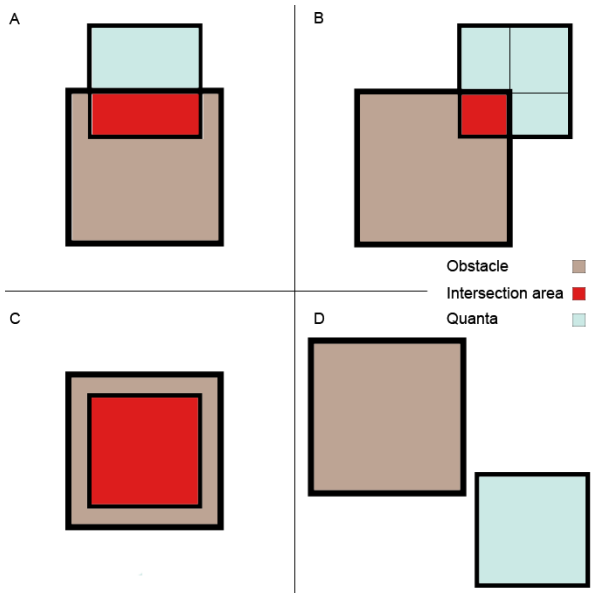

Figure 1. Cases of relationship between quanta and obstacles: A - quanta is separated into two parts, B - quanta is decomposed into four parts, C - quanta completely included on the obstacle surface, D - no interactions occur.

Snowflake and Quanta Formation

While the shapes of the ice crystals differ, their inherent volume can be constrained by a cylinder. The mass of a snow crystal can be computed from density and volume of the crystal. The computation of the mass takes also in consideration the vapor phase, the time factor [4]. The quanta are formed by spawning a set of particles at random positions enclosed in the volume attached to it (i.e. cube).

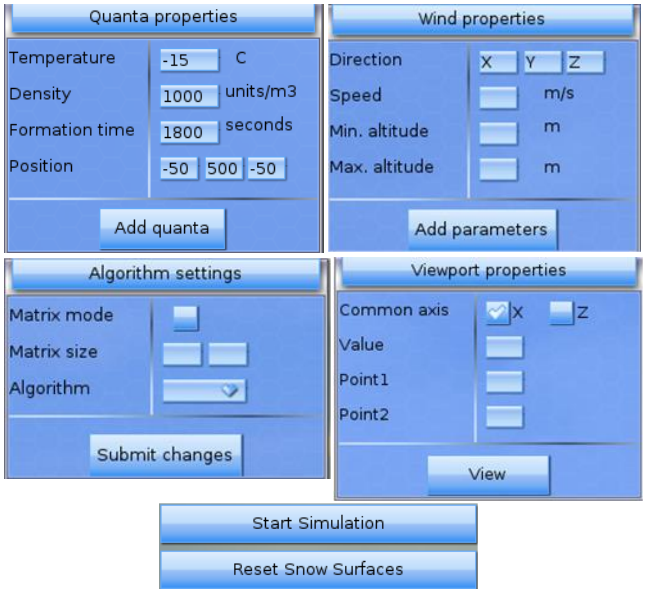

Figure 2. Application user interfaces for editing sessions for Quanta properties, Wind properties, Algorithm settings, Viewport properties, and Control options.

The volume is determined by getting the bounding box center, width and height and extracting the limits in world coordinates by mathematically deducing the half extents. The spawn positions are completely random inside the boundaries and do not conform to any rules. The snowflakes can spawn one on top of the other, as it actually happens in nature in case of the aggregation formation process.

Snow Accumulation

Snow accumulation is the most complex process of the entire system. It should accurately portray the growth of a surface according to the quantity of snow fallen. There are four primary steps in which the process can be decomposed: collision detection, mapping, growth calculations and animation.

Shaping the snow surface is done through either of two methods (according to user choice of algorithm): diffusive or recursive accumulation.

Diffusive accumulation takes all the points included on the mapped surface, one at a time. The diffusion coefficient is linked directly to the supersaturating of air with respect to water. It computes the corresponding vertex's Laplacian by calculating the mean value of all the adjacent vertices. The diffusion coefficient is multiplied by the computed mean value of vertices related to distance.

Recursive accumulation, while not directly linked to the climate, proposes a way to redistribute the snow surface in order to prevent surface spikiness and resemble the phenomenon. Due to the fact that it has no actual scientific basis, most of its underlying fundamental values can be tweaked by the user in order to study different effects.

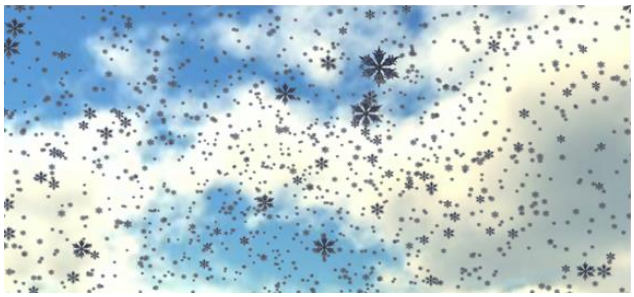

Figure 3. Particle system with varying levels of details.

EXPERIMENTAL EVALUATION

The experiments have used the following software configuration: the code has been developed in C++ under Windows 7 operating system, Ogre3D SDK 1.9 for graphics modeling and visualization [9] and [10], Bullet Physics 2.82 as physics engine [11], and MyGUI 3.2 for graphical user interface development.

Parameter Setting

There are a few parameters by which the user controls the experiments (Figure 2). They address the definition of quanta and wind, setting of the algorithms, definition of the viewport, and the options for controlling the snow simulation.

Snowflakes Visualization

Figure 3 presents the particle system with various levels of details for the snowflakes.

Visualization Modes

There are three visualization modes: (a) wireframe; (b) textures and colors, no lights but polygonal mesh; and (c) photorealism with lights and shadows.

The first of three visualization modes for the scene is the wireframe mode. While it does not portray lights, shadows, textures or any kind of colors it helps to better observe the vertices and the edges of the snow surface in order to properly study the accumulation behavior.

The second visualization viewed retains the textures and colors but disables lighting and maintains the vertices and edges on display. This has been the visualization mode that has been used for the majority of the development phase, because, contrary to the wireframe mode, it does not display hidden lines and points so it displays fewer details while retaining the useful information where the user can observe them.

The last mode and the most visually appealing is the realistic approach that has lighting and shadows enabled. Though not as relevant when checking for behavioral

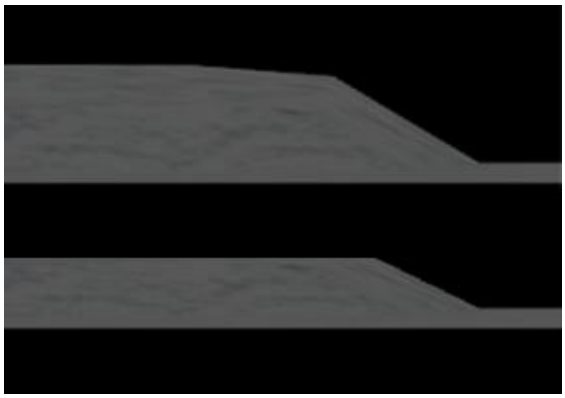

Figure 4. Snow accumulation at -15 degrees Celsius (top) and -5 degrees Celsius (bottom).

correctness, it does help to construct a parallel to the real world.

Temperature Impact

The next experiment focuses on the difference brought about by temperature. Figure 4 shows a comparison of the materialization of the theoretical concept, where 8 quantas have fallen in the same surface area. Because of the fact that the temperature of -15 degrees gives the highest super-saturation, it has the biggest snowflakes, so, in theory it should give a bigger growth while the slope should be a bit leaner because of the fact that snowflakes have more water between ice crystals. The other chosen value for temperature is -5 degrees Celsius, a choice that should result in rougher edges since the snow is ‚drier' and less volume to it because of the smaller snowflakes.

Diffuse Versus Recursive Algorithms

For the next experiment the long awaited comparison of the two snow accumulation algorithms has been sought out. The climate conditions are at -5 degrees with a formation time of half an hour. Each of the following scenes has 8

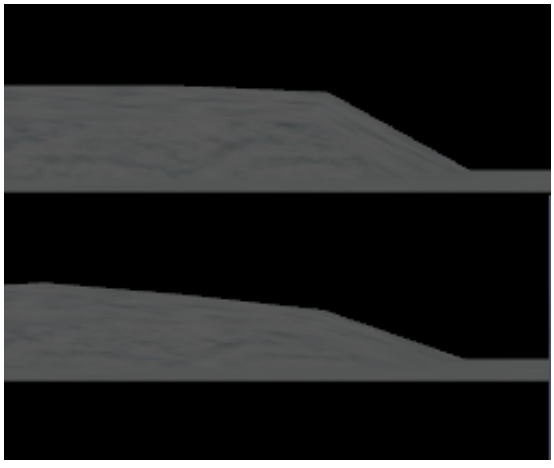

Figure 5. Snow accumulation by the diffuse (top) and recursive (bottom) algorithms.

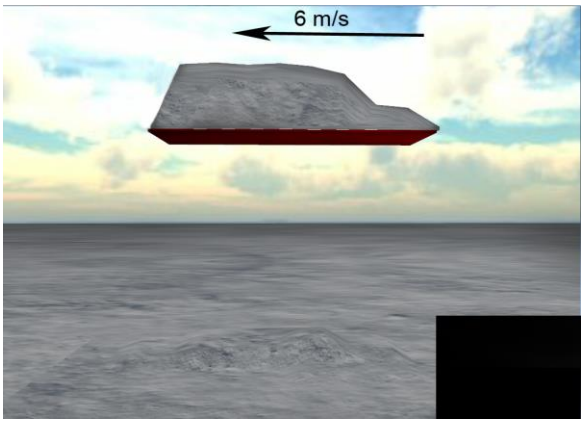

Figure 6. Windy snow accumulation of 8 quantas.

quantas deposited, with the recursive algorithm setting a threshold of 20 centimeters.

Though the recursive method does not depend on the water content of the snowflakes and chooses to portray the scene with less physics background information, it is quite obvious that it looks a lot more natural than the roughness of the diffusive method. The threshold is also quite an important factor of the recursive algorithm, but has no actual corresponding microphysics formula, so the results can greatly vary from one choice of a value to another.

Windy Snow Accumulation under Obstacle

The experiments consider the windy snow falling and accumulation under obstacles.

There are 8 quantas with a density of 1000 snowflakes, formation time is half an hour, at the temperature of -15 degrees Celsius, and the wind speed is 6 m/s toward direction specified in the Figure 6. The upper piece of snow represents just the snow accumulated and blown by the wind to the left.

Figure 7 shows the initial position of a group of quantas relative to the obstacle. The obstacle is located at 3 m above the terrain, and the wind is limited between 0.5 and 2.5 m relative to the ground.

Figure 8 presents the snow accumulation after all 8 quantas arrive on to the soil.

CONCLUSIONS

Contributions

Most of the concepts discussed in this paper have been used before in many other works. This is not the first time vertex morphing is employed or the discovery of the combination between a graphics and a rendering engine is made. The novelty of it all is the method in which they were all integrated together with the added novelties brought by the new quanta concept, decomposition technique idea and micro physics calculations. To have an application that can

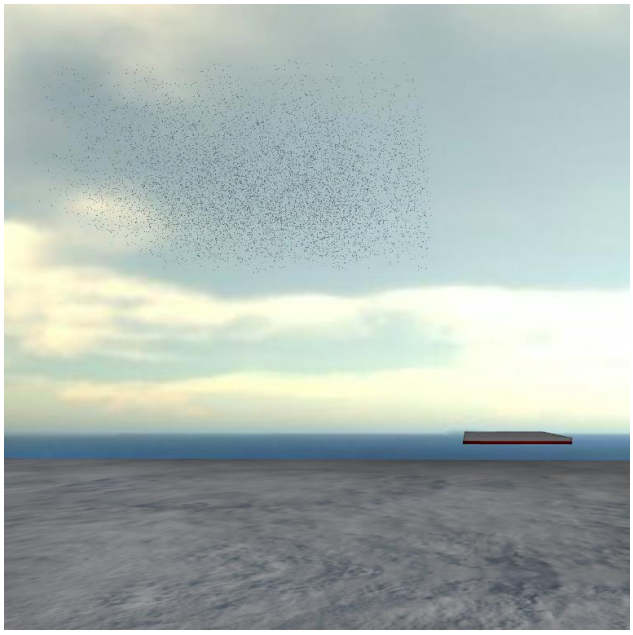

Figure 7. Initial position of a group of quantas relative to the right side obstacle.

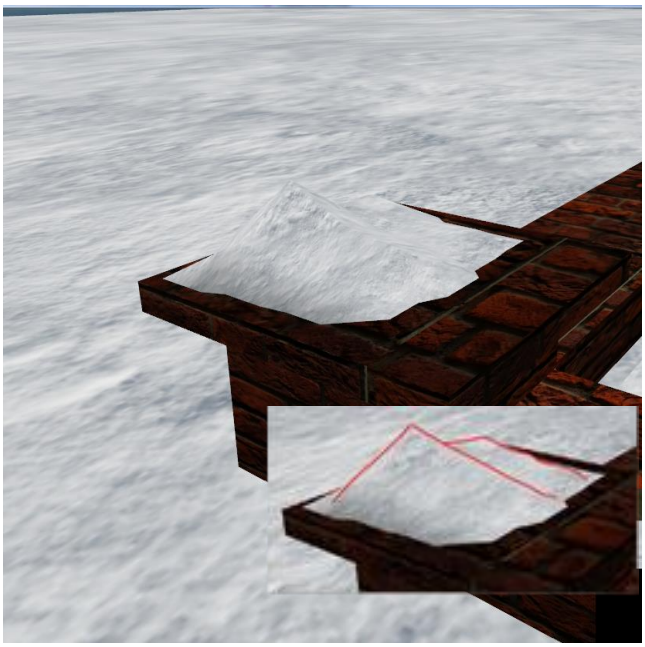

Figure 8. Result of snow accumulation after all 8 quantas arrive on to the soil.

help study the downsides or benefits of a certain implementation can speed up the process of deciding upon a case that yields fast results and can be applied universally.

The quanta decomposition method was not inspired by any algorithm, so it can be said that even though primitive, it is one of this paper's completely original contribution. A combination of methods that pertain to both statically accumulation and real time snow rendering for winter sceneries have been employed in order to test their likeliness to reality.

Another major contribution brought about was the inclusion of concepts found in microphysics to the process of forming appropriate sizes and weights for snowflakes. This not only helps keeping the climate under strict rules (e.g. temperatures of -1 degrees cannot produce snowflakes of great sizes), but also adds the benefit of practicality that can be applied to programs that model real world simulations in fields like architecture or civil engineering.

Critical Analysis of Results

Most of the objectives described at the beginning of the research work have been achieved. For the first part, there is no jitter or waiting time for particle system creation. The process runs fast independently of the actual parameters provided. Once the simulation has started, addition of particle systems does not cause any kind of problem or program interruption. All particle system properties are modelled exactly in the way that the user has specified and the difference between results can be observed as easily as visually analyzing the scene. The mass and radius for each snowflake is calculated through specialized formulas taken from scientific literature, so that particular objective is met throughout. Whenever the user starts the simulation all quanta's begin to move and their motion is constant over time. Due to the utilization of a physics engine, this motion environment (i.e. gravitation, wind) so it is as accurately portrayed as the implementation of the software integrated has allowed. Another purpose is fulfilled through the last statement, the one considering the motion in accordance to the wind parameters that are defined by the user. The movement of the particle system is heavily reliant on the properties of each snowflake, defining an added dimension of realism.

The predominant problem of the current state of the application is that no definition has been found yet for the programmed in margin, so while it is a great thing that it isn't hardcoded, the user has no actual basis on which to select this value. Furthermore, the threshold does not have any real world correspondence. The problem brought about by the diffusive method is that it does not yield realistic looking results in a dynamic environment (as can be seen in the figures from the testing chapter) and continues to accumulate on the same patch if the snow falls only on that portion. Also, the obstacle shapes have to be rather primitive in order for the decomposition to work. Because of the fact that the entire concept works on the presumption that the shape is encompassed by its bounding box, the method would not give realistic results for any shape that contains curved surfaces. The accumulation in concave shapes has to be developed as well.

Further Developments

From an implementation point of view the most important aspect would be to change from a bounding box approach to a rigid body check for quanta decomposition. While the technique would be altered a bit, the significant aspects of the algorithm would still remain.

Another development would be to include vertical accumulation, as can be seen on house walls in snowstorms or when the intensity and inclination of snow fall is greater than usual. This implies modelling the adherence to surfaces, and going into much more detail on the less insisted upon snow surface properties. So far, the quanta is the only one that has been fully parameterized. The actual melting factor has not been taken into account as well, and would be a great addition. Even though the difference between wet/dry snow and different temperatures is made, after creation there is no further impact of the atmospheric factors on the surface of the snow. This new feature would also help transition between snow and water and make it possible to increase the temperature range already available to house the creation of rain droplets and water flow.

There are two ways in which this application can be further developed from a strict usability point of view. Initially the idea of the study came about when planning for the development of a 3D application directed at the tourism field which aimed to portray Cluj-Napoca in the Middle Ages. Somewhere in the midst of development, the idea to actually alter the state of the city and be able to visualize it throughout the seasons surfaced, and with it, a dilemma arose. The fact that there was no way to realistically portray the scenery during winter was what drove the project to take shape. This project is the first path mentioned that the thesis can be integrated in.

A future expansion would consists of a complex scene , on the scale of representing a historical city center with details such as arcs or sculptures included, with the polished version of the algorithms implemented such that people can take advantage of technology in order to marvel at the lost times of the middle ages (e.g. visualization of Cluj-Napoca downtown).

The second research idea was to create an application that uses the developed algorithms in order to calculate structure resistance. Given the fact that the physics engine already supports mass and joints, all structures could be portrayed through the light of their respective materials. This could result in an interesting approach to modelling and designing constructions such as buildings or heavier formations. The experts would not have to calculate all the possible factors that come into play, but rather simply provide a set of parameters and see how the scene responds to them. It would test the points that are the most susceptible to fissures and tears and make sure that all buildings are solid and can withstand harsher environmental conditions. Again, the idea of melting the snow would be extremely helpful in this context, since it would deal with the problem of design that permits leakages to occur.

These are just a few of the directions that come to mind, both from an implementation and project purpose point of view. Due to the vastness of the virtual environments domain and its many uses in applications, many other uses could be found for the project at hand.

ACKNOWLEDGMENTS

The research has been carried out in the Computer Graphics and Interactive Systems Laboratory of the Computer Science Department, in the Technical University of Cluj-Napoca.

REFERENCES

1. Chris Willmore and Fermeglia Wilmore, „Snow Accumulation in Interactive Time", Advanced Computer Graphics, Rensellar Polytechnic Institute.
2. Deborah Michelle Patrick, „An Investigation of Snow Effects to Enhance the Graphics in Computer Games", Rhodes University, 2000.
3. Niels v. Festenberg and Stefan Gumhold, „Diffusion-Based Snow Cover Generation", TU Dresden, Germany, 2011, pp. 3.
4. John M. Wallace and Peter V. Hobbs, „Amospheric science", Second Edition, in *International Geophysics Series*, Volume 92, 2006, pp. 209-228.
5. Snowflake physical properties, http://www.noaa.gov/features/02_monitoring/snowflakes_2013.html
6. Gravity formal definition, http://whatis.techtarget.com/definition/gravity-or-gravitation
7. Wind theoretical considerations, https://eo.ucar.edu/basics/wx_2_c.html
8. Polygonal mesh theory, http://www.cs.mtu.edu/~shene/COURSES/cs3621/SLIDES/Mesh.pdf
9. Ogre 3D hardware specifications, http://www.ogre3d.org/tikiwiki/Hardware
10. Ogre 3D API specifications, http://www.ogre3d.org/docs/api/1.9/class_ogre_1_1_entity.html#details
11. Dynamic AABB trees http://www.bulletphysics.org/mediawiki-1.5.8/index.php/BtDbvt_dynamic_aabb_tree

Gamepad Vibration Methods to Help Blind People Perceive Colors

Vlad Trifănică
University "Lucian Blaga" of Sibiu
Faculty of Engineering
Victoriei Boulevard 10, Sibiu 55002, Romania
trifanica.vlad@gmail.com

Alexandru Butean
University Politehnica of Bucharest,
Faculty of Automatic Control and Computer Science
Splaiul Independentei 313, Bucharest 060042, Romania
alexandru@butean.com

Alin Moldoveanu
University Politehnica of Bucharest,
Faculty of Automatic Control and Computer Science
Splaiul Independentei 313, Bucharest 060042, Romania
alin.moldoveanu@cs.pub.ro

ABSTRACT
Visually impaired people, completely or partially blind cannot perceive the surrounding environment as normal healthy people can. Not knowing how colors and shapes look like is a serious penalty and they have to live with it an entire life. It is known that the human brain adapts and compensates with greater capabilities for hearing and touch when vision is low or missing, that is why blind people develop special skills and perform so good when it comes to music and kinetotherapy. Based on this information we are proposing a method that explains colors using vibrations. In this paper we would like to present a new way of transforming an image to vibration by using the capabilities of an Xbox gamepad. During the study on real blind volunteers, this kind of approach proved to be very effective and promising, allowing them to slightly understand how an image, painting or friend looks like.

Author Keywords
Vibration; colors, modern devices; visually impaired;

ACM Classification Keywords
H.5.m. Information interfaces and presentation - Human-Computer Interaction, Miscellaneous. ; K.4.2. Computers and Society: Social Issues - Assistive technologies for persons with disabilities.

General Terms
Human Factors; Measurement

INTRODUCTION

Color is an essential component of spatial recognition and there is a requirement to invest in systems that can help individuals who are blind or visually impaired to learn to perceive colors. Individuals who are color blind and endure different types of color blindness such as achromacy, tritanopia etc. encounter various problems in their daily lives. Blind individuals suffer from color deficiency as well. Individuals who are congenitally blind have to learn about color through artificial ways. A person that experiences blindness at later phases, has the sense of color but the absence of sensory data belonging to color, influences everyday activities. Individuals that suffer from such sensory and cognitive afflictions can be assisted by systems that permit color perception. Regarding the sense of touch, the development of haptic technologies have been growing lately and economical devices are becoming more broadly accessible.

Interesting results are obtained with haptic feedback using different sensors attached to gloves [6]. In this paper we present a first experiment using a novel method that aims to help blind people perceive colors using a gamepad.

Color Perception

In humans, color perception is an essential part of spatial processing. It allows decisive perception of the environment and helps in object detection, analysis, scene segmentation and other spatial tasks. Likewise, color perception plays an important role in social interactions. Color is not perceivable directly through the touch sensors in humans, because, as a feature, is entirely visual. This is one reason why presentation of color data to individuals who are visually impaired constitutes a complicated issue.

One approach regarding this problem is presented [4]. The innovative system which they proposed allows learning, presentation and examination of color information. Their system is based on a procedure that distributes colors as textures through a haptic device. The objective of the considered approach was to enable color perception and to supply a foundation for evaluating color correlation. Initially, users went through a learning phase, that allowed them to relate color to a realistically rendered texture. The testing of the rendering system has been done through a vibrotactile and a force feedback system. In both cases, after the learning phase, users could identify colors with high precision.

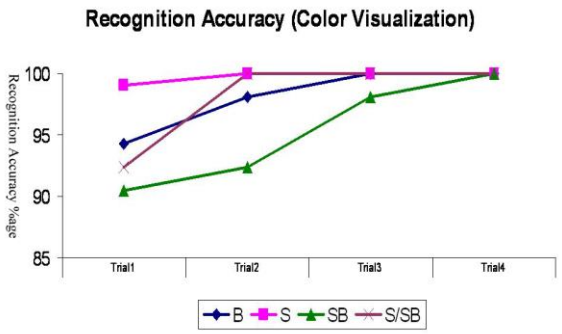

Figure 1. Learning and Recognition of Color

The examination of the learning system to establish the mapping and real-time system for color recognition proves that the proposed methodology may provide a reasonable basis for color perception.

Conceptual Systems

An economical, easily transported haptic device which can transmit 2-D texture-enriched graphical information to visually impaired individuals is presented in [1]. They developed a single finger, inexpensive, point contact haptic device that is extendable to more than one finger. This instrument (Figure 2) is used in order to convert an electronic visual diagram into is haptic form. This would provide supplementary information about part particularity and part direction. The device can identify a wide range of colors and can output a different variety of simulated texture in an appropriate way.

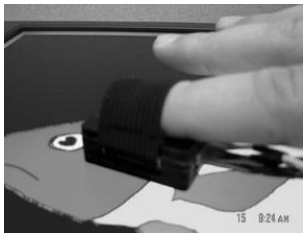

Figure 2. The device interacting with a computer screen

In [2] is introduced a new multimodal device scheme concept in consideration of helping individuals who are blind or visually impaired to be able to draw tactile pictures. The system resides in a multimodal representation able to activate the user's visual, haptic, and/or auditory sensory systems. Most notably, this system will permit the user to "paint" in textures, corresponding to colors, which can be tactually analyzed. Until now, a restricted form of the design has been implemented in MATLAB using a haptic device. A more thorough representation is being assembled in C#, with plans to integrate Microsoft ™ SDKs which allows the use of more than one pointing device. The finalized version will be subject to user testing.

Modern approaches

One example is presented in [3]. They focused on the representation of paintings through haptic displays from their digital image. The haptic display that they created perceives the location of the display on the virtual painting and then adds a tactile feedback in mechanical texture form. This procedure is done with the aim of illustrating the physical brushstrokes. In the near future, they will concentrate on developing a display system, display techniques and picture conversion mechanisms to represent paintings that can be found in public and private art galleries. The system could then be made usable in kiosks, Art Museums or ready for home use on the Web.

In the paper [5], there have been created and appraised two haptic and visual applications for acquiring information on geometrical concepts in group work in primary schools. There were to evaluated applications: The Static Application and The Dynamic Application. The Static Application consists of a 3D environment that encourages learning to distinguish between different angles and drawn shapes. The Dynamic Application is similar to The Static Application but supports cooperative learning of geometrical shapes such as cubes and notions such as volume and area.

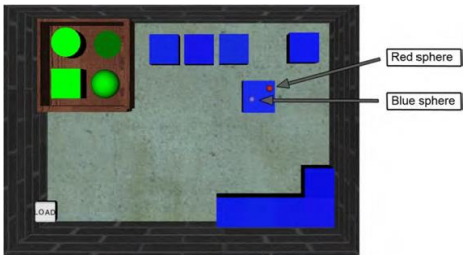

Figure 3. The dynamic application with two users represented by a blue and red sphere respectively.

The outcome of this study has demonstrated that sighted and visually impaired pupils who work in groups in a mutual haptic virtual environment can achieve and preserve a common ground of the physical arrangement and of the objects in a workspace. The most essential thing that the study has shown is the fact that the haptic informative functions can make a considerable improvement to both the overall cooperation and assignments solving method, as well as the involvement of the visually impaired pupil.

VIBRATION TOOL: XBOX ONE CONTROLLER

Microsoft invested over $100 million into refining the controller design for the Xbox One. The XBOX One Controller (Figure 4) is quite similar to the Xbox 360's Controller but with a different handle contour. It has off-set analog sticks, the A, B, X and Y face buttons, a directional pad, the Menu and the View buttons. The directional pad has a traditional design and the battery compartment is slimmer compared to previous XBOX controllers. The Xbox One controller includes a micro USB port: when

attached via a USB cable, the controller can operate without battery power.

Figure 4. XBOX One Controller

Gamepad vibration properties

The XBOX One Controller [7] has two vibration motors (Figure 5), a low frequency and a high frequency motor, one in the base of each grip. The functionality of these motors, is to supply force feedback effects to the user. A vibrating motor is essentially a motor that is improperly balanced. In other words, there is an off-centered weight attached to the motor's rotational shaft that causes the motor to oscillate. The amount of oscillation can be changed by the speed at which the motor spins.

Figure 5. Vibration Motors

VIBRATION EXPERIMENT

An experiment was designed in order to test the ability of the users to differentiate between 8 and 6 levels of vibration. For testing, 3 participants were recruited, having the age of 21, 35, respectively 65. At the beginning of the experiment, each user was instructed to press one of the buttons of the XBOX One controller, to feel the vibration. Afterwards, we provided 5 pairs of different levels of vibration, in order to test if the users can perceive which vibration is each pair is powerful. The results are shown in Table 1.

Users	Vibration levels	Accuracy
Participant 1 (Age: 21)	6	80%
	8	50%
Participant 2 (Age: 35)	6	100%
	8	80%
Participant 3 (Age: 65)	6	100%
	8	80%

Table 1. Result of the experiment

The results clearly suggets that is difficult to differentiate between 8 levels of vibration. A system that can help visually impaired individuals to perceive and learn just 6 basic colors is a more reliable / close to reality solution.

Color-Vibration Intensity

In order to implement the system, we chose 6 basic colors that needed to be perceived and learned: white, red, yellow, green, blue and black. Each color out of the 6 must have associated a level of intensity. In this early stage intensities are calculated with just a simple division, that means the difference between 2 consecutive colors is now the same but we plan to investigate in the future if would be better to consider another formula based on the intensity of the colors.

Level of Vibration	Color Associated
1	WHITE
2	RED
3	YELLOW
4	GREEN
5	BLUE
6	BLACK

Table 2. Vibration-Color Association

Converting RGB Images to the 6 Color Set

An important functionality of the system is the conversion of RGB Images to the 6 color set proposed by us. This is done by evaluating the similarity between every pixel in the image with one of the 6 basic colors, followed by the conversion of the pixel to the basic color.

Each pixel has 3 components: Red, Green and Blue. In the first step we used a standard 3-byte RGB conversion algorithm. In the second step we decided that cyan is close to blue and purple to red and this assumption helped us to reduce the number of colors from 8 (standard 3 bit) to 6 . Using this type of approach we decided if the pixel is more similar to red, green, blue or yellow. Another comparison is made between each component and two set values, 80 and 180, to check if the pixel is more similar to white or respectively black. After the testing is done by the algorithm is finished, the pixel is transformed in one of the 6 basic colors. In Figure 6 is presented a famous picture by artist William Henry Hunt that is processed using the method described above.

Original Image

Converted Image

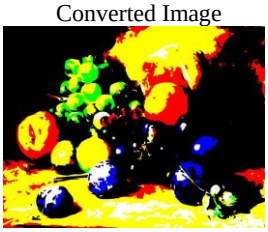

Figure 6. Image Conversion

Interaction with the system

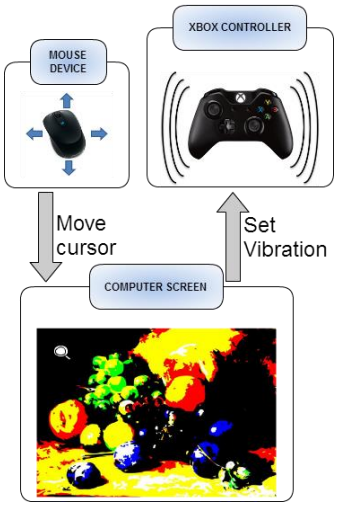

Figure 7. Interaction with the system

As seen in Figure 7, the user is instructed to move the mouse pointer on the processed image on the computer screen. Depending on the location of the pointer on the image, the XBOX controller will vibrate with the intensity associated with the color that is pointed by the mouse.

CONCLUSION

In this paper we have presented a new way of turning images into vibration using an XBOX Gamepad. This method turned to be promising because it was very well accepted by volunteers due to the easy interaction between the vibrating device and humans. A static image is converted to a 6 color scheme and then explored by the visually impaired using a mouse, generating different levels of vibrations.

Initially we decided to use 8 colors, corresponding to 8 different vibration levels, but after the hands-on study with several subjects, numbers have shown that at least in the beginning of this experiment would be better to stick with just 6 vibration settings corresponding to 6 basic colors.

The results of this research are encouraging us to continue the study and explore the capabilities of this method.

Future Development

In the near future, in addition to the XBOX controller, we will try to integrate the use of different types of standard PC controllers or other console controllers. Furthermore, we intend to add a new feature replacing the mouse cursor movement on the processed image by using the analog-enabled joystick of the controller. Since this was an early experiment to establish whether the concept might work, another experiment with more testing subjects and test cases is already in progress.

ACKNOWLEDGMENTS

This work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134398 and POSDRU/159/1.5/S/132395.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

The results presented in this research study were possible due to the help of National Association of the Blind People in Romania / subsidiary Sibiu and other subjects who have offered to be volunteers for testing and feedback sessions.

REFERENCES

1. Burch D.S., Pawluk D. T. V., "A cheap, portable haptic device for a method to relay 2-D texture-enriched graphical information to individuals who are visually impaired", *ACM SIGACCESS conference on Computers and accessibility*, 2009, pages 215-216
2. Headley P. C., Pawluk D. T. V., "A multimodal, computer-based drawing system for persons who are blind and visually impaired", in *ACM SIGACCESS conference on Computers and accessibility*, 2010, pages 229-230
3. Hribar V. E., Pawluk D. T. V., "A tactile-thermal display for haptic exploration of virtual paintings", *ACM SIGACCESS conference on Computers and accessibility*, 2011, pages 221-222
4. Kahol K., French J., Bratton L., Panchanathan S., "Learning and Perceiving Colors Haptically", *ACM SIGACCESS conference on Computers and accessibility*, pages 173-180
5. Moll J., Pysander E-L. S., "A haptic tool for group work on geometrical concepts engaging blind and sighted pupils", in *ACM Transaction on Accessible Computing (TACCESS)*, Volume 4 Issue 4, 2013, Article No. 14
6. Tapson J., Gurari N. Diaz J. Chicca E. Sander D. Pouliquen P., Etienne-Cummings R., The Feeling of Color: A Haptic Feedback Device for the Visually Disabled, BIOCAS 2008, pages 381 – 384
7. Xbox One Controller, https://goo.gl/6qIZkPaa

Analyzing Computer Game Strategies through Visual Techniques

Mircea Catalin Catana, Dorian Gorgan
Computer Science Department, Technical University of Cluj-Napoca
Str. G. Baritiu 28, 400027, Cluj-Napoca, Romania
mircea.catalin.catana@gmail.com, dorian.gorgan@cs.utcluj.ro

ABSTRACT
The paper presents and analyzes an application that can emulate different types of strategy games and provide significant specifications as to how the strategy impacts the outcome of the player experience. This application is meant to be used as a tool both by game developers, which can use it to calibrate their products and also by the game players, who will want to use it as a means to improve their strategies. The research explores the concept of strategy in games and tries to define a metric for evaluating this concept. Due to the complexity of videogames, the computation of every outcome for every possible decision a player can make is impossible, so the research proposes a method of combining the human familiarity with the game concept and the processing power of the computer to search the wide solution space. The paper analyses a concrete example by building a test bed for the tower defense type games.

Author Keywords
Game Strategy; Visual Analytics; Visual Techniques; Tower Defense Games.

ACM Classification Keywords
H.5.2 User Interfaces, I.3.7 Three-Dimensional Graphics and Realism, I.6.8 Types of Simulation.

INTRODUCTION

Video Games have been around for a long time, dating to the early 50's as part of computer science research, however it is not until the 90's when computers began to spread and be used by the general public that games have started to flourish. Where in the early days of game development a person could do most of the work by himself, highly specialized teams are now working on specific aspects of the game like: gameplay, user interface (UI), game design, etc.

Along the road games started to be categorized by their creators and players in types related to their content, play style and view of their simulated environment. Some of these categories include: 2D and 3D worlds, platformers, puzzles, role playing games (RPGs), etc. Out of all these, this paper focuses on games in which the player must devise a strategy as the core part of the gameplay in order to win the game. These are usually called strategy games and can be either turn based or real time.

In turn based games, each player takes turns in a predefined order to perform actions which have impact in the game world. A classic example of a turn based game is chess in which the white and black players move a piece per turn until an end game situation is achieved.

In real time strategy (RTS) games [9] the players interact with the world at the same time and practically "race" each other to the end of the game. These types of games are more challenging as players are not only forced to play a better strategic game than their opponent but they also have to make decisions fast in order to be quicker than the other players.

Our research mainly deals with tower defense games as a demonstration of the applicability of the visual analysis techniques used, but future releases could cover a wider array of game types. Tower defense games are a particular type of strategy game in which the user plays against a computer in single player mode. The player has to defend a safe point by building different types of turrets that shoot at the computer units. The computer's objective is to reach the player's safe point with its units. The units follow a predefined path from a spawn point to the player safe point and are generally generated in waves of increasing difficulty. The challenge and strategy lies in how the human player decides to spend his resources in order to build turrets which will defend his safe point. After each computer unit kill, the human player is awarded a small amount of resources which he will later be able to spend on building more powerful turrets. The game allows for a minimum number of computer units to reach the safe point before the player loses the game. If however the player manages to defeat all the enemy waves before that threshold is reached he wins.

In creating games, one frequent problem is that developers have to take great care and spend a lot of time balancing the properties of their software so that the experience of the player is enjoyable. This means tweaking all the variables in a way such that the game is not extremely easy to play, but also not very difficult or impossible. This process can be a daunting task, but an absolutely necessary one if the game is to be successful. Currently, there is no general purpose software which can aid in this process. Some gaming companies develop a small array of in-house tools that help game designers in this regard, but these are small

tools that are usually custom built for each game individually.

This paper proposes a method of addressing the previously mentioned issues by building a software tool which is able to simulate the conditions of a computer game, analyze the way a strategy is carried out and provide insight into ways of improving gameplay. This can be beneficial to both developers by providing them with a tool that can aid the process of calibrating the game difficulty levels and professional gamers which can use the software to improve the strategies they use in competitive games as well.

Videogames are highly complex systems with a lot of variables which means that we cannot simply compute every possible outcome for all the sets of inputs and just choose the best case. The method described in this paper proposes to combine the computation power of a computer with the analytical thinking and decision making capabilities of a human being. The main goal of this research is to develop a software application that can analyze strategies used in tower defense games and provide the user with visual guidance in developing a better strategy for a specific game input.

If we are to analyze all the possible combinations a user can choose to play during a game we have to think of the variables of the game as an n-dimensional solution space. Such a space where the axes have values like turret damage and position on the grid is very hard to be interpreted and visualized by a human being. A difficult task was to come up with a visual representation that would provide a view of this system from a perspective in which the user could easily understand the parameters and the optimal values for them. The paper proposes a visualization technique that makes use of the projections on the axes of the n-dimensional space.

The paper is structured as follows: Introduction section provides an overview of computer games and introduces the reader to the concept of strategy in games. This section also provides the motivation behind building this software product. Section 2 is a study of the work done by others which provides a base for this research. It treats strategy in video games and visualization techniques independently. Next section covers the theoretical analysis of the solutions, providing insight into the design of the application and the technologies used during development. The solutions are evaluated and validated by practical experiments. Last section draws conclusions and future directions.

RELATED WORKS

Game engines are highly complex software systems that are designed for the development of videogames. The concept is relatively new as games before the mid 90's were built from ground up every time. Jason Gregory marks in his "Game Engine Architecture" book [1] that the concept of game engines started to form around games like *id Software*'s Doom. The game provided a clear separation between core functionality and specific game aspects like custom assets or gameplay rules. By the late 90's game developers identified certain needs any game would have regardless of its contents and started building software that would manage these needs but that could also be reusable in other projects. Some of these components include a rendering engine, user input handling, a sound framework, and so on. A separation between the layers handling game logic and the lower layers handling the hardware and general purpose systems started forming in all the games since then. This eventually became known as the game engine. It is worth noting though, that not even today's modern game engines can provide a clear separation between them and the game.

An important point in the history of game engines was introduced by game moding. The concept of a game mod or modification was introduced with the rise of games like *id Software*'s Quake Arena and *Epic Games*' Unreal Tournament. In these games the engines were made highly customizable via scripting languages like Quake C. These moding capabilities allowed anyone who owned a license to build additional content for the original game or even build a new game. Shortly a large community was built around this concept and this facilitated the separation of the game engines from the actual games.

Nowadays there are a high number of different game engines that are licensed separately from any specific game. These engines have reached a point where anyone can pick them up and build any game they have in mind without worrying about how to implement every system that a videogame would need in order to run. There is enough free software to cover most of the needs of a developer, so the problem becomes more of choosing the right tool rather than not having access to it. An example of such commercially available game engines that are at our disposal now include Epic Games' Unreal Engine 4, which has features that mainly target FPS games and Unity Technologies' Unity3D engine.

Game engines are usually made from a runtime component and several tools that can work along it. The architecture is built in layers where upper layers depend on lower ones but not the other way around so as not to create circular dependencies and to promote low coupling between the systems.

One of the most important factors in whether someone keeps playing a videogame is the challenge level of it. The challenge level is directly linked to the engagement the user experiences while playing. Fraser et al. [2] presents data from an experiment in which they analyze how different game factors affect a player's performance in a game. Their paper directly relates to some aspects of this work, as they also build a test bed for a game with the goal of identifying which game factors affect gameplay. In the same way, the software application described in our paper can be used to examine in more detail these factors and their impact.

Besides tweaking parameters to different levels of difficulty, modern games are starting to use the data they gather during a play session and combine it with artificial intelligence (AI) to make the game difficulty level adapt in time. Every game has a learning curve and even though initially a low difficulty level might be suited to a certain player, he could learn the game and its mechanics rather quickly in which case a more difficult level should be ideal. So, if there is no difficulty adapting mechanism installed, the situation could rapidly converge to the initial problem.

In the early games, this need for increasing difficulty was mainly hard-coded by the developers in the design of the game levels. Each successive level would present a harder puzzle to solve or just more of the same difficulty obstacles to get pass by. This technique is still used today, but no longer enough. The racing game Forza, for example, uses the data gathered during online gameplay sessions to adapt the single player AI driving style. This means that as you climb higher in ranks during online play and get better, the single player AI will also be close to your ranking and driving performance. Moreover, if the player usually plays online versus his friends, the AI style will start to mimic his friend's driving style, thus enabling a user to have a connection with his friends even in offline mode.

Other games take an "in the middle" approach and predefine a set of parameters which will be used as a difficulty step during gameplay. If the player performs well and is starting to overcome the current level with ease, the difficulty level will be dynamically incremented to the next, if the player has trouble overcoming the current setting it will be dimmed down. Most games use this approach as more often than not, the complexity of the game is too great to build a fully adaptive AI system.

When discussing game difficulty and challenge in a strategy game, the paper "Exploring Design Features for Enhancing Players' Challenge in Strategy Games" [3] found that there are two important aspects one must consider: mental workload and physical effort. It is important to look closely and distinguish between these types of factors at the level of the game type and start working on building the game design around them. An action game for instance has great demand on the perpetual-motor skills of the player, so focusing the gameplay and building the game around this rather than on puzzle solving is essential.

In the context of strategy games, mental workload is directly linked with information availability. In short, the more information a player is given about the state of the game and the actions that are happening around him, the less mental work he has to do. The study has found that not only this is true, but also that mental workload tends to decrease in the case there is extremely few information. That is, players who are given too little information about what's going on tend to stop being challenged as they figure they cannot accomplish anything with what they have at their disposal. This is a very important piece of information for game designers as it tells them game balancing is the key factor for keeping players engaged.

Similarly with the case of mental workload, physical effort follows a similar path. In strategy games, the physical effort is measured in terms of the amount of resources a player has available. So again, in order for a game to be challenging, the amount of resources a game makes available to a player must not be too high neither too low.

Based on this study we can infer that in the case of tower defense games there is a strong relationship between the data the player has about the next waves of enemies (mental workload) and the resources available to spend on turrets (physical effort). Thus striking the right balance in a strategy for a tower defense game lies in the calibration of these two factors.

VISUAL BASED GAME STRATEGY SOLUTION

Game Strategy Solution Overview

The research aims to develop and experiment the techniques of Visual Analytics, through an application for analyzing and developing strategies in videogames. The software should be able to model a general type of game and based on the game's rules provide the user with information about how to improve his strategy. The application is meant to be used by players trying to get better at a game and also by game designers that are trying to calibrate the difficulty levels in the game they are developing. The solution treats videogames as an *n*-dimensional system in which each axis represents one of the variables in the game. Each such variable will have a value domain range defined by the game in question.

A set of equations describing the game rules and how these variables affect the strategy in the game will be devised based on a predefined metric. The metric will be built specific to the game and it will consist of a function that will score a set of fixed game parameters. Different parameter values will be given different scores by the metric function depending on how desirable each situation is compared to the other. This scoring will be done for a reduced set of values and the results will be displayed to the user through visual techniques. Based on the user's expertise in the field and the human ability to quickly analyze data, the user will chose a set of parameters from the ones evaluated and the computer will restart the computation for values in the vicinity of the chosen data. This way the process will repeat itself and converge towards an optimal solution. The stopping point of the process is determined by the user which can choose to go further until there is no visible change in the data, thus signaling that the optimum has been reached or he may opt to end the search earlier and have a partial solution available in a few iterations.

For the particular case of Tower Defense games we will analyze the following 3 aspects of implementation: building

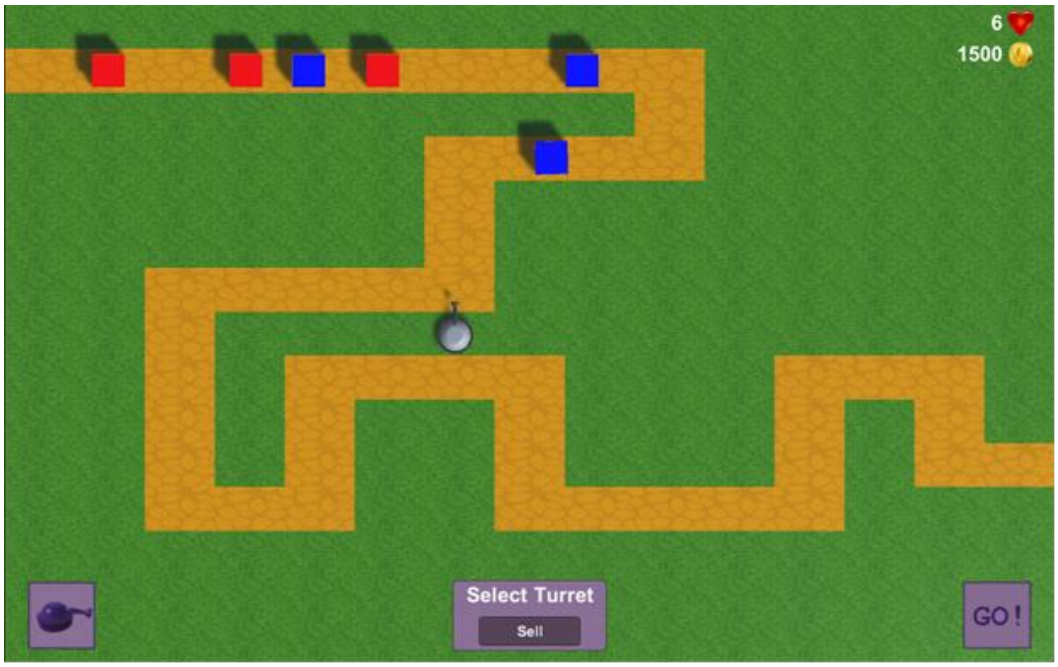

Figure 1. Game graphical user interface.

a simulation of the game, developing the underlying algorithms to analyze the strategy, and building a visualization system that transforms the output of the algorithms into visual data that is easily handled by a human user.

Building the Game

In order to analyze the strategies used, we must first have a working test-bed of the type of game we want to analyze. Since we want to use the application to support as many tower defense games as possible it is better to build a configurable base game ourselves than to try and make a tool which could integrate will all existing software. The implementation of the application is done using the Unity3D engine which allows for easy and quick prototyping of a game (Figure 1). The whole system is stripped to the very basics of tower defense games and is made to be able to model the entities of the game from a configuration file. This ensures that if we have the specification for the game, the system will be able to build a fully functional simulation of it that will come integrated directly with the analysis tools.

Game Evaluation Metrics

Before we start building the algorithms we must first analyze the tower defense specification and determine the metrics for evaluation. The following aspects have been identified as having the most significance in a tower defense strategy:

- Player health at the end of the game
- Resources left unspent
- Difference between total damage dealt and total enemy hit-points
- Total game time

Player health at the end because it directly translates into how many computer enemy units have managed to pass the built towers, thus having more hit-points at the end means a better strategy has been used. Resources left unspent influence the value of the strategy because if you can manage to achieve the same end result spending less resources, it means the strategy used has higher efficiency. The difference in enemy total hit-points and the damage dealt is a subtle way of fine tuning the efficiency of the strategy. Dealing more damage than necessary to the enemies translates into resources that could have been saved on that extra damage.

Finally, total game time is a straightforward unit of efficiency measurement, with the strategy that finishes earlier having the greater score. It is to be noted though, that total game time should be computed only by the sum of the time in which waves of enemies are active and not add the time in which the user is thinking about the strategy as this not affects its value. Thus the final efficiency of the strategy may be calculated as:

$$Efficiency = a*PlayerHealth + b*ResourcesUnspent + c*(TotalDamage - TotalEnemyHealth) + d*TotalTime$$

In order to be able to compute the efficiency then we must know all the above information, but in tower defense games enemies come in waves and you only have information about the waves you have already passed and the current

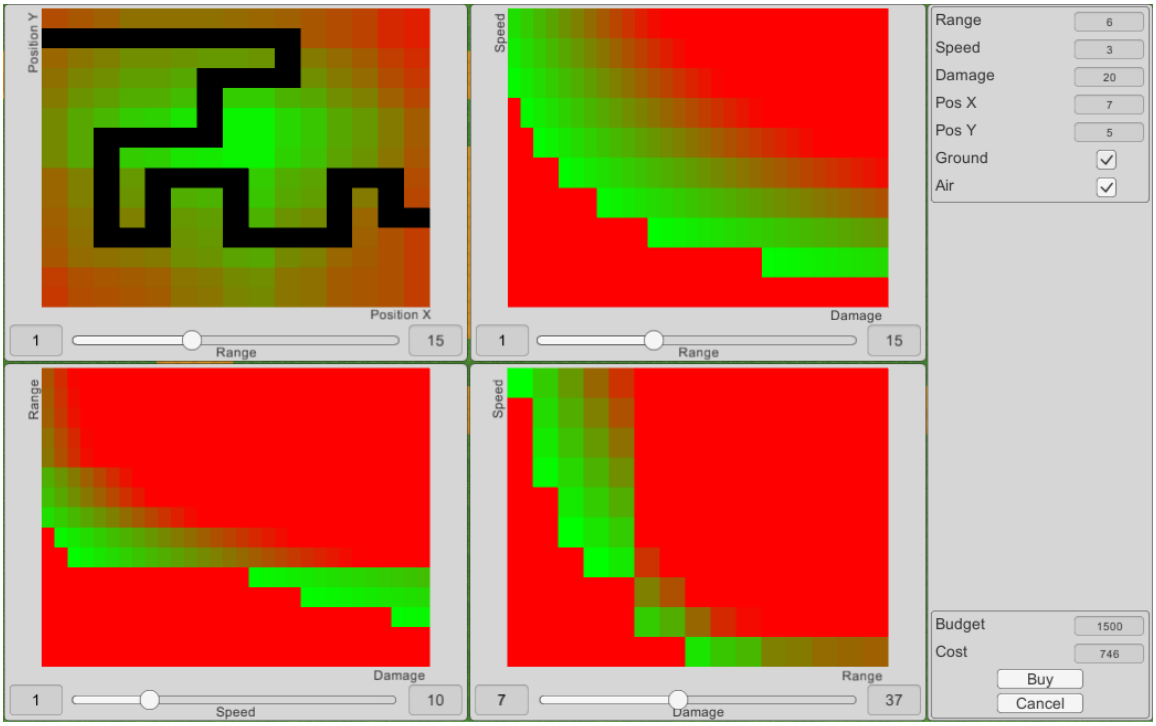

Figure 2. Turret Parameter Graphs.

one. This means that developing a strategy for the game is not possible from the beginning and that at least a play-through must be completed before we can do so.

Developing the Algorithms

The application then provides a method that will not only help the player get to the end of the game and learn the information about the waves as quickly as possible but in the process also learn enough information so that in the end a well-rounded strategy for the whole game can be computed. Thus the paper proposes 2 algorithms:

- Local Optimum – This algorithm will treat only the current wave of enemies with the resources at the player's disposal at that time. It will thus create a turret that has the optimal configuration in order to pass the current wave of enemies.
- Global Strategy – This algorithm will be run at the end of the game, once the information about all the waves is known. It will combine the local optimums developed along the way and generate a strategy considering the whole picture of the game.

To develop the equations for the algorithms we firstly need to know what is the n-dimensional problem space in the tower defense type of game. The following entities are what we must consider, they are split into fixed and variable to indicate that they are given by the specification of the game (fixed), or that they can be influenced by the player's decisions (variable):

Fixed:

Game_World {
    Map,
    Path
}
Enemy {
    Health,
    Speed,
    Damage
}

Variable:

Turret {
    Position,
    Range,
    Damage,
    Firing_Speed
}

The Local Optimum algorithm will try to balance the turret parameters in such a way that the turret cost is minimum, but the turret can also eliminate all the enemies in the current wave. The Range and Position parameters are strongly tied to the Game_World, and the perfect combination of these parameters ensures that the turret can fire for the longest amount of time possible on the enemies. The Range, Damage and Firing_Speed are tied to the Enemy component as they determine the total output damage the turret can achieve in the time the enemies are in the turret's range. The system probes a restrained domain of values for all of the parameters and displays the results to the user using the visual techniques described below.

Building the visualization tool

Up until now we established the metrics and a series of algorithms that can score a combination of the game parameters for a local case. The problem now is to build a

visualization tool that the human user can interact with, and from which he can guide the further search iterations of the algorithm.

It is very hard for people to reason about *n*-dimensional problem spaces especially when $n > 3$ and when the axes do not necessarily represent position. In order to make the user's experience easier we draw inspiration from 3D modeling software where you can see the 3-dimensional space not only from the perspective view but you can also chose to view from 2D viewports that display only 2 of the 3 dimensions at once.

In a similar fashion we choose to represent our problem space through a series of 2D projections where each axis represents the values for one of the turret's parameters (Figure 2). In addition to this representation each of our 2D viewports comes with a slider which is assigned to the domain of a 3rd parameter. By using this slider, the user can see how changes in the slider's parameter values affect the efficiency of the parameters displayed on the axes.

In order to further provide a better and faster understanding of the data, the points displayed in the 2D viewport are color coded with regards to the impact they have on the strategy's efficiency. Points which have a greater score are displayed in bright green and points on the graph which have a bad score are represented with red. In this way, the user can quickly scan the 2D space and find concentration points of green color indicating that is an area of convergence for the optimum parameter values.

As position is represented by the x and y coordinates on the grid map, it is not wise to separate these two parameters when displaying them in the viewports. The total number of possible combinations with 2 parameters on the axes and one on the slider for our model of tower defenses is 30, but most of them represent data that is not useful for the user as he cannot quickly reason about it. The identified useful viewports from which the user can choose the path of convergence are: a position viewport with the turret's x and y on the axes and the range parameter on the slider; and 3 viewports that cycle through combinations of turret damage, firing-speed and range on the axes and slider. With these 4 viewports, the user can view the turret parameters from different perspectives and using the coloring scheme can choose the best combination in order to create the local optimum turret.

Local Optimization and Global Strategy

By using the visual analysis tool at each stage (wave) of the game, the user ensures that he will reach the end of the game and have all the necessary information to construct a global strategy. We say global strategy and not global optimum because we want the game designers to have the possibility of generating strategies of different difficulties in order to aid the calibration of the difficulty levels in their games. This global strategy is driven from the choices made at the wave level, where the algorithm points out the optimal choices for that particular state, but the user has the possibility of choosing any values for the turret parameters he wants and proceed further with his selection. Thus the choosing all optimal locals will generate a "hard" difficulty setting while making not optimal choices will generate a lower difficulty setting.

Once the end of the game is reached, the application considers each local optimum as the threshold the strategy must pass for each stage. Also knowing the information about all the waves means that the algorithm can search for combinations of turrets that can achieve the thresholds of future waves at earlier stages. This reduces the necessity to build a turret at each wave and just pumps the parameters of the turret from the earlier stages to handle the later ones if this can be covered by the budget limitations. By analyzing the possible combinations to achieve all of the thresholds, the system can output a global strategy specification indicating to the user when to build a new turret and what its specifications should be.

The important factor in the whole process is that the global strategy bases its decisions on the local choices of the user. This allows users to create not only the best global strategy but also mediocre or weak strategies based on the local decisions and this is how game designers can calibrate their game difficulty levels, while player's will opt for the best local options to generate the best possible strategy.

Results

The finished application allows users to simulate in a test bed environment any kind of software which follows the standard tower defense game model. The modeling of specific game entities is done *via* the application's configuration file. Once the game is simulated, the application can be used as both a videogame providing entertainment value to the user, but also as a development or improvement tool.

Testing has been done on a personal tower defense project. The main reason behind this is that we didn't have access to all the data from a popular tower defense game in time or the game for which we had the data also contained a special feature which was out of the scope of this project which couldn't be modeled. The specification of our tower defense game however fits perfectly as it follows the basic model of such a game. It uses a grid like tile-map where the available tiles could be part of the path from the spawn location to the player's safe point or free spaces on which turrets can be built. The game also featured 2 types of turrets and multiple types of enemies.

We organized 2 types of tests in order to get results from both the perspective of a player improving on his strategy and from a developer trying to calibrate his game standpoint.

In the test regarding a player's strategy we first let the user play through our tower defense game as he would normally

do without having access to our analysis software. The end result of this stage was that he was able to beat the game but made some seemingly bad decisions along the way maybe spending more resources that he should have. He also lost a great deal of his life barely making it to the end of the game.

In the next stage we simulated the same tower defense game in the strategy analysis tool and let the player have another go at it. This time, by using the visualization tools for the turret parameters provided by the software, the player managed to defeat the game quicker and with no health loss.

Finally, after the player beats the game with our software, we let the program generate the strategy based on his decisions and played again through the original game following the instructions given in the generated text file.

We took the data from all three runs and measured the differences in terms of strategy efficiency based on the metric defined earlier. The data showed a big difference between the first run where the player didn't use the software and the second run where he could analyze the impact his decisions had on the game. The difference between the second run and the one where we followed the software generated strategy is not so great, but small improvements were spotted leading us to believe that in the case of other games which are harder than our example the extra computations may pay off.

In the second test scenario we assumed the role of a developer trying to configure difficulty levels for his game. We wanted to generate both an easy and hard level for a configuration of four enemy waves. The test started by configuring the existing level design in the test bed. After that in order to generate the hard difficulty we made the best decisions in turret parameters at each stage and wrote down how well the computer units did at each stage. After the end of this stage we compensated for the difference in the unit's current configuration and the damage output the ideal turrets could achieve. We ran the test another time and analyzed how well the turrets did against the new waves of enemies. After a small number of iterations we achieved a wave composition that could be fully destroyed only if the player made the best decisions in turret parameters and considered this the hard difficulty setting. Further, starting from this setting we ran some more iterations dimming down the specifications on the enemy wave compositions. This allows the user to make turret choices that are not ideal but still survive to the end of the game. With these configurations we achieved the easy difficulty setting.

Both test scenarios proved to be eventually successful which means that the tool can indeed be used for the described purposes. The second test however took more iterations than expected to achieve the desirable difficulty level so further improvement on this aspect should be done.

UNITY GAME ENGINE BASED DEVELOPMENT

Building a videogame or related software from scratch is a long and difficult process which requires a lot of pre-development for the tools and low level software to support it. The game engine chosen for this project is Unity3D because it is a mature software package and has a very active community providing great support [6] and tutorials [7]. Unity also supports coding in C# which is a high-level object oriented programming (OOP) language. This makes it easy to generate your own functionality and modify via scripting pre-existing assets.

Unity deals with objects as entities with a transform component to which multiple other components can be added. The transform specifies the object's location in 3D space, its rotation and scale. Unity treats everything as a component allowing for great modularity and plug and play behavior. Everything from mesh renderers to materials to code scripts are added to objects as components.

CONCLUSIONS

Visual Analysis based Strategy Development

To generate the strategy for the configured game you must first beat all the waves using the normal game mode. Once you finish finding out the individual strategy for each wave you will be presented with a game won screen which has a button labeled "Generate Strategy". If you ask for generating the strategy the system will start displaying a loading bar and begin computation of the strategy based on the choices you have made during normal gameplay. When the loading bar completes a message saying "The strategy has been generated!" will appear. From this point you may choose to close the application and open the text file strategy.txt found in the root directory of the application. The file will contain step by step instructions on how to execute the generated strategy for the tower defense configuration entered in the config.txt file. The strategy text file contains a clear specification of every parameter for every turret needed to execute the tactic and timing instructions as to when you may be able to build each turret. For each wave information about the total budget a player has at his disposal is presented in order to provide a checking mechanism while implementing the strategy. The steps presented in the file may be applied to any tower defense game which fits the configuration specification. This means it is applicable to both another run in the test bed application but also to released games.

Specific Future Development

The important thing to keep in mind is that the current state of the software and algorithms do not produce a **global optimum** solution for any tower defense game even if the user chooses only optimums for each of the waves of the game. The **global strategy** algorithm provides just a small optimization for the choices the player makes at each local stage in order to put locals into the global context. This is by no means the ideal solution. Future improvements of the

tool may include algorithms for devising a true optimal strategy and this could be done similarly as the local optimums through a visualization tool that will help the user search for it in the problem space.

Missing currently from the tower defense strategy analysis application is a dedicated user interface to modify the configuration file for the game. This makes it that each time you want to test bed a new tower defense specification you must enter its details manually in the configuration file and also make sure you respect the format of the file. From own experience this is an easy way to introduce bad data in the game and spend large amounts of time figuring out what went wrong. A special load and save mechanism would also improve the user experience as it would allow storing and sharing of different game configurations at a time.

Another feature that would come as a great improvement on the gameplay and usage of the software is the ability to replay the last wave. Often, mostly when trying to figure out difficulty levels for the game, the user would want to test different turret configurations. In the current setup the player would have to reach the same wave he wants to analyze from the beginning of the game if he wants to test out another configuration. This means reproducing the exact steps that led to that wave. A simple replay wave button would load the game state from the beginning of the previous wave. To support this feature a copy of the game state would have to be saved at the beginning of each wave, and a mechanism for restoring that game configuration when the user asks for replay wave, must be developed.

Finally, for the specific improvements the ability to interact with the global strategy algorithms would benefit any type of user greatly. Currently, the system runs the strategy analysis algorithm behind the scenes and generates what it considers the optimal approach based on the gameplay. However, the computer must go through a lot of data and a visualization scheme similar to the one for local optimums would also let the user guide in the search for global strategies. Another advantage into developing this branch of the application is the fact that a user might be able to generate multiple global strategies based on the previous gameplay.

Currently, the system choses one strategy which is based on calculations, but by performing this task the computer actually generates multiple feasible strategies that are simply dumped. Having access to all the generated strategies the computer develops may help game designers in the process of creating varied game experiences. This feature could be implemented by creating a secondary visualization module for the set of global strategy algorithms.

General Future Development

One direction the project may take includes the expansion of the application to provide its users with reverse engineering tools for game designers. This feature is more suitable for the development of games than for the normal player. It implies that by using the strategy profiling algorithms implemented in the software, a user could generate through the tool a specification for the design of a level in the game.

Another direction for further development is the creation of adaptive AI systems that use the strategy analysis module to change the way the computer plays the game based on decisions made by the human player. For example, on a hard difficulty setting in a game, the AI system could analyze what the player is doing through the strategy analysis module and adapt its units to be stronger or on the limit of the player's build. This way the computer can always change the way the game works offering strong gameplay and great chance for game replay-ability as the experience is never the same.

ACKNOWLEDGMENTS

The research has been carried out in the Computer Graphics and Interactive Systems Laboratory of the Computer Science Department, in the Technical University of Cluj-Napoca.

REFERENCES

1. Gregory Jason, *Game Engine Architecture*, CRC Press 2014, ISBN: 978-1-4665-6001-7.
2. Fraser James, Michael Katchabaw, and Robert E. Mercer, *A methodological approach to identifying and quantifying video game difficulty factors*, Entertainment Computing 5.4 (2014): 441-449.
3. Hsu Shang Hwa, Ming-Hui Wen, and Muh-Cherng Wu, *Exploring design features for enhancing players' challenge in strategy games*, CyberPsychology & Behavior 10.3 (2007): 393-397.
4. Alexander Justin T., John Sear, and Andreas Oikonomou, *An investigation of the effects of game difficulty on player enjoyment*, Entertainment Computing4.1 (2013): 53-62.
5. Aponte Maria-Virginia, Guillaume Levieux, and Stephane Natkin, *Measuring the level of difficulty in single player video games*, Entertainment Computing2.4 (2011): 205-213
6. Unity3D User Manual, [Online]: http://docs.unity3d.com/Manual/index.html
7. Unity3D Tutorials, [Online]: http://unity3d.com/learn/tutorials/modules
8. Uneal Engine official website, [Online]: https://www.unrealengine.com
9. RTS Games, [Online]: https://en.wikipedia.org/wiki/Real-time_strategy

Developing a Navigational 3D Audio Game with Hierarchical Levels of Difficulty for the Visually Impaired Players

Oana Balan
University Politehnica of Bucharest, Faculty of Automatic Control and Computer Science
Splaiul Independentei 313, Bucharest 060042, Romania
oanab_2005@yahoo.com

Alin Moldoveanu
University Politehnica of Bucharest, Faculty of Automatic Control and Computer Science
Splaiul Independentei 313, Bucharest 060042, Romania
alin.moldoveanu@cs.pub.ro

Florica Moldoveanu
University Politehnica of Bucharest, Faculty of Automatic Control and Computer Science
Splaiul Independentei 313, Bucharest 060042, Romania
florica.moldoveanu@cs.pub.ro

Alexandru Butean
University Politehnica of Bucharest, Faculty of Automatic Control and Computer Science
Splaiul Independentei 313, Bucharest 060042, Romania
alexandru@butean.com

ABSTRACT
The Binaural Navigation Game is a 3D audio game with hierarchical levels of difficulty that can be used by both normal sighted or visually impaired players. The purpose of the game is to test and train the sound localization skills, to entertain and to provide an alternative to the common video games available on the market nowadays. The sonification technique is based on the perception of 3D binaural sounds synthesized with non-individualized Head Related Transfer Functions (HRTFs) and on the inversely proportional sound intensity encoding of distance. Furthermore, we use an original method based on the simultaneous perception of two types of noise that aims to reduce the incidence of front-back confusions. This game has been tested in an experimental procedure (comprised of a pre-test, a haptic-auditory feedback based training and a post-test session) in which 10 visually impaired subjects (with a percent of residual vision ranging from 0% to 15%) have participated. This paper aims to present the design and development of the Binaural Navigation Game, the Game Editor (an application that allows the experimenter to set the layout of new sets of levels) and the Binaural Game Analyzer, a tool used to visualize and evaluate the players' performances..

Author Keywords
3D sound; HRTF; audio game; navigation skills; orientation and mobility skills.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION
Sound is an immersive and expressive medium that can convey a wide range of information on the meta-level of perception. In video or audio games, it provides powerful and reliable cues (concerning events and situations) that enhance the level of interactivity, engagement and the overall user game playing performance. Although in recent years many researchers and game developers have shifted their attention on the audio component, it is highly reduced in comparison to the visual content in most of the games available for PC or for mobile platforms [10]. Moreover, the audio-only games draw less interest than the video ones, being regularly developed by small groups of programmers or researchers and usually for experimental purposes [1].

The audio-only games provide accessibility features for the users who are partially or completely deprived of the visual sense. Furthermore, they should be designed with particular focus on the visually impaired players' particularities, skills and needs [4]. The audio games can be employed for educational purposes [2] or for training various abilities, for enhancing sound localization, improving the transfer of skills into real-world situations or for entertainment only [6, 14, 18, 21].

This paper presents the development of a navigational 3D audio game in which the entire navigation is based on the perception of 3D binaural cues that provide the listeners a complete spatial perception of the auditory setting. Although the sonification design is focused on the use of

3D binaural sounds as the only means of navigation, the acoustic environment is complemented by auditory icons and earcons [7] that enhance user interaction and provide a rapid response to the players' input. The purpose of the game is to find the location of several hidden targets while avoiding blocking obstacles. The number of targets and obstacles varies among levels, according to the difficulty.

The Binaural Navigation Game has been employed in an experimental procedure in which we tested the sound localization skills of 10 visually impaired people (5 women and 5 men, aged 27-63, with a percent of residual vision ranging from 0% to 15%) before and after a haptic-auditory feedback based training session. However, the focus of this paper is not on describing the sound localization experiment or its results (they will be discussed in another paper) but on introducing the developmental approach of the software tools used in our tests. The game level layout has been designed using the Game Editor software. Moreover, in order to assess the users' game playing performance, we have designed an application entitled Binaural Game Analyzer that provides visualization, real-time playback and statistical functionalities through an accessible and interactive interface.

THE SONIFICATION APPROACH

For encoding the direction of the target sound sources, we used a sonification technique based on the continuous and simultaneous perception of two types of noises (white and pink noise), in varying proportions, according to the direction of the sound source in space. To the front, the listener hears only white noise, while to the back he perceives solely the pink noise. In the right hemifield, the amount of white noise decreases (and that of pink noise increases), reaching equal levels at 90 degrees to the right. In the left hemifield, the relationship between the white and the pink noise is complementary to that from the right side, so that both noises record identical perceptive levels at 90 degrees to the left.

The aim of this encoding is to the help the listeners to differentiate the direction of the sound sources located to the front and to the rear and to avoid the very common reversal errors that occur in virtual auditory displays, generally known as front-back confusions [11, 20, 22] (situation when the listener encounters difficulties in determining whether the sound source is originating from the front or from the back). In this way, the perception of the spectral patterns of the noises, together with the directional cues provided by the 3D sounds convey a more accurate image of the location of the sound source in space. On the other hand, the obstacles have been encoded using a broadband alarm sound. Both types of sounds (the combination of white and pink noise and the alarm sound) have been processed using the Csound programming language [9] and synthesized with the non-individualized HRTFs from the MIT database [17].

The HRTFs are the basis of binaural rendering, as they represent the modifications that occur in the spectrum of the sound from the position of the source en route to the human subject's ears. These complex transfer functions describe the magnitude of the frequency response in correlation to the source position, defined by azimuth and elevation angle [3]. Non-individualized sets of HRTFs are generally recorded in anechoic chambers, using dummy head mannequins. However, as the HRTFs strongly depend on the morphological characteristics of the external auditory system (the pinna and concha have a non-linear frequency response [12]), the non-individualized (or generic) HRTFs introduce severe perception artifacts and localization errors [19]. On the other hand, recording personalized sets of HRTFs for each listener apart is a time-consuming and impractical task, so that most of the 3D audio based applications employ non-individualized HRTFs on a large scale. The simulation of the virtual position of a particular sound source in space can be achieved by filtering a given signal with the pairs of HRTFs corresponding to the left and to the right ear, creating thus left and right ear signals that are delivered through headphones [15, 19].

The HRTFs from the MIT database have been recorded using a dummy-head mannequin and a set of 7 loudspeakers placed at 1.4 meters distance from it. The elevation positions in the median plane range from -40 degrees (below the interaural axis) to +90 degrees (directly overhead). The azimuth measurement locations were disposed at 5 degrees angle increment all around the listener's head [17].

Csound is the first sound processing programming language, firstly developed in 1984 at the Massachusetts Institute of Technology and written entirely in C. Since then, it improved considerably, becoming today one of the most powerful and reliable music generating instruments. Moreover, it is available for the Linux, Macintosh and Windows operating systems, as well as for the new mobile platforms, such as Android [8].

In order to obtain 3D binaural stimuli for our game, we convolved the stereo sounds (the combination of white and pink noise in varying proportions and the alarm sound encoding the obstacle position) using the hrtfmove2 opcode (orchestra function, typical for the Csound programming language) [8]. The hrtfmove2 opcode generates 3D binaural sounds to be rendered over headphones for any location in the 3D space using the magnitude interpolation algorithm and a phase model based on the Woodworth's formula [13] for Interaural Time Difference (ITD) [13,22]:

$$ITD = \frac{r}{c}(\theta + \sin\theta) \quad (1)$$

Where r is radius of the subject's head, θ is the angle between the listener and the sound direction and c is the speed of sound, of approximately 340 m/s at sea level.

Magnitude interpolation means that the four nearest HRTF values (left, right, below and above) are linearly interpolated in order to obtain the most appropriate magnitude level for the given angular position [12]. The piece of code that processes the given audio input into a 3D binaural sound is the following:

aleft, aright hrtfmove2 aSignal, kAzimuth, kElevation, $HRTF_LEFT, $HRTF_RIGHT

where aleft and aright are the output signals for the left and for the right channels, aSignal represents the input stimulus, kAzimuth is the azimuth angle, kElevation corresponds to the elevation angle (set by default to 0 in our game), while $HRTF_LEFT and $HRTF_RIGHT represent the left, respectively the right HRTF spectral data files.

In addition to 3D sounds, we used the modifications in the perceived sound intensity to convey relevant information concerning the distance between the current virtual position of the player and the location of either target objects or obstacles. The inversely proportional relationship between the sound intensity level (SI) and the distance is the following:

$$SI=\begin{cases} 0, & d > dmax \\ SI_{MIN} + (SI_{MAX} - SI_{MIN}) * (1 - \frac{d}{dmax})^2, & d \leq dmax \end{cases} \quad (2)$$

Where d is the current Euclidean distance between the virtual position of the player and the target object/obstacle, dmax is the maximum recorded distance to the active target (in the case of obstacles, it has a fixed value of 150 pixels), SIMIN=0.05 represents the minimum perceptible sound intensity and SIMAX=1 corresponds to the maximum perceivable auditory level.

The auditory icons are "short, icon-like sound events that have semantic connections to the physical events they represent" [7]. They are largely employed in audio games, as they provide the player an easily understandable connection between their significance and the situations and events with which they are associated. Examples of auditory icons are the sound of footsteps that give clues about someone approaching, the sound of a door opening, a dog barking etc. In our game, we used the sound of a crash to indicate the collision with an obstacle.

The earcons are short, simple, abstract sounds (sometimes composed only of a few notes) that have associated a relationship with the events they represent [7]. As they do not have instantly recognizable characteristics that could be achieved from previous experience, the listener needs to learn first the meaning of the earcons before using them in auditory environments. The earcons we used in our game are a sparkling sound that can be heard when the player identifies the location of the hidden target sound source and the sound of a bell ringing that announces the end of the game.

THE BINAURAL NAVIGATION GAME

The Binaural Navigation Game (Figure 1) has been written in the C# programming language [5] and designed using the Microsoft Visual Studio Integrated Development Environment [16].

At the beginning of each game session, the player is required to introduce some personal information: name, age, sex, the number of years since he is suffering from the visual impairment and the number of the set level that he wants to play. The player can choose from different sets of levels that are created using the Game Editor application.

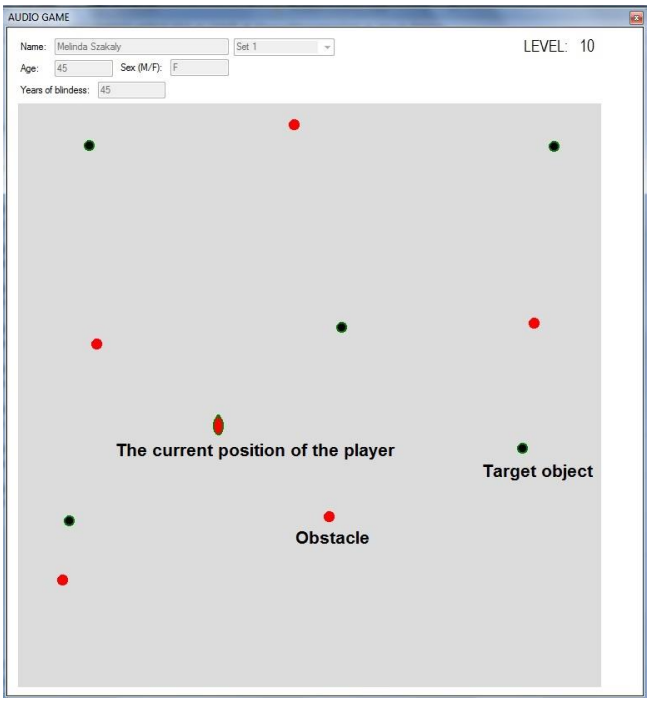

Figure 1 The Binaural Navigation Game

At each level, the player is required to navigate freely, using the mouse (or touchpad) movement as interaction method, in order to identify the location of the hidden target sound sources (only one target 3D sound can be heard at a moment of time), while avoiding blocking obstacles (the obstacles are audible within a 150 pixels range). After the location of the current target has been identified, it consequently turns silent and the next sound object becomes audible. If two or more obstacles are positioned at a distance smaller than 150 pixels, they are perceived simultaneously, although their corresponding 3D sounds vary in intensity and direction.

THE GAME EDITOR

The experimenter can design the layout of the game (the distribution of target auditory objects and obstacles) in a well-defined and interactive manner. Thus, he is required to introduce the level number and to place items (either targets or obstacles) on the canvas (Figure 2). The positions of the target objects need to be set first (by pressing the "Set

Targets" button from the user interface application), in ascending order, from the bottom side of the playing canvas to the top of it. Subsequently, the user is required to press the "Set Obstacles" button in order to start adding obstacles to the current level. In this case, no more than 5 obstacles per level can be introduced. The locations of the target objects or obstacles are set by clicking on the playing window in the positions where the items need to be placed. The game layout components are encoded by green (the target objects), respectively red circles (in the case of the obstacles).

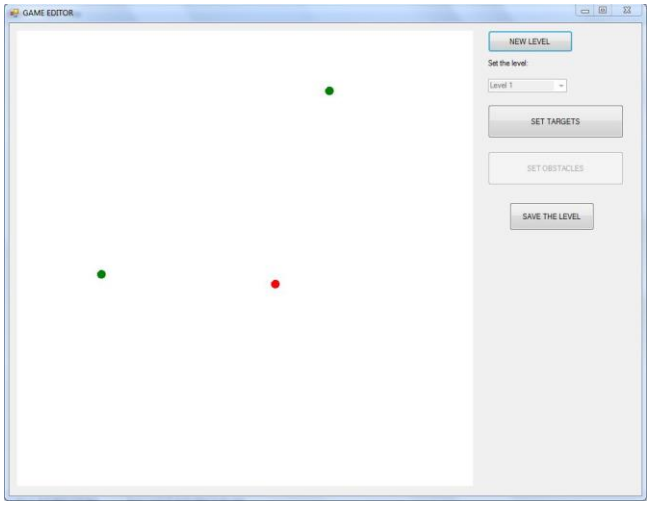

Figure 2 The Game Editor

The data for each level is stored in separate .txt files that comprise on distinct lines the items type ("t" for targets and "o" for obstacles) and the 2D coordinates in the horizontal plane for each of them. For example, the content of the "Level 3.txt" file, where we have 2 target sound sources and 2 obstacles is the following:

Level 3
t 127 481
t 627 36
o 201 559
o 630 125

THE BINAURAL GAME ANALYZER

In order to evaluate the game playing performance for each target object and for each level apart, we defined the following parameters:

P1 – the ratio of the distance travelled by the player from the starting position until he reaches the target sound source to the minimum possible distance between the starting position and the location of the target object (which is actually the Euclidean distance between these two points). For the first round of every level, the starting position is the center of the bottom side of the playing window, while for the others it is represented by the location of the previously identified target object.

P2 – the percent of correct travel decisions (movements performed towards the sound source, minimizing the distance to the target object).

P3 – the round completion time in seconds, i.e. the time needed to identify the location of the target source.

P4 – the number of obstacle collisions per round.

The Game Analyzer tool (Figure 3) allows real-time visualization and audio playback of the users' performances for each round (for each target object identification) and level.

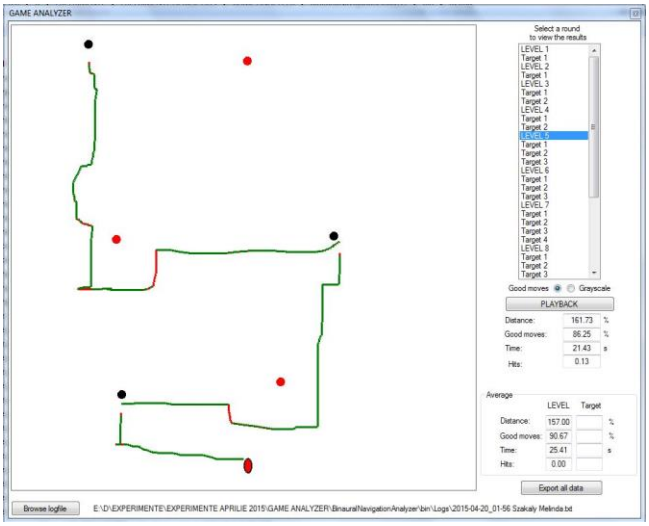

Figure 3 The Binaural Game Analyzer

This figure presents the player's performance for level 5 (which has 3 sound targets and 3 obstacles), using the "Good moves" visualization option.

By pressing the "Browse log file" button, the experimenter can upload the log file that has been created and updated during the gameplay session and that stores all the data concerning the user's performance. The experimenter can select a round or a level from the list box control on the right side of the interface. The "Good moves" visualization option displays the player's efficient displacements (mouse movements effectuated towards the sound source) as green colored segments, while the wrong movements that maximize the distance to the sound source are represented by red segments. The "Grayscale" visualization option presents the path travelled by the listener from the starting position until he reaches the target object of each round in progressive grayscale color tones, from light gray to black (Figure 4). By pressing the "Playback" button, the experimenter can see in real-time the performance of the player and simultaneously hear in the headphones the 3D sounds that the user perceived while navigating from the starting position to the target sound source. Furthermore, the Game Analyzer interface displays the mean values of parameters P1-P4 for each round and level, as well as for all the levels of the game. The "Export all data" button allows

the statistical information to be exported to an Excel file, together with the player's personal information, such as name, age, sex and the extent of the visual impairment condition.

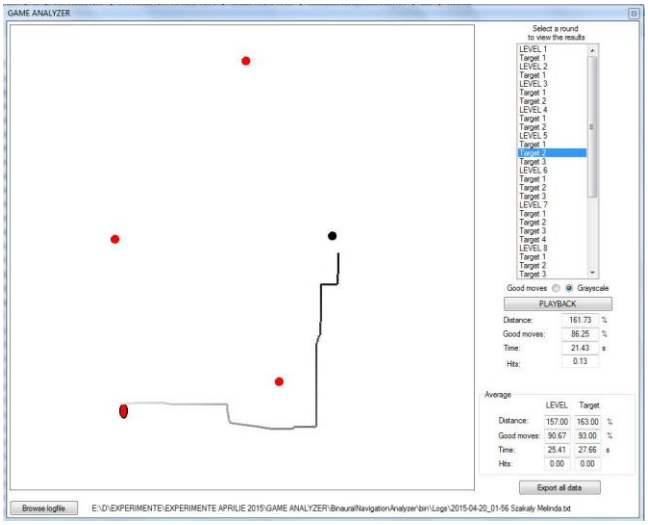

Figure 4 The "Grayscale" visualization option

This figure presents the player's performance for the second round of the 5th level of the game, using the "Grayscale" visualization option which paints the path travelled by the user in progressive grayscale tones

The purpose of the Game Analyzer application is to evaluate the game playing performances of the users through a complete statistical analysis (the mean values of the four studied parameters, distributed for each round, level and for all the levels of the game), as well as to provide a deeper understanding of the issues the subjects are confronted with during gameplay – the perception of 3D sounds, sound localization accuracy, front-back confusions, auditory-based spatial navigation and the interaction modality.

The Analyzer has been used as evaluation tool in another sound localization experiment, where the users were asked to identify the location of a hidden sound source by navigating freely from the starting point to the location of the target (randomly generated on the margin of a circle of fixed radius). In that case, the sound stimulus was a train of continuous white noise filtered with non-individualized HRTFs. The Analyzer instrument helped us to thoroughly investigate the occurrence of front-back confusions and the difficulties the players encounter when they have to discriminate the direction of the sources located in the front and in the back (Figure 5). These observations conducted to the idea of combining two types of noises (white and pink noise) with different spectral profiles in varying proportions, according to the direction of the sound source in space. The purpose of this sonification approach is to offer the listeners a more accurate perception of the sound sources originating from the front and from the rear hemifield.

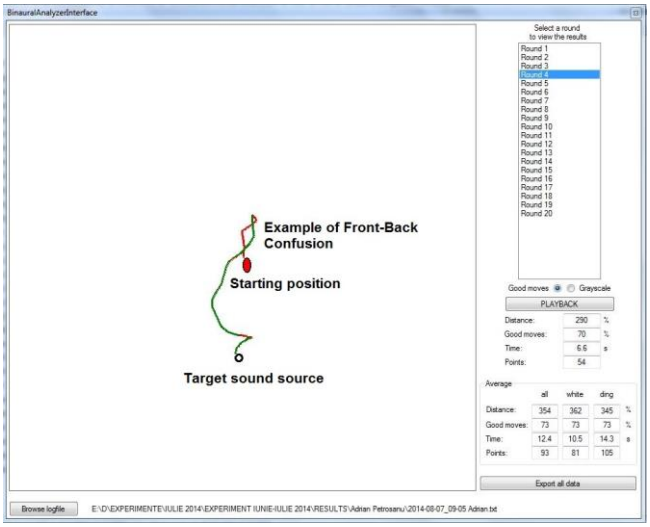

Figure 5 Example of front-back confusion

This figure shows that the listener, instead of going down to the target sound source, makes a front-back confusion and navigates initially up and then returns back on the correct path to the target

In order to play the Binaural Navigation Game or to operate with the Game Editor and with the Binaural Navigation Analyzer, the user is required to previously install the Csound kit [9] and the Microsoft .NET Framework.

CONCLUSIONS

To sum up, this paper provides a contribution to the design of navigational 3D audio games by presenting the development of the Binaural Navigation Game and that of its adjacent tools – the Game Editor, which allows the experimenter to simply and interactively define new sets of levels and the Binaural Game Analyzer, which offers statistical and real-time visual and auditory playback of the players' performances. The sound stimuli used were the combination of white and pink noise in varying proportions, according to the direction of the source in space that encodes the presence of target objects and the broadband alarm sound corresponding to the obstacles' positions. Both types of sounds have been convolved with the pairs of non-individualized HRTFs for the left and for the right ear, creating 3D binaural sounds delivered via headphones. The directional cues offered by the 3D binaural stimuli, together with the inversely proportional sound intensity encoding of distance aim to help the listener to easier identify the location of the auditory items (to reach the target obstacles and to avoid the blocking obstacles), to improve navigation and to offer the player a complete perception of the auditory environment. Other important sound cues are the auditory icons and the earcons that create a solid mapping between the acoustic perception and the events and situations from the game, enhancing thus the sense of presence and the level of player immersion.

Future plans involve the development of a portable version of the game, available on mobile platforms, that would have an additional haptic and vibro-tactile component.

ACKNOWLEDGMENTS

The work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/132395 and POSDRU/159/1.5/S/134398. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

REFERENCES

1. Bălan, O., Moldoveanu, A., Moldoveanu, F. Navigational Audio Games: An Effective Approach Towards Improving Spatial Contextual Learning for Blind People. The International Journal on Disability and Human Development, 14, 2 (2015), 109-118.
2. Bălan, O., Moldoveanu, A., Moldoveanu, F., Dascălu, M.I. Audio Games- A Novel Approach Towards Effective Learning in the Case of Visually-Impaired People. In Proceedings of the ICERI Conference, Seville, Spain, (2014), 6542-6548.
3. Bilinski, P., Ahrens, J., Thomas, M. R. P., Tashev, I., Plata, J. C. HRTF magnitude synthesis via sparse representation of anthropometric features. In Proc. IEEE ICASSP, Florence, Italy, (2014), 4501-4505.
4. Brieger, S.W. Sound Hunter developing a navigational HRTF-based audio game for people with visual impairments. In Proceedings of the Sound and Music Computing Conference 2013, Stockholm, Sweden (2013), 245-252
5. C# Programming Language, https://msdn.microsoft.com/en-us/library/67ef8sbd.aspx
6. Connors, E.C., Chrastil, R.E., Sanchez, J., Merabet, L.B. Action Video Game Play and Transfer of Navigation and Spatial Cognition Skills in Adolescents Who Are Blind. Frontiers in Human Neuroscience (Internet) (2014).
7. Csapó, Á., Wersényi, Gy. Overview of auditory representations in human-machine interfaces. Journal ACM Computing Surveys (CSUR), 46, 2(2013).
8. Csound Manual, http://www.csounds.com/manual/html/hrtfmove2.html
9. Csound Programming Language, http://www.csounds.com/downloads/
10. Friberg, J., Gärdefors, D. Audio games: New perspectives on game audio. Advances in Computer Entertainment Technology, 5, 3 (2004), 1-7.
11. Hill, P.A., Nelson, P.A., Kirkeby, O. Resolution of Front-Back Confusion in Virtual Acoustic Imaging Systems. Journal of the Acoustical Society of America, 108, 6 (2000).
12. Lazzarini, V., Carty, B. New Csound Opcodes for Binaural Processing. In Proc. 6th Linux Audio Conference, Cologne, Germany, (2008), 28-35
13. Letowski, T., Letowski, S. Localization Error: Accuracy and Precision of Auditory Localization (Chapter 4) in Advances in Sound Localization, 2011.
14. Merabet, L.B., Sanchez, J. Audio-Based Navigation Using Virtual Environments: Combining Technology and Neuroscience. AER Journal: Research and Practice in Visual Impairment and Blindness, 2, 3(2009),128-137
15. Meshram, A., Mehra, R., Yang, H., Dunn, E., Frahm, J.M., Manocha, D. P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound. In Proceedings of the International Symposium on Mixed and Augmented Reality, (2014).
16. Microsoft Visual Studio, https://www.visualstudio.com/
17. MIT HRTF Database, http://sound.media.mit.edu/resources/KEMAR.html.
18. Sanchez, J., Lumbreras, M. Usability and Cognitive Impact of the Interaction with 3D Virtual Interactive Acoustic Environments by Blind Children. In Proc. 3rd Intl Conf. Disability, Virtual Reality & Assoc. Tech., Alghero, Italy (2000).
19. Spagnol, S., Geronazzo, M., Rocchesso, D., Avanzini, F. Extraction of Pinna Features for Customized Binaural Audio Delivery on Mobile Devices. In Proceedings of International Conference on Advances in Mobile Computing, (2013).
20. Wallach, H. The Role of Head Movements and Vestibular and Visual Cues in Sound Localization. Journal of Experimental Psychology, 27, 4(1940), 339-368.
21. Westin, T. Game Accessibility Case Study: Terraformers- a real-time 3D Graphics Game. In Proc. of the Fifth International Conference on Disability, Virtual Reality and Associated Technologies, Oxford, UK, (2004).
22. Wightman, F.L. Resolution of Front-Back Ambiguity in Spatial Hearing by Listener and Source Movement, J Acoust Soc Am, 105, 5 (1999), 2841-53.

Text Generation Starting From an Ontology

Dragoş Alexandru Cojocaru
Politehnica University of Bucharest
Bucharest, Romania
dragos.cojocaru@cti.pub.ro

Ştefan Trăuşan-Matu
Politehnica University of Bucharest
and
Research Institute for Artificial Intelligence
Bucharest, Romania
stefan.trausan@cs.pub.ro

ABSTRACT

The subject of this paper is the development of an application which generates natural language text, starting from an OWL ontology. The Natural Language Generation, in the context of Semantic Web, represents a relatively new field of research, but due to the capabilities of the ontologies (central element of the Semantic Web) of being dynamically modified and completed with new information, the theme of the application is of great importance. The project employs Rhetorical Structure Theory to structure hierarchically the ontological content, resulting in a human-like discourse structure. Since the Semantic Web is continuously adding machine-readable content, the user can take advantage of this impressive database of knowledge transformed into coherent texts for human with the aid of our application.

Author Keywords

Semantic Web; ontology; Rhetorical Structure Theory; natural language generation

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

The Semantic Web aims at enriching the current Web with an additional layer of data which can be read and analyzed by intelligent agents [2]. Inside the Semantic Web, the central element is represented by the ontology. This element describes a domain of knowledge by introducing axioms which define the properties of a certain concept and the relations with other fields . Ontologies may be completed by adding new information regarding a notion and new knowledge can be deduced with the help of inference rules in the model of description logics [1].

The application presented in this paper has the objective to transform knowledge and information from an ontology into natural language text, in a manner as concise and expressive as possible, text dedicated to the human user. These characteristics of the resulting text are obtained with the help of the Rhetorical Structure Theory, which uses discourse markers to emphasize the relations between clauses.

Applications in the same field of text generation from an ontology are usually "verbalisers" which translate each axiom from the ontology into a sentence. The main disadvantages of this kind of applications are represented by the monotony, lack of expressivity of the resulting texts and the presentation of a great amount of information which might be either redundant or difficult to follow [2, 3].

A more advanced system in this field is "NaturalOWL", which generates coherent, fluent texts due to algorithms based on the "Centering Theory" [5]. In a similar way, the current application uses referring expressions for subjects and objects, sentence aggregation and techniques to prevent information repetition. On the other hand, the current application intends to obtain a more structured text with the aid of Rhetorical Structure Theory.

The application is functional for any ontology, but in order to obtain the final text, domain dependent data is required. Therefore, at present, only information in the context of the Wine ontology is constructed, which defines the way to treat each property in order to build sentences.

In the next sections there will be presented the application architecture, the process of generating text, suggestions for future work and the conclusions of this paper.

SYSTEM ARCHITECTURE

The text generation is developed in several steps (see Figure 1), where the result of a stage represents the input for the next stage. We will now discuss the process behind each stage separately.

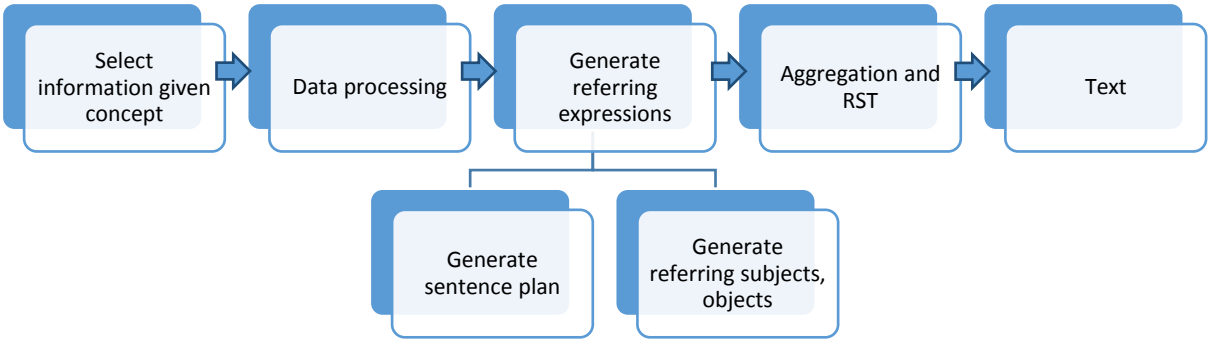

Figure 1.Application architecture describing the main stages of development towards reaching the final natural language text

Information selection

The application receives the name of a class or individual from a specified ontology. Using the Jena framework, the ontology is loaded and consequently we have all the available knowledge from the chosen ontology, as well as from imported ontologies.

Considering the name of the class or individual, we select the axioms which refer this concept. For instance, assuming the input is the name of a class, we begin by extracting the superclasses, subclassses and equivalent classes from the graph of the ontology.

Next, we obtain the other RDF triples which indicate the properties of the class, whether it is an intersection, reunion class or not and its constituent elements. The following step represents the extraction of data related to objects connected to the initial class through properties, in order to provide their description, too. If we present the subject S and we have the RDF triple <S, P, O>, where O is also a class, we would also describe the object O in a similar manner as we would do for S. The stage of axioms selection does not include the description of property classes, or further links in the graph, because they either bring uninteresting information for the user (presenting property characteristics such as: transitive, functional), or they result in a text difficult to follow [5].

Data processing

The selected axioms are transformed into their own form of triplets <Subject, Predicate, Object>. Because the Object in a RDF triple may be an unnamed class representing an union or an equivalent class, we break the object, building new triplets in which the object is a named class, individual, or data value of a property.

The object O of a triplet has to be recursively transformed as we further explain, until we reach a named class. If subject S is an intersection class of objects O_i we produce multiple triplets <S, isA, O_i>. Similarly, if S is a union of other classes, we obtain a disjunction of triplets <S, isA, O_i>. If we encounter an object which is a complement of a class the resulted triplet will be <S, isNotA, O>. Considering the following fragment of ontology:

```
<owl:Class rdf:ID="Meursault">
  <rdfs:subClassOf>
   <owl:Restriction>
    <owl:onProperty rdf:resource="#hasBody"
/>
    <owl:hasValue rdf:resource="#Full" />
   </owl:Restriction>
  <owl:Restriction>
   <owl:onProperty rdf:resource="#hasFlavor"
/>
    <owl:hasValue rdf:resource="#Moderate" />
   </owl:Restriction>
  </rdfs:subClassOf>
  <owl:intersectionOf
rdf:parseType="Collection">
   <owl:Class rdf:about="#WhiteBurgundy" />
   <owl:Restriction>
   <owl:onProperty rdf:resource="#locatedIn"
/>
    <owl:hasValue
rdf:resource="#MeursaultRegion" />
   </owl:Restriction>
  </owl:intersectionOf>
 </owl:Class>
```

we obtain the following triplets:

<Mersault, hasBody, Full>

<Mersault, hasFlavor, Moderate>

<Mersault, isA, WhiteBurgundy>

<Mersault, locatedIn, MersaultRegion>

A special case is represented by the constraints over the properties: restriction value, cardinality. As we can observe from Table 1, if we have a cardinality restriction on a property we transform in the triplet <S, Property , n>. In the case of a "AllValuesFrom" or "SomeValuesFrom" restriction we obtain an intersection and union respectively, of triplets, <S, P, O_i>, where O_i belongs to the set of restriction values.

Property constraints	**Resulted triplet**
<owl: Class A> <owl:onProperty rdf:resource=#Prop /> <owl:maxCardinality> n </owl:maxCardinality>	<A, maxCard(Prop), n>
<owl: Class A> <owl:onProperty rdf:resource=#Prop /> <owl:allValuesFrom> <owl:oneOf Collection <owl : Thing rdf:about=#B> <owl: Thing rdf:about=#C>	<A, Prop, AND(B, C)>
<owl: Class A> <owl:onProperty rdf:resource=#Prop /> <owl:SomeValuesFrom> <owl:oneOf Collection> <owl : Thing rdf:about=#B> <owl: Thing rdf:about=#C>	<A, Prop, OR(B, C)>

Table 1. The resulting triplets for property restrictions: cardinality and value

Generate sentence plan

The obtained triplets represent the base for generating the sentences, using a template for each property. This template specifies certain patterns and parts used in the specific sentence for each property(verb, prepositions), the sentence structure and allows the filling of fields with information received from the triplets [5]. A model of sentence plan is the following:

Subject_Expression(article, number) – Verb(voice) – Nouns / Adjectives(optional) – Object_Expression.

Property: locatedIn

Template: Subject_Expression – "to locate"(passive voice) – in(preposition) – Object_Expression.

In the case of properties which also hold constraints, it is generated a plan according to the type of restriction. For instance, using the same property "locatedIn", but also having the maximum cardinality restriction, the result will be:

Subject_Expression – "to_locate"(passive voice) – in(preposition) – "at most" – n – "regions: " – Object_Expression. ("LaneTannerPinotNoir is located in at most 2 regions: SantaBarbara and Sonoma.")

Generate referring expressions

The subject and the object of a triplet can have different forms inside a sentence. Initially, they are obtained through the tokenization of the subject or object of the triplet or using the ontology field rdfs:label, a string which represents the natural language form of that concept. Then, in a similar manner to sentence plan we can provide a template for the subject and the object. This plan indicates what kind of article should be used with this noun(definite, indefinite), its number and optional adjectives. Moreover, to avoid the repetition of the subject, the noun can be replaced in the following sentences with a pronoun, a determiner or a demonstrative pronoun:

<Zinfandel, hasColor, Red>

<Zinfandel, hasFlavor, Strong>

We obtain: "A zinfandel has red color. It/ This wine has strong flavor."

Sentence aggregation

An important step towards reaching the final text is represented by the text aggregation and the use of Rhetorical Structure Theory. The text aggregation is utilized to obtain concise text and can be applied at the identification of certain patterns [1]. The aggregation is possible due to the order of the resulted triples. We can group sentences which express axioms at the same level of relation compared to the main class. Therefore, this stage can only occur if the triplets are sorted before, according to the same subject, or property. In the case of triplets of the form: <S, P, O_1>, <S, P, O_2> , they are replaced with the triplet <S, P, O_1 and O_2>.

If we have the triplets <S, P_1, O_1> and <S, P_2, O_2> and P_1 and P_2 are equivalent, they are replaced with <S, P_1, $O_1$and O_2>.

Aggregation can also occur in the situation in which we have triplets with different properties and objects, but with the same subject: <S, P_1, O_1>, <S, P_2, O_2>, <S, P_3, O_3>:

<Muscadet, hasBody, Light>

<Muscadet, hasFlavor, Delicate>

<Muscadet, hasSugar, Dry>

We obtain: “Muscadet is light, it has delicate flavor and it is dry.”

In the case we have 2 triplets which refer one property and then a retriction of the same property, the 2 triplets combine in the manner suggested by the following example:

<StEmilion, exactCard(madeFromGrape), 1>

<StEmilion,madeFromGrape, CabernetSauvignon>

Result : “StEmilion is made from exactly 1 grape: Cabernet Sauvignon.”

Rhetorical Structure Theory

The Rhetorical Structure Theory is employed to create a tree hierarchy of sentences or part of sentences linked with rhetorical relations to ensure the coherence of the resulted text [4]. The Rhetorical Structure Theory helps to emphasize the rhetorical relations between clauses or sentences, using “discourse markers”. These relations describe the intention of the clauses on the reader and can be: Condition, Antithesis, Justification, List, Contrast [8, 9]. However, considering the content of an ontology and its informational purpose, only part of these relations can be applied on the ontological content.

The “list” relation can be used to suggest equality in the text hierarchy between different members or axioms. The “elaboration” relation may indicate additional information about a class, whereas the “concession” relation, indicating a contradiction between statements, may be used to link with the description of related classes. Finally, the “condition” relation is applied in the case of restricted properties [6]. Considering a given concept, this will become the root of the tree structure and the main nucleus of the Rhetorical Structure Theory schema. Next, we discover the subclasses and superclasses of this concept, which will be placed in a relation of “elaboration”. Since superclasses and subclasses occupy the same level in the hierarchy, they will be grouped in a rhetorical relation of “list”, highlighted by the markers: “,”, “and”. Then, we describe the properties of that class and in the case of restriction on properties, we place them in a relation of condition, using the demonstrative pronoun “that” or the pronoun “which”.

The elements of the “allValuesFrom” restriction on a property are in a relation of “elaboration” to the main sentence, using the marker “:”. Lastly, to add information about a linked class, the following markers are used: “Additionally”, “Furthermore”, placing these statements in a relation of “elaboration” with the main nucleus [6].

SUGGESTIONS FOR FUTURE WORK

The main drawback of this project is represented by the static adding of domain dependent information for each ontology we want to obtain natural language text. This operation depends on the number of properties in the ontology and has to be updated manually if the ontology is modified or improved with new knowledge. A future improvement could be the integration of a global database called lexicon to obtain different forms of verbs, nouns [5]. Additionally, each template for a property could be saved in a new ontology which can be used in similar projects in the same field.

Because there is no precise technique to evaluate the output obtained, we should show samples of outputs to different users to rate the resulted texts, according to how useful and easy to follow they consider the text.

Additionally, we can produce better outputs if we would consider trials with different methods involved, and based on the feedback received from users we would choose the proper combination of techniques.

CONCLUSIONS

We have provided a description for a natural language generation application which aims at producing fluent, coherent text describing a class or individual from a selected OWL ontology. In order to achieve accessible texts our application uses different techniques such as: Rhetorical Structure Theory, domain dependent information, or sentence aggregation.

The field of Natural Language Generation is a relatively new research topic, but of great perspective in the context in which the Semantic Web is in a continuous expansion, and the user could benefit from new-created or modified ontologies to receive knowledge in a easy to assimilate manner.

REFERENCES

1. Kalina Bontcheva. 2004. Automatic Report Generation from Ontologies: the MIAKT approach. *9th International Conference on Applications of Natural Language to Information Systems.*
2. Nadjet Bouayad-Agha. 2012. From ontology to NL: generation of multilingual user-oriented environmental reports. *Proceedings of the 17th international conference of Applications of Natural Language Processing and Information Systems.*

3. Roger Evans, Paul Piwek and Lynne Cahill. 2002. What is NLG?. *Proceedings of Second International Natural Language Generation.*
4. Eva Forsbom. 2005. *Rhetorical Structure Theory in Natural Language Generation.* GSLT:Natural Language Generation Lecture.
5. Gerasimos Lampouras, Dimitrios Galanis and Ion Androutsopolous. 2013. Generating Natural Language Descriptions from OWL Ontologies. *Journal of Artificial Intelligence Research*.
6. Shao Fen Liang, Donia Scott, Robert Stevens and Alan Rector. 2011. *Unlocking Medical Ontologies for Non-Ontology Experts*. Proceedings of BioNLP Workshop.
7. William C. Mannn and Sandra Thompson. 1987. *Rhetorical Structure Theory: Toward a functional theory of text organization*. ISI: Information Sciences Institute
8. Daniel Marcu. 2000. *The theory and practice of discourse parsing and summarization*. MIT Press.
9. Nick Nicholas. 1994. *Problems in the Application of Rhetorical Structure Theory to Text Generation.* M.Eng.Sc.Degree *Thesis, University of Melbourne*

A Design Framework for Foreign Language Learning Applications

Mihaela Colhon

University of Craiova
Department of Computer Science

mcolhon@inf.ucv.ro

ABSTRACT

In this article we present a method for generating and applying syntactic motivated patterns in order to develop a foreign language learning mechanism. The patterns have been extracted from a parallel corpus that has been automatically annotated for morpho-syntactic descriptions and syntactic constituents. The proposed language learning framework is not designed around the well-known list of words. Using this application, the user does not necessarily practice the foreign language lexicon, which is supposed to be known at a medium level. Instead, with this application, the user is guided to learn the so-called "translation knowledge".

Auhor Keywords

Machine Translation, Learning Schema, Human-Computer Interaction.

ACM Classification Keywords

I.2.1: Natural language interfaces; H.5.2: User-centered design.

INTRODUCTION

The task of natural language translation from one language to another is attributed to human intelligence. Besides the knowledge of the two languages, it requires an "understanding" of the source language text and to transform the mental picture created through understanding of the source text into its target language representation.

The design of the systems developed for foreign language instruction needs to be grounded on what we know about human learning, language processing and human-computer interaction. Nevertheless, machine translation provides a good starting point for foreign language learning giving a rough understanding of the natural language constructions, generating alternative translation choices and creating proper resources of example translations.

Human-Computer Interaction (HCI) deals with all the relevant aspects concerning the design, the implementation and the evaluation of interactive systems. From the HCI point of view, the interactive applications of these days have to face the following challenges:

- Speech/Text Recognition and Understanding
- Speech/Text Generation
- Interactive Machine Translation

Natural language processing (shortly, NLP) can be done on several levels. The first level implies phonology and phonetics data for speech recognition and understanding or morphology data in case of natural language texts recognition and understanding. The next levels imply syntactic, semantic and pragmatic based processing.

For the applications with natural language interfaces, a number of problems are encountered due to the ambiguity of natural language constructions, to the huge amount of involved lexical knowledge and of natural language utterances.

The main goal of this study is to develop a design framework for a foreign language learning software product. Because learners vary both in how they learn and what they want and need to learn, we can not say which is the best way of learning. When creating an application like this it is important to keep in mind that learning is a highly subjective process, and for this reason the user characteristics and ergonomy must to be carefully treated [9]. Trăușan-Matu [12] considers that a user friendly application must allow easy, effective, safe and without risks usage and must to be useful, easy to learn and easy to remember.

A possible solution consists in developing applications that can follow the user's knowledge and needs [8]. In this paper we give a possible design framework for foreign language learning applications.

The remainder of the paper is organized as follows: in Section 2 are presented the previously research works based on which the proposed foreign language learning mechanism is developed. The proposed framework for foreign language learning is detailed in Section 3 together with a particular case exemplification concerning nominal phrase learning mechanism. The manner in which the learning schema can be modeled upon the user knowledge level is presented in Section 4. Finally, the conclusions and future work guidelines can be read in Section 5.

AUTOMATIC TRANSLATION SYSTEM AS TARGET LANGUAGE LEARNING APPLICATION

Given a source-language (e.g., Romanian) sentence, the problem of machine translation is to automatically produce a target-language (e.g., English) translation. We found that translating lemmas and morpho-syntactic descriptors of the source language sentence words and then generating, the corresponding word-forms in the target language achieves better results than the baseline phrase-based translation model. In view of this, we have

developed a symbolic Machine Translation program [1] with English as Source Language and Romanian as Target Language, the eRoL System[1].

In any automatic translation system, getting syntactic data with the scope of producing linguistic information about the source sentence structure involves a pre-processing step of the source-language sentences also known as *parsing*. The resulted structure of a sentence has to indicate the relationships that exist between the words of that sentence or how the words are grouped into syntactic phrases like noun phrases (NPs), prepositional phrases (PPs), verb phrases (VPs), etc. Usually, all these information are stored in a tree representation.

LINGUISTIC RESOURCES

Parallel corpora can generate extremely valuable linguistic knowledge for machine translation studies such as: in *direct approaches*, parallel corpora are used to extract information about lexical units (how a particular word is translated in a certain construction), in *transfer-based approaches*, parallel corpora are used to extract transfer rules while in *statistical approaches*, these corpora are used to extract translation rules and to assign probabilities to possible translations.

Linguistic resources upon which the translation data are created are based on parallel natural language constructions extracted from multilingual corpora. Such a corpus, called JRC-Acquis, is the compiled part of the parallel texts from the Acquis Communautaire legislative documents. The Acquis Communautaire is a collection of parallel texts in 22 official European Union languages, including English and Romanian.

Two segments of texts from a pair of parallel texts which represent reciprocal translations make a *translation unit* [13]. Phrase based machine translation techniques work with pairs of phrases, the so-called translation units, which are consistent with respect to the inner word transation-alignment: the words of a phrase are aligned to words of the other phrase and not to the outside words.

The current practice in phrase-based translation has shown that creating large syntactic phrase tables allow the learning of "translation knowledge". Indeed, most of the phrases syntactical motivated are expected to be translated without interleaving with other phrases/words. In general, noun phrases tend to obey the above rule to a much greater degree. Conversely, verb phrases usually suffer modifications in structure during translation, caused by adjunct movement [4].

A *parallel treebank* is a special type of parallel corpus that has been grammatically annotated in order to identify and label different syntactic information about the text. Such syntactic information usually implies incorporating into the text markers which indicates the syntactic dependencies relations or the phrase-based structures[2].

Techniques that were applied on the corpus sentences include tokenization, part-of-speech tagging and lemmatization. Part-of-Speech (POS) tagging, describes the annotated words in terms of grammatical tagging (Noun, Verb, Pronoun, etc.) and morphological information (sequences of codes about the inflectional features of the words such as gender, number, person, case, etc.). Often, POS tagging can include lemmatization, by indicating the lemmas of the words. A POS tagset generated during the MULTEXT-East project [5] includes *morphosyntactic descriptions* (*MSD*) for all the languages of the project (including Romanian).

The application described in this paper uses 4389 English-Romanian parallel patterns extracted from a English-Romanian Treebank [2] with syntactic constituents.

The Treebank was constructed upon 1420 sentences of the English-Romanian corpus developed at the Alexandru Ioan Cuza University of Iaşi by the Natural Language Processing Group of the Faculty of Computer Science. For this bilingual corpus, the English and Romanian parts of the JRC-Acquis corpus were used.

In the considered Treebank, each translation unit has representations at several levels:

- at lexical level (sequences of words);
- at POS level (part of speech of the annotated words);
- at phrase-based level (syntactic constituents)

Such a collection was designed to be used in a translation automatic mechanism with the scope of moving from words to phrases as the basic unit of translation [3].

PATTERN-BASED TRANSLATION MODEL

The knowledge enclosed in the eRoL system development is represented by syntactic patterns defined in terms of morpho-syntactic specifications and phrasal tag specifications in the form introduced by the Penn Treebank project[3]. In this manner, the system uses the so-called *informed language model* which is described by word-forms but also by MSD specifications or POS data.

The representation of the parallel English-Romanian parse trees of the Treebank are flattened into linear string form following the bracket representation for syntactic trees with constituents [11]. From the Treebank, the system considers the syntactic level representations of the parallel phrases by focusing mainly on POS tags instead of real words.

In Figure 1 it is given a screen shot of the eRoL Web interface.

[1] There is a web demonstrator of this system available at http://www.mcolhon.ro/erol/eRoLsystem.html.

[2] Traditionally, phrases markers are taken to be syntactic constituents of a sentence.

[3] The web address of the project is http://www.cis.upenn.edu/~treebank/.

Figure 1. The eRoL System Web Interface

In what follows we will exemplify the implemented translation mechanism, by considering the following English sentence:

A brown and beautiful cat plays with a big ball.

The input sentences are parsed with Stanford Parser[4] tool [6] in order to mark the syntactic phrases and the POS data of the sentences' words. For the considered example, the representation obtained with Stanford Parser has the following syntactic tree form:

```
[S [NP DT/a [JJ brown] CC/and [JJ beautiful] [NN cat]] [VP [VBZ plays] [PP IN/with [NP DT/a [JJ big] [NN ball]]]]]
```

The system finds in the English-Romanian Treebank the Romanian phrase corresponding to the last constituent of the given phrase, which is:

```
[PP IN/with [NP DT/a [JJ *] [NN *]]
```

and, after translations word-to-word are made, the generated Romanian phrase is:

```
[PP IN/(împreună)cu [NP o: Tifsr minge:Ncfsr voluminoasă:Afpfsm]
```

where `IN/(împreună) cu` represents the translation of `IN/with` and `[NP o: Tifsr minge:Ncfsr voluminoasă:Afpfsm]` is the translation of the English noun phrase `[NP DT/a [JJ big] [NN ball]]`.

FOREIGN LANGUAGE LEARNING FRAMEWORK

Any learning process is schema-based. The recent interests in language learning schema match up closely with an ongoing theme from cognitive psychology that bases all long-term learning on the construction of schemas. This view sees the learning of new material as involving integration into old material [7].

Using the linguistic resources constructed for the eRoL system, an application for English language learning can be developed for Romanian users. The presented learning mechanism can be applied to any other pair of languages by replacing Romanian with the name of the user native language and English with the foreign language name.

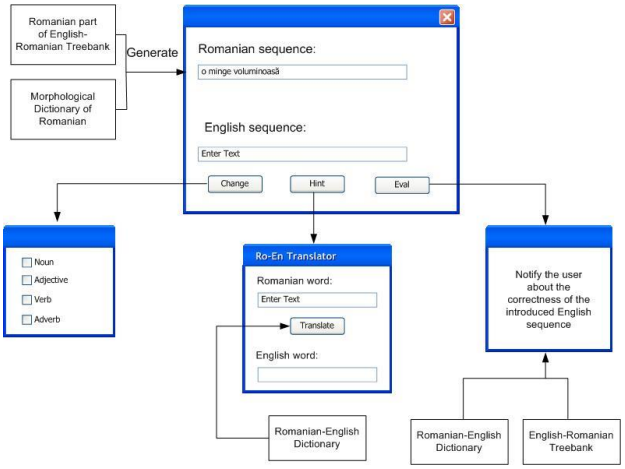

Figure 2. The Design Framework of the Application

In Figure 2 is shown the design of the proposed learning application. Basically, it consists of a simple main window where the Romanian sequences are displayed with the scope of receiving the user translations. The main window has three controls designed for the following tasks:

- the "Change" button let the user to decide which kind of patterns wants to practice: Noun Phrases, Adjective Phrases, Verb Phrases or Adverb Phrases;
- the "Hint" button uses a Romanian-English dictionary resource in order to help the user with the Romanian to English word based translations;
- the "Eval" button starts the evaluation of the English sequence introduced by the user and highlight the user's errors (if it is the case). If the evaluation ends successfully, another Romanian sequence is generated in order to be translated.

In what follows we give the algorithm that guides the learning process by means of the patterns extracted from the English-Romanian Treebank.

Foreign Language Learning Algorithm

```
1. listPOS_CW <-{Noun, Adjective, Verb, Adverb}
2. CW <- listPOS_CW.pop()
3. while (!user.wantsToStop())
3.1. extract all patterns including CW
3.2. do
3.2.1. generate sequences of words based on extracted patterns
3.3. until (user.wantsToStop() ||
       user.wantsToChange() ||
       user's translations are correct)
3.4 if (user's translations are correct)
3.4.1 CW <- listPOS_CW.pop()
3.5 endif
```

[4] Because the English sentences are processed using Stanford Parser (web page: http://nlp.stanford.edu/software/lex-parser.shtml), PENN Treebank parse trees are generated. As a direct consequence, the English words are annotated with PENN POS tags, as this is the tagging standard used by Stanford Parser.

```
3.6. if (user.wantsToChange())
3.6.1. CW <- user's CW choice
3.7 endif
4. endwhile
```

This learning schema, we consider, will enable users to practice the foreign language learning in a gradual manner, starting with easy sequences formed by few words and continuing with larger sequences. Also, the order in which the patterns are generated ensures a proper language learning method as the application starts by generating noun phrases and ends with the verb phrases (more complex than the nominal ones). This order is ensured by the `listPOS_CW` stack. Obviously, the user can change the category of natural language constructions he wants to practice (by selecting the "Change" button from the main window).

Exemplification. Noun Phrases Learning Mechanism

In order to exemplify the manner in which the learning process is ensured in the proposed framework, in this section we will present the NP learning design.

The simplest patterns for Noun Phrases consist of a single noun. In a first phase the application will generate sequences for `N*` patterns like `Ncmsoy`. Here are several entries of the Morphological Dictionary of Romanian [10] corresponding to this MSD sequence: "fratelui" (in En. "of the brother"/ "to the brother"), împăratului (in En. "of the king"/ "to the king"), "miliardarului" (in En. "of the billionaire"/ "to the billionaire").

The user is passed to the next level of learning process if its translations for the current level are correct. A next level for NP learning is generated by using patterns which include noun tokens but also extra function words tokens like in the following examples:

- "această chinezoaică" (in En. "this Chinese woman") or "această antilopă" (in En. "this antelope") corresponding to the pattern `Dd3fsr/această Ncfsrn,`
- "unui cortegiu" (in En. "of a procession"/ "to the procession") or "unui idol" (in En. "of an idol"/ "to the idol") corresponding to the pattern `Timso/unui Ncms,`
- "în conformitate cu materia" (in En. "in accordance with the matter") or "în conformitate cu decizia" (in En. "in accordance with the decision") corresponding to the pattern `Spca/în_conformitate_cu Ncfsry,`
- "astronava acestuia" (in En. "its spaceship") or "căruța acestuia" (in En. "its carriage") corresponding to the pattern `Ncfsry Pd3mso/acestuia.`

If the user matches the translations for the received sequences, more complex patterns will be generated. The larger the sequences are, the fewer the possible translations will be. This is determined by the fact that a sequence complete from the meaning point of view, usually has a unique translation.

ADJUST THE LEARNING SCHEMA UPON THE USER KNOWLEDGE LEVEL

The manner in which the application is designed permits automatic adjustments upon the user's level of expertise. Indeed, the level of difficulty for the generated sequences that must be translated can be automatically adjusted by restricting the data involved in the generation phase. More precisely, the application could use:

- only a part of the Romanian Morphological Dictionary which, in this case, means fewer words that will be used in the generated Romanian sequences
- patterns limited to medium size for beginners or maximum size for advanced users

Using these simple restrictions users with less expertise could be trained using simple sequences made from several words (that will be repeated if the user fails with his translations) while the advanced users will be trained using the whole lexicon – meaning, a large vocabulary that has to be passed and using complex sequences made upon the largest patterns from the Treebank.

CONCLUSIONS AND FUTURE WORK

In this paper, we present a design framework for foreign language learning applications. The proposed learning mechanism ensures the development of a proper learning schema using the existing linguistic resources for the involved languages, in our case Romanian and English. Our proposal addresses the usability issues concerning this kind of applications by proposing a progressive learning schema. Our permanent concern is to improve the set of parallel patterns and also the entries of the used English-Romanian dictionary in order to cover more utterances in the considered languages.

ACKNOWLEDGMENTS

The author wants to thank the Natural Language Processing Group of the Faculty of Computer Science, Alexandru Ioan Cuza University of Iaşi, for providing the English-Romanian corpus upon which the presented study was made.

REFERENCES

1. Colhon, M. 2013. *eRoL: Automatic Voice Translator for Romanian. Building Resources for a Symbolic Machine Translation Program*, Universitaria Publishing House, Craiova 2013, 151 pages, ISBN: 978-606-14-0648-7.
2. Colhon, M. 2012. Language Engineering for Syntactic Knowledge Transfer, In *Computer Science and Information Systems Journal*, 9(3), ISSN 1820-0214, 1231-1247.
3. Colhon, M. 2012. Acquiring Syntactic Translation rules from a Parallel Treebank, *Journal of Information and Library Science INFOtheca*, 2, XIII (Dec. 2012), 19-32, Serbian Academic Library Association, ISSN: 1450-9687 (e-magazine)
4. Colhon, M. 2011. A Contrastive Study of Syntactic Constituents in English and Romanian Texts, In

Proceedings of the Workshop "Language Resources and Tools with Industrial Applications", Eds. Iftene A., Trandabăţ D.-M., 11-20
5. Erjavec, T., Krstev, C., Petkevič, V., Simov, K., Tadić, M. and Vitas, D. 2003. The MULTEXT-East Morphosyntactic Specifications for Slavic Languages, In *Proceedings of the 2003 EACL Workshop on Morphological Processing of Slavic Languages*, 25-32, Budapest, Hungary.
6. Klein, D., Manning, C. D. 2003. Accurate Unlexicalized Parsing", In *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, 423-430
7. MacWhinney, B. 1995. *Evaluating foreign language tutoring systems*. In: Holland, Kaplan and Sams, 317-326.
8. Pribeanu, C. 2001. *Human Computer Interfaces Design*, MATRIX ROM Bucureşti Publishing House (in Romanian).
9. Pribeanu, C. 1999. *Human Computer Interaction*, Didactic and Pedagogical Publishing House, Bucureşti (in Romanian).
10. Simionescu, R., Cristea, D., and Haja, G. 2012. Inferring diachronic morphology using the Romanian Thesaurus Dictionary. In *A. Moruz, et al. (eds.). Proceedings of the 8th International Conference "Linguistic Resources And Tools For Processing Of The Romanian Language"*, "Al.I.Cuza" University Publishing House, Iaşi, 85-92, ISSN 1843-911X.
11. Taylor, A. 1996. Bracketing Switchboard: An Addendum to the Treebank II Guidelines, http://www.seas.upenn.edu/~jmott/prsguid2.pdf
12. Trăuşan-Matu, Ş. 2000. *Evolved Human-Computer Interfaces*, MATRIX ROM Bucureşti Publishing House (in Romanian)
13. Tufiş, D. and Ion, R. 2007. Parallel Corpora, Alignment Technologies and Further Prospects in Multilingual Resources and Technology Infrastructure. In *Proceedings of SPED 2007*.

APPENDIX A. GLOSSARY OF NOTATION

The following table gives the notation used in this paper:

MSD tag	The meaning of the notation (according to MULTEXT-East lexical specifications)
Afpfp(s)	Adjective qualifier positive feminine plural(singular)
Dd3fpr-	Determiner demonstrative third feminine plural direct
Tif(m)sr(o)	Article indefinite feminine(masculine) singular direct(oblique)
Ncf(m)sr(o)y(n)	Noun common feminine(masculine) singular direct(oblique) +definiteness(-definiteness)
Pp(d)3mso	Pronoun personal(demonstrative) third masculine singular oblique
Spsa	Adposition preposition simple accusative

Opinion Summarization for Hotel Reviews

Bogdan Cristian Marchis
University Politehnica of Bucharest
313 Splaiul Independentei, Romania
bogdan.marchis@gmail.com

Alexandru Tifrea
University Politehnica of Bucharest
313 Splaiul Independentei, Romania
alex.tifrea93@gmail.com

Mihai-Cristian Volmer
University Politehnica of Bucharest
313 Splaiul Independentei, Romania
mihaivolmer@gmail.com

Traian Rebedea
University Politehnica of Bucharest
313 Splaiul Independentei, Romania
traian.rebedea@cs.pub.ro

ABSTRACT
This paper presents a new approach for finding the best n-grams that efficiently summarize a large set of reviews. The proposed unsupervised method uses a readability score and a representativeness score to select those n-grams that best convey the main opinions contained in the processed reviews. In order to further refine the selected n-grams, we use sentiment analysis and part of speech (POS) tagging to impose certain requirements that the n-grams that we are looking for should meet. Furthermore, the best n-grams were classified into several topics, which allowed a better prevention of redundancy among the summarizing n-grams. Therefore we offer an unsupervised, mostly non-aspect based, unstructured opinion summarization algorithm that can be easily implemented for any web platform that accepts reviews, due to its genericity. In order to assess the results of our algorithm, we summarized hotel reviews extracted for the TripAdvisor[1] website. The algorithm produces readable results that convey relevant opinions about the hotels that we used for testing.

Author Keywords
Opinion summarization, opinion mining, natural language processing, n-grams, micropinions.

ACM Classification Keywords
H.5.2. User Interfaces: Natural language, I.2.7 Natural Language Processing.

INTRODUCTION

Almost every online platform that offers services to customers has a reviews section nowadays. This is one of the building blocks on which e-Commerce relies, but it also raises a few problems. The main issue is that popular services like Amazon, eBay or TripAdvisor can end up having hundreds or thousands of reviews for a single product. Of course, this makes it very hard for users to go through all the reviews and extract the relevant opinions from all of them. Another problem, which makes it all the more difficult for users to form an opinion about a product, is the fact that quite a few of the reviews do not talk about a concrete aspect of the given product, but rather use very general phrases to describe it, such as '*The hotel was great*' or '*I didn't like it at all*'.

Generating an opinion summary starting from a very large text is something that has been extensively researched in the last few years [1, 7, 8, 9, 11, 12] as the number of opinions on the web is always increasing, thus making it harder and harder for users to take into account every opinion when deciding upon a product, a service or, in our case, a hotel. Opinion summarization can vary from simply giving an overall rating [14] based on all the ratings for a single hotel, or generating a rating based on what each user had to say about that hotel, to generating a new review that will be a summary of every other review [2, 4, 6, 17]. Generating an opinion summary can be compared to normal text summarization - finding a small set of keywords or key phrases in the text that best describe the overall text [10] - with the difference that not all keywords or key phrases can be considered an opinion. This makes opinion summarization more difficult than generic text summarization.

In this project we seek to generate a few sentences that best represent a summary for hundreds or thousands of reviews. In order to do this, we thought that the best approach was not to use the sentences that were already in the reviews, but to create new sentences, based on the text, and then check how representative each sentence is for the entire set of reviews. We chose this approach because it is often the case that a sentence covers more than just one topic and contains more than one opinion. By creating new sentences instead of extracting sentences from the reviews, we no longer have to make sure the sentence that we selected from the original text is appropriate (i.e. the sentence that exactly matches the opinion that the algorithm decides to output). Generating a new sentence is a demanding task, not only because the sentence has to be grammatically correct, but also because it has to convey an opinion that is relevant to the large set of reviews. Moreover, with such an approach, that only uses a representativeness score to filter opinions, it is very hard to keep among the selected n-grams those which contain negative opinions, since these are, based on our observations, far less common than the positive ones.

[1] www.tripadvisor.com

We take this aspect into account and offer certain bonuses in order to keep a good balance between positive and negative reviews, regardless of what the general opinion about the given hotel is. For instance, a generally well-seen hotel, can still have flaws. This is why we want to make sure the reviews mentioning these flaws stand a chance in the face of the more numerous positive reviews.

Our approach is mostly non-aspect based, meaning that we don't look for opinions related to particular aspects of the product/hotel. All the resulting n-grams that summarize the reviews are built and selected only based on their readability and representativeness scores. We also give bonuses to n-grams that are formed around certain syntactical structures as will be explained in the following sections. It is only after all the n-grams are computed and sorted by their scores that we use some predefined topics to extract only those summaries that are relevant to our product. This is the aspect based part of the algorithm.

RELATED WORK

State-of-the-art algorithms for review summarization are usually aspect based, looking for opinions related to certain features of the product that is reviewed [8]. This approach is highly dependent on the way the features are chosen. Syntactic tree parsing, POS (part of speech) tagging or a supervised approach are just some of the methods that are very common when trying to extract the features that need to be analyzed.

In generic text summarization, most attempts use extractive summarization which consists of extracting relevant and representative fragments from the given text. Some important contributions in this direction are [7, 11, 12]. Abstractive summarization is considerably harder than the extractive approach, since it raises the problem of generating readable phrases that can also convey a relevant opinion, which is also representative for opinion summarization.

Our work builds on the results obtained by [1]. They propose an abstractive unsupervised algorithm that extracts micropinions and generates short sentences to present them to the users. Unlike their work, ours is a hybrid approach, using a non-aspect based algorithm to generate the candidate sentences and then using some predefined topics (i.e. features) to select only sentences relevant to the product which is being reviewed. In addition to that, we use a slightly different algorithm for generating the candidates, which allowed us to use some tools more efficiently (i.e. Microsoft Ngram Service, CoreNLP etc.) by traversing the solution space in a breadth-first fashion. Moreover, our version of the algorithm lends itself to working with an adaptive pruning mechanism, which allows us to select at each step only valuable n-grams. Thus we can improve the efficiency of the algorithm up to the level that we desire by heuristically adjusting the thresholds used for the readability and representativeness scores. Of course, this has a direct impact on the quality of the results, but can prove useful when running an implementation for a system with limited resources, like a mobile device.

Another work that is similar to ours is [9]. They offer both an abstractive and an extractive algorithm. However, in their abstractive approach, they generate whole paragraphs, using complex natural language processing tools in order to achieve this. Rather than doing this, we aim to create small, concise sentences, not longer than 10-15 words.

OPINION SUMMARIZATION SYSTEM

We wanted to create an algorithm that would find a few sentences that would best summarize the opinions of a few hundred people. In our opinion, the best solution was to find short sentences, between three and eight words, that would summarize the set of reviews as good as possible. The main reasons for selecting this approach are the following. Firstly, it is fairly easy to construct short sentences based on the initial text and it also means that it will be easy to check the sentences for their readability and their representativeness, which we are going to discuss a bit later. Secondly, having a short sentence also means that it is going to be easier to compare it with another sentence, which helps us reduce the run-time considerably.

In order to generate the sentences that best summarize the opinions of hundreds of users, we decided that the best solution would be to use a bottom-up approach and create new sentences based on the words from our initial text (which is made by concatenating all the reviews of a given hotel). Although we are using words from the initial text, the words in the newly created sentences do not have to be in the same order as they were in the original sentence, making sure that we have a better chance of creating a sentence that summarizes the opinion of more than one user. If we consider a sentence to be represented as s_i, our final result should be a set of sentences, $R=\{s_i\}_{i=1}^{k}$ and each word from s_i should be a word from our original text $T=\{s_i\}_{i=1}^{n}$, which is also a set of sentences.

Since we earlier stated that the sentences we create might be different from every sentence in the original text, we must create some functions that will assure us that the result is readable and is relevant to our initial set of reviews. The first functions that we need to take into consideration are a readability score [5], which tells us how readable a sentence or a small group of words is, and a representativeness score, which tells us if the sentence is relevant for our initial set of reviews. After creating these two functions, our job should only be to determine which sentence has the best score of representativeness and of readability:

$$R = max_{R=\{s_1,s_2,\ldots,s_k\}} \sum_{i=1}^{k} (S_{rep}(s_i) + S_{read}(s_i))$$

where:

1. S_{rep} is a function for scoring the representativeness of a sentence
2. S_{read} is a function for scoring the readability of a sentence

Once we have a set of sentences, R, we have to take into account that two or more of those sentences might have the same meaning. To address this problem, as we mentioned in the paragraph above, we have to define a similarity function, S_{sim} which will give us the similarity score between two sentences. If the result returned by the similarity function, $S_{sim}(s_i, s_j)$, is above a similarity threshold, which was determined heuristically, we will consider the two sentences, s_i and s_j, to be similar. If we find that two sentences are similar, we have to eliminate the one that yields the lowest result for the formula: $S_{read} + S_{rep}$.

Once we have these functions, we can determine the best sentences by using a threshold for each function and a threshold for the sum of the representativeness function and the readability function:

$$R = \{s_i\} \text{ for } i = 1 : k \text{ where } S_{rep}(s_i) \geq \sigma_{rep}, S_{read}(s_i) \geq \sigma_{read} \text{ and } S_{sim}(s_i, s_j) \leq \sigma_{sim} \forall s_i, s_j \in [1, k]$$

where:

- σ_{read} is the minimum readability accepted for a sentence
- σ_{rep} is the minimum representativeness accepted for a sentence
- σ_{sim} is the maximum similarity between two sentences

Readability

The purpose of the readability function is to tell us if a sentence is grammatically correct and to make sure that the sentence makes sense for a user. This function is especially important in our case, since we are creating new sentences by combining every possible word from our initial text. At this point, we have to realize that when we put together two words from a text, most of the results will make no sense at all. Since most of the results will make no sense, we can drastically reduce the number of word tuples by setting a high enough σ_{rep} value.

In our implementation of the readability function, we chose to use the Microsoft Web N-gram Service. This cloud-based platform provides a joint probability score for a given sentence. The cloud platform provides a readability score for a sentence and uses as training set all documents indexed by Bing in the en-us market.

To access this service we used the Python module provided on the Microsoft N-gram blog from MSDN. On top of the module we implemented a server-like interface. Thus, the Python module runs in parallel with the main process. Queries are sent from the main process to the Python module, which resolves them with the Microsoft N-gram server.

Representativeness

The representativeness function determines whether a newly created sentence is found in the original reviews, how many times it is found and in which combination of words. We will not take into account how readable a sentence is when computing its representativeness score, because for this part we only care if a sentence appears in a form or another in the original text.

The first thing considered when computing the representativeness score is whether or not the words that form the new sentence are actually closely related in the original text. In order to make sure the words are related in the original text, we first check whether the words are in the same sentence, and if so, how many times they appear in the same sentence and whether the distance between the two words, inside the sentence, is not greater than a window size, C. The first condition makes sure that two words that have a high appearance rate in the initial text, but almost never appear together, will not be able to form a new sentence. The second condition makes sure that strongly connected words have a better chance of forming a new sentence, then words that just happen to appear in the same sentences a lot.

We will define S_{rep} as:

$$S_{rep}(s_i) = S_{rep}(w_1 w_2 ... w_n) = \frac{1}{n} \sum_{j=1}^{n} pmi_{local}(w_j)$$

Where pmi_{local} is defined as the local pointwise mutual information (PMI) function:

$$pmi_{local} = \frac{1}{2C} \sum_{k=j-C}^{j+C} \log_2 \frac{p(w_j, w_k) \cdot c(w_j, w_k)}{p(w_j) \cdot p(w_k)}$$

Where:

- C is the size of the window within which we look for closely related words
- $p(w_j, w_k)$ is the frequency of two words co-occurring in the same sentence
- $c(w_j, w_k)$ is the frequency of two words co-occurring in the same sentence and in the same window
- $p(w_j)$ is the frequency of a word in the text

Let us consider the following example: if we have a text made of twenty sentences and two words, w_j and w_k, which appear in the text ten times each. The words also appear five times in the same sentence and twice in the same window, C. The values for the relations above, will be $p(w_j, w_k) = 5/20$, $c(w_j, w_k) = 2/20$ and $p(w_j) = p(w_k) = 10/20$, which means that we get the final value: $pmi_{local} = 1/(2C) * \log_2(0.1)$.

In order to speed up the implementation of the system, a hash table has been used, with the word as a key, and a vector as a value, in which we can find the exact position of every word. Using this structure, we know exactly in which sentence we can find a word, so we only need to iterate through that sentence when computing the pmi. This structure gives us a worst case time complexity of O(n*m*l) where n is the total number of unique words, m is the number of sentences in the initial text and l is the maximum size of a sentence, in words.

The representativeness value is computed for each n-gram before the readability score because the readability score is much more time consuming, since the score is not computed locally, but rather on a remote web server. By computing the representativeness first, more than three quarters of the original n-grams were eliminated, since most of the n-grams are not strongly related.

Similarity

Looking for the best sentences to represent the initial text we noticed that it might be the case that most of the sentences have the same meaning. This might happen if every user says something like '*The hotel was very clean*', but in many different ways. In order to prevent this from happening, we need a good similarity function. We use the Jaccard index to compare two n-grams:

$$J = \frac{|w_i \cap w_j|}{|w_i \cup w_j|}$$

Of course, this only has to do with the sets of words of the n-grams, so no semantic analysis is performed (this may be seen as possible further work). At all times, the queue containing n-grams waiting to be processed only contains n-grams that are not similar to one another.

It is rather important what course of action is taken when a new n-gram that is about to be added to the queue does not pass the similarity test. There are two different scenarios that are considered:

i) the new n-gram is similar to just one n-gram in the queue; in this case, we simply compare the readability and representativeness scores of the two n-grams and only keep the one with the highest scores;

e.g.: the queue is [..., *bathrooms are ok*, …] and we try to add *bathrooms are nice*

ii) the new n-gram is similar to more than just one n-gram in the queue; this means that some part of the new n-gram is similar to one existing n-gram, and some other parts are similar to other n-grams (because otherwise it would mean that the same fragment of the new n-gram is similar to all the existing n-grams with which it doesn't pass the similarity test which leads to a contradiction since we stated before that all none of the n-grams in the queue are similar to one another). When this happens it usually means that the new n-gram conveys information about more than just one topic (since it overlaps with two existing standalone n-grams). In this case, we drop the new n-gram.

e.g.: we try to add *bathrooms are nice and staff is helpful to the queue*: [..., *bathrooms are nice*, *staff is helpful*, ...]

General algorithm

As stated earlier, the easiest way to generate sentences that would best summarize the reviews is to generate the sentences using words from the initial text. If we generate new n-grams based on the words that are in the original reviews, we have a better chance of generating sentences that better summarize the opinion, than we would have if we were to start from scratch.

The first step of the algorithm is to look for the most frequent unique words in the text. The final n-grams generated by the system will only have words that have a frequency higher than a threshold, σ_{freq}. This threshold was determined heuristically and is usually dependent on the number of reviews that are in the dataset. Once we have the most frequent words, we use backtracking to generate every possible bi-gram. However, we employ the pruning algorithm described in the previous section, which eliminates every bigram that has no chance of existing in the original text. We also have to mention that we check every possible permutation of a bigram, hence, we will compute the representativeness score only once, but we will determine which permutation is the best by using the readability score. There is no point in checking the similarity at this point since we only have bigrams and we will only use them as a seed to generate new n-grams. In order to make the algorithm faster, these bigrams are stored in a hash table, for which the key is the first word of the bigram, so that it will be easy to add the bigram to an existing n-gram.

The values for the thresholds that we used in order to select only those n-grams that meet certain requirements (e.g. high readability score, high representativeness score, limit redundancy using the similarity score etc.) were chosen heuristically, by considering both the execution time of the algorithm and the quality of the resulting summaries. The window size C used in the representativeness formula was also chosen by testing different values for it, between 5 and 15.

In the process of generating (n+1)-grams we use the n-grams that we already have at this stage and try to concatenate them with every possible bigram that we have not eliminated after the first stage of the algorithm. A new (n+1)-gram can be formed only if the last word of the n-gram matches the first word of the bigram. We use Breadth-First Search (BFS) to generate the (n+1)-grams since the requests for the Microsoft Web N-gram Service are the most time consuming parts of the algorithm. By using BFS, we can generate every possible (n+1)-gram, eliminate the

worst ones using the representativeness score and only afterwards interrogate the cloud-based platform.

Generate (n+1)-grams
1: Input: queue in which we have every n-gram
2: Output: new_queue in which we will have the (n+1)-grams
3: FOR every n-gram in queue
4: FOR every bigram matching n-gram
5: new_ngram ← JoinNgrams(n-gram, bigram)
6: if new_ngram.representativeness > σ_{rep}
7: new_queue.push(new_ngram)
8: ENDIF
9: ENDFOR
10: ENDFOR
11: GetReadability(new_queue)
12: FOR every new_ngram in new_queue
13: IF new_ngram.readability > σ_{rep}
14: FOR every ngram in all_ngrams
15: IF GetSimilarity(new_ngram, ngram) > σ_{sim}
16: eliminate the worst ngram
17: ENDIF
18: IF new_ngram is not similar with any other existing ngram
19: all_ngrams.push(new_ngram)
20: ENDIF
21: ENDFOR
22: ELSE
23: new_queue.delete(new_ngram)
24: ENDIF
25: ENDFOR

If the similarity function finds that two n-grams are similar, we are going to keep only the best n-gram, which will be the one with the highest sum between the readability score and the representativeness score. We could check the similarity only at the end of the algorithm but we found that the number of similar n-grams generated this way was quite large, so a lot of time is wasted to generate and assess new n-grams that were just (partial) copies of other n-grams. Once we have the (n+1)-grams in *new_queue*, the algorithm proceeds recursively to generate new n-grams until we reach the maximum size of an n-gram, eight, or until none of the newly created n-grams pass the pruning stage.

Sentiment Analysis

The best way to summarize the opinion of a user is to capture his or her sentiment [13, 15] towards a hotel. Since the purpose of the algorithm is to summarize how good or how bad a hotel is, the stronger the sentiment expressed by a user is, the better. The tool that we found fit to analyze the sentiment of a sentence was CoreNLP [3]. CoreNLP provides a rating for a given sentence, using the following classes: very negative, negative, neutral, positive and very positive. Based on the rating that is returned by CoreNLP, a bonus is added to each n-gram. The bonus added for an n-gram that has a very positive or a very negative rating is higher that the bonus added for an n-gram rated as positive or negative.

After early testing, we found that most of the sentences that we were creating had small grammar problems. The sentences were readable, but the words inside the sentence were not quite in the correct order. For example we would get '*nice the staff was*' instead of '*the staff was nice*'. The first thing we did was to find the readability of every permutation possible for the initial n-gram and choose the one with the best readability. The second solution was to rearrange the words inside the sentence until the words are in the write order. To change the order of the words inside a sentence, we only need the part of speech for each word, which we could find by using CoreNLP. Some of the rules that we were using to rearrange the words were: adjective before noun and adverb before verb.

We realized that we could use the fact that we already have the part of speech of each word to improve the generation of new sentences/n-grams. There is no way that we could generate a sentence that would express a strong sentiment without it having at least a noun and an adjective. We thought that we could speed up the generating phase of the algorithm by adding a bonus if an n-gram had a noun or an adjective. We also figured it would be a good idea to penalize a sentence if it contains an interjection, a determiner, a preposition or a number.

Topic detection

After generating every possible sentence and choosing the best ones using the algorithm described above, we have to choose the sentences that are most relevant for each hotel. The best way to determine if a sentence is relevant for a given domain is to find the topic of a sentence and match it with some predefined topics relevant for that domain. In this case, the selected topics were: kitchen, room, staff, noise, location and price.

The tool we used is Word2Vec[2], which is an efficient implementation of the continuous bag-of-words and skip-gram architectures for computing vector representations of words. This tool provides an algorithm that, for a given word, creates a list of words having the highest cosine distance related to it. Based on that algorithm, we were able to obtain the cosine distance of two given words.

We created lists of words for each predefined topic. To obtain the topic of an n-gram, we calculate its cosine distance towards each topic. The topic with the highest cosine distance towards the n-gram is the matching one. However, if that distance is not above a minimum threshold, the n-gram is considered to be out of topic and can be dropped from the final output. We established the value for the threshold heuristically, by evaluating the

[2] https://code.google.com/p/word2vec/

performance of the algorithm for several different values, the same way we did with the other thresholds mentioned throughout this paper.

RESULTS

For testing purposes, various datasets for reviews were used. These were collected from the TripAdvisor website and there was a focus on hotels with at least several hundred different reviews. The average number of requests sent to the Microsoft N-gram Web Service is 195 per review. We make an average of 55000 requests per hotel. Obviously, the server requests are one of the most time consuming part of the application. Hence, we limited the number of bigrams at 5000. The number of n-grams with a size greater than 2 is limited at 1500. This limit was never reached, since the number of (n+1)-grams is usually considerably lower that the number of n-grams. This decrease is caused by the pruning of the n-grams which takes place during the BFS traversing.

As stated before, each n-gram is given a score based on readability, representativeness and POS bonus. As expected, during the course of each run, the mean readability value decreases and the mean representativeness value rises as the n-gram size increases.

The readability and representativeness values are constant throughout every output and are not affected by the number of reviews, as seen in Figure 1 and Figure 2. The spike is caused by a hotel that has poorly written reviews. As a consequence, the generated n-grams have lower representativeness score. This forces the algorithm to finish in an early stage.

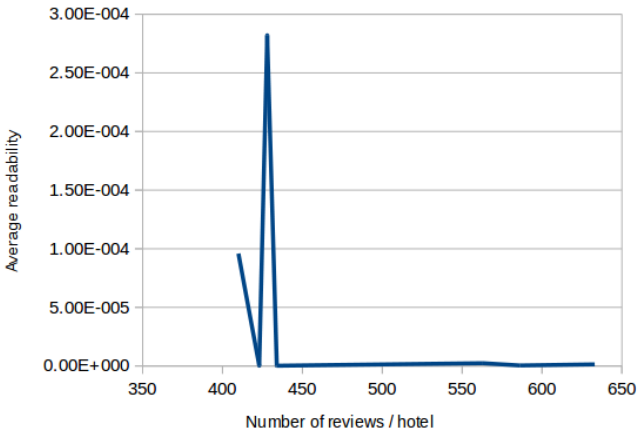

Figure 1: Average readability of the summaries from multiple hotels

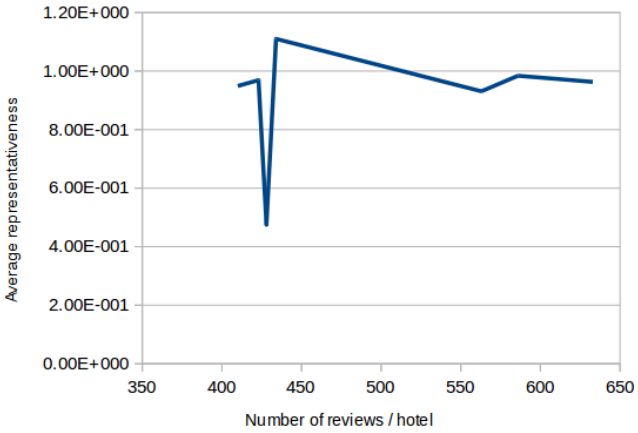

Figure 2: Average representativeness of the summaries from multiple hotels

In what follows, there are some output examples from two different hotels. The application generates for each hotel, on average, 20 summaries. The output is formatted as follows: [<n-gram>, <readability>, <representativeness>, <POS bonus>, <sentiment polarity>] <topic>

Output for Hotel Christina:

[romanian red glass wine and welcome cheese, 1.44092e-09, 1.0253, 0.11 NEUTRAL] food
[bucharest at the hotel i had, 9.28771e-08, 1.09938, -0.092 NEUTRAL] none
[metro bus 5 minutes walk, 4.1583e-07, 0.830261, -0.101 NEUTRAL] location
[accommodating staff very, 8.7571e-10, 0.763112, -0.091 NEUTRAL] staff
[smooth cab she said, 6.06846e-07, 0.528349, 0.11 NEUTRAL] location
[food was excellent, 0.000520357, 0.52147, 0.11 POSITIVE] food
[helpful staff, 7.7023e-06, 0.440511, 0.11 NEUTRAL] staff
[street noisy noise, 4.23309e-05, 0.419229, 0.11 NEUTRAL] noise

Output for hotel Palm Opera:

[will definitely book again soon next, 7.45793e-09, 1.23048, 0.109 POSITIVE] none
[welcoming attentive staff and professional at rest, 9.25606e-09, 1.17844, 0.109 POSITIVE] staff
[location we were good, 2.90115e-07, 1.07646, 0.109 NEUTRAL] location
[available water supply area a lounge touch, 2.2848e-09, 1.07143, 0.109 NEUTRAL] room
[not disappointed me fortunately my wife, 8.25395e-08, 1.03246, 0.11 NEGATIVE] none
[free lounge drinks they offer, 5.15874e-08, 1.02223, 0.108 NEUTRAL] food
[floor bath was lovely had got, 8.14061e-10, 1.0202, 0.109 POSITIVE] room
[children just loved ikea play kitchen, 8.23065e-10, 1.21441, -0.09 POSITIVE] food

In order to assess the quality of the resulting n-grams (e.g. the n-grams that are shown above) we created a survey where summaries could be rated on scale from 1 to 5 (5 being the highest). The survey consisted of 187 summaries from 7 hotels and was taken by six fellow students coming from different backgrounds, namely not just computer science. We have instructed them to focus on how useful a summary is and not necessarily on whether or not it is grammatically correct. After aggregating the data we have obtained an average of 3.12. We have also noticed that this score fluctuates in a rather narrow window of just 1 (i.e. between 2.5 and 3.5) when we vary the number of reviews that we process, as seen in Figure 3.

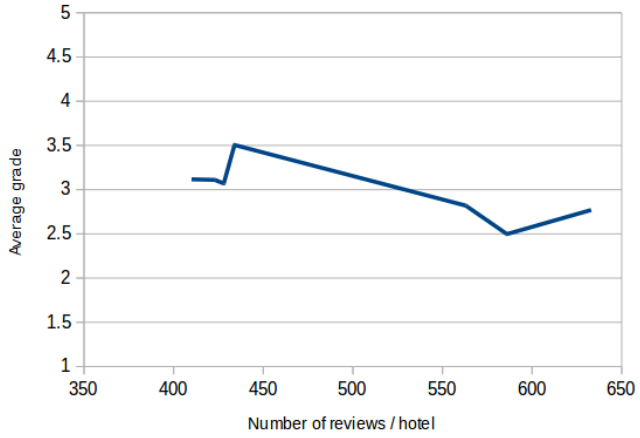

Figure 3: Average grade for hotels with different number of reviews

CONCLUSION

To sum up, our approach offers a fast solution to the problem of finding a set of n-grams that can summarize a given set of reviews without losing the recurrent aspects that appear throughout the reviews. Using the readability and representativeness scores as well as other useful metadata like the POS tags and the sentiment conveyed by an n-gram, the proposed algorithm attempts to use as much information as possible in order to select only the most suitable n-grams. On top of all this, the topic selection functionality makes sure to get rid of n-grams that do not cover a relevant subject as far as the reviewed product is concerned.

As far as other similar attempts go, our approach provides simple, unsupervised and unstructured way to summarize reviews while taking advantage of as much of the available information as possible. Since most of the algorithm is non-aspect based (except for the final part where we start looking for reviews with high scores that also fit a certain topic) it can be changed rather easily to work on different web platforms and different products, even though our implementation uses hotel reviews crawled from www.tripadvisor.com. As for further work, more optimization, including parallelization of some of the operations, will allow for more data to be analyzed in a decent amount of time and, therefore, will greatly increase the quality of the yielded n-grams.

ACKNOWLEDGEMENTS

This work has been partly funded by the Sectorial Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/ 132397.

REFERENCES

1. Carenini, G. and Cheung, J. C. K. Extractive vs. nlg-based abstractive summarization of evaluative text: the effect of corpus controversiality. In *Proc. of the Fifth International Natural Language Generation Conference 2008*, (2008), 33-41.
2. Di Fabbrizio, G., Stent, A. and Gaizauskas, R. A Hybrid Approach to Multi-document Summarization of Opinions in Reviews. In *Proc. The 25th International Conference on Computational Linguistics 2014*, 1 (2014), 23-29.
3. Flesch, R. A New Readability Yardstick. *Journal of Applied Psychology 32*, 1 (1948), 221-233.
4. Ganesan, K., Zhai, C and Viegas, E. Micropinion Generation: An Unsupervised Approach to Generating Ultra-Concise Summaries of Opinions. In *Proc. of the 21st international conference on World Wide Web 1*, 1 (2012), 869-878.
5. Hu, M. and Liu, B. Mining and summarizing customer reviews. In *Proc. of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, (2004), 168-177.
6. Kim, H. D., Ganesan, K., Sondhi, P. and Zhai, C. Comprehensive Review of Opinion Summarization. Technical Report University of Illinois at Urbana-Champaign (2011).
7. Kupiec, J., Pedersen, J. and Chen, F. A trainable document summarizer. In *Proc. of the 18th annual international ACM SIGIR conference on Research and development in information retrieval 1995*, (1995), 68-73.
8. Lin, C.-Y. and Hovy, E. A trainable document summarizer. In *Proc. of the 18th conference on Computational linguistics 2000*, (2000), 495-501.
9. Liu, J., Cao, Y., Lin, C., Huang, Y. and Zhou, M. Low-Quality Product Review Detection in Opinion Summarization. In *Proc. of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning 2007*, (2007), 334-342.
10. Lu, Y., Zhai, C. and Sundaresan, N. Rated aspect summarization of short comments. In *Proc. of the 18th*

International World Wide Web Conference 2009, (2009), 131-140.

11. Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proc. of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations 2014*, (2014)
12. Paice, C. Another stemmer. *ACM SIGIR 24*, 3 (1990), 56-61.
13. Pang, B. and Lee, L. Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval 2*, 1–2 (2008), 1-135.
14. Pang, B. and Lee, L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proc. of the First Workshop on Graph Based Methods for Natural Language Processing 2006*, (2006), 45-52.
15. Pang, B., Lee, L. and Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. In *Proc. of the ACL-02 conference on Empirical methods in natural language processing 10, 1-2* (2002), 79-86
16. Velikovich, L, Blair-Goldensohn, S, Hannan, K. and McDonald, R. The viability of web-derived polarity lexicons. In *Proc. of the Annual Conference of the North American Chapter of the Association for Computational Linguistics 2010*, (2010), 777-785.
17. Wang, H., Lu, Y. and Zhai, C. Latent aspect rating analysis on review text data: a rating regression approach. In *Proc. of Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining 2010*, (2010), 783-792.

Automated Paper Annotation with *ReaderBench*

Ionut Cristian Paraschiv, Mihai Dascalu, Stefan Trausan-Matu
University Politehnica of Bucharest
313 Splaiul Independentei, Bucharest, Romania
ionut.paraschiv@cti.pub.ro,
mihai.dascalu@cs.pub.ro,
stefan.trausan@cs.pub.ro

Philippe Dessus
LSE, Univ. Grenoble Alpes
Grenoble CEDEX 9 France
philippe.dessus@upmf-grenoble.fr

ABSTRACT
The annotation of articles from a given domain and the generation of semantic metadata can be considered a reliable foundation for creating a paper recommender system. Within this paper, the models from other previous researches are extended with the capability of visualizing articles and the most important concepts from a domain within imposed timeframes. This can be very useful for researchers to check out the most important publications from a given period, to view which are the trends and how a domain has evolved. Our previous analyses used the articles to build a paper graph and to suggest the most relevant articles, given a user defined query in natural language. This research contains a use case and creates visual graph representations to enhance the overall perception of the evolution of a domain.

Author Keywords
Scientometrics; paper recommendation system; time analysis; discourse analysis; semantic similarity.

ACM Classification Keywords
I.2.7 [**Natural Language Processing**]: Discourse, Language parsing and understanding, Text analysis.

INTRODUCTION
A researcher's daily activities usually involve the study of new papers, as to use the information in building solutions and observing how the domain evolves. Since the retrieval of documents from the Internet can lead to large data flows, it is important to consider other approaches for a more comprehensive analysis of the domain. In this context, a paper annotation system that automatically retrieves papers on a given topic and tags them can be critical and can make the exploration phase of the research literature easier.

We propose a model that takes a large set of paper abstracts and tags them, later on annotating the results within a semantic database. The database can be queried for user defined texts, and can enable the researcher to explore the resulting graphs using different timeframes, as to retrieve the most important articles and concepts within a period. Moreover, a list of similar topics with the user's query is shown with the intention of stimulating the user in his/her research tasks.

The initial part from this paper will concentrate more on similar studies that discuss on how to build network graphs for scientific papers. We continue with the methods used behind the current model that demonstrate its potential and extensibility, as well as how they are used, and we finish with possible future improvements of our system.

RELATED WORK
A paper annotation system can be built with database software that uses keyword matching such as *Mendelev* or *DevonThink*, or other complex methods [11]. However, the current model does not rely on information retrieval [13] as we describe an alternative of annotating a dataset of documents. A different, older approach of indexing papers relies on co-citation analysis [6; 7], but that method is out of scope in terms of the current research.

Information Retrieval
Information Retrieval [13] techniques aim at finding materials of an unstructured nature, usually text, that satisfy information needs from within large textual collections. This process is usually concerned on how to store and structure data in such a manner that it will facilitate the retrieval of information based on a given query and in a relatively small amount of time. Being a text recognition tool, the query, consisting of a Boolean combination of keywords, is usually mapped with the collection; therefore, no semantic meaning is associated to the query or to the set of documents. In this way, the more complicated or complex the query is, the smaller are the chances of finding relevant results.

Semantic Similarity Analysis of Paper Abstracts
Our paper annotation model relies on two widely used methods. The first one, Latent Semantic Analysis [6; 7], is a natural language processing method that is used for analyzing relationships between documents and their terms, in our particular case – abstracts [9]. The method builds a document-term matrix that basically assigns for every word its corresponding number of occurrences within the document. After applying a Singular Value Decomposition, the dimensionality of the matrix is reduced, while keeping its similarity structure with a marginal error. At the end, the documents are compared by computing the cosine similarity between their associated vectors within the semantic space.

The second method, Latent Dirichlet Allocation [2], uses topic distributions among documents, and in combination with LSA can give an aggregated cohesion scores [4] that can be more accurate for computing semantic distances. In the end, in order to compute the semantic distance between two words, our model also relies on semantic distances extracted from the lexicalized ontology WordNet, together with LSA and LDA semantic models [3].

SYSTEM'S IMPLEMENTATION

Using the approaches described in previous subsections, the papers from the initial dataset were tagged using the content from their abstracts, that usually contains the main ideas [8], as seen in our previous research [14]. In this manner, three different views on the dataset of articles have been built. The first one, the document similarity view [14], generates a graph with the papers which are connected if their semantic similarity exceeds a threshold. The second one, the concept map view [14], extracts the most important concepts from the subset of papers, and builds a graph where the words are connected using their relevancies [4]. In both these two views, the nodes are sized depending on their centrality, and the links between them are enforced based on the similarity. In this way, the user will see at the end which are the most important words and documents from the set of papers. The third view displays the document space for a particular paper [14], and can help researchers when they want to read semantically related articles.

After this first implementation phase which was previously presented [14], the improvements presented in this paper are focused on enabling the user to define his/her own queries. In the end, similar graphs to the document similarity view [14] are displayed, but with the papers that have a high semantic cohesion with the text introduced by the user. Moreover, the user can also check semantically related concepts with the query terms (concepts that do not specifically appear inside the query), thus stimulating his/her imagination with ideas for new queries. This is the first exploratory extension of our system.

The text given by the user enters inside the same pipeline as the abstracts: text preprocessing, lemmatization, part-of-speech tagging, syntactic dependency analysis and topic extraction [4; 12]. The next step is to represent the query using LSA and LDA vectors, and to compute its distance with every document from the dataset. The LSA query vector is obtained by summing up normalized occurrences of each constituent lemma vector representation, whereas for LDA the Gibbs inference tools is applied on the query in order to deduce the topic distributions based on pre-trained models. In order to increase the user's control, the final view shown in Figure 1 also contains a threshold used for displaying the links between the papers and the query, which can be manually adjusted. The documents are also shown inside a table, ordered by their similarity with the input query.

The second extension presented in this paper consists of enabling the user to select a timeframe for the displayed papers and concepts. In this manner, a researcher can check the evolution of a domain, the most important articles in a period and the most central concepts. As many domains evolve in a dynamic manner, this is definitely the way to check past and current trends, as well as concepts that become more important.

From a technical point of view, the paper annotation system uses the core components from *ReaderBench* [4; 5], a versatile tool for text and discourse processing. With a fully functional natural language processing pipeline incorporated, LDA and LSA semantic pre-trained models, WordNet and semantic distances, as well as Social Network Analysis [1], *ReaderBench* is an extensible tool for most of the undergone text processing steps.

These two extensions have the purpose of stimulating a researcher in his daily tasks by suggesting papers, similar concepts, as well as the modeling of the evolution of a domain within a period of time. Together, they can definitely support anyone interested in learning a domain. The next section is centered on specific use cases of how the system can be used with a real dataset of paper abstracts.

USE CASE

Dataset Description

The used dataset of abstracts consists of papers extracted from the citation index Web of Science, from the Education and Educational Research [10] domain, taken between the years 2000-2004. From this dataset, a subset of paper which contained within their abstracts the keywords "*IT"*, "*technology"* or "*computer science"* was extracted.

User Queries

Given the database of annotated abstracts, the user inputs an initial query "*electronic learning and information technology*", with the intention of finding papers related to informational systems that implement learning facilities. Using a threshold value of 70%, Figure 1 displays a sub-graph with the most related articles with the query. Table 1 displays what is the content of these articles.

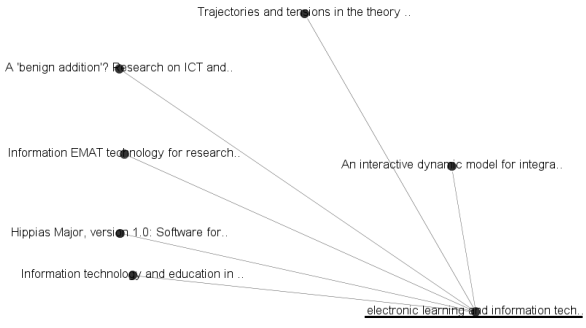

Figure 1. Semantically related articles to the input query

From Table 1, we can observe that the most related abstracts with the user's query are semantically related, and unlike standard information retrieval systems, the results don't necessarily have common words with the query. This can impact the retrieval of enhanced search results, and can definitely help anyone in finding documents about a subject.

For the example query in this subsection, the researcher could be interested to check some related concepts with his/her query. The system is capable of displaying the related keywords in a graph, which can be explored by

the user. In this context, the model suggests words such as "learner", "teacher", "science", "curriculum", "research", "process", and "development" (see Figure 2), which are quite relevant given the input query. Moreover, we can check the underscored words as being from the input query, while the others are being automatically introduced as semantically related to them. All the words with a threshold over .5 in terms of semantic similarity are shown, and they are grouped together using the same stem; in the end, the shortest lemma is being displayed.

Paper title and abstract	Relevance
Trajectories and tensions in the theory of information and communication technology in education "For largely historical reasons, information and communication technology in education has been heavily influenced by a form of constructivism based on the transmission and transformation of information. This approach has implications for both learning and teaching in the field. The assumptions underlying the approach are explored and a critique offered...."	.78
Information technology and education in the information age "This paper attempts to gain an understanding of current and potential impact of information technology (IT) on education in the information age. First, it attempts to highlight that integration of information technology in teaching is a central matter in ensuring quality in the educational system. ..."	.74
Hippias Major, version 1.0: Software for post-colonial, multicultural technology systems "The first half of Plato's Hippias Major exhibits the interfacing of the first teacher (Socrates) with the first version of a post-colonial, multicultural information technology system (Hippias). In this interface the purposes, results and values of two contradictory types of operating system for educational servicing units are exhibited to.."	.72

Table 1. Samples from the related papers

Unlike standard information retrieval systems, this model performs better when it comes to user queries, as in this case, more complex queries usually mean richer semantic content, and thus more accurate results.

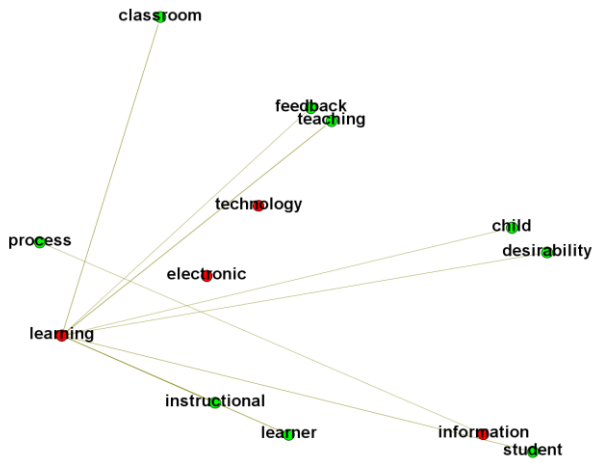

Figure 2 Semantically related concepts

Timeframe View

Although not a computationally demanding task, the ability to filter articles and important concepts using different timeframes can be very appealing to researchers, as they can check the trends and the evolution in time of a specific domain. In Figure 3, a subgraph from the article similarity view from the year 2002 is shown, displaying the central article having a relevance of 0.81.

Paper title and abstract
A collaborative, investigative recombinant DNA technology course with laboratory "A recombinant DNA technology course was designed to promote contextual, collaborative, inquiry-based learning of science where students learn from one another and have a sense of ownership of their education. The class stressed group presentations and critical reading and discussion of scientific articles. The laboratory consisted of two research projects: random cDNA …."
Technology in the first two years of collegiate mathematics "We present several roles of technology and suggest various ways that technology could have a lasting and significant impact upon the quality of mathematics courses being taught in the first two years of collegiate mathematics. Overcoming some mathematicians' anxiety and reluctance to address applied problem solving so as to take full advantage of the opportunities remains a challenge for the future…"
Maximising the educational affordances of a technology supported learning environment for introductory undergraduate phonetics "New technologies afford a range of opportunities that can transform teaching techniques and offer enhanced possibilities for learning. This potential is often not grasped by the technologist or the educationalist when introducing new technologies into the learning situation and a situation arises which can be described as 'New technology, no new pedagogy.'…"

Table 2. Central article summaries in 2002

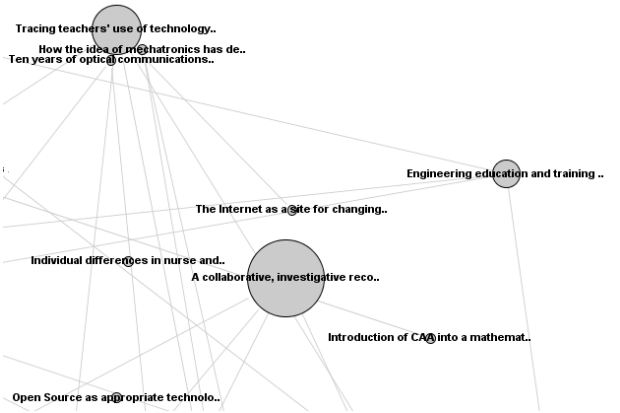

Figure 3. Subgraph of the article similarity view – Year 2002

The graph of the most important concepts from 2002 is displayed in Figure 4.

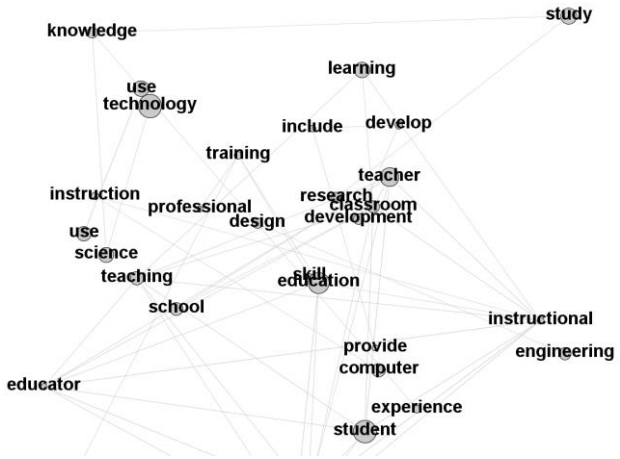

Figure 4. Central concepts in 2002

As displayed in Figure 4, the most important concepts from the dataset in 2002 are *educational, classroom, training and instruction*. It will be very interesting to check how the central article's topics and the corresponding concept graph have changed in 2004.

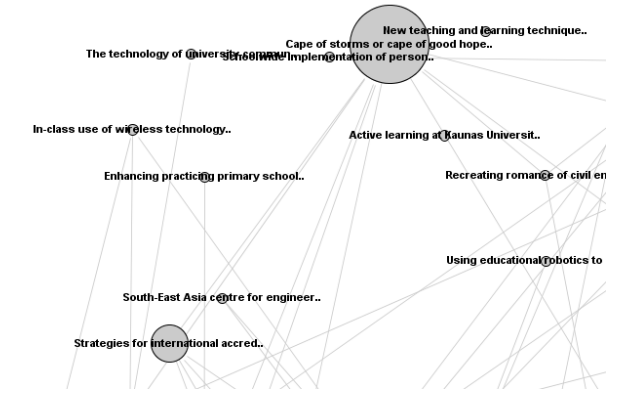

Figure 5. Subgraph of the article similarity view – Year 2004

Paper title and abstract
Cape of storms or cape of good hope? Educational technology in a changing environment "This article locates and describes the work of the Multimedia Education Group (MEG) at the University of Cape Town (UCT). This work is contextualised by three national and international challenges, these being (1) the need to increase access to new technologies and overcome the digital divide, (2) the need to respond to a new communication order, and (3) the urgency of transforming higher education…"
Technology and curricular reform in China: A case study "This article reports on a 5-year study of a technology-enhanced educational reform initiative at a university in eastern China. A facility team attempted pedagogical and curricular reform to better prepare English Majors to use new technologies for international communication, collaboration, and research. The team developed several project-based courses and incorporated technology into traditional lecture courses..."
Web-based curriculum development of a manufacturing engineering technology programme "The aim of this paper is to present the use of the Internet in developing the curriculum of a manufacturing engineering technology programme in Turkey. The programme was implemented in the curricula of 15 two-year colleges over six months to provide seamless progression from vocational high school to two-year colleges and meet the needs of Turkish and global industry…"

Table 3. Central articles summaries in 2004

Overall, it becomes clear that there is a difference in terms of semantic meaning between the central articles from 2002 and the ones in 2004. As technology has evolved in that period, the central articles are more related to information technology applied inside learning environments, which supports the idea that the user can check using our system how the domain has evolved in time.

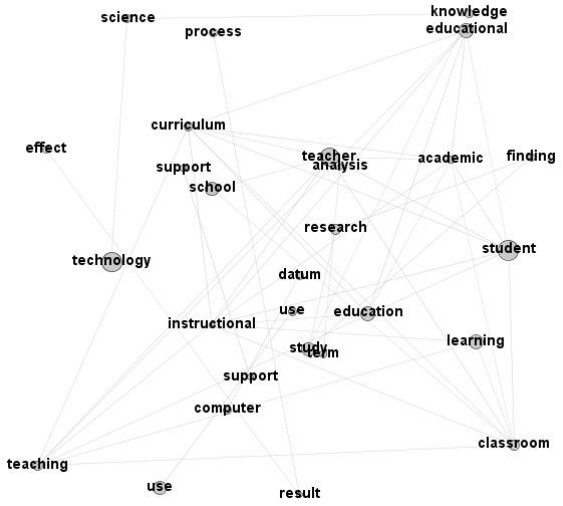

Figure 6. Central concepts in 2004

To sustain this idea, Figure 6 displays some of the most important concepts from 2004, where the most preeminent lemmas are "*technology", "education", "teacher", "study", "learning", "science"* and *"computer"*. They demonstrate the idea that the domain has evolved as new technologies appeared and were applied inside the educational research domain.

Another interesting experiment on the two datasets of papers was to find out which are the most similar concepts. In this manner, we have extracted the most important 100 concepts from the papers written in 2002, Dataset 1 (DS1), and from those written in 2004 (DS2). Every concept from DS1 was compared with every concept from DS2, and Table 4 displays the most similar pairs from the two subsets. In this manner, the researcher can get additional information regarding the domain and how it has evolved.

Concept from 2002	Concept from 2004	Sim.
Classroom	Student	.92
Educator	Student	.86
Design	Engineering	.75
Knowledge	Experience	.71
Educational	Student	.71
Professional	Practice	.66

Table 4. Concept similarities from the subsets

Further on, Table 5 depicts the most similar abstracts between the two subsets, being a good metric to see how the content has evolved in the two years. The idea of the table is that, by checking on the articles that are semantically related in different periods, a researcher can observe a domain's trends and what emerging solutions have appeared regarding certain problems.

Article pairs (marked as X.a and X.b)	Score
1.a *Introduction of CAA into a mathematics course for technology students to address a change in curriculum requirements* **(2002)** The mathematical requirements for engineering, science and technology students has been debated for many years and concern has been expressed about the mathematical preparedness of students entering higher education. This paper considers a mathematics course that has been specifically designed to address some of these issues for technology education students …	.955
1.b *Standardized test outcomes of urban students participating in standards and project based science curricula* **(2004)** Considerable effort has been made over the past decade to address the needs of learners in large urban districts through scaleable reform initiatives. We examine the effects of a multifaceted scaling reform which focuses on supporting standards based science teaching in urban middle schools. The effort was one component of systemic reform efforts …	
2.a *Integrating algorithm visualization technology into an undergraduate algorithms course: ethnographic studies of a social constructivist approach* **(2002)** Algorithm visualization (AV) software graphically illustrates how algorithms work. Traditionally, computer science instructors have used the software as a visual aid in lectures, or as the basis for interactive laboratories. An alternative approach, inspired by Social Constructivist learning theory …	
2.b. *Classroom use of multimedia-supported predict-observe-explain tasks in a social constructivist learning environment* **(2004)** This paper focuses on the use of multimedia-based predict-observe-explain (POE) tasks to facilitate small group learning conversations. Although the tasks were given to pairs of students as a diagnostic tool to elicit their pre-instructional physics conceptions, they also provided a peer learning opportunity for students. The study adopted a social constructivist perspective...	.913
3.a *Conditions for classroom technology innovations* **(2002)** This article reports on a study of the complex and messy process of classroom technology integration, The main purpose of the study was to empirically address the large question of "why don't teachers innovate when they are given computers?" rather than whether computers can improve student learning. Specifically, we were interested in understanding the conditions …	
3.b. *New teaching and learning techniques facilitated by information technology* **(2004)** A wide variety of classroom techniques are being advocated to increase learning: active learning, collaboration, integration of assessment and feedback, and the use of concrete physical manipulatives. These techniques must be transformed into practical tools and be infused with content from the subject area. At the same time, the information technology revolution has provided new tools …	.898

Table 5. Most similar articles from the subsets

This section presented in detail a generalizable use case that can be easily extrapolated on any dataset of papers and can enable researchers to better understand a domain. Moreover, the results clearly indicate that the evolution of a domain can be better understood by analyzing the semantic content of the articles within certain timeframes.

CONCLUSIONS

As more and more research communities appear and they are more dynamic than ever, it is becoming quite hard for a researcher to keep up-to-date with this fast growing information. In this context, a paper annotation model and viewer can be a good alternative to better visualize the papers from a certain dataset. Moreover, the support for user defined queries, graphical visual representations and timeframe filtering increase the overall understanding of a domain and, in the end, the productivity of researchers.

As future developments, the timeframe snapshot will be displayed in an interactive and animated manner, not just as a static graph within a period of time. Moreover, a current drawback must be addressed: the long preprocessing time due to the NLP processing pipeline, and a relatively small number of possible papers that can be loaded directly into the system's memory. Therefore, some improvements must be done for our model in terms of memory and CPU consumption. In order to address these issues, clusters of papers will be created and the search will be conducted within a multi-hierarchical structure of documents..

ACKNOWLEDGMENTS

The work presented in this paper was partially funded by the FP7 2008-212578 LTfLL project and by the Sectorial Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreements POSDRU/159/1.5/S/134398 and POSDRU/187/1.5/S/155420. We also thank Pablo Jensen and Sebastian Grauwin for providing the initial corpus of paper abstracts.

REFERENCES

1. Bastian, M., Heymann, S., and Jacomy, M., 2009. Gephi: An open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media* AAAI Press, San Jose, CA, 361–362.
2. Blei, D.M., Ng, A.Y., and Jordan, M.I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, 4-5, 993–1022.
3. Budanitsky, A. and Hirst, G., 2006. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics 32*, 1, 13–47.
4. Dascalu, M., 2014. *Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence*. Springer, Switzerland.
5. Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., and Nardy, A., 2014. Mining texts, learners productions and strategies with ReaderBench. In *Educational Data Mining: Applications and Trends*, A. Peña-Ayala Ed. Springer, Switzerland, 335–377.
6. Deerwester, S., Dumais, S.T., Furnas, G.W., Harshman, R., Landauer, T.K., Lochbaum, K., and Streeter, L., 1989. Computer information retrieval using latent semantic structure In *4,839,853*, Uspto Ed. Bell Communications Research, Inc., USA.
7. Deerwester, S., Dumais, S.T., Landauer, T.K., Furnas, G.W., and Beck, L., 1988. Improving Information Retrieval with Latent Semantic Indexing. In *51st Annual Meeting of the American Society for Information Science 25*, 36–40.
8. Ding, Y., Song, M., Wang, X., Zhang, G., C., Z., and Chambers, T., 2014. Content-based citation analysis: The next generation of citation analysis. *Journal of the American Society for Information Science & Technology 65*, 9. DOI= http://dx.doi.org/10.1002/asi.23256.
9. Gordon, M.D. and Dumais, S., 1998. Using Latent Semantic Indexing for literature based discovery. *Journal of the American Society for Information Science 49*, 8, 674–685.
10. Grauwin, S. and Jensen, P., 2011. Mapping scientific institutions. *Scientometrics*. DOI= RoCHI_v2.0.doc DOI 10.1007/s11192-011-0482-y.
11. Joeran, B., Langer, S., Genzmehr, M., Gipp, B., Breitinger, C., and NüRnberger, A., 2013. Research Paper Recommender System Evaluation: A Quantitative Literature Survey. In *Workshop on Reproducibility and Replication in Recommender Systems Evaluation (RepSys) at the ACM Recommender System Conference (RecSys'13)* ACM, Hong Kong, China.
12. Jurafsky, D. and Martin, J.H., 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Pearson Prentice Hall, London.
13. Manning, C.D., Raghavan, P., and Schütze, H., 2008. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, UK.
14. Paraschiv, I.C., Dascalu, M., Trausan-Matu, S., and Dessus, P., 2015. Analyzing the Semantic Relatedness of Paper Abstracts - An Application to the Educational Research Field. In *2nd Int. Workshop on Design and Spontaneity in Computer-Supported Collaborative Learning (DS-CSCL-2015), in conjunction with the 20th Int. Conf. on Control Systems and Computer Science (CSCS20)* IEEE, Bucharest, Romania, 759–764.

Conversational Agent that Models a Historical Personality

Adrian Bogatu, Dorin Rotarescu, Traian Rebedea, Stefan Ruseti

University Politehnica of Bucharest
313 Splaiul Independentei,
Bucharest, Romania

Autonomous Systems
22 Tudor Vladimirescu,
Bucharest, Romania

bogatu.adrian@gmail.com, dorinrotarescu@yahoo.com, traian.rebedea@cs.pub.ro, stefan.ruseti@cs.pub.ro

ABSTRACT

In this paper we discuss the current approaches in question answering (QA) and their applicability in building a conversational agent that models a historical figure that gives informative and relevant answers to user questions about the life of that personality. We analyze two main methods: one in which we use an ontology to build our knowledge base and one where we don't have a knowledge base and we solve the *answer sentence selection* problem for question answering. We observed that the first method is better for answering more general questions and the second method can deal with more specific and complex lexically and syntactically questions. The implementation of the conversational agent relies on the two combined approaches, the second being a fallback option if the first method is not able to provide an answer.

Author Keywords

Conversational agent; natural language processing; information retrieval; question answering; ontologies; answer sentence selection.

ACM Classification Keywords

H.3.4 Systems and Software: Question-answering (fact retrieval) systems.

INTRODUCTION

More than 60 years ago, Alan Turing raised the question "Can machines think?" in the book with the same name [10] where he devised the "Imitation Game" test, which is similar to the now known Turing test. Since then, computer scientists have tried to create programs that can interact with humans and maintain a human-like conversation.

Cleverbot[1] is one of the programs that accomplished that and also obtained a score of being 59.3% human from the jury of the competition, while humans achieved a 63.3% "humanness" score. Cleverbot uses a database of saved conversations updated constantly and replies with an answer by matching the user input to previous phrases found in the database.

Other conversational agents simulate the personality of a certain person or typology of person and usually use rule-based systems. An example is the bot ELIZA that tries to match the user input to a rule and output the answer associated with that rule [12]. Another example is Freudbot, a chat-bot similar to ELIZA that tries to impersonate the psychologist Sigmund Freud and talk about his theories and biography in the first person [6]. A different category is open domain question answering software systems like IBM's Watson, which uses Natural Language Processing (NLP), information retrieval, machine learning and other techniques to provide an answer to a question [2].

This paper presents the current approach towards the implementation of a conversational agent that models a historical character using basic NLP concepts, information retrieval methods and question answering techniques. The conversational agent can be used in museums to guide and inform visitors or in schools as an e-learning tool.

Building a conversational agent using NLP is not an easy task, firstly because we do not always speak in a grammatically correct fashion. Secondly, in a conversation, the speaker assumes that the listener knows and understands the details of the ongoing conversation. A big challenge is to determine the context in which a question is posed and to understand the meaning beyond the lexical structure of the question.

From a programming point of view, the state-of-the-art chat-bot relies on a set of files containing rules in the form of question-answer pairs, usually defined in AIML [11] or similar languages, which are constructed based on the way the answer is expressed.

However, in order to build a robust conversational agent it is expected that an input (question) to have multiple rules that can be matched against the question and provide a relevant answer. This implies that a large number of various rule-answer pairs are needed. On the other hand, it is hard to predict the interaction between rules when adding new rules to an existing set.

In order to model a historical figure, we started from that person's biography on Wikipedia and the associated DBpedia[2] page. Consequently, we can identify the way a certain property is expressed starting from the properties and their values from DBpedia correlated with the Wikipedia text.

The paper is organized as follows: the next section discusses some existing approaches on building a

[1] Cleverbot, online at http://www.cleverbot.com/.

[2] DBpedia, online at dbpedia.org.

conversational agent, using either an ontology or a text matching method in order to answer questions. The next section describes the steps we took to create a knowledge base and generate rules for our conversational agent. In the end of the implementation section the results using ontologies are presented. After that, the integration of the answer sentence selection approach, as a fallback method for the first approach, is analyzed. Finally, the conclusions are presented.

RELATED WORK

Our project is inspired from conversational agents implemented using ontologies as well as conversational agents that do not use a knowledge base and rely only on the question asked and a corpus from where the answer will be extracted. Next we will present some background on these two approaches.

Ontology approach

As defined in [5], an *ontology* is an explicit specification of a conceptualization, where *conceptualization* is defined as an abstract, simplified view of the world that we want to represent.

The Intelligent Verilog Compiler Project [9] is a tutoring system used for teaching the Verilog language. It is said to be intelligent in two ways: "it helps check the syntax and the semantics of the learner's program and it finds a technical or English definition, comparison or example suitable to the error being reported in the context of the piece of the code. It displays the information next to the incorrect code and errors in order to 'scaffold' learning without directly providing the answer" [9]. This method of interaction is accomplished using an ontology of the Verilog language.

Another dialog system where ontological resources are used is one personifying the author Hans Christian Andersen. The domains of discourse contain his fairy tales, his life and the user [8]. It is stated that the reasons to use ontologies are: faster development because of the shared ontology over different conversation domains; the fact that the application can be easily extended to support new domains of conversation.

A technical approach on using ontologies for question answering is described in [4] and given an implementation in [3]. The authors thought of *matching* as an operation on two graph-like structures that "produces a mapping between elements of the two graphs that correspond semantically to each other" [4]. Starting from the said concept of *matching*, the authors imagined the next step: semantic matching, which also analyzes the meaning behind nodes in the two graphs.

As described in [4], this approach has two main features:

- search for semantic correspondences by mapping meanings (concepts), and not labels, as in syntactic matching. As the rest of the paper makes clearer, when mapping concepts, it is not sufficient to consider the meanings of labels of the nodes, but also the positions that the nodes have in the graph.
- use semantic similarity relations between elements (concepts) instead of syntactic similarity relations. In particular, we consider relations, which relate the extensions of the concepts under consideration (for instance, more/less general relations).

In the case of ontologies, this approach works if we can construct an equivalent graph-like representation of a given ontology.

Answer sentence selection approach

Answer sentence selection is the task of finding a sentence from a set of candidate sentences that best answers a given question.

The method we rely on the most is the one described in [1]. In [1], there are three main approaches analyzed, but we are only interested in the first one: the approach that uses algorithms that rely on "sophisticated syntactic/semantic processing" [1]. The authors present three main types of extracting and using knowledge for the answer sentence selection problem. The first one is to determine the type of answer (also known as "Qtarget") to a given question. For example:

Question: What is the duration of the song "Hey, Jude"?
Qtarget: TEMPORAL-QUANTITY
or
Question: When were you (John Lennon) born?
Qtarget: DATE

After finding the type of answer needed, the set of possible answers can be filtered according to the semantic type of the question, and we can remove all the sentences that have a type that does not match the found Qtarget.

The second type of knowledge contains the semantic relations between constituents that appear in the question. The correct answer should preserve these relations. Example from [1]:

Question: Who killed Lee Harvey Oswald?
Text: Jack Ruby, who killed John F. Kennedy assassin Lee Harvey Oswald.

The explanation is given in [1] immediately after the example: "Even if 'John F. Kennedy' is textually closer to the question terms 'killed' and 'Lee Harvey Oswald', the system will choose "Jack Ruby" because its logical subject relation to the verb matches that of the interrogative in the question."

The third type of knowledge relies on the use of paraphrases. Because the wording in a potentially correct answer is not always similar with the one in the question, immediate textual matching does not always work. The idea is to generate alternate formulations of the question (but preserve the meaning) in order to increase the matching chances for a good answer.

An example of reformulation of a question from [1]:

Question: How deep is Crater Lake?
Reformulation patterns:

- Crater Lake is <what distance> deep?
- depth of Crater Lake is <what distance>?
- Crater Lake has a depth of <what distance>?
- <what distance> deep Crater Lake?
- Crater Lake's depth is <what distance>?

OUR METHOD

ChatScript

ChatScript [3] is a chat-bot engine, a tool that helps build conversational agents that are based on rules. It uses a scripting language to build these agents and process natural language. ChatScript represents the state-of-the-art for conversational agents and helped with the transition "from matching patterns of words to matching patterns of meaning." [13]

A chat-bot is modeled through a set of script files that contain rules. A rule is formed from a pattern and a response. The response represents the output that a ChatScript bot will provide if the input matches the pattern. For example:

```
u: ( Where * you * born ) In the capital.
u: ( When * born  ) This century.
```

The elements in the parentheses constitute the pattern and the sentence after the pattern represents the answer. The star symbol is a wildcard that can match none, one or more words.

In our case, the input is represented by the question asked by the user, therefore we want to create the best patterns for each possible answer we have retrieved from a person's biographical text. Example for our generated rules:

```
u: (vb marry) I married Elsa Löwenthal on 2 June 1919 , after having had a relationship with her since 1912 .
u: (vb die in in) I died in Princeton Hospital early the next morning at the age of 76 , having continued to work until near the end .
```

In the examples above, in the parenthesis we find the part of speech (in our examples, this is a verb) returned by Stanford NLP and the lemma of the word from the expression found by the algorithm.

Because our program needs to support a large number of different historical figures, we needed to automate the creation of ChatScript specific files. The method of generating the scripts is described in the *Pattern generation* subsection.

Stanford CoreNLP

Stanford CoreNLP [7] [4] is a tool for analyzing and processing text. It integrates some useful modules that we used for Wikipedia articles. The part-of-speech tagger was used to differentiate between verbs, nouns, adjectives etc. The tokenizer was used to split paragraphs into sentences and sentences into words, while the lemmatization tool was employed to find the canonical form for a word. In the end, the co-reference resolution system was used to link the subject of sentences to the anaphoric proper name in the context of a paragraph, if there exists such an anaphoric element.

Methods

Expressing a Property

In order to understand what is trying to be expressed in a sentence, we start from the DBpedia properties of a large set of people. Subsequently, for every property, we try to determine how that property is expressed in the Wikipedia corpus. For the information extraction, we use Stanford CoreNLP.

To find the manner in which a property is expressed, we searched the value of that property in the sentences from Wikipedia where the person which the article is about appears in. Having the desired sentence identified, we annotated it using Stanford CoreNLP and obtained a syntactic parse tree. Analyzing this tree, we determined that the root is the verb directly connected with the subject. Having the parse tree, we considered that the best way to express the property is the path from that property to the root verb.

Applying this algorithm to a large set of people, we managed to build a big, but extensible knowledge base by introducing the most relevant output expressions in a knowledge base.

Examples

In Table 1 we present some entries in our knowledge base, where the property is extracted from DBpedia and the lexicalization represents an enumeration of ways the given property appears to be expressed in the Wikipedia articles.

DBpedia property	Lexicalizations
birthDate	born; born in
almaMater	receive in; graduate as
award	award; receive
college	graduate from; attend
deathPlace	die in
profession	serve in; become
spouse	marry; marry to

Table 1: Examples of how specific DBpedia properties are most often expressed.

[3] ChatScript, online at http://chatscript.sourceforge.net/.

[4] CoreNLP, online at http://nlp.stanford.edu/software/corenlp.shtml

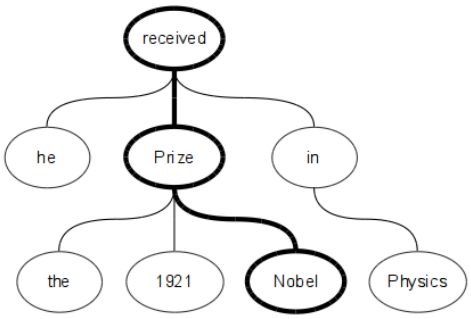

Figure 1: Parse tree for the phrase "*He received the 1921 Nobel Prize in Physics*"

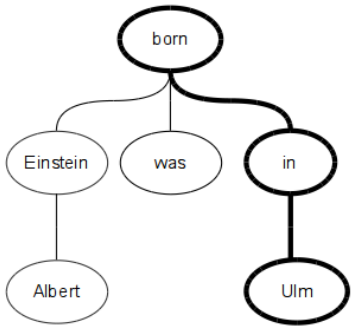

Figure 2: Parse tree for the phrase "Albert Einstein was born in Ulm."

In Figure 1 is shown the syntactic parse tree for the sentence "He received the 1921 Nobel Prize in Physics" without the corresponding syntactic elements. The DBpedia property that connects "Albert Einstein" and "The Nobel Prize" and for which we want to identify a lexicalization is the "award" property. These are the characteristics of the search:

Subject: Albert Einstein
Object: Nobel Prize
Property: Award
Output: receive

In Figure 2 is presented a similar parse tree from which we can extract the following information:

Subject: Albert Einstein
Object: Ulm
Property: BirthPlace
Output: born in

We can see that in the second example the found expression of a property is the path from the verb to the value of the DBpedia property (in this case "BirthPlace") excluding its value.

Pattern Generation

In order to generate ChatScript files for a specific person, we fetch that person's Wikipedia page, split it into phrases and keep only those that have the person as a subject. This filtering was done using the Stanford Deterministic Co-reference Resolution System. After extracting all the phrases referring to the current historical figure, we select only those that express a property from DBpedia matching the expression against the knowledge base. We then create a rule-answer entry to add to the ChatScript files. The rule is represented as an expression of a property that appears both in the analyzed sentence and the knowledge base. The answer is the analyzed sentence from Wikipedia which is converted to be expressed in the first person. The conversion from the third person to the first person of the sentence is accomplished with Stanford's Part-of-Speech Tagger and CoreNLP. All these patterns are written in ChatScript's file hierarchy. For a fast and easier way to find the answer, we arranged ChatScript's files by the properties of the person.

INTEGRATION OF ANSWER SENTENCE SELECTION

Because it is impossible to build an exhaustive rule-based system, a secondary approach to this problem has to be taken into consideration in order to give a good answer. In addition, ChatScript has its own limitations coming from the fact that it ignores a rule after it first matches it. Therefore, a fallback option is needed in case the former approach fails to provide an answer.

Considering the fact that the former approach gives better answers the simpler and more common questions are asked, we observe that either it fails to match questions that are more complex or it has too many matches for a question that uses a common verb (like "to be" or "to have") and the results will be inaccurate or noisy. The solution to avoid this is to try and find the answer directly from the source of the previously described knowledge base with an ad hoc approach considering every sentence from that respective source.

Answer sentence selection

Following the goal of having to answer a question for a certain historical figure, the set of possible answers is reduced to a set of sentences from that person's biography. This leaves us with the task of identifying a sentence from a biography that has the highest probability of correctly answering the question at hand.

Considering what was previously stated, that this approach tries to find the answer to a more complex question, we can assume that, at least for now, there is a great deal of semantic information embedded in the form of the question (lexically and semantically) so that the chances are a part of the answer textually lies in the question. Therefore, what we can do is actually search for the question (or paraphrases of the initial question) in the reference text.

To get the best results out of this approach, we need to follow a number of steps.

First, we need to remove unnecessary words, including stop words, the interrogative words (what, when, where, who, why and how) and irrelevant verbs ("to be", "to have" and

other similar verbs as described above) in order to remain only with meaningful words, i.e. the kernel of the question.

Second, we want to use the Stanford CoreNLP software to lemmatize the question (i.e. to convert every word to its appropriate canonical form) because, as described later on, the corpus used for a historical figure will be lemmatized too. This will help in the search step because words will more likely match if they are in their base form.

Third, we try to find alternative ways of expressing the input question and attempt to search for all these variants in the biographical text. We do this by trying different synonyms for the words in the question so that we can get more results, even if the initial question is formulated in such a way that it does not contain the exact words that might appear in the sentence representing the correct answer.

Next, the top paragraphs from the Wikipedia article are filtered based on the textual matching score between the question and the respective paragraph given by Apache Lucene[5], a specialized text indexing and searching tool. Then, we apply the same algorithm at the level of sentences instead of paragraphs. In short, to get the best answer the corpus is divided in separate paragraphs and a small set of paragraphs where the answer sentence might be part of are selected. Next, we attempt to find an even smaller set, made of sentences that are the best candidates to answer the question.

After we have a set of sentences that passed the lexical filtering, we want to eliminate those in which the subject of the sentence does not match the subject of the question. This mechanism is similar to maintaining the semantic relations as described in [1], and presented above in the Related Work section. To achieve that, we want to use Stanford CoreNLP, and in particular the Stanford Deterministic Co-reference Resolution System, to determine who is the subject of a given sentence.

After the syntactic filtering we are left only with the semantic filtering. This means we want to filter out all the sentences that do not have the type as the one expected by the question. For example, questions starting with "When" expect an answer sentence that contains a numerical value.

Finally, we choose the first sentence in order of the previously gathered relevance scores. Using the aforementioned approach, we manage to answer more complex questions.

Algorithm

1. fetch the question for a historical figure
2. fetch the biography text for that person
3. split the text into paragraphs
4. lemmatize the question and all the paragraphs
5. find a set of top paragraphs against which the lemmatized question matches
6. reduce this set to a set of sentences from these paragraphs against which the lemmatized question match
7. eliminate the sentences that do not refer to the subject of the question
8. filter out the sentences that don't respect the expected answer type
9. return the best answer found

RESULTS

Knowledge base approach

Because a conversational agent is an interactive system, verifying and validating it is not an easy task, therefore the only way of testing the chat-bot was by asking questions and analyzing the returned results. Most of the test questions were conceived by the development team. The first time a historical person is chosen to interact with, we generate all the rules (questions and answers) that can be extracted from the knowledge base which are written in ChatScript format. After this step, each time this historical person is selected, we use the already generated data from the ChatScript. Because of this the response time is less than 1 ms.

So far, the conversational agent provides answers to simple questions that express properties from DBpedia. Some examples follow below.

Examples for Albert Einstein bot

Q: Where were you born?
A: I was born in Ulm, in the Kingdom of Wurttemberg in the German Empire on 14 March 1879.
Q: When did you die?
A: I died in Princeton Hospital early the next morning at the age of 76, having continued to work until near the end.
Q: What award did you receive?
A: I received numerous awards and honors, including the Nobel Prize in Physics.
Q: What schools did you attend as a child?
A: I attended a Catholic elementary school from the age of 5 for three years.
Q: Who were you married with?
A: I married Elsa Löwenthal on 2 June 1919, after having had a relationship with her since 1912.

Personality	Different answer
John F. Kennedy	146
Abraham Lincoln	141
Ronald Regan	131
Adolf Hitler	129
George W. Bush	125

Table 2: The number of different answers given by the bot.

[5] Apache Lucene, online at https://lucene.apache.org/.

A way of validating this method is done through the number of different answers we can provide, which are presented in Table 2.

Answer sentence selection method

The testing of the answer selection methods was done for a set of 5 personalities and about 20 questions for each personality, some of them general, others more specific. What was observed during the testing is that, in the majority of cases, a correct answer was somewhere in the top 10 sentences outputted by the lexical pipeline.

Personality	P@1	P@2	P@3
John Lennon	40%	60%	80%
Albert Einstein	30%	60%	80%
Napoleon	30%	40%	40%
Charlie Chaplin	42%	57%	57%
Adolf Hitler	50%	83%	83%

Table 3: The percentage of correct answers for various personalities in case of the first (P@1), the first two answers (P@2) and the first three answers (P@3).

From the tests done, it is noticeable that all the answers took less than 700 ms. Here we present several examples from a discussion with the conversational agent using the answer sentence selection method. The chat-bot impersonates John Lennon:

Q: What was your debut album?
A: My emotional debut solo album, John Lennon/Plastic Ono Band (1970), was received with high praise.
Q: Who shot you?
A: At around 10:50 pm on 8 December 1980, as me and Ono returned to their New York apartment in the Dakota, Mark David Chapman shot me in the back four times at the entrance to the building.

CONCLUSIONS

This paper presented the two methods used together to implement a conversational agent that models a historical figure: the first method of generating ChatScript files using ontologies extracted from DBpedia and the fallback method using answer sentence selection with textual matching. The advantage of implementing a chat-bot that answers trivia questions from the perspective of a historical personality is that for each person, there is a small amount of information to be processed. In addition, the questions are easy to predict, unlike questions in a general purpose open-domain question answering system.

ACKNOWLEDGMENTS

This work has been partly funded by the Sectorial Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of European Funds through the Financial Agreements POSDRU/159/1.5/S/132397 and by POSDRU/155420 – PROSCIENCE.

REFERENCES

1. A. Echihabi, U. Hermjakob, E. Hovy, D. Marcu, E. Melz, and D. Ravichandran, "How to Select an Answer String?," *Adv. open domain Quest. answering*, pp. 383–406, 2006.

2. D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, and C. Welty, "Building Watson: An Overview of the DeepQA Project," AI Magazine, vol. 31, no. 3. pp. 59–79, 2010.

3. F. Giunchiglia, M. Yatskevich, and P. Shvaiko, "Semantic matching: Algorithms and implementation," *J. Data Semant.*, 2007.

4. F. Giunchiglia and P. Shvaiko, "Semantic Matching," *Knowl. Eng. Rev. J.*, vol. 18, no. 3, pp. 265–280, 2004.

5. T. R. Gruber, "A translation approach to portable ontology specifications," *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, 1993.

6. B. Heller, M. Procter, and D. Mah, "Freudbot: An investigation of chatbot technology in distance education," *Proc. World Conf. Educ. Multimedia, Hypermedia Telecommun.*, pp. 3913–3918, 2005.

7. C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard and David McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit", *ACL 2014*, 2014.

8. M. Mehta and A. Corradini, "Developing a Conversational Agent Using Ontologies," *HCI'07 Proc. 12th Int. Conf. Human-computer Interact. Intell. multimodal Interact. Environ.*, pp. 154–164, 2007.

9. K. Taylor and S. Moore, "Adding question answering to an e-tutor for programming languages," in *Applications and Innovations in Intelligent Systems XIV*, 2007, pp. 193–206.

10. A. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

11. R. Wallace, "The elements of AIML style," Alice AI Found., 2003.

12. J. Weizenbaum, "ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1966.

13. B. Wilcox, "Beyond Façade: Pattern Matching for Natural Language Applications ", Telltale Games, 2011.

Disruptive Technologies – UV Protect – Smart Watch Application

Andrei-Bogdan Baran
"Alexandru Ioan Cuza" University,
Faculty of Computer Science
General Berthelot, No. 16
andrei.baran@info.uaic.ro

Adrian Iftene
"Alexandru Ioan Cuza" University,
Faculty of Computer Science
General Berthelot, No. 16
adiftene@info.uaic.ro

ABSTRACT
Health care industry is one of the most interesting domains where we can bring innovative and smart technologies that can have a big positive impact on our health. These days people are starting to be more and more concerned about their health and they are interested in having small wearable devices that allow them to be able to get real time data through different sensors regarding the environment or different activities they are involved in.

In this paper we will show how new disruptive technologies can be used to monitored the level of the ultraviolet radiation and then how the user can be notified about that.

Author Keywords
Disruptive Technologies; Smart Watch; Samsung Gear S; Ultraviolet Radiation

ACM Classification Keywords
H5.2. Information interfaces and presentation; J3. Life and Medical Sciences – Medical Information Systems

INTRODUCTION
The first step in protecting against any factor that could have a negative impact on our health is prevention. This paper will focus on the impact of UV on our health and how we can prevent different health conditions caused by excessive UV exposure using the smart watch app UV Protect.

Exposure to the ultraviolet (UV) radiation is one of people's many concerns, especially for those from Nordic countries, because the UV radiation can still harm you even if it is not sunny or warm outside. There is no link between UV levels and the temperature, because many other factors could affect the amount of UV radiation that reaches the Earth such as ozone layer, latitude, altitude, time of day, cloud cover or air pollution. Excessive UV exposure can cause eye damage, premature skin aging, sunburn or skin cancer. It is important to understand the risk we expose ourselves and to take action in order to protect our body.

UV Protect is a standalone application that runs on Samsung Gear S smart watch and it helps user to protect against UV exposure by automatically detecting the ultraviolet levels variations and by informing him about the potentials risk that he is exposed.

STATE-OF-THE-ART
The idea is not new to smartphones. There are applications that can indicate the UV index level, predict how long we can stay in the sun before burning, etc. such as Wolfram Sun Exposure Reference App [1].

There are also some custom build in wearable devices such as band wrist (UveBand) [2] or Violet [3] gadget that are sync with a mobile phone application in order to monitor the user real-time UV exposure.

But the smart watches have not any complex application that could offer us a variety of features such as automatic detection of UV level changes that are harmfully to our body, offer detailed information about past records, offer help and custom advices based on the UV level detected.

For example, the integrated application S Health from Gear S has a small application that indicates the UV index and keeps the most recent records. But this application requires user interaction in order to find out if he is at risk of sun burn or not.

In what follows, we present an application created by us, UV Protect, which will have a considerable positive impact on our health by informing us about potential excessive UV exposure.

The main advantages of this app comparing with the existing ones are: a standalone app that does not require another device like a mobile phone, which is used in similar apps, to do more complex actions. All the data is collected and processed using the smart watch where the UV Protect is installed. Another important feature comparing with other existing apps is the automatic mode that can be enabled just with two taps. Using this feature the app can automatically reads UV data from time to time, having configured a recurrence in minutes and a threshold UV Index value that indicates the minimum exposure the user wants. The app will notify the user via short wrist vibrations if the maximum value was reached and will display proper advice. Moreover, this app also can run offline, because it does not require an explicit Internet connection to store the data. The app can store the data locally and when it has an Internet connection, all the data from the queue will be synchronized. More details about these features and more will be explained in the next section.

SYSTEM ARCHITECTURE

UV Protect is a standalone application compatible with the wearable devices that have Tizen Wearable profile of the Tizen OS [4] and have support for UV and light sensors, such as Gear S or above.

Tizen is an open-source operation system that could run on different devices area like mobile, TV, In-Vehicle-Infotainment and wearable.

The Tizen Wearable architecture, offers a web runtime engine that allows Web application to run outside the browser. The Web Runtime provides the following Tizen Web APIs to web application: W3C/ HTML5 APIs and Tizen Device API [5].

W3C/HTML5 offers various features that you can use in creating Tizen Web applications such as: DOM, HTML5, communication, storage, security, performance and optimization, location and UI. Tizen Device API based on JavaScript provides advanced access to the device's platform capabilities such as sensors data, human activity monitor and other system information.

At the application level, UV Protect application consists from 2 main modules: user interface (UI) module and services module.

The UI module is responsible for offering a rich user experience by displaying the ultraviolet information gathered from UV sensor with a simple tap or two, using vivid graphics, rich text and flexible content layouts. This module offers pages for displaying the current info about ultraviolet index, pages for displaying history data in charts with different filters and pages for user profile and different settings. Also it provides notification alerts that inform the user in case the sensors detect UV radiation levels that are higher than the threshold set by the user (see Figure 1).

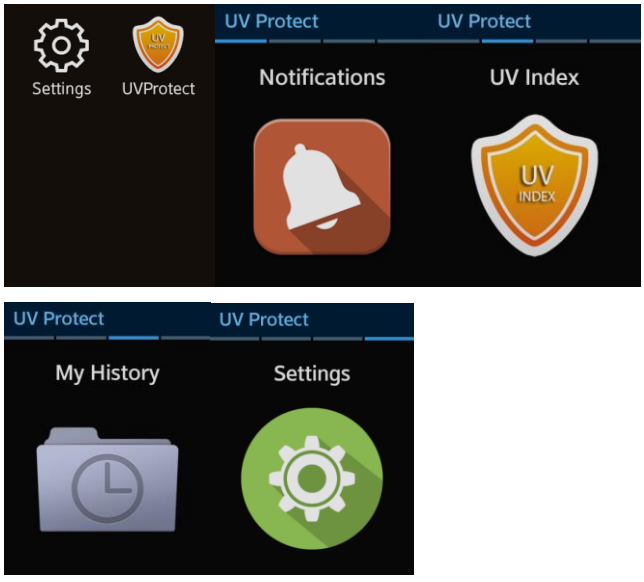

Figure 1. The Main UV Protect App features: Notifications, UV Index, My History and Settings.

The services module has 4 main other subservices: one for listening for UV sensors and retrieving the data, one for sending the data to a remote REST API for storing the details, one for synchronization, in case no Internet connection is available at the moment when the service reads data sensor and one for sending notification by wakening the app and showing the details.

The listener service reads data from the ultraviolet sensor that is integrated in the smart watch and it compares them with a well know index level. According to the United States Environmental Protection Agency [6] the UV index scale is from 1 to 11 (see Figure2).

Figure 2. The UV Index scale used in the United States conforms to international guidelines for UVI reporting established by the World Health Organization[1].

On this scale there are 5 levels and according to these, UV Protect automatically send notifications on the watch and in the same time the application informs the user how to protect him via custom recommendations based on which interval the UV index is.

The user can customize the notification settings as you can see in Figure 3, by setting a recurrence in minutes and a new sensor read will occur repetitive at that recurrence. Also the user has the ability to set a threshold UV Index value that indicates the minimum value witch will activate the alerts and the notifications. Moreover if the user activates the Vibration, the notification will be followed by a wrist vibration. In this way the app will have more success in getting the user attention.

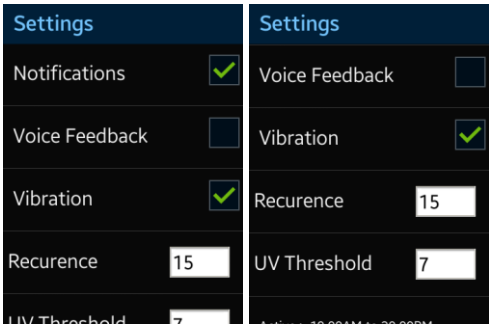

Figure 3. Settings panel.

After the notification service has been called and the user got notified, the data is stored locally if no Internet connection exists or it is uploaded to a remote Restful REST API [7] service that store all user records. The data stored locally is synchronized when an Internet connection is available via a transaction.

[1] World Health Organization: http://www.who.int/uv/en/

Also the app has integrated a history feature that allow the users to see the maximum UV Index values per days, starting with a day selected through a UI control. More details will be shown in Case Study 2.

The application is written using the Tizen IDE for Wearable and the technologies used are: JavaScript, HTML5 and CSS and Tizen Advanced UI (TAU) Framework that provides tools, such as UI components, events, effects and animations for wearable app development.

The Restful API is writing in C# using the ASP.NET Web API 2 [8] and the resources follows the HAL (Hypertext Application Language) [9] format that helps keeping an easy way to hyperlink between resources.

CASE STUDIES

In the next section we will present 2 case studies regarding the usage of the application.

Case Study 1

Let's say that a family with kids decided to take a small walk in the park, but it's almost 11:15 AM. One of the parents wants to check the UV radiation level and he uses his smart watch and the UV Protect App to detect if there is any possibility that his kids are under UV exposure risk.

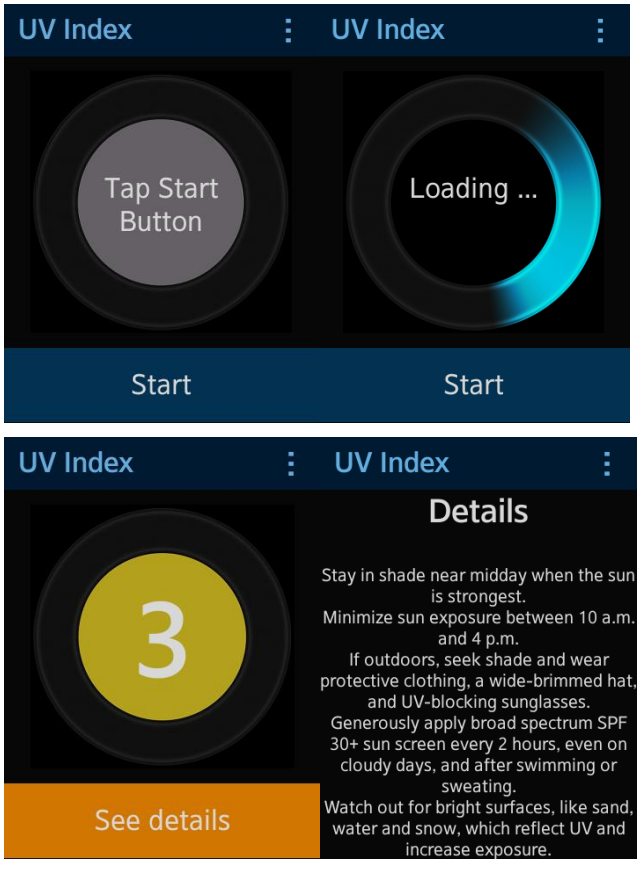

Figure 4. An example on using the application manually (the user has to do all the process).

How we can see in Figure 4, the user receives a list of recommendations.

Case Study 2

Let's say a person decides to take a walk on the beach at 1:00 PM, because it was in and he configures his UV Protect App to read the UV sensor data at every 1 minute, with a threshold of 4. After he finishes the configuration, he could enjoy the beach and the warm day, knowing that he does not need to worry about getting a sun burn, without knowing it.

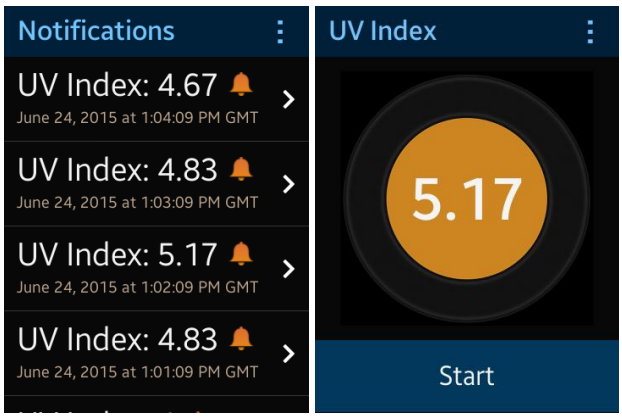

Figure 5. An example of using the application notifications.

The application has done its job and after a while his watch recorded the UV radiation levels and started vibrating, letting him know that UV exposure is too high and it is recommend seeking shade and wearing protective clothing, a wide-brimmed hat, and UV-blocking sunglasses. Also it suggested him to generously apply broad spectrum SPF 30+ sunscreen every 2 hours, even on cloudy days, and after swimming or sweating.

Moreover, the user has access to his history logs, as you can see in Figure 5.

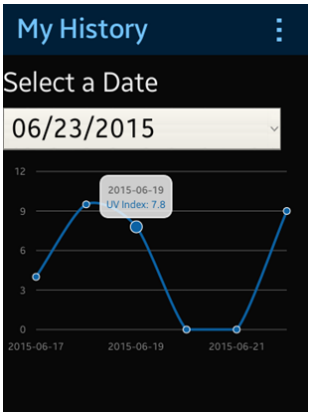

Figure 6. My History chart

The chart shows the maximum values per day, in the last 7 days from the day selected. The line chart is first computed with local data, if exists, otherwise the data is fetched from the REST API. This graph is usefully because it can show

the user relevant information about how the UV levels changed during a specific period of time and what were the maximum UV Index values on the period selected.

EVALUATION

The amount of UV radiation can be affected by different factors. According to the Australian Government – Bureau of Meteorology [10], one of the main important factor still remain the ozone layer, because lower ozone values in the atmosphere can increase the UV levels. For example Australia is more exposed to UV radiation because it is located closer to the ozone hole over the Antarctic and this means much higher and harmful radiation get through the ground level. In addition, Australian people have the highest rates of skin cancer in the world. Every year, around 1.200 Australians die because of different types of skin diseases even if this can be prevented with the proper actions. Another factors that have a considerably impact on the UV radiation levels are: latitude, time of the day, time of the year (season), altitude, cloud cover, air pollution and land cover. In the following line there is some brief info about the factors that we mentioned.

Up to 50% of the UV radiation is received around midday when the sun is at the highest point, between 11:00AM and 2:00PM. The angle of the sun influence the amount of UV radiation because solar energy must travel a greater distance through atmosphere when the sun is low in the sky and the UV may be absorbed by water vapors or other atmospheric components. Also the air is thinner and cleaner on higher altitudes; therefore more UV radiation reaches on top of the mountains. For example on altitudes near 2000m, the areas will receive up to 25% more than locations situated at the sea level.

Moreover the clouds cover and air polluting such as urban smog can influence the UV radiation levels by absorbing or reflecting back towards space. Also different surfaces can reflect the UV and have a big impact on our health. For example the snow reflects up to 88%, sea surf: 25 to 30% or sand: 7 to 18%.

In the following lines we will present some statistics regarding different scenarios where we tested the UV Protect App.

On a day between 11:00AM and 12:00AM, having the temperature 17°C and with high cloud coverage and small precipitation, being exposes to direct natural light the values recorded were very small, between 0 and 1.22 out of 15 (this is the maximum value that the UV sensor can read). In this interval we measured the UV levels in a public transport, but the index was 0, because the glass from the public transport reflected any UV radiation that existed at that moment on a direct light contact.

On the next day, between 1:32PM and 2:00PM, with a temperature of 24°C, with a clear sky with some clouds, on direct sun light the UV levels where between 5 and 5.84. Under an umbrella or a tree, the UV index decreased down to 0.61. Also the same test we conducted under a bridge and at a train station and the registered values were between 0.22 and 0.44. Inside the building, no matter the day time, if no direct sun light was preset the values always were 0.

Between two reads, at the same moment of the day and same conditions, the values could different with a maximum of 0.5. This is due to the position of the hand; on witch hand (right or left) do you wear the watch etc.

CONCLUSION

UV Protect is an application for everyone that wants to be informed and avoid harmful exposure to UV radiation. It is an application that is easy to use by any person, due to its user orientated interface. It offers different features such as: UV Index with general recommendations to protect against excessive UV exposure, notifications and history logs.

This application could help you to avoid unnecessary sun burn and to prevent different skin conditions by keeping you away from too much UV radiation.

ACKNOWLEDGMENTS

The research presented in this paper was funded by the project MUCKE (Multimedia and User Credibility Knowledge Extraction), number 2, CHIST-ERA/ 01.10.2012.

REFERENCES

1. Wolfram Sun Exposure Reference App https://itunes.apple.com/us/app/wolfram-sun-exposure-reference/id452253352?mt=8&ign-mpt=uo%3D4
2. UveBand http://www.uveband.com/
3. Violet http://www.liveultrahealthy.com/
4. Tizen https://www.tizen.org/
5. Tizen Web Runtime API https://developer.tizen.org/dev-guide/2.2.1/org.tizen.web.appprogramming/html/basics_tizen_programming/web_app_runtime.htm
6. United States Environmental Protection Agency http://www2.epa.gov/sunwise/uv-index-scale
7. RESTFull Web Service http://www.restapitutorial.com/lessons/whatisrest.html
8. ASP.NET Web API2 http://www.asp.net/
9. Hypertext Application Language http://stateless.co/hal_specification.html
10. Australian Government – Bureau of Meteorology http://www.bom.gov.au/uv/faq.sht

Introducing Basic Geometric Shapes to Visually Impaired People Using a Mobile App

Bogdan Troanca
University "Lucian Blaga" of Sibiu
Faculty of Engineering
Victoriei Boulevard 10, Sibiu 55002, Romania
troancab@gmail.com

Alexandru Butean
University Politehnica of Bucharest,
Faculty of Automatic Control and Computer Science
Splaiul Independentei 313, Bucharest 060042, Romania
alexandru@butean.com

Alin Moldoveanu
University Politehnica of Bucharest,
Faculty of Automatic Control and Computer Science
Splaiul Independentei 313, Bucharest 060042, Romania
alin.moldoveanu@cs.pub.ro

Oana Balan
University Politehnica of Bucharest,
Faculty of Automatic Control and Computer Science
Splaiul Independentei 313, Bucharest 060042, Romania
oanab_2005@yahoo.com

ABSTRACT
The problem of blindness and other eye disabilities is highly important as it affects millions of people on the globe. In the same time, the number of modern touchscreen devices is increasing fast. They are widely spread and becoming accessible for education purposes. For those who suffer from congenital visual disabilities is very difficult to imagine the world as it is and the struggle begins from the early school stages and continues for the rest of their lives. Modern touchscreen devices, also called smart devices, represent a viable solution for people with sight problems due to their rich multimedia experience. Thus, they can be used as assistive devices. This paper describes how learning basic geometric shapes using a smartphone or a tablet can improve several important aspects like: general feeling of using touchscreen devices, perception of a basic figure's shape by simply touching the screen, intuition on how to follow a GPS map. This methodology represents only the initial stage from a long-term learning process that aims to help visually impaired people to understand and use the power of smart devices in order to perceive the world around them as normal people do.

Author Keywords
Sound; vibration; mobile devices; visually impaired;

ACM Classification Keywords
H.5.m. Information interfaces and presentation - Human-Computer Interaction, Miscellaneous. ; K.4.2. Computers and Society: Social Issues - Assistive technologies for persons with disabilities.

General Terms
Human Factors; Design; Measurement

INTRODUCTION
Around the world there are over 285 million people with sight problems, 39 million being completely blind and 246 with low vision. [14]. Visual impairment is unequally distributed across age groups. About 82% of all people who are blind are 50 years or older, although they represent only 19% of the world's population. Due to the expected number of years lived in blindness (blind years), childhood blindness remains a significant problem, with an estimated 1.4 million blind children below age 15 [16].

The market trends [17] reveal that modern devices like smartphones and tablets are becoming cheaper and thus accessible to everyone. In schools these devices are becoming popular and modern learning tools [5] are being developed using modern technology advantages.

Throughout our research we tried to understand how blind people describe, understand and learn basic geometric shapes as well as how they relate these shapes with the surrounding environment. As a proof of concept, an Android app was developed to guide them in the process of learning 2 basic geometric shapes using a touchscreen device.

HOW THE BRAIN OF BLIND PEOPLE WORKS
A very hard thing to imagine for people with no sight problems is what visually impaired people [14] think about geometric shapes. Because of the fact that they never saw how a line or a curve looks like, they simply cannot imagine how these shapes actually are.

While normal people have this major disadvantage, blind people have other important abilities in their favor. These

capabilities cannot be developed by people without visual disabilities, no matter how hard they try. Based on the brain's capabilities to rewire and distribute resources from affected areas, the sensors' migration from vision towards touch and hear, balances the scale and importance of these senses [4] [10]. Therefore, the brain area responsible for touch and hearings performs way better. In order to benefit from their unique aptness and obtain maximum results, the entire learning process [11] of visually impaired people should focus mainly on sound and haptics.

For people with low vison suffering from various eyesight problems, there are different methods that can be used to help and guide them. These methods are based on the contrast between light and dark [13], which in their case can still be perceived using their eyes. This is the reason why all geometric shapes developed for the experimental system are programmed using this kind of color contrast.

HOW BLIND PEOPLE LEARN THE GEOMETRIC SHAPES IN A TRADITIONAL WAY

Figure 1: Traditional 3D shapes [2]

Despite the fact that we live in a modern era full of computers, mobile devices and industrial robots that can do almost anything, the learning process is still based on old fashioned techniques that are still great. These techniques are advancing very slowly because of the investments required in this field. People with good sight are still easy to educate, thanks to the use of books or computers dedicated to help them read information and perform tasks. On the other hand, visually impaired people cannot use these methods; therefore other procedures must to be developed. They learn Braille [3] for reading, use audio books for gathering information easily and the most important, they develop the tactile sense, through which they can feel and learn shapes, objects and much more.

Regarding shapes, the recommended method for learning them is by guiding the person, using the tactile sense, which is extremely good developed in the case of educated people with sight problems. In the learning process, they use 3D shapes, as shown in Figure 1. This way, they try to imagine how the shape looks like. Another way to learn basic geometric shapes (ex. rectangle, circle, and triangle) is using plain paper, having the edges emphasized. This way, the edges can be easily distinguished. Regardless of the learning method, the problem of learning more complex shapes remains. Not all objects can be reduces in size so that visually impaired people can perceive them. In this case, they learn the objects based on combination of basic shapes. For example a house is a triangle placed over a square, or a pyramid over a cube. But learning the exact design of a house, or other complex object is impossible. They can only understand a small part from an elaborate and rather complicated design.

THE IMPORTANCE OF HEARING AND TACTILE FEEDBACK WHEN USING A MOBILE DEVICE

The technological progress is a big advantage for all people. The fact that medicine evolves, putting at everyone's disposal complex assistive devices, represents a big step forward for visually impaired people. In some cases, however, these people can face difficult problems while trying to adapt to touchscreen mobile devices. The challenge is to get used to phones and tablets that do not have any buttons. The buttons help them coordinate and do simple tasks, such as calling or sending a message. Although modern devices are developed in order to perform complicated actions in a smart, simplistic way, they are not yet fully adapted to blind people necessities; therefore they cannot easily use all the important features [4].

The most important aspect that blind people rely on when using a device is the feedback or notification sent by the device. Thanks to the hearing and tactile abilities that visually impaired people can achieve, when the accessibility option is activated [1], the device becomes more user-friendly. Depending on the action performed by the user, sounds and vibrations will be enabled when the user presses a specific element. For fulfilling his purpose, the element needs to be long pressed or double tapped, depending on how the system is implemented. The vibration is controlled by the vibration motor inside the device, which can be controlled by the programmers during the app development stage. The motor can vibrate different periods of time having different intensities or it can follow a specified vibration pattern.

THE IMPORTANCE OF LEARNING SHAPES USING A DEVICE WITH TOUCHSCREEN

In order to benefit from the technologic development, the devices need to be adapted to the visually impaired people's necessities. The results depend also on the effort these persons make in order to learn how the devices work. A good starting point is learning the basic geometric shapes. The same way children learn how to draw lines, so that they can write letters and numbers afterwards, blind people have

to learn how to draw lines and basic shapes in order to further develop other important abilities. Learning these shapes on a touchscreen will help the user track more complex objects. An important advantage of this learning method is that users are able even to understand a map through tactile and audio guidance. This way, people with sight problems will analyze the basic geometric shapes on the screen and will rapidly identify the route they have to follow. This solution could be a great add-on for the map application [9] for blind people. Figure 2 exemplifies this situation.

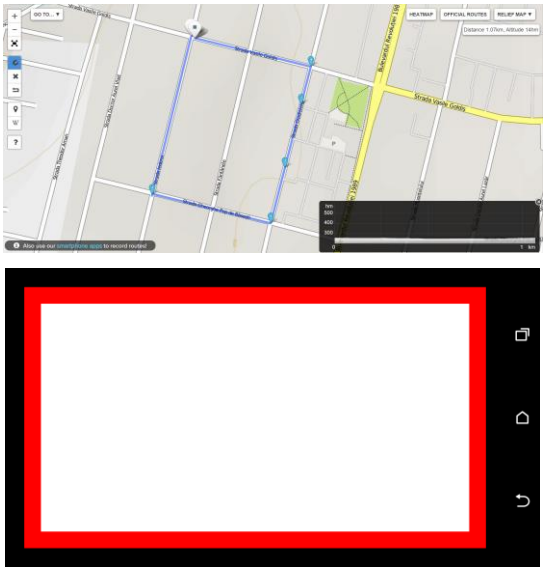

Figure 2: Understanding a map

BASIC GEOMETRIC SHAPES

In this application, there are two different basic geometry shapes that need to be learned by the visually impaired people. These are the rectangle and the circle. For successfully learning these shapes, the activities were implemented in such way, that they will always stay in full screen. This way, the user will not be able to touch the notification bar, which would interrupt the normal interaction. The application was programmed so that at the beginning of each activity there will be a HELP button as large as the screen. This button will force the user to listen to the instructions. The device must be held all time in landscape mode, if not, pressing the HELP button will tell the user to turn it on. The HELP button can be observed in Figure 3, together with the message that the device is held properly. By tapping the screen, the user will hear what to do in the activity. To close the HELP button the user must do a long press action on the screen.

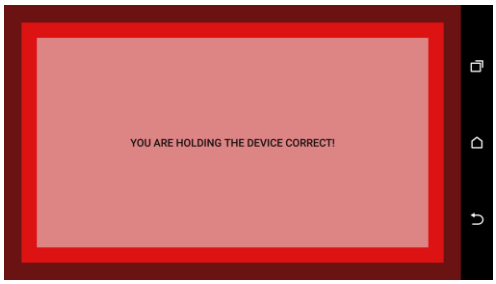

Figure 3: Help button

The activity layout [8] was created so that both geometric figures are drawn using white color on a black background, and having a red outline. These colors were used in programming the entire application because this color contrast can easily be detected by people with sight problems. In order to listen to all the instructions from HELP option as well as the rest of the system, Google Text-to-Speech engine [7] has been used. Another important feedback method used here is the vibration. Thanks to the vibration motor inside the device, the user will feel different vibration patterns depending on the activity and the buttons he presses.

Both activities are being chosen after selecting the difficulty level the user wants.

RECTANGLE ACTIVITY

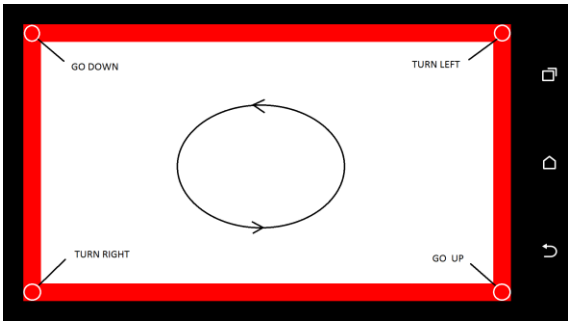

Figure 4: Rectangle Activity

In Figure 4 there is the rectangle activity. The chosen level in this picture is beginner, so that the vibrating outline will be wider and easier to touch. The user is guided by the HELP option to find the vibrating line and to follow it counterclockwise. Touching the line will make the device vibrate. Once the user touches the inner section of the rectangle, the vibration will stop, so that the operator knows he made a mistake and he must find back the vibrating outline of the rectangle. If the shape is drawn correctly, each time a corner is touched, the device will indicate the next direction the user must followed. A feedback will be given when the user completes the entire shape.

CIRCLE ACTIVITY

In figure 5, there is the circle activity with medium level. The vibrating line will be smaller this time, being harder for the user to follow it. This means he already knows how a circle looks like, but needs to practice. Best practice would be to follow this line until perfecting the technique.

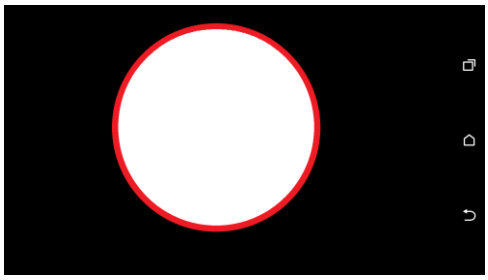

Figure 5: Circle Activity

CONCLUSION

A tool like this could be successfully used in the learning process of blind children from specialized schools. By learning these shapes, they can understand how different objects look like and in the same time will also start to get familiar with touchscreen devices.

By learning geometric shapes on a touchscreen device, visually impaired people could even start to understand and use the routes from different standard mobile mapping systems, like Google Maps [6].

The circle and the rectangle activities represent only the first phase of a comprehensive learning process that uses the power of mobile devices. After several successful testing sessions with blind people volunteers we plan to take the user experience to the next level. The first step will be to implement vibrations, audio guides and menus for other basic shapes like triangle, star, and serpent line.

After the subjects will start to adapt and perceive the shapes faster and faster, a good approach would be to combine basic shapes with more complex shapes (ex: house - rectangle and triangle). If the study goes well on 2D shapes we can try to help them perceive 3D shapes [12].

In this early stages of the research we implemented the proof of concept app using Android devices, but the study goes on and we will build a similar model also for iOS.

ACKNOWLEDGMENTS

This work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134398 and POSDRU/159/1.5/S/132395.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

The results presented in this research study were possible due to the help of National Association of the Blind People in Romania / subsidiary Sibiu and other subjects who have offered to be volunteers for testing and feedback sessions.

REFERENCES

1. Accessibility Tools on Android Devices SDK. https://goo.gl/x2NWkH
2. American Printing House for the Blind, Inc. http://goo.gl/59wYZ7
3. Braille. https://goo.gl/ItLKfc
4. Butean A., Bălan O., Moldoveanu A., Moldoveanu F., Touchscreen Based Audio and Vibro-Tactile Applications as Assistive Systems for People Suffering from Eye Disorders, ICSV22, 2015
5. Butean A., Moldoveanu A., Morar A., From Classic Math School Books To Interactive Gamified Elearning, eLSE, Vol. 1, ISSN: 2343–7669, ISSN–L: 2066–026X, 2015
6. Google Maps Navigation. https://goo.gl/HzQkrN
7. Google Text-to-Speech. https://goo.gl/TwE0Id
8. Layouts in Android. http://goo.gl/cso9HO
9. Nisha K.K., Pruthvi H.R., Hadimani S. N., Reddy Mohana G.R., Ashwin T.S., Domanal S., An Android GPS-Based Navigation Application For Blind, VINCI'14, pp. 240, 2014.
10. O'Connor A., Really? The Brain Gets rewired if one of the Senses Is lost, Well - New York Times, http://goo.gl/TdIcPq
11. Suggested Teaching Strategies: blind and visually impaired students – Newcastle University. http://goo.gl/SBi8fM
12. Salamin P., Thalmann D., Vexo F., Visualization Learning for Visually Impaired People, Technologies for E-Learning and Digital Entertainment , Volume 4469, 2007, pp 171-181
13. Tips for Making Print More Readable. http://goo.gl/BftSs5
14. Visual impairment. https://goo.gl/4iHcDT
15. Visual impairment and blindness. http://goo.gl/1eHNlY
16. Worldwide Smartphone Average Selling Price Will Plunge To $265 By 2017 : IDC http://goo.gl/IapfvR
17. John Jill Technology - Blindness and Visual Impairment: http://goo.gl/GW88au

Indoor Localization and Navigation Using Phone Sensors and a 3D Model of the Building

Szabolcs Orban
Computer Science Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
orbanszabolcs11@gmail.com

Teodor Stefanut
Computer Science Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
teodor.stefanut@cs.utcluj.ro

ABSTRACT
This article presents an innovative approach in offering indoor navigation assistance based on mobile device sensors and a 3D model of the building. We will discuss the three main stages that are required in order to implement the proposed solution: (1) creation of the 3D model, (2) definition of the navigation attributes (built to also support localization) and (3) implementation of all the required functionalities in an Android application. One of our main concerns is related to the identification of the current location of the user in an indoor environment efficiently using the device's sensors. For this purpose we will describe an algorithm that relies on the accelerometer and the compass to provide accurate location information and to simulate the movement of the user into the environment, without consuming too much energy.

Author Keywords
indoor localization, mobile device sensors, indoor navigation, virtual representation

ACM Classification Keywords
D.2.2 Design Tools and Techniques: User interfaces; J.m Computer Applications: Miscellaneous; H.5.2 User Interfaces: Interaction styles.

INTRODUCTION
In this paper we propose a solution to provide helpful user guidance through different faculty buildings. Indoor localization based on smart phones is a popular research field nowadays because of the popularity and advanced functionalities of these devices. There are three main approaches that can be used to establish user location: based on GPS (Global Positioning System), based on RSS (Received Signal Strength - the attributes of the WiFi signal received by the phone) or based on the use of data provided by other device sensors (accelerometer, gyroscope, compass, etc.).

The most famous localization technology is the Global Positioning System (GPS). Its main problem is that it requires a high amount of energy to function and drains the phone's battery rapidly. In addition, GPS signals are transmitted in a frequency that cannot easily penetrate walls and other barriers, so it can hardly be used indoor.

Another popular approach, much more energy efficient than the previous one, refers to the use of Received Signal Strength (RSS) of the WiFi signal, which can be used for location estimation based on the previously known location of the access point. This is currently considered to be a valid metric for location calculation and the vast majority of mobile devices support and implement it. The drawback of this method is that its accuracy is not good enough for guidance inside a building, and that is why it is considered to be unreliable for our purposes [10].

The newest method is related to the new internal sensors which were integrated to the current smart phones hardware. Namely, they are the *accelerometer* and the *compass*. The accelerometer measures the current acceleration of a device compared to free-fall acceleration. A compass is a sensor that indicates the north direction by measuring the three axes computed from the magnetic field strength. There are a lot of applications that use these sensors. For example, smart phone games use the accelerometer to detect the angular movement (tilting) of the device.

In his article, Moustafa Youssef [6] suggested that a good way to compute the location of the device relative to a previously known point is to use the functionalities of the accelerometer and the compass. These sensors work well using the device's resources efficiently (e.g. battery).

In previous research the accelerometer has also been used to determine human position in an environment [2]. The possible positions were classified in three main categories: walking, sitting or standing. By also using the location, the system can determine whether the person was sitting at a pub or walking in the park.

Other systems use accelerometers to detect the number of human steps. For example the pedometers are devices that count each step a person takes by detecting the motion of their hands or their hips. These devices are widely used by people in order to motivate themselves into exercising routines or just to increase their physical activity [9]. Pedometer based applications count human steps using different movement patterns and thus are able to determine the distance traveled. The information registered is later matched with the direction obtained from compass data.

In a similar way, our proposed approach aims to combine the data from the accelerometer and from the compass in order to define the movement of the mobile device in the indoor environment and to reflect these changes into our 3D building model. While we use the compass to establish the direction of motion the accelerometer allows us to estimate the distance traveled by the user on the detected direction.

RELATED WORK

Indoor Localization

Although GPS-based approaches can provide a good performance in outdoor scenarios, in the case of indoor localization the system will have some issues regarding the accuracy of the result.

In [3] Kaiqing Zhang described a positioning method based on WiFi RSS that enables the estimation of indoor location using smart phones. In order for this approach to work as expected, map information about the environment is also necessary for measurements and real-time indoor localization.

WiFi RSSI (Received Signal Strength Indicator) approach indicates the distance between the user and the access point to which the device is connected. The advantage of this approach is that WiFi technology is found almost everywhere and no additional infrastructure is needed for localization. There are two main approaches based on this technology: fingerprinting [4] and ranging-based methods [5]. The localization algorithms use combined location information from IMUs, maps and WiFi RSS. In order to improve the performance, the map is represented using area states and for each state a different localization approach is implemented.

3D Navigation

This is considered to be a new way of using a normal map, as 3D technology offers not only a representation of the real-world, but also gives users relevant information as advanced city models, 3D landmarks and icons or a digital elevation model [8].

In order to create a 3D Navigation model, there are three main aspects that need to be taken into consideration: to represent the navigation environment, to support path finding and to provide context information about the location. Each of these requirements can be fulfilled by using a special 3D model which contains both topology and semantics of the building. Next to the model created, the system also requires a representation of the topology of the building in the form of a network and information about how is this related to the geometrical model.

In order to handle the movement of a user inside a 3D model there are some problems that need to be taken into consideration. One solution proposed by W. Xua and M. Kruminaite [11] is to define a pattern so that the accelerometer measurements can be interpreted as movements. The solution also uses step detection methods to monitor how the acceleration value drops when the user is balancing the phone while walking. Also the turning can be determined by checking the abrupt change of the azimuth angle given by the compass. By using these types of sensors the location data will not be affected by the electrical appliances and magnetic materials from indoor locations.

There are no standard representation rules for defining all elements from an indoor environment, so the system needs to have its own rules. In other words the map should be represented using intuitive and easy to understand representation rules. The model should contain semantics that represent the navigation environment and the context information.

A good way to achieve this type of model is the method used by OpenStreetMap (OSM) [1] which is an open source geographical database. The structure of the network used consists of nodes and relationships (edges). Nodes are spatial components and edges represent the connections between different components (see Figure 1). Each relation can get an annotation and a label by using a key-value-pairs structure. The tags are used to define the elements from the map that are represented in the 3D model and provide a better understanding on how navigation is performed.

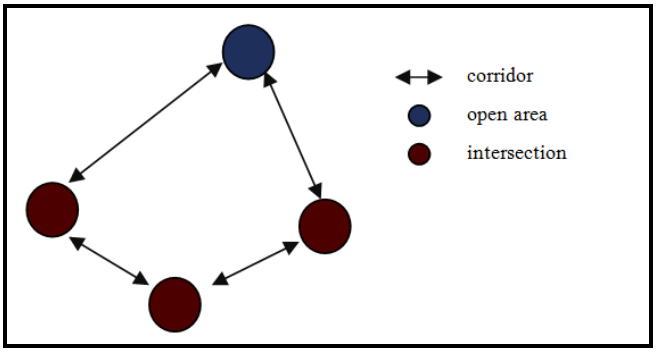

Figure 1: The graph representation of indoor elements.

Each node can contain different elements in order to be classified. For example an intersection point can be defined as a "door" or a "turning point". In indoor representations, these elements show an entrance to an open area or the link between two corridors. A corridor can have doors and walls defined within its elements [3].

There are four different actions taken into consideration depending on the above defined model:

1. *Walking in Corridor*: because the width of one corridor can be relatively small, the user moves in one dimension while walking in the corridor.
2. *Turning*: the user is at a turning point. When he is close to a turning point on the map, an abrupt change of the compass angle is detected.
3. *Verifying the turning*: it is possible that the chosen corridor may not be the right track.

4. *Walking in Open Area*: this action is determined by the fact that the user walks into a "Corridor door" that connects a door with a room.

The network is bound with the 3D geometrical model using different IDs for nodes stored in the object's attributes.

PROPOSED METHOD

The proposed application has the role of a guide in the faculty buildings. It allows any user that possesses an Android phone to search for different classrooms, administrative locations or laboratories.

As an initial step, the user must specify to the system his current location and the name of the classroom he wants to reach. The application displays the 3D model of the building marking the user's current position and an optimal path to the destination.

In order to provide accurate information and synchronize the virtual movement to the real one, the application has to detect the progress of the user and navigate him to the classroom. He/she has the possibility to move freely in the building while the application maps his position from the physical world to the virtual world model in real time.

3D Modeling and Loading the Model

Creating a 3D replica of a building is not a very simple process. First of all the exact blueprints of the structure are needed and must be loaded into a 3D modeling program. After the prints are loaded all the lines of the walls must be traced keeping gaps for the doors.

After the tracing of the walls on a 2D plane, the lines must be extruded to a specific height to form a 3D wall. The next step is to fill the upper side of the doors until the roof is reached to form a door-like gap that is smaller than the height of the wall.

In the case when we have to model more than one floor, except for the base of the upper and lower levels we also have to model the stairs connecting these or even an additional semi-level between the two floors. This part also needs to be modeled with the exact data from the blueprints.

Finally texture must be mapped onto the model and exported into a format which can be read by Android Studio.

Graph Path

One of the core functionalities of the application is to compute the shortest path from the current location of the user to the indicated destination. In contrast with the graph building strategy presented in [3], our approach does not classify the nodes of the graph depending on the type of the room. In the application there is no need to classify them, because there is no variety of rooms.

In order to compute the shortest path a predefined graph containing the coordinates of the rooms and some connector nodes are used. After the optimal path is computed the result is shown on the loaded 3D model as shown in Figure 2. Furthermore the user is free to walk wherever he/she pleases, as the current location on the graph is always computed to the closest node.

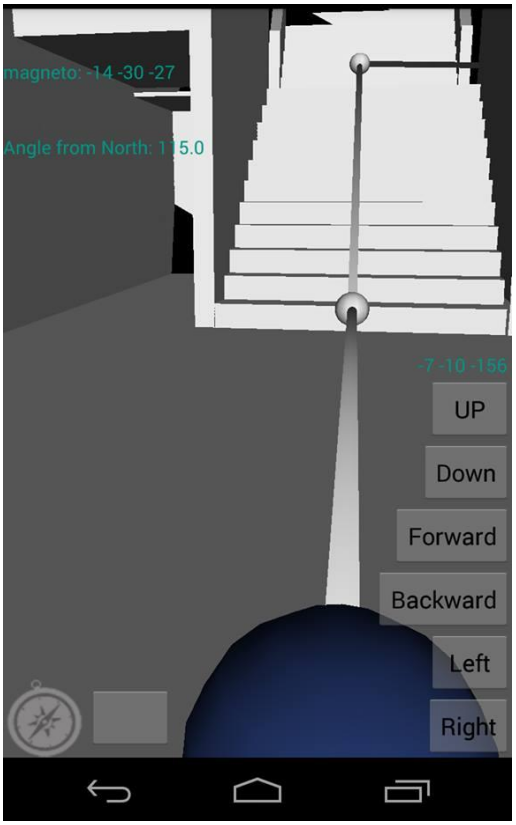

Figure 2: Example of the 3D model and the graph path

From time to time, the system may accumulate some positioning errors. These errors appear when the movements performed by the user in the real world are not correctly detected by the device. For correcting these errors the user can adjust his present location by jumping to the previous or to the next node and continuing the navigation from that position.

Identifying the correct location in the virtual model can be achieved by the user based on the similarity of the model to the real world. By "looking around" in both environments, the user should be able to identify the closest marker to his real position or to realize he/she leaved the path indicated by the system. A very important role in the accuracy of user decision is played by the photorealism of the 3D model.

Rotation Interaction

The main interaction types with a 3D object involve observations and rotations in all the available directions. On a smart phone this actions are usually achieved by swiping to a direction.

The physical screen has only two coordinates: X and Y (see Figure 3) while the system of the model has three coordinates: X, Y and Z (see Figure 4). The main challenge is to map the actions from the 2D screen to the 3D space.

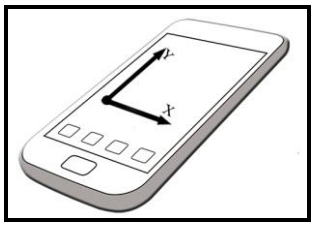

Figure 3: Coordinate system of the mobile screen

Figure 4 shows that the model can be rotated in three directions, but because of the representation constraints rotating the model around the Z axis is not recommended (as it would allow the user to turn the buildings upside down). Therefore for the rotation only the two virtual axes, X and Y are used. On the other side, the coordinate system on the screen of the phone is a Cartesian coordinate system: X is the horizontal and Y is the vertical coordinate (see Figure 3).

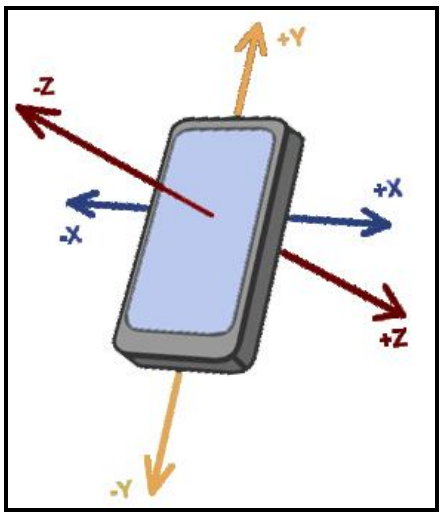

Figure 4: Coordinate system of the virtual model

Touch detection is checked and saved periodically. The mapping of the physical coordinates to the virtual ones has the formula:

$$R_X = R_X + (touched_Y - current_Y) \times K$$

$$R_Y = R_Y + (touched_X - current_X) \times K.$$

R_X and R_Y are the Euler rotation angles of the virtual X and Y axis of the model. The two float values $touched_X$ and $touched_Y$ are the coordinates of the point where the touch began, respectively the values $current_X$ and $current_Y$ compose the current position of the user's touch on the screen. The value K is a constant which is responsible for the correct use of scale.

Gyroscope Rotation

Another possible way of rotating the object is to use the built in gyroscope of the device. The gyroscope is a sensor that detects the rotation around the three axes. The value returned is linear to the change rate of the angles and is expressed in deg/s.

The degree can simply be computed from

$$degree = \frac{G_1 - G_0}{t_1 - t_0}.$$

The gyroscope value is G_1 at time t_1 and G_0 at time t_0. Finally the result can be added to the rotation properties of the object.

This feature can be used without swiping or touching the screen and allows the user to rotate and to look around. To correct the possible detection failures of the gyroscope, the user can easily adjust the angle with the swiping interaction.

Accelerometer

For computing the distance traveled by the user in the real environment, the use of the built in accelerometer is needed. This sensor detects the force directed to the opposite direction of the acceleration vectors.

To get the sum of the three vectors

$$V_X, V_Y, V_Z,$$

the following formula must be used:

$$V = \sqrt{{V_X}^2 + {V_Y}^2 + {V_Z}^2}.$$

If the device is situated in a motionless state the force of gravity still applies to the Z axis of the accelerometer. This force must be constantly decreased from that vector.

In Android there are two types of accelerometers already implemented. One is the TYPE_ACCELEROMETER, which is a hardware type sensor used for motion detection as tilt or shake of the device [7]. This is the regular sensor already presented. The other one is called TYPE_LINEAR_ACCELERATION, which is a hardware and software type sensor. This sensor returns the acceleration force applied to all the three axes excluding the force of gravity. It is used for monitoring the acceleration.

The second sensor is suitable for the application and therefore this one is chosen. In the analog world continuous math is used to determine the distance travelled which is:

$$Velocity = \int Acceleration$$

$$Distance = \int Velocity$$

In the digital world discrete math can be used and the integration becomes summation:

$$Velocity = \sum Acceleration$$

$$Distance = \sum Velocity$$

Step Detector

For illustrating the movement of the user in the virtual representation of the building, the implementation of a step detector is required. As mentioned before, this can be achieved using the accelerometer.

Each time the acceleration is triggered the detector decides from the force of the movement whether it was a step made by the user or just a hand movement. If a step is detected the application performs the translation of the 3D model according to the movement direction.

To eliminate the possible noises, a low-pass filter is applied to the detector. This hinders data from small movements of the hand. At the same time, a high-pass filter is also implemented to prevent shacking and sudden movements from being considered as steps.

Compass

In order to bind the movement of the user with the movement in the 3D model an orientation method is needed. A simple and efficient way to achieve this is to use the position of the North Pole associated with a fixed point in the 3D world. The orientation of the virtual user can be computed from the angle between the phone's orientation vector and the point of the North Pole.

Using the device's geomagnetic field sensor the changes of the earth's magnetic field can be monitored. This sensor returns the azimuth, pinch and roll values. For the compass to work only the azimuth value is needed. Azimuth indicates the rotation around the axis that points towards the center of the Earth.

Knowing the angle towards the North Pole and the distance travelled from the accelerometer's data, the device is able to synchronize the position of the user in real world with the position in the virtual 3D world, independent of the movement trajectory or the virtual looking-at direction.

Phone Holding Anomalies

The previously shown solutions work only if the user is holding his device straight vertical to the ground because of the orientation of the used vectors. In real usage scenarios however, the average user holds the phone oblique around the X axis as shown in Figure 4. Additional tilting of the phone during the use causes the device to register incorrect values along the Y vector, affecting the computed results.

For most of the scenarios, because the Y vector is not part of the main component of the straight movement, it can be usually ignored.

Stair Problem and Solution

Ignoring the Y vector does not lead to problems when walking on a single floor. However, it causes problems in case of going up the stairs. As the device cannot detect the vertical movement of the user, it cannot correctly represent the height position in the virtual world.

However, if the optimal path contains a staircase, then the application can access the 3D positions of the nodes before and after the stairs. A quick and easy solution for positioning the user correctly into the 3D model is to compute the height of the corresponding position on the stairs based on the two known coordinates and assign it to the height of the user's position.

The algorithm starts when the user has reached in his/her path a segment defined by two nodes that have different height values:

1. Get the closest point on the segment from the user's current position using only the X and Z values
2. Compute the height of the point as a linear interpolation between the segment's end points
3. Assign the height of the point to the height of the users position

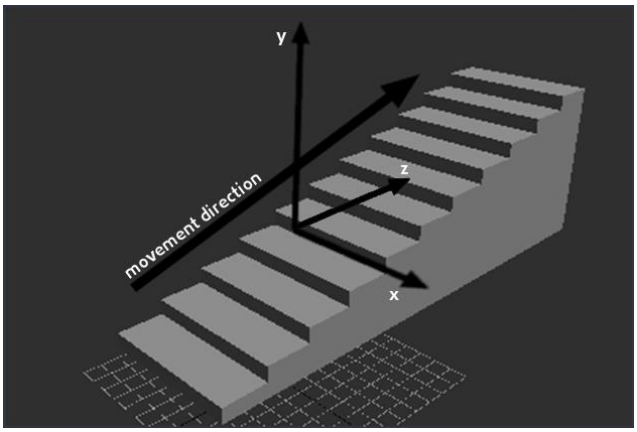

Figure 5: Computing the height position on the stairs

This way the positioning of the user in the virtual world always corresponds with the real position on the stairs.

USER SCENARIOS

At the beginning, the user must indicate his current location from a predefined list and the room number he seeks, as shown in Figure 6. This step is necessary due to the implemented localization approach, which (as mentioned in the previous sections) requires a previously known location for being able to correctly compute and map the user movement.

After retrieving the information from the user, the application will search for the optimal path between the two locations, based on the physical distance. Finally the user will be presented with the 3D model of the building and the highlighted route to be followed (see Figure 2). The initial position of the user into the virtual world is established at the predefined coordinates associated with the initial location selected.

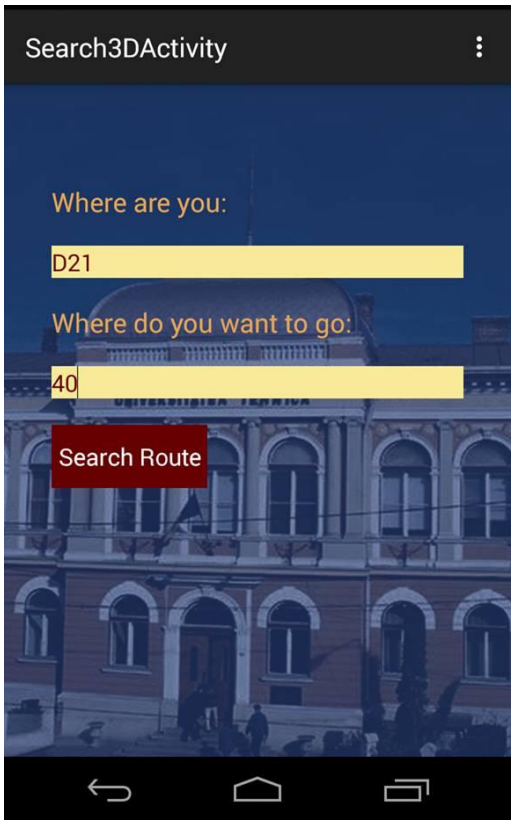

Figure 6: The initial step in finding the desired route

After the map is displayed, the user can interact with the virtual world through swipes and dedicated UI controls. The application functionalities are represented in an interface created using the graphic layout tool for Android so that anyone can easily use it. In Figure 7 the main flow of the application is represented graphically.

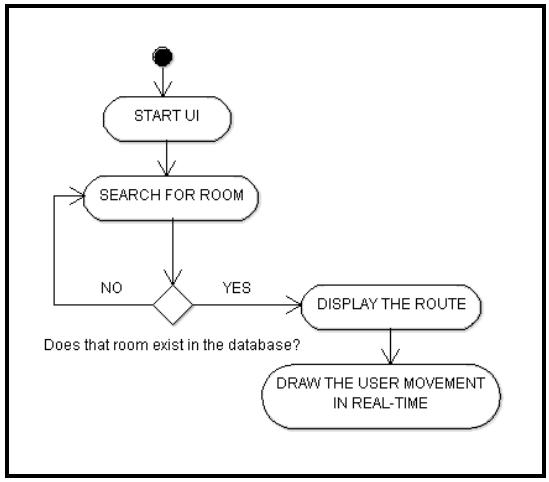

Figure 7: Main flow of events when searching a room

CONCLUSION

The result of our work is a stand-alone application that is used for real-time indoor localization and navigation based on a 3D model improved with map-like information.

Also, in order to get the location accurately and without consuming that much energy of our mobile device we implemented a method which uses only the phone sensors (accelerometer and compass) to determine the location and movement of the user. This is considered to be a new approach in handling indoor positioning.

Specific user interactions have been implemented to enable easy corrections of the computational errors related to the synchronization of the user location in the virtual and in the real world. The method by which the user interacts with the application is a good example of a prototype and can be used in future similar applications.

REFERENCES

1. *Indoor Mapping – OpenStreetMap Wiki*, Sept. 2014. [Online]. Available:

   http://wiki.openstreetmap.org/wiki/Indoor_Mapping

2. Jim Scarlett, *Enhancing the Performance of Pedometers Using a Single Accelerometer*
3. Kaiqing Zhang. Hong Hu. *Indoor Localization Algorithm For Smart phones*
4. P. Bahl and V. N. Padmanabhan, *RADAR: An in-building RF-based user location and tracking system*, in *IEEE Annu. Joint Conf. on Comput. and Commun. Soc.*, vol. 4, 2000, pp. 775–84
5. P. Bahl and V. N. Padmanabhan, *A comparative study on indoor localization based on RSSI measurement in wireless sensor network*, in *IEEE* Int. Joint Conf. on Comput. Sci. and Software Eng., Nakhon, Pathom, May 2011, pp. 1–6
6. Moustafa Youssef, Mohamed Amir Yosef, *GAC: Energy-Efficient Hybrid GPS-Accelerometer-Compass GSM Localization*
7. Sensors Overview, Android Developers

   http://developer.android.com/guide/topics/sensors/sensors_overview.html

8. Tomtom 3D Navigation

   http://automotive.tomtom.com/en/real-time-maps/3d-navigation

9. Tudor-Locke, *Taking Steps toward Increased Physical Activity: Using Pedometers To Measure and Motivate*
10. Ubejd Shala , Angel Rodriguez , *Indoor Positioning using Sensor-fusion in Android Devices*
11. W. Xua, M. Kruminaitea, B. Onrusta, *A 3d Model Based Indoor Navigation System For Hubei Provincial Museum*

An Approach for Detecting ID Frauds in a Traditional Voting System Using a Smartphone Stand

Paul Strimbeanu
University "Lucian Blaga" of Sibiu
Faculty of Engineering,
CS3E Student
Victoriei Boulevard 10,
Sibiu 55002, Romania
paulstrimbeanu@gmail.com

Alexandru Butean
University Politehnica of Bucharest,
Faculty of Automatic Control and Computer Science
Splaiul Independentei 313,
Bucharest 060042, Romania
alexandru@butean.com

Diana Butean
The KPI Institute
European Division
Marketing and Online Platforms Department
Harbour Esplanade, Suite 606,
Melbourne 198, Australia
diana@butean.com

Florica Moldoveanu
University Politehnica of Bucharest,
Faculty of Automatic Control and Computer Science
Splaiul Independentei 313,
Bucharest 060042, Romania
florica.moldoveanu@cs.pub.ro

ABSTRACT
All over the world, the voting systems are very important tools that contribute to the basic principles of democracy and equality. In every developing country, the problem of online voting becomes a necessity, therefore solutions must be found. Nowadays mobile smart devices are equipped with high performance lens and cameras, therefore they can capture high quality pictures in almost no time. Optical character recognition methods (OCR) are getting 99% accuracy on standard font documents with little or no noise at all. Having this in mind we propose a system that verifies in real-time the ID of an elector, right before a traditional voting system procedure. This solution is based on a stand containing a mobile device that takes a quality picture of that ID and sends it to a server to extract the data. The use of such a system does not affect the actual voting process and does not interfere or relate with the voter's choice at all.

Author Keywords
Mobile devices, E-government; voting system;

ACM Classification Keywords
H.5.m. Information interfaces and presentation - Human-Computer Interaction, Miscellaneous. ; J.1 – Computer Applications – Administrative Data Processing - Government; I.7.5 Document and text processing– Document Capture – Optical Character recognition

General Terms
Human Factors; Design; Security;

INTRODUCTION
Electronic election systems are used since 1960 when punched cards were introduced. The first country to use them on a large scale was United States of America, who implemented the system during the 1964 presidential elections. As a result, a new automated system to verify the authenticity of the voters was needed.
Online voting systems discard any physical evidence of the process, relying solely on the electronic infrastructure. During our research, we have developed a transition model between the traditional and online voting systems. In this paper we propose an architecture where mobile technologies and devices can be used to check in real-time a voter's IDs before the actual vote.

ONLINE VOTING SYSTEMS
Internet voting systems proposed in research literature [1, 2] use cryptographic techniques to get to a property called end-to-end (E2E) verifiability [3]. This feature ensures that the ballots have been counted accurately without trusting the computers or officials to behave honestly. The Estonian voting system is the largest Internet voting system in the world. It was introduced in 2005, being used by 30% of participating voters to cast their vote at the most recent elections [10] and because of this it can be used as an example. The system is based on the Estonian national ID infrastructure. These smartcards have the ability to perform cryptographic functions, which in combination with card readers and client software allow Estonians to log in to websites and make legally binding signatures on documents [5]. Based on this, they are used to authenticate and sign the ballots. As an extra security measure, the smartcards are

associated with a PIN code to authorize each operation. The source code from the server is published to a GitHub repository 2-3 weeks before the elections and the infrastructure is configured one week before the election in a public ceremony. It consists of four machines: Vote forwarding server (VFS), Vote storage server, Log server and Vote counting server. Before each election, a set of voting applications is published by the election authority.

In order to vote, a person has to launch a client application, insert his ID and enter the PIN code. By doing this, a secure connection with the VFS is established. After the server verifies the voter`s eligibility, a list of candidates for the voters district is returned [4]. After the voter selects his choice, he enters his PIN code again to sign the vote. The client pads the choice using RSA-OAEP and randomness, encrypts it with a 2048-bit encryption public key and signs the encrypted vote with the private key of the voter. The result is sent to the server. Voters are allowed to vote multiple times but only the last vote counts. Earlier votes are revoked and while the system indicates if the user voted previously, it does not show the number of times. The vote can also be overridden by voting in person on the Election Day. If the voter wants to confirm that his vote was recorded correctly he can use a smartphone app provided by the election authority [6]. The verification can be done three times per vote and up to 30 minutes after casting.

The storage server processes the encrypted votes to verify the signatures and removes any invalid or revoked votes after the online voting has ended. Officials export the set of valid votes in a public counting session, making sure only the anonymous encrypted votes remain after the signature is stripped away. These votes are burned to a DVD and transferred to a counting server. The counting server decrypts each vote and the results are combined with the totals from the in-person polling stations and published as overall results for the election.

The problem with the system is that it is vulnerable to denial-of-service attacks against the voting process. By sending many specially created requests containing fields with long names, an attacker can exhaust the server`s log storage, thus blocking it from accepting new votes. Another problem is a shell-injection vulnerability in the user interface of the server. It would allow operators to execute arbitrary shell commands on the election servers with root privileges. This can be very dangerous and proves the fact that open source doesn`t guarantee the absence of vulnerabilities [7].

TRADITIONAL VOTING SYSTEMS

In the United States, conventional voting systems are formed by joining many state-wide elections conducted independently by local election jurisdictions. States can use different voting systems, the decision being made by each county. The systems vary [11] from paper ballots and punch cards to mechanical lever machines, optical scan and direct recording electronic devices. In addition, a variety of voting processes are employed throughout the nation. Traditionally, people cast their vote on the Election Day. However, some alternative methods do exist:

- Absentee ballots, which allow people to vote-by-mail before the election and are available to voters who prove that are unable to get to the polling place;
- Vote-by-mail, available to everyone who registers as a voter, the person only has to fill-in the ballots and return it by mail, thus removing the need of polling places;
- Satellite voting, allows early voting from sites around the county for a period of time (several weeks to a few days) prior to elections.

While all these methods allow people to cast their votes in any condition, increasing turnout and convenience, a big problem appears because the process is very hard to manage and the counting of the votes is slow. These systems require voters to register before voting, making the process even more complicated. People need to make sure they don`t make a mistake when they cast their vote because they are not allowed to use multiple paper ballots, which leads to many votes being canceled.

However, there are advantages to paper ballot voting [8]. They give a reliable audit trail and perhaps more importantly, they guarantee the protection of electoral neutrality.

PROPOSED ARCHITECTURE

The idea behind this concept is that simplicity is the ultimate sophistication. Trying to solve a complex problem using a complicated solution only makes the matter worse. Because of that, the architecture of the system is as follows: using a smartphone’s camera a photo of an ID is taken; the photo is then transferred to database storage; the processing platform uses a secure connection (see Figure 1). After the information contained in the photo is processed, the data is stored online in a database so it is available anytime to anyone who has access to the system.

Data security is a big concern because we are dealing with personal information. As a result, a secure connection to communicate between the phone stand, the cloud processing system and the government database is used. All the data transferred is secured with a 128-bit encryption keys. This guarantees that the transfer is safe and no information is disclosed without approval.

In order to verify the IDs and determine if they are indeed genuine the system tracks several fields. It uses a combination of 3 unique identifiers from the ID to create a criterion on which each entry will be verified. Each time a new person tries to vote, the document is first verified against the existing items in the database so that no duplicates exist. This gives us the assurance that each entry is indeed unique and no ID can be used to vote multiple

times. In order to get the information from those specific fields, this project uses an online OCR processor [9]. The OCR processor takes the captured photo as an input and converts the information into text with a high success rate, 99.8% [9]. The information provided by the OCR engine contains personal data (name, surname, place of birth and address) and identification data (serial number, ID expiration date and PNC) that are stored in a database.

To eliminate the possibility of someone using a fake ID that has a unique combination of fields, the system can compare all the data gathered after the voting process has ended to a government database. This database contains all the relevant information regarding each person that has the right to vote. Because of this, every time an invalid ID appears it is flagged as illegal voting and subsequent actions should be taken to verify the validity of it and the people who voted using such IDs.

A 2-step verification is implemented because of this. The first step and the one that gives instant notifications ensure no one votes twice. After each entry is verified so there are no duplicates in the database, a notification is sent back to the phone if the confirmation went through or if there is a duplicate and the validation failed. This first notification provides real time verification. The second notification is implemented as a collective response from all the IDs used in the process after they have been compared with the government database and proved to be fake. Using these two steps ensures that no person votes using a fake ID.

ADVANTAGES OF THE PROPOSED ARCHITECTURE

The proposed architecture grants a simple and secure way to verify each person who wants to vote. By eliminating pre-registration, privacy and neutrality are guaranteed. Also the use of a smartphone reduces the cost and grants portability. Client or server fraud attempts are eliminated as no crucial vote related information is stored and cloud processing ensures the process moves fast and without any delays. Using a secure internet connection coupled with data encryption guarantees no personal information is accessible to anyone. By keeping paper ballots in combination with online ID verification the architecture combines the advantages from both online and traditional voting systems. It doesn`t require major modifications to polling places. Because of that and the accessible price of the components, it can be seen as a first step in a migration process from a traditional system to an online system.

PROOF OF CONCEPT

In order to make a proof of concept for the proposed model we have implemented an alpha version of this system. Testing was made using standard plastic ID's (released after 2002) and two Android devices for multi-point data acquisition. Each devices has the applications installed and connected to the secure server which was established on Google AppEngine. The entire scenario made for one voter lasted between 34 – 56 seconds depending on the availability and speed of the cloud system. Since cloud scalability is not a problem nowadays, we assume that the number of simultaneous voters is not a problem.

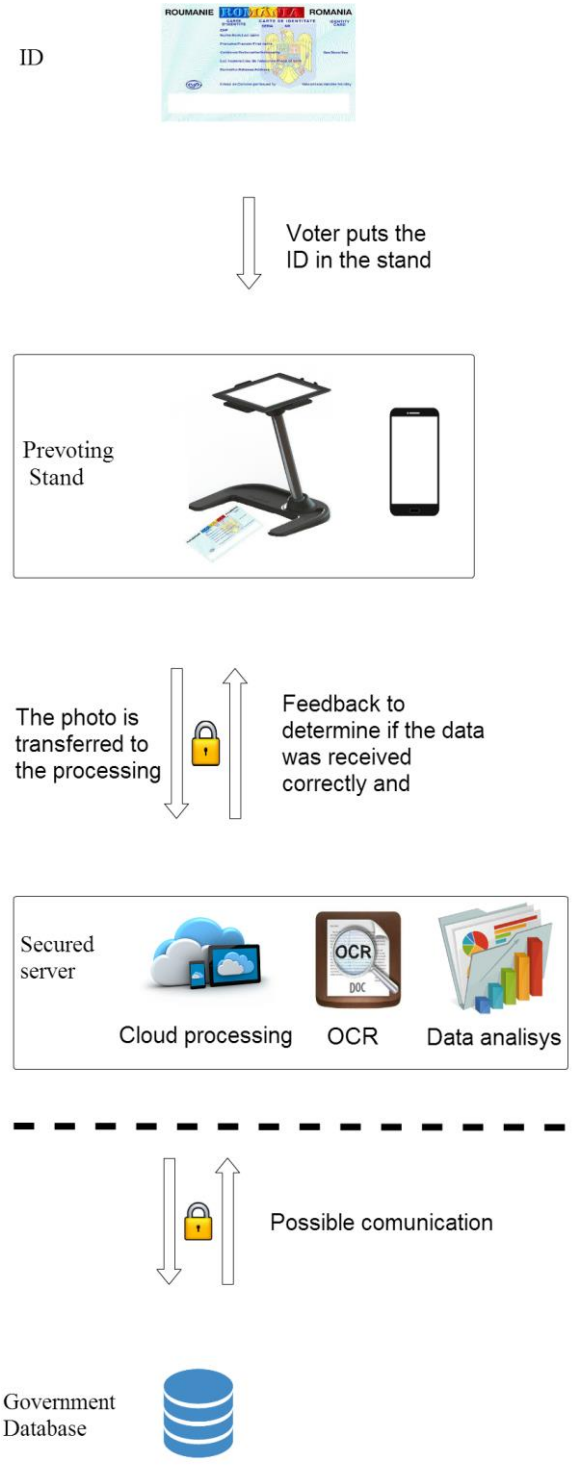

Figure 1: System Architecture

CONCLUSION

In this paper we have presented a solution that offers the possibility of checking an ID before a traditional voting procedure. Notifications are triggered in real-time concerning the legitimacy, validity and uniqueness of the vote. The procedure requires positioning the ID in a special stand that contains a smart device with a good camera that can capture an image and send it to the cloud server. The server processes the image and extracts the data using OCR. Every information is stored encrypted in the database for later use or for immediate notifications.

After the development of an alpha version just for proof of concept purposes we have identified a series of problems: blurring of the lens needs to be detected and a notification has to be triggered; the distance between the stand and the ID needs to be calibrated for every device type; there are at least 4 slightly different types of plastic card ID's and the OCR match must be made individually for every card type; internal flash of the camera cannot be used because on different types of plastic outputs different mirror-light effects.

FUTURE DEVELOPMENTS

Making the system available in the entire voting area is a very important goal. Taking that into consideration, it would be a good idea to implement local temporary databases in polling places that don`t have internet connection. This allows the architecture to work in the most remote places. The processing and checking of the information would be done after the voting process has ended and fake IDs and duplicate votes are flagged. We believe this would reduce the necessity of a real-time synchronization between an existing database and the government database with valid ID's.

Another improvement that can be implemented to allow operating in locations without access to internet is local phone OCR processing. This means that the information authentication should be done on the device, requiring more powerful smartphones so that the time needed for each ID verification is reasonable.

These features will raise the costs of the system but still offer the possibility of real time ID validity notifications, local vote duplication detection or later on synchronizing and statistics. The most important aspect is that they would allow system usage in extreme conditions.

ACKNOWLEDGMENTS

This work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134398.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

REFERENCES

1. Adida B., Helios: Web-based open-audit voting. In *Proceedings of the 17th USENIX Security Symposium*, Aug. 2008.
2. Ansper, A., Buldas, A., Oruaas, M., Priisalu, J., Veldre, A., Willemson, J., Virunurm K.m E-voting concept security: analysis and measures. Technical Report EH-02-01, Estonian.
3. Chaum, D. Secret-ballot receipts: True voter-verifiable elections. *IEEE Security & Privacy*, 2(1):38–47, 2004.
4. Cybernetica AS. Internet voting solution, 2013. http://goo.gl/62qv2h
5. Estonian Certification Authority. Mis on ID-tarkvara? In Estonian. https://goo.gl/lXKNF0
6. Estonian National Electoral Committee. Valimised: Android Apps on Google Play, Oct. 2013. In Estonian. Accessed: May 13, 2014, https://goo.gl/GcXkxe.
7. Kitcat, J., Source availability and e-voting: An advocate recants. *Commun. ACM*, 47(10):65–67, Oct. 2004.
8. Mote C.D. Jr., Report of the National Workshop on Internet Voting: Issues and Research agenda. Proceedings of the 2000 annual national conference on Digital government research, Pages 1-59, 2000
9. OCR (ABBYY) Processor: http://goo.gl/ZS8YJc.
10. Springall, D., Finkenauer, T., Durumeric, Z., Kitcat, J., Hursti, H., MacAlpine, M., Halderman, J.A., Security Analysis of the Estonian Internet Voting System. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Pages 703-715. 2014
11. Voting system: https://goo.gl/sEB9WG.

Socialization Techniques in Virtual 3D Space

George Marian Hotca, Teodor Stefanut, Dorian Gorgan
Computer Science Department, Technical University of Cluj-Napoca
Str. G. Baritiu 28, 400027, Cluj-Napoca, Romania
hotca_marian@yahoo.com, {teodor.stefanut, dorian.gorgan}@cs.utcluj.ro

ABSTRACT
This paper presents the LivingRoom platform for online socialization within the 3D virtual space. The user is represented by an avatar that is moving around the virtual space and can communicate with other avatars. The audio and video communication among avatars is triggered by their interactions similarly with the human action within the real world.

Author Keywords
Virtual Space; Visual Techniques; User Interaction Techniques; Audio Communication; Video Communication; Avatar.

ACM Classification Keywords
H.5.2 User Interfaces, I.3.7 Three-Dimensional Graphics and Realism, I.6.8 Types of Simulation.

INTRODUCTION

The saying of Eric Hoffer "Creativity is the ability to introduce order into the randomness of nature" states the basic idea of the LivingRoom platform.

A socialization platform is an ensemble of web services that allow users to build a public profile within a limited system, to create a list of other users sharing a connection, and to visualize and navigate through the list of connections. LivingRoom is a socialization platform concerning mainly on providing users with the possibility to make new connections in the social life, even with unknown persons.

The main characteristic of the social platforms is the user profile [6], which is public and includes a list of friends (i.e. connections) that are users of the same system as well. In the classic social networks the user adds and enhances new personal information to his profile, by the answers fulfilled during the creation of the account. The LivingRoom platform does not impose the enrichment of the user's profile with irrelevant information or a list of human connections. This is just for allowing a high level of information confidentiality.

The interaction with others [10] is natural, and actually is the most preferred way of using an application, motivated by competition, collaboration desire, organizing virtual meetings with others and self-exploration. Because in the LivingRoom platform the competition is not a priority, the last three motivations are mainly considered.

The collaboration [10] is the opposite way to competition through which we can interact with another user. The collaborative environments are preferred because they involve more users.

The games and the interactive applications allow us to join the group of friends even though this activity is accomplished within the virtual space. The interactive applications are ways through which two persons could consolidate or start a friendship.

The paper is structured as follows. The next section is an exploration of the literature, highlighting the related achievements. Then three sections detail the LivingRoom solution as a collaborative platform within the 3D virtual space. The last section concludes on the development and evaluation of the LivingRoom platform, and makes recommendation for the future research directions.

RELATED WORK

Instant Messaging Systems

Instant messaging systems (Instant Messaging or "chat") are systems that allow instantaneous exchange of text messages or multi-media content with one or more people via personal computers. The advantage of such a system is the fluency in messaging, with no significant delays, creating the impression of a real conversation but without direct interaction between the interlocutors.

Instant messaging systems are based on client-server architecture. The client (or the user) connects to a central server sending messages, which are interpreted by the server and then redirected to the destination user. Most systems involve the use of contact lists. Each user is forced to add a contact to the list, his request needs approval from the other contact, after which it is allowed to submit messages.

Web Chat Applications

A web chat system [9] allows real-time inter-user communication through an easy-to-use web interface. Online messaging gives the user instant access, the only requirement being the use of a web browser, which is an advantage because it will always access the latest version of the application, without requiring installation of updates.

Omegle [4] is the most representative web chat application. The application allows users to communicate without the need of creating an account. It randomly connects two users

in a one-to-one chat through which they can communicate anonymously. If it was originally created to be able to communicate only through text messages, subsequently a video mode was implemented which allows users to communicate using the device's microphone and webcam.

At the end of the conversation, users have the option to save their conversation history and to share it using a generated link. The conversation is not fully private as the user's personal information could be distributed over the network without his approval.

Chat Systems based on Virtual World

The virtual world known under the term MMOW (Massively Multiplayer Online World) [3] is a computer simulated environment, populated by users who are represented by avatars and can explore the virtual world simultaneously and independently, participate in activities or communicate among them.

The user accesses a simulated world that presents perceptual stimuli [13] that can manipulate elements of the world and thus offers a high degree of telepresence experience. Telepresence refers to a set of technologies that enable a person to feel that it would be physically present.

"3Dchat" application is the most representative in the niche of virtual worlds. It combines social networking concepts with online virtual world concepts, allowing users to interact with each other through avatars in a three-dimensional space. Although interaction is achieved largely through mouse and keyboard, the system allows voice communication within a certain area in the scene.

SOCIALIZATION PLATFORM IN VIRTUAL 3D SPACE

"LivingRoom", unlike other social platforms, aims to use the concept of communication cluster. The communication cluster is a group composed of all users who adhere to it. The application allows the user to change the cluster dynamically. If the cluster to which the user wants to be connected does not exist, it is created.

The purpose of the cluster is to group the users who wish to belong to particular geographical settlements (village, city, country, continent, etc.). Thus, a user can only interact with users in the same cluster, the interaction with users in other clusters is not allowed.

Each user is represented by an "avatar", a term with a wide range of meanings. In the Hindu religion, "the avatar is an incarnation of divine beings in the form of man or animal. Usually, a deity that exists at a higher level of reality chooses to manifest physically in the reality of people's lower level for specific purposes." [1].

In IT, the avatar is "a graphical representation of the user or the user's alter ego (character). They may have a three-dimensional form, as in games or virtual worlds, or a bi-dimensional form, such as an icon for Internet forums or online communities "[1].

In the socialization platform, the user is represented by an avatar, which is a three-dimensional character. Each avatar is placed in the 3D space corresponding to the cluster he is associated with.

A 3D scene is a group of objects and entities in a well-defined virtual three-dimensional space [16], and can be characterized by three geometric parameters of the physical universe (excluding the fourth dimension - time). These parameters can be labeled by a combination of three elements chosen among the terms: length, width, height, depth and breadth.

The initial position of the character in the scene is obtained by placing it in a direction perpendicular to the vertical axis, and amplitude which will tend to zero (irrespective of the metric used) to the origin of the plane orthogonal to the vertical axis.

Camera in the field of gaming is the point of view of a character or the world view, in the context of the application scene. Most applications use one or more types of camera.

A view frustum [7] is defined as the volume of space that contains all the visible elements of interest in the three-dimensional scene. The X axis is oriented to the right and Y axis is pointing upwards. Z-axis direction depends on the graphics system used.

"LivingRoom" uses two types of perspective: "First Person" and "Tracking".

The "First Person" perspective [5] combines the camera with the character point of view, which means that the user sees through the eyes of the avatar (Figure 1). The user's viewing perspective corresponds to the viewing perspective of its avatar. This increases the realism of the application, even within the virtual environment.

The "Tracking" perspective [5] involves the tracking of the character along a predefined line in the 3D space. The camera can perform operations such as rotation, speed up and slow down the movement relative to the point that it aims, or change its position synchronously with the tracked character. The user can dynamically change the viewing direction by using specific control actions.

Animation [12] is the concept that visually brings to life the character in the application. Changing the position involves an animation of the avatar. Character animation can be seen as a state machine, and depending on the state of the character it will play an appropriate animation. Characters will have two types of animation: "idle" and "walking". The "idle" animation will show how the character is at rest, and the "walking" animation will occur when the character is in a motion state.

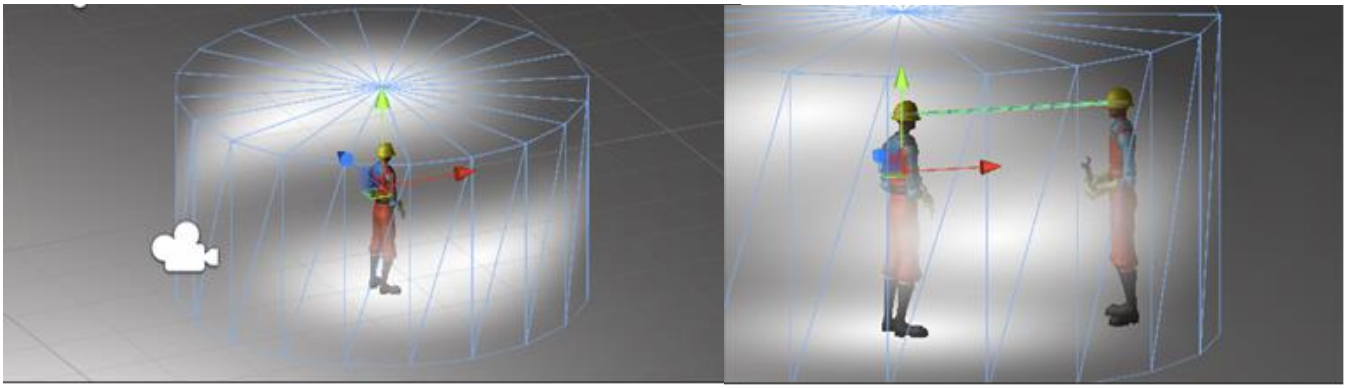

Figure 1. Audio and video communication techniques: (left) audio collision, and (right) video collision.

Most interactions between persons may take the form of verbal communication, such as talking, or non-verbal such as gestures or body language. The avatars are representations that take human characteristics in appearance, interaction manner and decision making, being used in a large number of applications that require interaction as natural interfaces.

One of the difficult problems is the development of a character that looks and behaves like a human, without another behavioral failure known as “Uncanny Valley" [14]. "Uncanny Valley" is a principle defined by Masahiro Mori in 1970 to describe robots and characters that resemble human creatures but do not behave like them. The feeling of frustration and mistrust is increasing proportionally with the non-realism of the human aspect and behavior.

The interactions between characters are classified into: audio interaction and visual interaction (Figure 1). Audio interaction between avatars is achieved by fulfilling the condition that an avatar is situated at a predetermined distance to another avatar, similar to the real world. Conceptually, this can be seen as a sphere centered in the avatars position.

Visual interaction between avatars is achieved by fulfilling the condition that an avatar "looks" into the direction of the destination avatar’s eye. It means that the camera orientation is toward the eye of the other avatar, and the

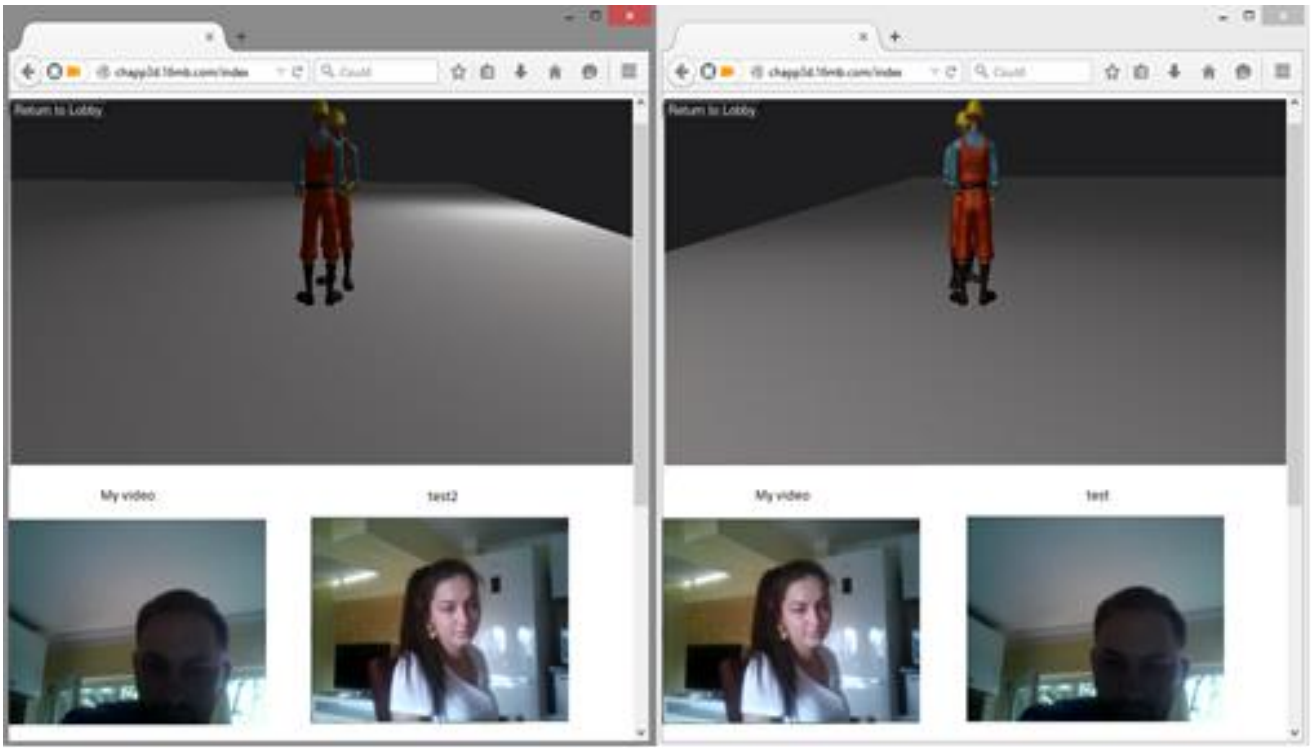

Figure 2. Video communication between two users.

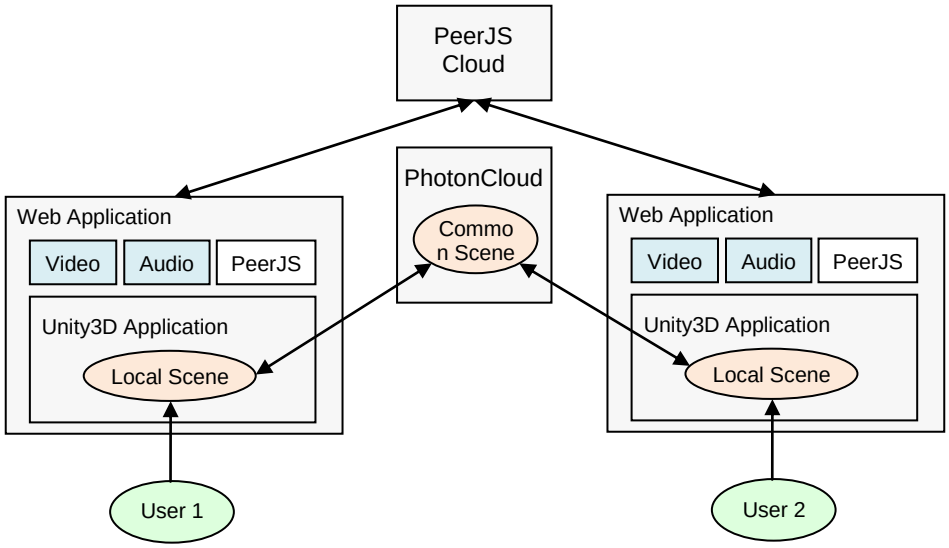

Figure 3. LivingRoom system architecture.

distance between them is equivalent or less than the threshold value. The Ray-Casting algorithm is used to identify the visual interaction.

The visual collision assumes that there is already an audio collision. The distance constraint between the two avatars has to be satisfied as well, which makes communication more naturally and similar to human communication mode.

AUDIO AND VIDEO COMMUNICATION TECHNIQUES

To perform actions within the system, the user must be authenticated and have specific permissions of his user type.

The main functionality of the system consists of the audio and video calls.

The scenario to start a video communication (Figure 2) is the following:

Start video-communication

1. The user controls his character in order to reach the audio collision with the other actor;
2. The user directs his visual perspective toward the eye of the other character;
3. The system detects intention to initiate a video call;
4. The system checks if the two users have the required resources;
5. The system asks for permission to use the hardware video resources;
6. The emitter sends a request for video transmission to the receiver;
7. The receiver accepts the video call;
8. The emitter responds with the video transmission;

8.1. One of the two users closes the communication

End video-communication

SYSTEM ARCHITECTURE

The application requires the use of two independent servers: Photon Cloud [8] and Cloud PeerJs [2]. PhotonCloud is used to synchronize characters (position, rotation) and all components attached to them in common scene (Figure 3). The scene is a common 3D place where all users of a cluster are represented by their characters. Each cluster has an individual scene.

This server works most of the time in emitter mode, sending to all Unity3D clients [11] within the same cluster, the positions of all characters. Otherwise, it works as a listener (receiver), respectively, expecting information about the position and rotation of a user's character. Then, it broadcasts them to all clients in that cluster.

PeerJs Cloud server is used for peer-level networking allowing the users to achieve the audio and video calls. Each user gets a peer connection after a successful authentication into the application. The user accesses the application through a Web browser. The application consists of two main components: the front-end application (client-side web scripting) and Unity3D application (plugin).

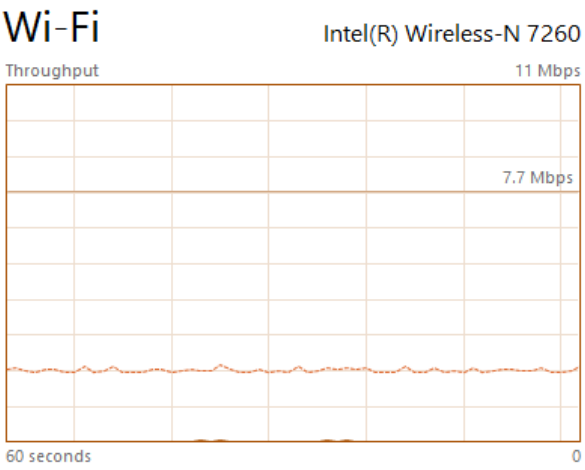

Figure 4. ChatRoulette bandwidth consumption

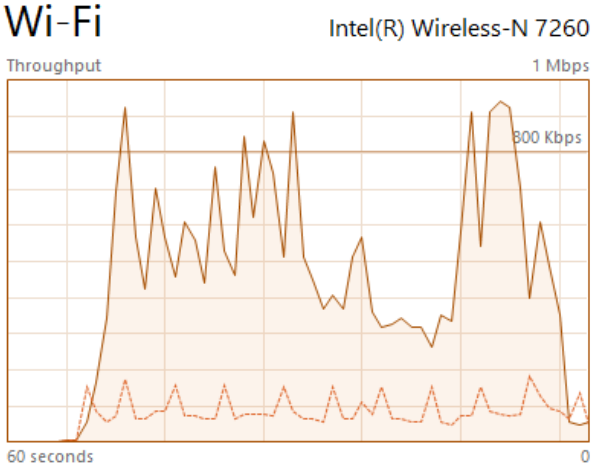

Figure 5. LivingRoom bandwidth consumption

The Unity3D application renders the graphics - local scene and the character within the scene. Through this application, the user is able to control his character by a set of predefined controls. Conceptually, the local scene displayed for each user is a copy of the common virtual 3D scene. The user controls his character within the local scene, and the PhotonCloud server synchronizes the characters' position in all the other local scenes.

EVALUATION

Hardware resources

Running an instance of the application and having a audio transmission in progress requires about 100-120 Mb of RAM, with 25-30% more than the "SmallWorlds" application which is one of the most representative virtual chat software, but this surplus is justified by the advanced graphics that the "LivingRoom" has.

LivingRoom uses about 0.6-0.7% CPU because of the Unity3D application which uses hardware acceleration. This number is about double than the CPU power required by the "SmallWorlds" application which has a 2.5D graphic system, avatars represented as sprites and basic interaction, by far less interactive.

Bandwidth

"ChatRoulette" application (Figure 4) is similar to "Omegle" (presented in this paper) and for a video resolution of 640x480 (VGA) requires 500-550 Kbps bandwidth.

"LivingRoom" application (Figure 5) requires about 2Mbps of bandwidth for video transmission through WebRTC, the video resolution being 1280x1024 (~1.3 megapixels). The result is very good if we consider that the bandwidth required includes PeerJs and PhotonCloud communication as well.

The transmission efficiency is about equal if we consider that VGA resolution is about 4 times smaller than a resolution of 1.3 mpx specific to the HD web camera used.

CONCLUSIONS

One of the most difficult tasks was synchronizing the characters in the local scenes, since each user specifies the commands in an asynchronous mode. The experimental results have proved the effectiveness of collision detection in Unity, considering that these computations are executed every frame, and the application runs up to the value of 50-60 fps. This means that all the computations of the collision run 50-60 times per second, so every 1.5 milliseconds.

The audio and video transmissions are effective, clear and comparable with the ones offered by applications such as Skype (Figure 2). The experimental results are satisfactory since the free versions of the PeerJs Cloud and PhotonCloud servers offer some limitations of the number of connections (e.g. 50 peer connections, and respective 20 concurrent users), and lower bandwidth.

In conclusion, the LivingRoom platform meets the objectives of a social platform by inter-connecting users in groups called clusters, and by implementing a multimedia communication system. The users initiate audio and video communications by interaction techniques in natural manner.

FUTURE DEVELOPMENT AND RESEARCH DIRECTIONS

Google Chrome announced the withdrawal of support for NPAPIs (Netscape Plugin Application Programming

Interface), which means that Unity Web Player applications will no longer be supported.

The main objective is to create a cross-browser and cross-platform application, in order to solve this technology issue. One possible solution is to use WebGL, a graphic programming interface for the Web (OpenGL for the Web), as a rendering platform due to its compatibility in all HTML5 browsers.

Furthermore, the social interaction between the participants can be further developed through additional gestures and interaction techniques that can be implemented.

ACKNOWLEDGMENTS

The research has been carried out in the Computer Graphics and Interactive Systems Laboratory of the Computer Science Department, in the Technical University of Cluj-Napoca.

REFERENCES

1. Avatar in computer applications, http://en.wikipedia.org/wiki/Avatar_(computing)
2. Badass Javascript, ”PeerJs : A peer to peer networking library in Javascript using WebRTC”, 2013, http://badassjs.com/post/43090030238/peerjs-a-peer-to-peer-networking-library-in
3. Bartle Richard, “Designing Virtual Worlds”, New Riders, 2003
4. BeWebSmart.com, ”What is Omegle? Is Omegle okay for kids?”, 2015, http://www.bewebsmart.com/internet-safety/what-is-omegle-is-it-okay-for-kids/
5. Camera Comes First [Game Design], http://www.whatgamesare.com/2011/10/camera-comes-first-game-design.html
6. Danah M. Boyd, Nicole B. Ellison, "Social Network Sites: Definition, History and Scolarship", Journal of Computer-Mediated Communication, pp.210-215
7. Eric Lengyel, ”Mathematics for 3D Game Programming and Computer Graphics” Third Edition, Course Technology, a part of Cengage Learning, 2012
8. Exit Games, ”Photon Unity Networking”, https://www.exitgames.com/en/PUN.
9. Gao Kevin, "5 Best Practices for Increasing Sales with Live Chat Software", SalesForce, 2014.
10. Jesse Schell, ”The Art of Game Design – A Book of Lenses” Second Edition, CRC Press
11. Jon Brodkin, ”How Unity3D Became a Game-Development Beast”, 2013, http://insights.dice.com/2013/06/03/how-unity3d-become-a-game-development-beast/
12. Kyle D’Aoust, “Unity Game Development Scripting”, Packt Publishing, 2014
13. Marvin Minsky, “Telepresence”, Omni, 1980
14. Patrick G. Kenny, Thomas D. Parsons, Albert A. Rizzo, ”Human Computer Interaction in Virtual Standardized Patient Systems”, Institute for Creative Technologies, University of Southern California
15. Stay In Touch With Yahoo Pager, 1998, https://web.archive.org/web/20101220214553/http://yhoo.client.shareholder.com/press/releasedetail.cfm?ReleaseID=173501
16. Three-dimensional space coordinate system, http://en.wikipedia.org/wiki/Three-dimensional_space
17. Yahoo! Messenger, http://ro.wikipedia.org/wiki/Yahoo!_Messenger

Gestural Recognition by a Four-Kinect Module in a CAVE "Le SAS"

Saleh Salous saleh.salous@citu.fr
Julien Newton soosaine@ece.fr
Laure Leroy Laure.Leroy@citu.fr
Safwan Chendeb safwan.chendeb@citu.fr

CiTu paragraphe, Paris 8 university
2 Rue de la Liberté, 93200 Saint-Denis, France
http://www.citu.info/, http://www.univ-paris8.fr/

ABSTRACT

Among the various types of interactions in a virtual reality (VR) room, physical gestures such as moving one's hands are one of the ways to communicate with the virtual environment. It is possible to capture those gestures with sensors, in our case the Microsoft Kinect. In the context of a large CAVE, not one but four Kinects are used to track the user and capture its gestures. However, the question of the detection of specific gestures by this system comes into play as the CAVE requires recognizing the user's gestures in order to interpret them into an input. In this paper we focus on the gestural recognition ability of our CAVE applied to three specific gestures: Raising the right hand, raising the left hand and short hopping. Moreover, we provide experimental results showing that a Four-Kinect system provides more effective gesture recognition than a single sensor.

Author Keywords

Multi-Kinect; gestural interaction; gesture recognition; virtual reality; CAVE.

INTRODUCTION

In the field of Virtual Reality, one of the most important aspects is the interaction between the human user and the computers processing the virtual environment. In order to create a fluid and seamless user experience, it is necessary to properly track the user, recognize its inputs and successfully apply them to the VR scene.

In our context, we have a CAVE named "Le SAS" with a two-screen set-up that includes a front screen and a ground screen. (see Figure 1)

We discussed the topics of multi-sensor configurations, user tracking and camera calibration in [7][8]. The goal of this article is to propose a method to recognize specific gestures from the user in order to translate them into inputs.

Indeed, the number of potential interactions is high, and this step is essential to provide proper interaction between the users and the VR environment.

Figure 1 : Representation of "Le SAS"

This paper is organized as follows. Section 0 provides a literature review. Section 0 will outline our approach and explain the data collection and posture recognition processes. Section 4 will focus on experiments designed to test our approach. The test results are also available for reference. Finally, Section 0 will conclude on the acquired data.

STATE OF THE ART

Gestural recognition is a known topic in computer vision and analyses of gestures in a VR context have already been conducted [1, 3, and 4].

Several notions have previously been established in this field, such as types and categories of gestures.

For instance, in his survey of gestural interactions, [4]defined types of gestures as well as a large amount of specific movements and explained their use in various fields such as medical science or entertainment.

[3] Provided an extensive depiction of the state of the art in 3D gesture recognition, and explained the upsides and drawbacks of the technology.

Other researchers such as [1] chose to focus on the gestures in themselves and analyze the potential risks of physical exhaustion, muscular fatigue and pain that some of the gestures can provide if done repeatedly. This kind of analysis is useful in order to determine the least uncomfortable gestures to use in an interactive environment such as the SAS.

Gestural recognition was successfully used in other experiments; for instance [5] developed a gesture recognition application that can use a low-cost webcam to recognize hand gestures and translate them into interactions with digital objects in a virtual environment.

[6] are also able to track gestures, in the context of Augmented Reality applications. However, their approach focuses on using finger-mounted tracking targets to provide a higher level of accuracy and individual finger tracking.

Among the fields in which gestural recognition can be used, history is the one chosen by [9] with a VR application that is implemented in a museum and allows visitors to explore a 3D re-creation of an ancient ship called Vrouw Maria using a hand gesture-based interface.

In [2], the authors created and presented a complete system with a stereo camera and a microphone to receive inputs from the user, a projector to display the virtual world with which the user interacts, as well as software developed to link recognized gestures and speech with the VR system's visual output.

PROPOSED APPROACH

In this section, we will explain the methodology we follow to recognize inputs in our CAVE. This part will first explain the collection of joint data and then discuss the posture detection technique in the CAVE.

Posture data collection and posture detection

The environment in which the data is collected is a CAVE named "Le SAS" with two screens. Each is 3 meters high and four meters wide. Data about the user's location and its joints is measured by four Kinects in a multi-sensor system that was notably discussed in [7].

The objective of this system is to cover the maximum amount of space in the "SAS" and track the user's gestures regardless of its position inside the CAVE.

As the user is supposed to be able to move around the "SAS", the multi-sensor system tracks the user's coordinates for each frame and an algorithm [8] is used to determine in real-time which of the Kinects have the more accurate joint data.

The actual algorithm used by our module to detect gestures can be observed in Figure 2. Each Kinect sends joint data to the gesture recognition module in real-time. As all Kinects may not be very close to the user, the module applies a filtering mechanism discussed in [8], where only the most relevant Kinects for the desired measured joints are taken into account.

In our case, the specific joints required for the detection of the three gestures we wish to test (cf Section 0) are both elbows, both hands and both knees.

Once the user joint data is properly recorded by the CAVE's multi sensor-system, the potential gestures reflected by those coordinates have to be detected as such before being translated into inputs to interact with the VR environment. The second part of the algorithm is tasked with analyzing the filtered joint data and detecting gestures.

In order to recognize a raised hand, the algorithm compares the X and Y coordinates of the hand and elbow joints. If the hand's two coordinates are higher than the elbow's, the module returns a gesture notification.

The method used to detect the short hopping movement is to compare the differences in the X and Y coordinates of both knees between a previous frame and the current frame. If both knees' joint data provide a difference lower than a chosen threshold, the system signals a short hop.

It is also possible to develop other specific posture detection calculations related to new gestures we may intend to implement in our CAVE. The reason why the hand-raising gestures have been chosen for this experiment is related to another input device previously implemented on the "SAS". Interaction with the "SAS" is also possible with a joystick that allows the users to move on the left or the right. As a consequence, the tested gestures aim at replicating those movements.

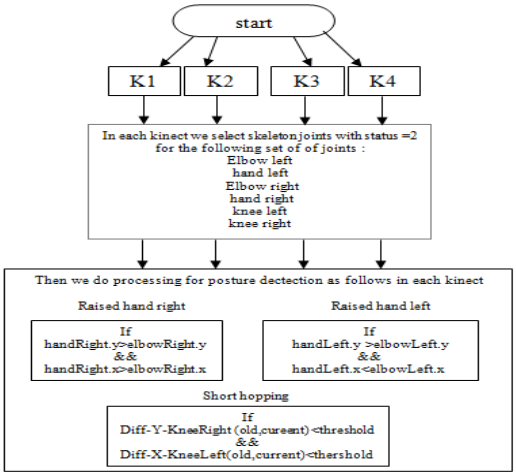

Figure 2: Posture detection algorithm

The technical context related to the data transmission and the software libraries is as follows. Accessing the Kinect joint data from the multi-sensor system and using the values to perform recognition operations was possible thanks to the Microsoft Kinect SDK.

The data transmission itself used the UDP protocol. This protocol was chosen as to send and receive data in real-time.

Constraints of the CAVE

While gesture recognition on its own requires complying with a few constraints, the environment in which we wish to deploy the gestural recognition abilities of our sensors provides new challenges.

Indeed, the gestures are used to interact with the VR environment. Therefore, the user should do movements that will not interfere with his or her experience within the CAVE. For instance, asking the user to move backwards in order to make his virtual self-move backwards is potentially dangerous as the user can fall off the ground screen if he or she is not cautious. A similar issue can happen with moving forward as the user may collide with the front screen.

Those constraints have to be taken into account when designing what gestures will be required for our VR applications. However, it is still possible to test gestural recognition with gestures that may not be implemented in future works.

EXPERIMENT

Posture detection

Experiment methodology

Our starting hypothesis is that four Kinects working together are more accurate and cover more potential user locations than a single sensor and are therefore more suited to capture gestures in a CAVE.

In order to obtain proper feedback on the effectiveness of our methodology and to provide conclusive data on our hypothesis, we set-up experiments to test whether the four Kinects system properly tracked the users' gestures and whether it was more reliable than a single Kinect.

Participants

The experiment was conducted as follows: 23 participants were invited to the "SAS" and were not informed beforehand about the specific objective of the experiment. This was done to avoid creating bias in the users' behavior during the tests. All of the participants are adults. 12 of them are male and 11 are female.

Detected Postures

Three gestures were tested in this experiment: raising the right hand, raising the left hand, and doing a short hop as seen in Figure 3. We asked each of the participants to do each of these gestures four times at random locations inside the "SAS" while a single Kinect was tracking them. This procedure was repeated with the four Kinects system instead.

Overall, each person made 3 different types of gestures. Each specific type of gesture was repeated four times by each participant for both Kinect set-ups (single Kinect and four Kinects). As 23 people took part in the experiment, the total number of recorded gestures is 276 for each Kinect configuration inside the "SAS" for a total of 552 gestures.

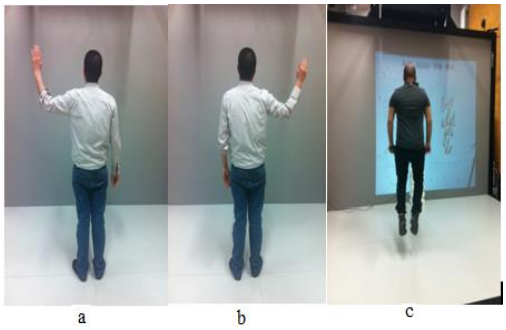

Figure 3: Participants in the middle of the experiment performing the three gestures a) Raising the left hand, b) Raising the right hand, c) Short hopping

The system analyzed the data in real-time to recognize the participants' gestures using the algorithm discussed in Section 0. We recorded this data to provide the results of our experiment.

False detection

As our approach aims at providing an accurate motion sensing system, there is another variable that has to be taken into account before interpreting the experiment results.

Indeed, the phenomenon known as "False Detection" can occur and provide a layer of inaccuracy to the data captured from the user's movements.

False detection happens whenever one of the sensors registers a motion that was not actually performed by the user, thus losing in overall accuracy. In our case, we measured the false detection issue regarding our three tested gestures within our SAS.

The testing protocol is as follows. A random user chosen from the 23 participants is asked to move as he or she wishes inside the SAS during 15 minutes without performing any of the three gestures: raising either hand or short hopping. We then analyze the data provided by our gesture recognition system and find the number of false detections.

The aforementioned experiment was conducted twice: once with a single Kinect, and once with the Four-Kinect set-up.

Another type of false detection also exists. False Detection can also occur when the system exhibits erroneous gesture recognition and does not detect the proper gesture when a user performs it. This type of false detection can be measured from the same data pool than the one used to determine the sensors' gesture recognition accuracy.

Experiment results

Detected postures

Figure 4 shows the results for the three gestures (raising the right hand, raising the left hand and short hopping). As each gesture was repeated four times by each of the 23 participants, the data pool

from which the experiment results were extracted contains 92 recordings for each combination of gesture and Kinect set-up. The results of our experiment are as follows. While the right hand raising gesture was successfully recognized by the four-Kinect system 93% of the time, only 77% of the gestures were detected by the single Kinect. The success rate for the left hand raising gesture is similar to the right hand one, with the four-Kinect system accurately capturing 91% of the gestures and the single Kinect only 75% of them. Short hopping proved to be a slightly more difficult gesture to detect according to the results. Indeed, the four-Kinect system's success rate is 87% and the single Kinect's is 68%.

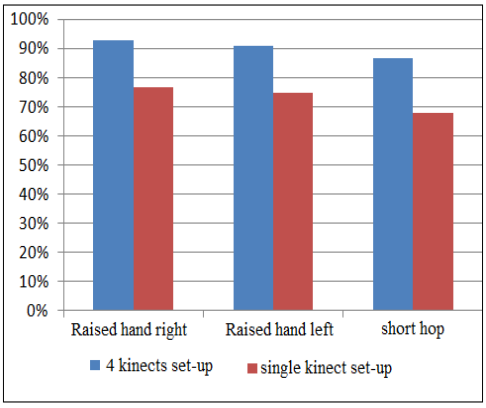

Figure 4: Level of accuracy of the gesture recognition module for each of the three tested gestures and both Kinect configurations

False detection simulation

As explained in Section 0, our experiment also measured the risk of a false detection occurring while using the SAS.

The results are as follows. The single Kinect set-up returned 17 false detections during the 15 minute-long test. The Kinect detected a short hop 9 times, recognized a raised right hand 5 times and returned a raised left hand 3 times.

However, data fusion has proved to significantly decrease the number of false detections in a multi-Kinect set-up. Once the 4 Kinects' joint data was fused into a single skeleton, only 6 false detections were returned by the system. 3 short hops, 2 raised right hands and 1 raised left hand were detected, thus providing a lower amount of false detections than a single Kinect.

CONCLUSION

Gestural recognition is a technical field with several constraints and its implementation in a Virtual Reality environment adds more issues that have to be considered when developing gestural recognition solutions.

Our experiment demonstrated that the use of a multi-sensor system provides more accuracy than a single sensor. As shown by the experiment results, the use of a multi-Kinect system increased the accuracy of the gesture detection by 16 percentage points for arm-raising gestures and 19 percentage points for short hopping. Furthermore, the four-Kinect set-up is suited to the size of a CAVE and complies with the specific constraints of the "SAS" as it provides a larger detection range.

ACKNOWLEDGEMENTS

This research and its results are made possible thanks to the members of the CiTU-Paragraphe lab. Thanks to Sobhi AHMED in particular for his assistance on the subject and his support and advice.

REFERENCES

1. Ahmed, S., Leroy, L., Bouaniche, A., (2014). Approches de Réduction des Impacts Négatifs des Interactions Gestuelles sur le Corps Humain. Reims Image 2014, p113-121.
2. Demirdjian, D., Ko, T., & Darrell, T. (2005). Untethered gesture acquisition and recognition for virtual world manipulation. Virtual Reality, 8, 222–230
3. LaViola Jr., J. J. 3D Gestural Interaction: The State of the Field. ISRN Artificial Intelligence 2013, pp. 1 – 18.
4. Majchrzak, A., Wagner, C., & Yates, D. (2006). Corporate Wiki Users : Results of a Survey. Wall Street Journal, pp. 99–104.
5. Rautaray, S. S. (2012). Real Time Hand Gesture Recognition System for Dynamic Applications. International Journal of UbiComp, 3(1), 21–31. doi:10.5121/iju.2012.3103
6. Reifinger, S., Wallhoff, F., Ablassmeier, M., Poitschke, T., & Rigoll, G. (2007). Static and dynamic hand-gesture recognition for augmented reality applications. Human-Computer Interaction, Pt 3, Proceedings; 4552, 728–737.
7. Salous, S., Ridene, T., Newton, J., and Chendeb, S,. "Study of geometric dispatching of four-kinect tracking module inside a Cave," In Proc. 10th International conference on Disability, Virtual Reality and Associated Technologies, Gothenburg, 2014, pp. 369-372.
8. Salous, S, Newton, J., Leroy, L., and Chendeb, S,. "Dynamic sensor selection based on joint data quality in the context of multi-kinect module inside the CAVE "Le SAS".IJCTE, V9N1, Feb. 2017.
9. Şen, F., Díaz, L., & Horttana, T. (2012). A novel gesture-based interface for a VR simulation: Re-discovering Vrouw Maria. Proceedings of the 2012 18th International Conference on Virtual Systems and Multimedia, VSMM 2012: Virtual Systems in the Information Society, 323–330.

Immersive Virtual Reality application using Google Cardboard and Leap Motion technologies

Mihai Chifor
Computer Science Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
chifor.mihai21@gmail.com

Teodor Stefanut
Computer Science Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
teodor.stefanut@cs.utcluj.ro

ABSTRACT
Virtual Reality applications have known an accelerated development in the past few years stimulated by the advancements in the hardware capabilities and also by the benefits from the similarity with the real world. However, the available solutions for immersive VR are still very expensive and often lack the friendliness expected on user interaction techniques. In this paper we will present a low cost, immersive, VR capable setup based on two new technologies: Google Cardboard and Leap Motion. The proposed solution is showcased using an educational application that allows its users to virtually visit Elmwood Park Zoo, Pennsylvania and learn more on the wild animals in the process. Most of the user interactions are performed through hand described gestures inspired from the real world, helping the users to accommodate more easily and perform naturally in the virtual environment.

Author Keywords
Virtual Reality; Google Cardboard; Leap Motion; immersive; educational

ACM Classification Keywords
H.5.1. Information interfaces and presentation (e.g., HCI): Multimedia Information Systems.

INTRODUCTION
Virtual reality (VR) represents today a promising media platform that allows the users to experience different computer-generated worlds in a very similar approach to the real one. One key aspect in feeling these worlds as real is the user's ability to interact with and influence the virtual environment. Therefore, a simulation of the interaction between a human being and the real environment has to be implemented into the virtual one, giving to the user the illusion of reality.

Up until recent years, one of the main issues concerning VR has been related with the very limited possibilities of the devices that allow the users to interact with the scene. Currently, there are two main categories of haptic interfaces: the *on-body* (attached to the user's body) and *off-body* (positioned independently of the user, somewhere in the real environment – on the floor, walls, desk, etc.).

Nowadays, most commercially available devices are off-body, allowing the user to move freely in the real environment. A few devices inspired from game-optimized ones have been used as alternative inputs for the VR environments, but only with limited success as they proved to feel unnatural. Due to the similitude with reality, one of the first user instincts when immersed in a virtual world is to use his/her body to move around and interact with the elements of the environment. While walking around the virtual scene using user's physical body is currently a popular research direction that is complementary to the one presented in this paper, using the hands-free motion in virtual interactions has proved to be a more natural approach and represents our choice of interaction.

Continuous development of the virtual simulations and their integration with VR environments has encouraged the implementation of learning scenarios that could benefit from the advantages of this technology. Through the use of detailed 3D models and virtual reality environments as part of the training process, one can experience an entirely new side of training. Some of the main benefits would be:

- the learning process has a higher level of interactivity being able to transform the user experience from passive to active
- while immersed into the virtual environment, the user is protected from distractions from surrounding elements
- VR based learning can represent a great solution for situations where the teaching materials are very expensive or the conditions of training are very hard to reproduce in real world; a scenario that has been modeled with a high level of detail allows learners to interact with it, understand and follow best practice procedures or carry out complex scenarios with very reduced cost / trainee (as all the resources can be reused as many times as needed).
- the interactive VR scenarios help the user gain knowledge, test his/her reactions in dangerous situations and test possible harmful scenarios in a safe environment, without being put at risk.

The purpose of this article is to describe the user experience into a 3D immersive environment created with the use of 2 particular pieces of hardware, namely Google Cardboard [4] as a head mounted display (HMD) and Leap Motion Controller [10] for hand tracking recognition. We developed our application based on a real life zoo model,

namely Elmwood Park Zoo, Pennsylvania [14]. The virtual world was developed using a 1:1 scale to increase the realism by maintaining the proportions of the objects, animals and even user's virtual hands.

RELATED WORKS

Continuous development of mobile phones has allowed in the last few years the development of more complex 3D mobile applications than ever before. Joining this trend, Google has proposed a new visualization technique that allows these applications to immerse their users into Virtual Reality environments, providing spectacular results at very low costs. This new approach is named Google Cardboard, and is actually composed out of a headset to support the display and two optical lenses, a mobile device (which is the actual display), a magnet (for user interactions) and specific display settings of the 3D application.

Google Cardboard has been showcased in applications like:

- Earth: fly where your fancy takes you on Google Earth
- My Videos: Watch your videos on a massive screen
- Exhibit: Examine cultural artifacts from every angle
- Photo Sphere: Look around the photo spheres you've captured
- Windy Day: animated short story

Probably the most popular application that engages the user in a virtual tour using Google Cardboard is Google Maps for Android [5]. The application has been extended to allow its users to visit different places (like Versailles) in virtual reality (VR) mode and with a virtual guide [4]. It also provides the user with directions, interactive maps, and satellite/aerial imagery of many countries. However, the available user interaction techniques are limited to the ones based on the phone's magnetometer, which only allows the user to confirm specific actions or options into the virtual environment.

Similar to Google Maps for Android, in our application we have focused on a particular place and we have created a detailed 3D model of the environment. The main goal has been to enable the user to interact in different manners with the elements of the VR medium, using natural gestures described with his/her own hands. For movement tracking and gesture recognition we have used the Leap Motion device, presented in detail in the next section of the paper.

Leap Motion is a promising input device that, for now, is under heavy development. Its capabilities have been showcased in several demo application, one of the most complex being "Form and Function 3D" app. In this educational tool the users have the possibility to compare hearts of different animals and learn anatomy in the process. The 3D accurate anatomical models have been created based on real-world specimens and can be manipulated by the user through hand gestures through Leap motion device. The application functionalities include surface and internal views of the organs, blood flow simulations and quiz mode for knowledge evaluation.

Currently, only a few applications that combine both virtual reality and the hand tracking capabilities of the leap motion controller are available on the Leap Motion App Store [9]. Most of these applications are actually games, while our example is more concerned with educational aspects.

HARDWARE USED IN IMPLEMENTATION

Leap Motion

The Leap Motion Controller represents a new and important improvement in consumer finger/object and gesture tracking input technology. The device has been made available to the public in summer 2013, more as a prototype than a final product. Therefore until now only a few applications have been developed with optimizations for this type of user interaction and very few scientific works have been published. The device has the dimensions of a USB stick (80 mm length and 13 mm width) and allows for precise and fluid tracking of multiple hands, fingers or small objects in free space.

There are very few details known about the Leap Motion Controller's inner structure and its basic operational properties [8]. One clear fact is that it uses infrared imaging for object tracking, containing two cameras and three infrared LEDs to improve lighting conditions. Figure 1 shows an image of the controller's hardware setup.

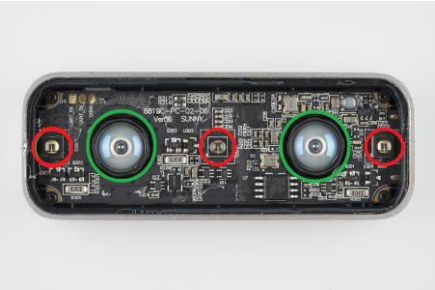

Figure 1. The 3 infra-red sensors are marked with red, while the 2 cameras are marked with green

According to the manufacturer, the sensors accuracy in position detection is about 0.01mm. However, recent research [6], [13] has shown that an accuracy of roughly 0.02mm can be obtained in realistic scenarios. In order to get objects positions from the stereovision images, all the calculations are performed on the host computer. This can result in a high machine load: about 20% processor usage for Intel Core I3, 2.4GHz laptop [12]. The gestures have an almost instant response time because the controller detects any movement with 300 frames per second rate. The data is transmitted through USB cable to the computer, which process and parses it into objects.

The developers of the Leap Motion Controller have released until now SDKs only for Windows and Macintosh platforms. Therefore a computer is needed to process all the data received from the controller and pass the results to the

smartphone. An Android SDK, which is currently under development, would streamline the application development and would significantly improve the user experience.

Google Cardboard

As briefly mentioned before, Google Cardboard is a Google project to make an affordable, cardboard made, Virtual Reality headset that uses a smart phone as the screen. It is intended as a low cost system to encourage interest and development in VR. The reason why we chose to use this headset is because of its simplicity, being a do-it-yourself project, which means you can built your own cardboard using the plans from the official website. A Google Cardboard HMD is built using the following parts: a piece of cardboard cut into a precise shape, two 40 mm focal distance lenses and a ring-magnet combo for triggering actions. The lenses are set back from where the phone is placed at a precise distance so that they can each observe only half of the smart phone screen.

How does Google Cardboard work?

In [7] it is explained that: "Most human beings use what is known as binocular vision to perceive depth and see the world in 3D. The binocular vision system relies on the fact that we have two eyes, which are approximately 3 inches apart. This separation causes each eye to see the world from a slightly different perspective. The brain fuses these two views together. It understands the differences and uses them to calculate distance creating our sense of depth and ability to gauge distance."

The assembly works by presenting two slightly different images, one to each eye. The application that runs on the phone is splitting the screen in two and is displaying the same view twice, but with a slight translation on one of the sides. Because each of the users' eyes is looking at a slightly different image than the other, the 3D effect is created. Furthermore, as the rest of the view field is blocked by the cardboard, the environment visualized by the user becomes immersive.

The sense of immersion in this Virtual Reality is further improved through head tracking technology that uses phone's sensors to detect any head movement and adjust the view accordingly. The accelerometer and gyroscope provide all the necessary data for the head tracking effect, so the user can freely look around and have that virtual reality experience.

In the initial Cardboard design, basic user interaction is provided using the magnet placed on the side of the cardboard, which acts as a click or a press. Using the magnetometer, the phone is able to sense magnetic changes caused by the magnet movement, and can transmit the input further to the application. However, this type of interaction is very limited and enables the user to only confirm or indicate a specific element. All the possible actions must be therefore described through two states controls (free state / selected).

Smartphone

Our pilot application has been tested on 5 different smartphones, namely: Samsung Galaxy Note 2, Sony XPeria Z1, HTC One M7, LG G2 and Samsung Galaxy Note 4. The application was developed to support versions of Android of 2.3 or greater.

Samsung Galaxy Note 2 has Mali-400MP GPU, which is pretty old piece of technology, compared to the other smartphones mentioned, which all have incorporated at least Adreno 320 GPU. These differences are visible in terms of performance, because the GPU is the responsible hardware for 3D processing.

Also, the screen resolution plays an important role in terms of image quality. Again, Galaxy Note 2, with a screen resolution of 720 x 1280, rendered poor images compared to the other smartphones.

Why is the screen resolution so important? Because the two lenses inside the Google Cardboard act as magnifiers, thus making the pixels visible for lower screen resolutions. Samsung Galaxy Note 4 produced outstanding results, with a resolution of 1440 x 2560 pixels with no visible pixel looking through lenses. But there is a trade-off between image quality, screen resolution and application performance: while you do get super quality images using quad-hd screens, the performance and responsiveness of the application is affected. The best overall results were obtained with the mid-range smartphones, with a resolution of 1080 x 1920 pixels, for which we recorder a good balance between image quality and performance.

CASE STUDY

As virtual reality and hand tracking are both emerging technologies, creating a natural user experience through their combination is not an easy process. In real life it comes natural to any human being to use the hands in order to interact with and control objects. The ability to use identical actions in virtual worlds, avoiding gestures that don't make sense in real environments, would make the interaction as intuitively as possible.

The application that we developed tries to demonstrate this new way of interaction with a computer by taking advantage of the two affordable devices: Google Cardboard and Leap Motion Controller, in the context of a virtual tour of a zoo. The application also has educational purposes, providing an active experience for children who want to learn more about animals in an interesting and fun way. Three different ways of interaction with the environment have been implemented: using a Gamepad (Moga Controller), using Google Cardboard's magnet and through the Leap Motion Controller.

In a VR experience, where your orientation is provided by the movement of your head, we found out that there is no

place for "clickable" devices, such as the game controller, keyboard or a mouse. These devices are not capable of providing an immersive experience, as there is no genuineness in interaction with the scene. Because the orientation in the scene is controlled through the movement of users' head there is no other use for the previously mentioned devices than a "final" action over any virtual object. Google Cardboard's magnet makes sense only for interaction within a menu, whereas for any real-world object that has an equivalent in the virtual environment the Leap Motion Controller emulates perfectly. This is the main reason why we concentrated more on the interaction through this device.

The lack of an Android SDK not only limits the experience because the Leap Motion Controller cannot be placed on the head mounted display, but also makes a computer absolutely necessary in order to be able to use the data from the controller. Beside this hardware necessity, another software responsible for forwarding the data from the controller to the application is needed.

For generating the virtual world and manipulating its objects, the Unity IDE has been used along with its Game Engine, which provides a lot of useful built in features such as the physics engine. For head tracking capabilities we have made use of the Durovis Dive SDK [3]. This functionality could have been easily implemented using the raw data from the smartphone's gyroscope, or from its accelerometer, but the reason why this plugin was used is because it implements certain algorithms to reduce the motion sickness.

The hand tracking capabilities, on the other side, were implemented using the Leap Motion Controller SDK [10]. Trinus Gyre was used to glue together the data coming from application, controller and smartphone's sensors and project it on the phone's screen.

Figure 2 shows the menu scene that is opened when the application is launched. The image was taken in Unity, the upper part representing the perspective view of the scene, visible from the Editor. The interactive objects are the menu items, with the camera objects (there are two cameras, for left and right eye, even though from the image appears to be only one) providing orientation for them. Another object that can be seen is the leap motion controller, which obviously is placed in front of the cameras. The controller and the cameras are only visible in the Editor to give the developers a preview about their positions when the application will actually be run.

The bottom part of the image is what the user actually sees. Again, there are two perspectives (one for each eye) and with a closer look it can be seen that the images are slightly distorted providing some differences between them, so that the depth can be perceived. The interaction with the menu consists of two simple gestures that are combined into one, namely *KeyTap* and *Push*. These two gestures can be viewed in Figure 3.

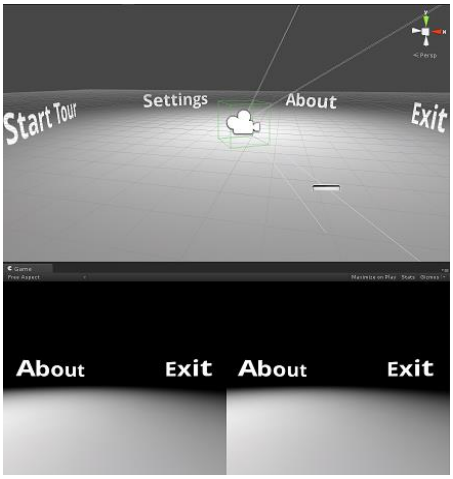

Figure 2. Menu Scene

A *KeyTap* is recognized when the tip of the index finger rotates down toward the palm and then springs back to approximately the original position, as if tapping. The tapping finger must pause briefly before beginning the tap. A *Push* gesture is recognized when your hand, while in the field of view of Leap Motion Controller, reaches the necessary downwards velocity (the hand needs to be parallel to the controller).

Because the SDK lacks a production level implementation, the tracking has some errors now and then. Even a gesture as simple as KeyTap is not recognized all the time. To improve the interaction, we decided that the selection of an item is done using a combination of KeyTap and a Push gesture, this way increasing drastically the chances that the selection will be recognized. The Push gesture was configured so that the minimum length of the hand movement is fairly small, and thus the combination of the two gesture results in a single one, very similar to KeyTap. The selected item is the one that the user is looking at when the gesture is performed (there is a small dot that indicates a ray projected from the center of the screen).

When the user selects "Start Tour", the main scene is loaded. After the scene is loaded, the character controlled by the user is placed outside the zoo, where he/she is able to see the available interaction gestures. All the gestures available in this scene can be observed in Figure 3.

In our virtual environment the walking action is achieved using the **Point** gesture (Figure 3). When the user points with one finger above the controller, and as long as that gesture is performed, the character keeps moving in the direction that the user is looking at. To stop the character from moving, one either removes the hand from the Leap Motion's field of view, or simply closes the fist.

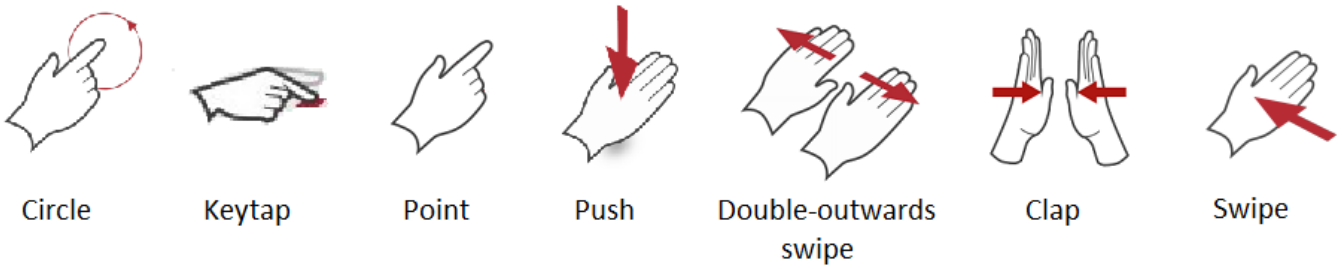

Figure 3. Interaction gestures available in the main scene

In order to enter the zoo, two actions need to be performed. The first one represents pressing a button, for which the system responds with an audio playing that gives information about the second action needed to be performed, that is opening the doors. This is accomplished using the *Double Outwards Swipe* gesture (Figure 3): the hands are placed initially above the controller, with the palms parallel to it and with their thumbs touching, and then perform a swipe with each hand in opposite directions. The character has to be placed in front of the doors for them to actually open, as it would make no sense to be able to open the doors from a far distance (see Figure 4).

Figure 4. Single eye's view during the double outwards swipe gesture

The size of the virtual world is pretty big, because we used a 1:1 scale to make virtual hands and objects look as realistic and natural as possible. Because of this, the animals, which are able to move inside their cages, can get sometimes pretty far away from the fence. In order to "call" an animal to get a closer look of it, you need to look at it and perform a *Clap* gesture (Figure 3).

As the name of the gesture suggests, you have to clap your hands once and the animal will start walking towards you until it reaches the fence. From there, you can perform a *Circle* gesture (Figure 3) to get details about the animal. A screen appears in front of the camera with all the information, which you can dismiss by performing a *Swipe* gesture.

One of the biggest problems in VR is simulator sickness, which is caused by a conflict between different sensory inputs (ear, visual field and body position). Being able to control movement reduces the experience of motion sickness. The display responds to user's movements all the time, even in menus. This is very important because, as in real world, you need to be able to change your orientation any time, regardless of the situation. Also, the virtual cameras rotate in a manner consistent with head and body movements.

Interaction challenges using Leap Motion in VR environments

Because the Leap Motion Controller is set on the table, it was virtually attached to the cameras so that the user is conscious about hands position with respect to the controller, so wherever the user is looking he/she is able to see his\her hands. If the controller would have been placed on the head mounted display, it would have made more sense to attach the controller to the "body" of the character, because our solution does not represent the way human body works.

While the software is constantly getting better at tracking potential hand poses, some will track better than others. In order to achieve the most reliable VR tracking pose, the users need to keep their fingers splayed and hands perpendicular to the Leap Motion's field of view. Many development recommendations can be found in "Leap Motion VR Best Practices Guidelines" [11].

CONCLUSION

Virtual Reality environments allow the development of complex and very detailed learning scenarios that can be experienced by the users in a very similar way to the real world. Through this approach, harmful and dangerous situations or hard to create conditions can be thoroughly analyzed as many times as necessary, without endangering the user.

One of the main limitations in the spread of this technology in educational domain is related to the high costs usually involved in developing a VR system. Google Cardboard represents an affordable and very promising virtual reality experience for the masses. Although it is a "do it yourself" headset, the results obtained are very impressive and there are no changes required in the architecture of the mobile

device. Any smart phone on the market, with decent resolution and graphical acceleration capabilities, can be transformed in a virtual display. Creation of a VR application is also an easy to do process if you're using an existing engine, such as Unity3D. The most important part for generating the depth sensation is the use of two cameras that present a slightly different image to each of the user's eyes. Another important step to achieve virtual reality is to enable the head tracking so that the user is immersed into the experience.

Until recently, most of the input devices were unfit to be used in immersive virtual worlds, feeling highly unnatural to the users. Leap Motion is a possible solution to this problem, but for now it feels more like a prototype. Even though the hardware setup seems very robust, it lacks a good software implementation for hands tracking. At the moment the device seems to be more intended for developers than for general public. The feedback provided by the users of our application leads to only one conclusion: for the interaction with the scene through Leap Motion, the functionality is only obvious after playing around with the device for some time.

In [1] Baudel and Beaudouin-Lafon explored the limitations of the gesture based interaction systems. One of the key findings was that gestural communication used more muscular activity than simple keyboard interaction, mouse interaction or speech. As a consequence, in order to avoid fatigue in users, the gestures must be chosen so they require minimal effort, be concise and fast to describe. Most of the times, the long use of this input approaches may induce fatigue in the user [2].

One other drawback we noticed is related to Leap Motion Controller and the inability to place the device directly on the head mounted display. As already mentioned, this limitation occurred because the only supported platforms for SDK are Windows and Mac. Due to this constraint the user experience is limited, he/she not being aware of his hands position relative to the controller. This also implies that a computer is required to receive the data from the leap motion controller and forward it to the application running on the smart phone.

REFERENCES

1. Baudel, T., Beaudouin-Lafon, M. Charade: Remote control of objects using free-hand gestures. *Communications of the ACM - Special issue on computer augmented environments: back to the real world*, 36 (7), 1993, pp. 28-35
2. Christova, P., Kossev, A., Kristev, I., Chichov, V.: Surface EMG recorded bybranched electrodes during sustained muscle activity. *J. Electromyogr Kinesiol* 9, 263–276 (1999)
3. Durovis Dive SDK : Available online: https://www.durovis.com/sdk.html
4. Google Cardboard. Available online : https://www.google.com/get/cardboard/ (accessed on 19/05/2015)
5. Google Maps. Available online : http://en.wikipedia.org/wiki/Google_Maps
6. Guna, J., Jakus, G., Pogačnik, M., Tomažič, S., Sodnik, J. *An Analysis of the Precision and Reliability of the Leap Motion Sensor and Its Suitability for Static and Dynamic Tracking*. Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, Ljubljana 1000, Slovenia
7. Gurman. J.B. STEREO Home Page http://stereo.gsfc.nasa.gov/classroom/how.shtml
8. Jakus, G., Guna, J., Tomazic, S., Sodnik, J. Evaluation of Leap Motion Controller with a High Precision Optical Tracking System, *Human-Computer Interaction Advanced Interaction, Modalities, and Techniques*, 16th International Conference, HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II, pp.
9. Leap Motion App Store. Available online : https://apps.leapmotion.com/categories/vr-beta
10. Leap Motion Controller. Available online: https://www.leapmotion.com (accessed on 19/05/2015)
11. Leap Motion VR Best Practices Guidelines. Available online: https://developer.leapmotion.com/assets/Leap%20Motion%20VR%20Best%20Practices%20Guidelines.pdf
12. Leitão, P.M.O. Analysis and Evaluation of Gesture Recognition using Leap Motion, *Proceedings of the 10th Doctoral Symposium in Informatics Engineering - DSIE'15*, 2015, pp. 83 – 94
13. Weichert, F.; Bachmann, D.; Rudak, B.; Fisseler, D. Analysis of the Accuracy and Robustness of the Leap Motion Controller. *Sensors 2013*, 13, pp. 6380–6393.
14. Winner, L. Do artifacts have politics? In *The Social Shaping of Technology* (2nd. ed.), Donald MacKenzie and Judy Wajcman (eds.). Open University Press, Buckingham, UK, 28-40.

Real Time Visualization of Crowd Dynamics Scenarios

Dan Razvan Ilies
Technical University of Cluj-Napoca
Str. G. Baritiu 28, 400027, Cluj-Napoca, Romania
danilies92@gmail.com

Adrian Sabou
Technical University of Cluj-Napoca
Str. G. Baritiu 28, 400027, Cluj-Napoca, Romania
adrian.sabou@cs.utcluj.ro

Dorian Gorgan
Technical University of Cluj-Napoca
Str. G. Baritiu 28, 400027, Cluj-Napoca, Romania
dorian.gorgan@cs.utcluj.ro

ABSTRACT
This paper presents an approach to real-time simulation of crowd dynamics on GPU enabled computing architectures. We discuss challenges with parallelization of agent-based models, implementing parallel simulation algorithms, visualization and interaction with the simulated scene and, most importantly, ensuring communication and synchronization between all these processes. Our main objective is to provide interactive simulation of realistic models such as pedestrian dynamics, in which large crowds move and interact among themselves and with the environment. Simulation parameters like scene complexity, scene composition, as well as the number of agents are varied in order to simulate different scenarios and to assess the impact on performance.

Author Keywords
Visualization; Crowd dynamics; Interactive simulation; Social forces; Social models; Graphics Processing Unit.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Algorithms; Human factors.

INTRODUCTION

Applications in computer animation and simulation vary from simple scenarios like physically-based simulations, which focus on visual quality, perceived realism and computation efficiency, to complex scenarios which require an accurate reproduction of measured parameters. The latter are usually encountered in critical systems' simulation, based on heavy numerical computation, with the aim of ensuring exact calibration with real-world conditions. Such examples can be encountered in Computational Fluid Dynamics, modelling turbulent flows or crashworthiness tests. All these applications require a huge amount of processing power and utilize powerful techniques borrowed from fields such as applied mathematics, numerical analysis, computational physics or mechanical engineering in order to satisfy the ever increasing expectancies that users have with regard to simulation systems.

Over time, sequential computing devices such as the Central Processing Unit (CPU) have evolved significantly. At first, there was little change in algorithms and techniques from one version to another, the performance mostly being improved by raising the operating frequency. Once the thermal barrier was reached, the parallel era began. Thus CPUs evolved into powerful multicore architectures and this direction led to creating a new type of computing device, dedicated to massive parallel computation, the Graphics Processing Unit (GPU). This new computing direction brought forth the need for efficient parallel techniques and algorithms in order to fully utilize existing resources.

GPUs were at first used to process massive parallel data involved in Computer Graphics and complex graphical algorithms, but, because of their massively data parallel architecture, they drew the attention of the High-Performance Computing (HPC) community and researchers began using them for general purpose parallel computation. Thus the General Purpose Computation on the GPU (GPGPU) current began. However, programming GPUs for general purpose computation proved difficult to learn, so GPU vendors began making their devices more flexible and easier to program through interfaces such as CUDA and frameworks such as OpenCL.

Among applications to benefit from GPU-based acceleration are discrete, particle-based or agent-based simulations, their inherently parallel nature making them perfectly suitable for GPU-based implementations. Thus, in the last decade, research work in this domain has produced new, highly efficient techniques for simulations, especially for the most time consuming parts. However, when adding the real-time attribute that such simulations usually require, along with the need to transform raw processed data into a form visually meaningful to the users, we end up with probably one of the most complex class of applications attempted to be implemented on such architectures.

This paper presents an approach to real-time simulation of crowd dynamics on GPU enabled computing architectures. We discuss challenges with parallelization of agent-based models, implementing parallel simulation algorithms, visualization and interaction with the simulated scene and, most importantly, ensuring communication and synchronization between all these processes. Our main

objective is to provide real-time simulation of realistic models such as pedestrian dynamics, in which large crowds move and interact among themselves and with the environment. Simulation parameters like scene complexity, scene composition, as well as the number of agents are varied in order to simulated different scenarios and to assess the impact on performance. The rest of the paper is organized as follows: the following section highlights on existing related works, the third section presents an overview of the model used for our simulations, section four focuses on techniques used for achieving interactivity and real-time visualization, while section five presents some experimental scenarios and performance measurements.

RELATED WORKS

The term of crowd dynamics refers to a system of behaviors and psychological processes that arise in a social group or between several such groups. Studying the dynamics of crowds can be useful for understanding behaviors in decision making, information diffusion or disease spreading in society, as well as in many other contexts. Such applications can be seen in domains such as: psychology, sociology, political sciences, business or education.

The idea that macroscale behavior can be deduced from microscale interactions between individuals has become key concept in understanding human behavior in crowds. Examples of such collective behaviors have been represented in several ways, one of them being the social forces model.

A social forces model sufficiently complex to allow for realistic computer-based simulations of human crowds has been introduced by Helbing and Molnar [3]. They claimed that pedestriam motion can be described in such a way as to be the result of social forces. These forces are not directly exerted only by the environment, but are rather a measure of the internal motivation of each individual that arrises in order to guide the individual towards certain actions or movements.

After the initial introduction of this model, different improvements and variations started to emerge, since it was found that it could also be applied to a variety of domains other than pedestrian dynamics, like biology, for simulating the behavior of microscopic particles. Different variations of attraction and rejection forces lead to new models being created that described the behavior of large crowds of particles even better than the initial one and that would realistically reproduce several observed phenomena. Thus several important works emerged which present collective patterns, [8] [10] [6] and [4].

However, the assumptions that the above mentioned model relies on and the exact form of the presented social forces has never been measured or validated empirically, even if the functions describing the interaction between individuals could definitely influence the behavioral patterns that were the outcome of the simulations. This has been proved by a series of studies on different animal species [1]. The most accurate and correct studies were limited to calibrating the assumed parameters for the interaction forces in order to minimize the errors in predicting individual behaviors.

Regarding simulating social models on high-performance architectures, Joselli et al. [5] present a case study in which they evaluate the performances of a simulation system with one CPU and two GPUs. The model that they use is an agent-based one, in which each individual has several properties and interact with their environment. They used a series of data structures and algorithms in order to accelerate the computation, most of them done in the GPU. Among the used techniques are spatial hashing, in which each agent is assigned a hash code based on its location in the simulated scene. The authors have chosen to split the scene among the two GPUs in such a way as to allow each one to process part of the scene and to synchronize common areas. Their experiments show a speedup up to 1.8 for a very large number of agents. However, the authors do not specifically discuss real-time visualization and techniques that might improve performance when interactivity is required.

Sabou et al. [9] present an extension of particle models with regard to their initial purpose and propose a solution for simulating sociophysics models interactively using a particle-based visual approach. An existent agent-based "small-world" model is mapped on a particle-based grid and its evolution in time is simulated on a high-performance graphics cluster in order to model technology adoption and consumer behavior. Several experimental scenarios validate the initial hypothesis that particle-based models can be extended beyond their original scope and evaluate the system's performance and scalability.

SIMULATING CROWD DYNAMICS

As seen in the previous section, agent-based models contain individuals that interact in a given environment. The agents may be either distinct computer programs or distinct parts of the same program, with the purpose of representing social actors – persons, organizations or nations. These agents are programmed to react to the computing environment that they are placed in, a model of a real environment in which the agents would interact.

One crucial aspect of agent-based modeling is that agents need to be able to interact, i.e. exchange information carrying messages and to behave accordingly with what they learn from these messages. The messages can be, for instance, specific dialogs between persons, but also indirect means of transferring information such as observing another agent or detecting the effects of another agent's actions. The possibility to model such interactions is the main way in which agent-based modeling differs from other computational models.

Of course, keeping in mind that this is a visual simulation, the agents will have to be visually represented in one way or another. In our particular case, we have chosen a representation in a bi-dimensional virtual scene in which each entity is described by its *X* and *Y* coordinate. This way, users can observe in real-time the movement of agents and their interactions.

We will now describe the construction of the simulated scenario and the different types of forces that act on agents.

Defining the simulation scenario

Besides the main actors of the simulation, which are the agents, in order to obtain a realistic model, we had to introduce several other elements. Thus, we will have:

Scene boundaries

It was decided that all simulations were to take place in a restricted environment which can be defined to the user's best suiting. This way, the one that initiates the simulation has the possibility to choose the horizontal as well as the vertical limits that will constitute the boundaries of the simulated scene. Agent movement is restricted between these user defined limits, providing for a more controllable scenario and a better visualization experience. The boundaries will be drawn as simple straight walls through which agents cannot pass.

Obstacles

Besides the aforementioned boundaries, the interior of the scene will contain different obstacles that will influence the agents' trajectories. The obstacles can have various shapes, ranging from simple walls to complex polygonal objects. Same as with boundaries, the users can control obstacle placement as they see fit, both their position and their shape.

Social forces

Each agent has a clearly defined objective during the simulation. Furthermore, during their movement towards their objective, there will be a series of interactions, both agent-agent and agent-other elements. The model for the forces that act upon agents and the equations of movement are the same originally proposed by Helbing and Molnar [3]. As they said, it is often believed that human behavior is chaotic or unpredictable, but, for relatively simple situations, certain behavior patterns can be created, among which, the social forces model. Due to the fact that pedestrians are already used to a multitude of situations, their reactions are most often automatic, based on their similar previous experiences. Thus, the velocity and the direction of each pedestrian could be represented as a vector quantity $\overrightarrow{F_\alpha}(t)$, which is the so called social force and which represents the effect of several other forces that the environment and other pedestrians generate. In what follows, we will briefly describe the types of forces that influence the pedestrians' movement.

Attraction forces towards the objective

This is the main force that drives the agent towards its goal. Normally, agents will take the shortest route, which is a straight line, unless they encounter obstacles, in which case they will temporary modify their objective in order to avoid them. The formula that computes these forces is:

$$\vec{F}_\alpha^{\,0}(\vec{v}_\alpha, v_\alpha^0 \vec{e}_\alpha) := \frac{1}{\tau_\alpha}(v_\alpha^0 \vec{e}_\alpha - \vec{v}_\alpha) \quad (1)$$

where $\vec{v}_\alpha$is the current speed, $v_\alpha^0 \vec{e}_\alpha$ is the desired speed and $\tau_\alpha$is the relaxation time (i.e. the delay in agent acceleration).

Repulsion forces from obstacles

On their way towards the objective, an agent ca encounter different types of obstacles. Normally, even in real life, pedestrians keep their distance to obstacles such as walls or other kind of objects, thus we require a formula to express a force of repulsion coming from obstacles. This formula is the following:

$$\vec{F}_{\alpha B}(\vec{r}_{\alpha B}) := -\nabla_{\vec{r}_{\alpha B}} U_{\alpha B}(\|\vec{r}_{\alpha B}\|) \quad (2)$$

where $\vec{r}_{\alpha B}$is the distance between the agent's current position and the obstacle and $U_{\alpha B}(\|\vec{r}_{\alpha B}\|)$ is a monotonic decreasing potential that scales the repulsion force with the distance between the agent and the obstacle.

Repulsion forces from other agents

Similar to the previous case, when a repulsion force is generated by the obstacles, repulsion forces are generated from other agents. When agents are in proximity, this repulsion force appears that is intended to keep agents from violating each other's "personal space". The repulsion depends on the distance between agents and their relative velocities. For great distances, this force is negligible, but as agents get closer, it will increase exponentially. The formula to compute these forces is:

$$\vec{F}_{\alpha\beta}(\vec{r}_{\alpha\beta}) := -\nabla_{\vec{r}_{\alpha\beta}} V_{\alpha\beta}[b(\vec{r}_{\alpha\beta})] \quad (3)$$

where $V_{\alpha\beta}[b(\vec{r}_{\alpha\beta})]$ is a monotonic decreasing function with the same role as before and *b* is the small radius of the elliptic shaped personal space of the agent.

Acceleration structure based on hash codes

In order to speed-up the neighbor search when dealing with a large number of agents, we apply a spatial hashing technique. The simulated world is split into a grid of cells (Figure 1), each agent belonging to a single such cell at any given time. Thus, we have to establish a relation between the agent and the cell, based on the agent's position.

The basic idea is to firstly determine the position of an agent in the scene and secondly we must search for the cell that contains that position. To speedup calculations, each agent is assigned a hash code, computed using the following formula:

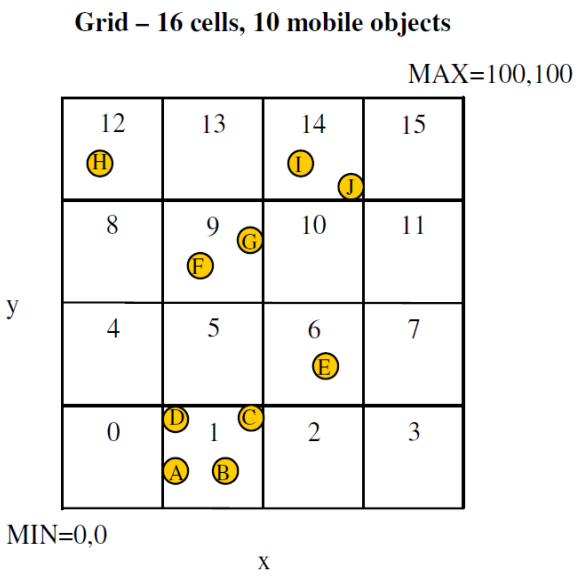

Figure 1. Example grid for 16 cells and 10 agents [2].

$$HashCode(agent) = \big((x * p1)\ xor\ (y * p2)\big)\ mod\ p3 \quad (4)$$

where

$x = agent.x/cell\_size$

$y = agent.y/cell\_size$

and *p1*, *p2* and *p3* are large prime numbers.

The role of this formula is to assign a code to each agent based on its position in the scene. The purpose of these codes is to use them to obtain an agent-cell association, which explains why the code computation depends on the size of the cells dividing the simulation space. Code computation for each agent must occur at each simulation step since agents modify their positions as long as their objective was not reached. Moreover, to be able to associate agents and cells, the latter need to have an assigned code, computed using the same formula. Since cells do not modify their positions, their codes can be precomputed.

Thus, after computing the aforementioned codes, each agent will be assigned the hash code of the cell that it belongs to, making it easy to quickly determine the neighbors for each agent during the simulation process. In order to guarantee a unique code for each cell, the prime numbers must be much larger than the total number of cells. Code computing is done using the specified formula for better performance and computing speed.

REAL-TIME SIMULATION AND VISUALIZATION

Accelerating computation using the GPU

GPUs are electronic components specially designed to execute a huge number of operations in parallel. Their initial purpose was to create raster images in a framebuffer to present through a display device, but they have recently started to be used more and more for applications and systems designed to offer a huge degree of parallelism.

GPU-accelerated computing is a technique that uses a GPU together with a CPU to accelerate scientific, analytic or engineering applications. This path was opened by NVIDIA in 2007 and a level was reached in which GPUs power entire data centers, especially power efficient ones serving Universities and small and medium enterprises [7]. GPUs accelerate a wide range of applications, from applications in the auto industry to mobile phone apps, drones or robots, offering superior performances.

The way this acceleration works is by taking over the intense workloads from the CPU and running them in the GPU, while the rest of tasks continue running in the CPU.

In our case, the main and most complex entities are the agents. Most computation is done around them, the rest of the scene being mainly static. Thus, fast processing of agents would lead to better overall simulation performance. Seeing as all agents require the same set of operations at one time, processing them on the GPU is the best choice. Each agent will be processed by a different GPU thread and necessary data will be transmitted between host and device.

After deciding on all elements required for host-device communication, the simulation can be attempted. We must firstly configure all simulations and visualization parameters and after that we can start an infinite loop. This simulation loop contains two main phases: the computation phase and the visualization phase.

The computation phase is done in a distinct function which is called from inside the infinite loop and contains a series of commands to be executed, either in the CPU or on the GPU. Computing new agents' positions is done in three distinct steps:

1. The first kernel computes the hash codes for all agents using their positions in the scene.

2. The second step consists of computing the hash structures that allow for fast discovery of all agents inside of a specified cell, using the second kernel.

   This way, we get two data structures with the same dimension as the number of cells, the first one indicating which is the first agent inside a certain cell and the second one indicating what is the last agent in that cell, with regard to a data structure in which agents are ordered by their hash code.

3. The third step is the actual computing of forces, new positions and velocities for all agents, using the third kernel.

Once computation is finalized, the scene must be prepared for rendering.

Rendering is done as a two-step process. The first step involves rendering all elements in the scene besides the agents (scene boundaries, obstacles, grid of cells, etc.). The second step involves rendering the agents themselves. As the number of agents becomes sufficiently large, the computing process is no longer the only issue that has an impact on performance, the sheer size of the crowd imposing penalties upon the rate at which visual information is rendered. Seeing as the positions of all agents are computed on the device and stored into OpenCL buffers, they would normally have to be copied back and forth between host and device in order to render them, generating a large number of memory transfers. The solution is to combine the GPU computing and the GPU rendering by using an interoperability mode between OpenCL and OpenGL.

Combining computation and visualization

In order to be able to utilize the GPU for both GPGPU computation and as a traditional rendering pipeline, we must avoid unnecessary memory transfers while switching between operating modes, since the general purpose computation and the rendering process basically use the same data, namely particle positions.

This calls for an interoperability solution between OpenCL and OpenGL that can be achieved through a special data structure called a Vertex Buffer Object (VBO) and which allows for OpenGL data manipulation by OpenCL, without the need to transfer data back and forth (Figure 2). The VBO is an extension for OpenGL intended to improve performance by providing benefits of vertex arrays and display lists while avoiding downsides of their implementation. VBOs allow vertex array data storage in high-performance graphics memory on the server side and efficient data transfer. Using VBOs, the number of function calls and redundant usage of shared vertices can be reduced.

TEST SCENARIOS

In this section we present in detail certain test scenarios that were executed to validate and evaluate our solution. The experiments were carried out by varying different elements in the scene such as the number of agents, the number and size of obstacles or simply by creating some special scenarios.

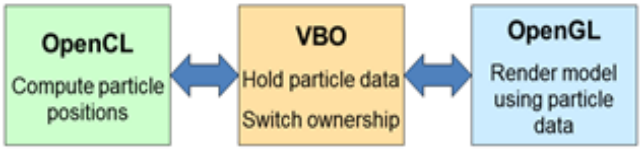

Figure 2-OpenGL/OpenCL interoperability.

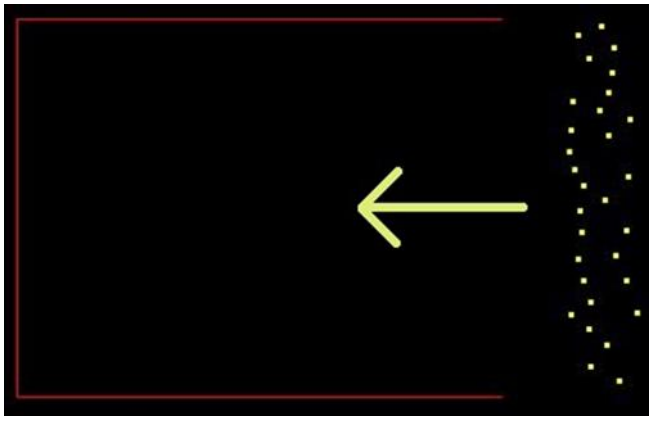

Figure 3 - Experiment 1

Besides the proper functioning of our application, we wanted to emphasize the differences in performance that appear when executing just on the CPU executing on hybrid CPU/GPU architectures, the differences being quite notable. In order to be able to run these tests, we implemented a version of our application that runs entirely on the CPU, this way avoiding all the GPU configuration overhead.

Testing with regard to the number of agents

The first type of test and one of the most important for noticing the differences in performance was done by increasing the number of agents in the scene.

The number of agents has the greatest impact on application performance, because agents are the main actors and all processing is done around them.

Except for computation for splitting the scene in a grid of cells and computation strictly regarding the scene, which are quite few, all other computation is done in order to compute forces, velocities and new positions of each agent. Thus, a huge impact on overall performance was to be expected.

In order to best emphasize the performance gain obtained on GPUs, besides observing the simulated scenario and the evolution of the simulation, we ran the simulation both on the CPU and on the GPU and compare the results. The first test consists of a simple simulation scenario with few agents placed on the right side (Figure 3). We computed the time required for all agents to migrate to the left side of the scene.

After several successive runs using both application versions, a significant increase in processing time is noticed on the CPU-based one. Table 1 shows the computed execution times.

Number of agents	CPU execution time	CPU+GPU execution time
10	3 s	5 s
100	6 s	5 s
250	8 s	6 s
500	14 s	7 s
750	43 s	8 s
1000	63 s	9 s

Table 1 - Execution times

These test easily prove the performance gain obtained by parallelizing the application. We will also compare performance results between the two versions of the application in subsequent tests.

Testing with regard to scene partitioning

When dividing the scene into a grid of cells in order to speed up the neighbor searching procedure, the number of cells (and thus their size) can vary, depending on scene complexity and size, or number of processed agents. There is no general formula to determine the optimum number of cells, thus we shall try to determine them empirically.

When using smaller cell sizes (Figure 4) we obtain the advantage of processing only a small part of the scene when computing agent interactions. Due to the fact that a smaller region surrounding each agent is taken into account, the number of neighboring agents is relatively small, thus generating less computation. Since computation for each agent is handled by a single GPU thread, this should count pretty much.

On the other hand, when using a larger cell size (Figure 5), even if the neighborhood is larger, the data structures holding hash info is considerably smaller. Thus, searching

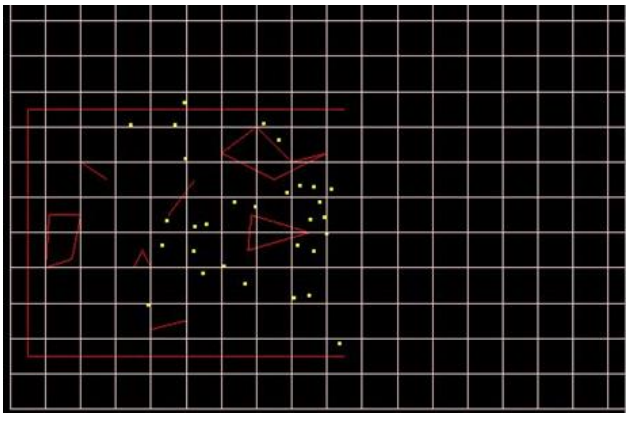

Figure 4 - Experiment 2 - Small cell size

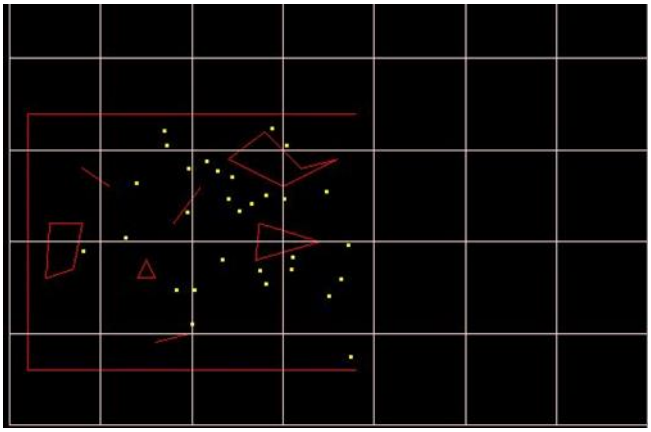

Figure 5 - Experiment 2 - Large cell size

for neighboring cells as well as neighboring agents is faster, which should account for a performance gain, even if the number of neighbors for each agent is larger than in the previous case.

Experiments showed that the best performance results are obtained for a balanced partitioning of the scene, which means that the relative dimension of the cells with regard to the scene should be chosen in a way as to ensure that the entire scene is covered by approx. 50-100 cells. Even if for a small number of agents the differences are not obvious, for large number of agents, this will impact on execution times.

Testing with regard to agents' trajectories

The third types of tests were carried out in order to validate the correctness of the implemented model and of interactions between simulation elements. This was carried out by generating the agents in several ways and varying their objectives.

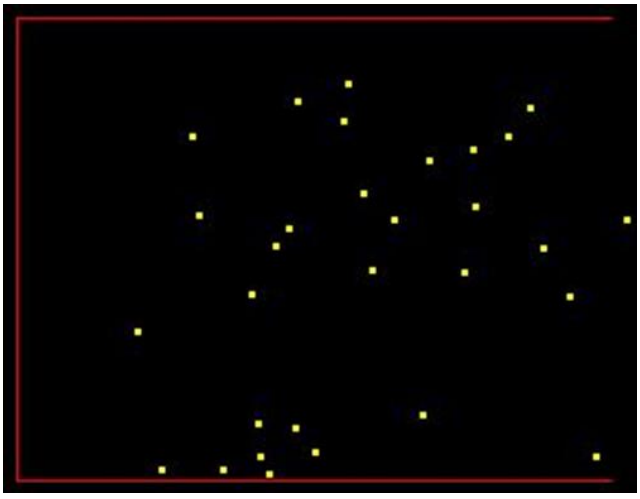

Figure 6 - Experiment 3-1 - Random agent generation

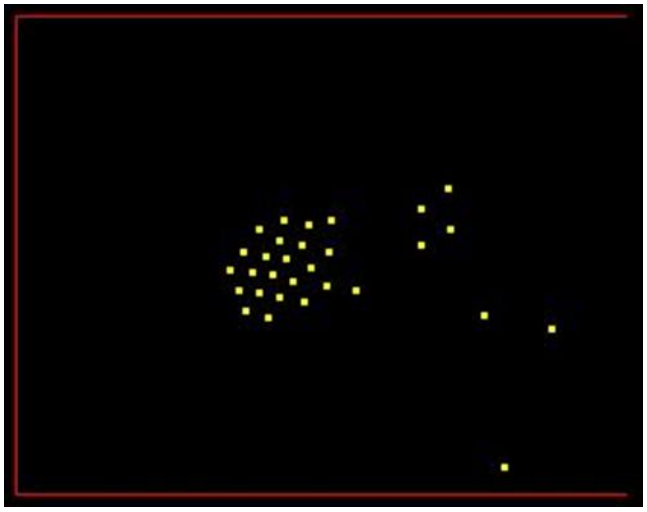

Figure 7 - Experiment 3-2 - Clustered agents

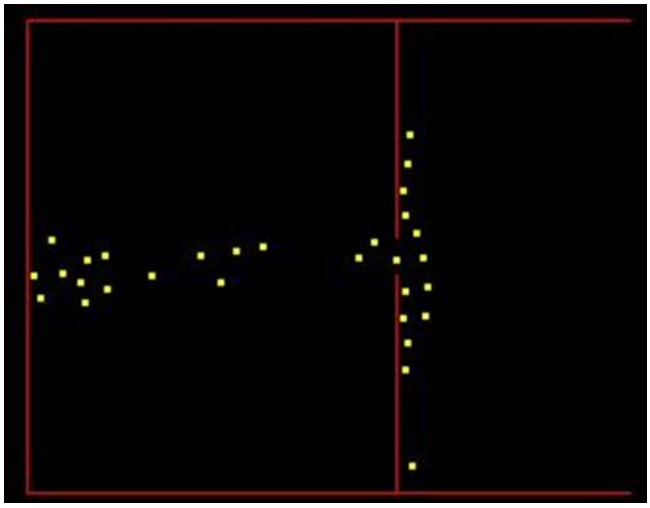

Figure 8 - Experiment 3-3 - Agents forced through a small opening

For the first test, we generated all agents at random positions in the scene with random objectives (Figure 6). Although the movement in the scene was chaotic, no interaction problems were detected.

For the second test, the agents were programed to cluster in the middle of the scene (Figure 7) in order to test their behavior in a crowded environment, but without supplementary difficulties posed by obstacles, with many agent-agent interactions. The simulation was once again without problems, although a small decrease in performance was noticed when all agents were in close proximity.

For the third test, all agents were forced through a tight opening (Figure 8). This test extends the previous one, but this time with difficulties posed by obstacles. The agents' behavior was consistent with the previous test.

It is worth mentioning that all three tests were carried out on both the CPU and the GPU version of the application and the results were consistent with performance measurements in the first experiment, namely the simulation time for the CPU version increases considerably with the number of agents, while for the GPU version, the increase in execution time is considerably smaller.

Testing with regard to scene complexity

Last but not least, we followed the impact that the scene

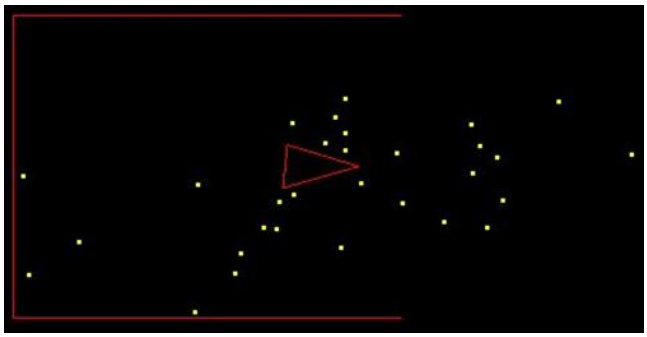

Figure 9 - Experiment 4 - Scenario 1

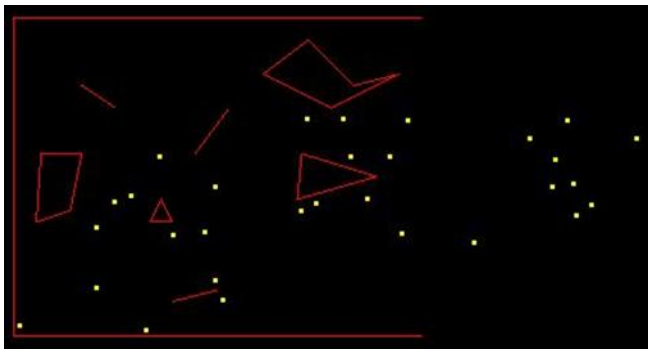

Figure 10 - Experiment 4 - Scenario 2

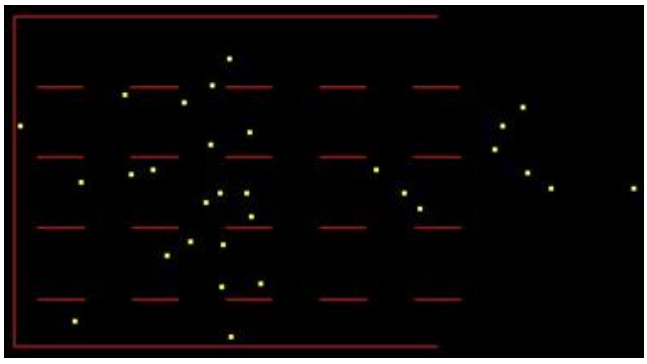

Figure 11 - Experiment 4 - Scenario 3

complexity has on overall simulation performance. Several scenarios were generated, ranging from simple ones, with 1 or 2 obstacles (Figure 9) to complex ones containing a much larger number of obstacles of different shapes (Figure 10, Figure 11). All agents were generated in the same positions for all scenarios in order to observe just the influence of scene complexity.

Results show that even if the scene is far more complex, there is an insignificant increase in simulation times when compared to differences in times for the previous experiments, when we were varying the number of agents.

CONCLUSIONS

Agent-based models are one of the best methods for realistic simulations of a complex environment or system which are usually non-linear and for which no simple and intuitive solutions exist that can offer precise results. Even so, the complex computation during the simulation remains an issue, due to the fact that the complexity of the studied system leads to heavy calculations in order to obtain results close to the real world.

Simulation techniques based on GPUs are an excellent solution to these problems, due to their parallel architecture, capable of executing hundreds and thousands of difficult operations at once. Test results showed that there is a considerable improvement in performance when using at least a GPU for simulations. However, as the complexity of the studied systems increases, a single GPU ceases to be sufficient, thus requiring more powerful architectures such as GPU clusters in order to carry out these simulations.

This paper presented key concepts for designing a real-time crowd dynamics simulation and visualization system that works in a hybrid CPU/GPU architecture, as well as having the potential to be extended for multiple GPU equipped nodes.

The simulation system that was presented proved the advantages that such a hybrid CPU/GPU architecture can have over traditional CPU-based architectures. Test results show significant improvements in all simulated scenarios, with the most significant one occurring when the number of agents was large. This is good news, since, with these simulations, the main entities are the ones that matter and that are wished to be present in a large number. This confirms that the approach presented is a promising one for interactive simulation of large crowds. Future development plans include porting the application to a GPU cluster in order to accelerate the computation process even further and to allow the user to interact in real-time with the simulated scenario.

ACKNOWLEDGMENTS

The research has been carried out in the Computer Graphics and Interactive Systems Laboratory of the Computer Science Department, in the Technical University of Cluj-Napoca, and partially supported through PN-II-PT-PCCA-2013-4 project funded by MEN-UEFISCDI, Contract no. 344/2014, PECSA - Experimental High Performance Computation Platform for Scientific Research and Entrepreneurial Development.

REFERENCES

1. Iain D. Couzin, Jens Krause, Richard James, Graeme D. Ruxton and Nigel R. Franks. 2002. Collective Memory and Spatial Sorting in Animal Groups. *Journal of Theoretical Biology, vol. 218, no. 1*.
2. Erin J. Hastings, Jaruwan Mesit and Ratan K. Guha. 2005. Optimization of Large-Scale, Real-Time Simulations by Spatial Hashing. *In Proceedings if the 2005 Summer Computer Simulation Conference 37 (4), 9-17*.
3. Dirk Helbing and Peter Molnar. 1995. Social force model for pedestrian dynamics. *Phys. Rev. E 51, 4282*. http://dx.doi.org/10.1103/PhysRevE.51.4282.
4. Anders Johansson, Dirk Helbing and Pradyumn K. Shukla. 2007. Specification of the Social Force Pedestrian Model by Evolutionary Adjustement to Video Tracking Data. *Advances in Complex Systems, vol. 10*.
5. Mark Joselli, Jose Ricardo Da Silva and Esteban Clua. 2014. An Architecture for Real Time Crowd Simulation Using Multiple GPUs. *In Proceedings of the 2014 Brazilian Symposium on Computer Games and Digital Entertainment (SBGAMES)*.
6. Taras I. Lakoba, D. J. Kaup and Neal M. Finkelstein. 2005. Modifications of the Helbing-Molnár-Farkas-Vicsek Social Force Model for Pedestrian Evolution. *Simulation, vol. 81, no. 5*.
7. NVIDIA. CUDA. 2015. Retrieved July 21, 2015 from http://www.nvidia.com/object/cuda_home_new.html.
8. Alessandro Pluchino, Cesare Garofalo, Giuseppe Inturri, Andrea Rapisarda and Matteo Ignaccol. 2014. Agent-Based Simulation of Pedestrian Behaviour in Closed Spaces: A Museum Case Study. *Journal of Artificial Societies and Social Simulation 17 (1) 16*.
9. Adrian Sabou, Dorian Gorgan and Ioan Radu Peter. 2014. Interactive Particle-based Simulation of Sociophysics. *In Proceedings of the 2014 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*.
10. Andreas Schadschneider, Ansgar Kirchner and Katsuhiro Nishinari. 2002. CA Approach to Collective Phenomena in Pedestrian Dynamics. *In Proceedings of the 5th International Conference on Cellular Automata for Research and Industry*.

Real-time Video Processing in Web Applications

Cristian Ionita
Academy of Economic studies
Bucharest, Romania
crionita@ie.ase.ro

Alexandru Barbulescu
Academy of Economic studies
Bucharest, Romania
alexbarbulescu@ie.ase.ro

ABSTRACT

The OpenGL ES standard is implemented in modern desktop and mobile browsers through the WebGL API. This paper explores the potential for using OpenGL ES hardware acceleration for real time video processing in standard HTML5 applications. We analyze the WebGL performance across device types and compare it with the standard JavaScript and canvas performance.

Author Keywords
WebGL; GLSL; OpenGL ES; kernel; convolution matrix; JavaScript

ACM Classification Keywords
H.5.2 User Interfaces

INTRODUCTION

The present paper proposes a method of using hardware acceleration available in existing devices in order to enable real time video analysis and processing in standard web applications. It measures the performance of the proposed solution, compares it with existing techniques and determines the range of devices and algorithm types that are feasible using the current methods.

HARDWARE ACCELERATION IN WEB APPLICATIONS

WebGL (Web Graphics Library) is a JavaScript API (Application Programming Interface) which offers access to 2D and 3D rendering in any compatible web browser. One of its advantages is that no add-on or plugin is necessary. This is used in HTML <canvas> elements. WebGL was initially developed by Vladimir Vukićević ([1]) working for Mozilla Foundation [3]. Initial release was in March 2011 but a stable release has been made two years later in March 2013. WebGL is a cross platform library also based on OpenGL ES [4]. The first experiments of the author were made in 2006 and in 2007 Mozilla and Opera did their individual experiments in order to prove the usability of the concept. In 2009 the Kronos Group started the WebGL Working Group and in 2011 the first version of the standard was published. The standard was adopted by the major web browsers developers in 2013.

Although the standard was adopted in 2013 the technology was widely adopted by the industry only in the last year. The rapid adoption in the last 12 months (May 2014 – May 2015) confirms the industry's interest in the technology across all device types: desktop 62% → 90% and mobile 29% → 78% ([5]). The Kronos Group also published the WebCL – Web Computing Language standard in March 2014, but currently no browsers natively support it.

One important concept is that WebGL allows using low level programming and the power of the hardware graphics processing unit. The library makes use of two buffers: frontbuffer which is the image currently visible and backbuffer which is the image being rendered. The browser can move the backbuffer to the frontbuffer at any time except during the execution of JavaScript code.

WebGL can draw points and lines but the basic shape for 2D and 3D drawings is the triangle. This is a primitive figure for 3D drawings because a plane is uniquely defined by 3 points. WebGL uses the parallel power of the GPU (graphics processing unit) and the integrated or shared memory to compute the final image.

In the last ten years the evolution of computer graphic cards transformed them from a component that renders images processed by CPU to one that delivers graphics processing units being able to make real time processing and transformation. In programming concepts, these moved from Fixed Function Pipeline (FFP) to Programmable Pipeline (or shaders) as in [2]. In a programmable pipeline with GPU the programmer establishes the vertex transformation and fragment processing using high level programming languages like GLSL. The compiled programs are loaded to the GPU and the tasks are automatically executed. This is done by a complex system consisting of hardware and drivers. The drivers are exposing the hardware functionalities to the programmers and graphics libraries.

The modern GPU exposes fully programmable hardware units (also named shaders units). The name came from the idea that the units are connected in a pipeline and the output of one shading unit is the input for the next. The Vertex Shader is the programmable Shader stage in the rendering pipeline that handles the processing of individual vertices. According to OpenGL documentation Vertex shaders are fed Vertex Attribute data, as specified from a vertex array object by a drawing command. A vertex shader receives a single vertex from the vertex stream and generates a single vertex to the output vertex stream. A Geometry Shader (GS) is a Shader program written in GLSL that governs the processing of Primitives. Geometry shaders reside between the Vertex Shaders (or the optional Tessellation stage) and the fixed-function Vertex Post-Processing stage. The geometry shader is optional and does not have to be used. After creation of vertices geometry we need to create pixels. This is done by the rasterization process. This process takes all the primitives and splits them into

individual fragments which are colored by fragment shader and are turned into frame buffer. The programmable fragment shader unit takes the fragments produced by the rasterization process and executes an algorithm provided by a graphics developer to produce the final color, depth and stencil values for each fragment. This part can be used to achieve special visual effects, including post-processing filters.

One special part of the image processing is the use of parallelization. Processing the images involves many repetitive operations that are very time consuming. The architecture of modern GPUs is based on many parallel execution units and is appropriate for data parallel algorithms.

KERNEL IMAGE PROCESSING

The bitmap image is represented by a matrix of pixels. Many image processing algorithms can be expressed using a convolution process to apply a kernel to an image. This is done by applying a mask also named filter matrix or kernel. It involves performing matrix operations in order to calculate the result matrix. In the result matrix every pixel value is calculated from the initial matrix multiplied with the kernel matrix. In most of the cases the kernel matrix consists of 5x5 or 3x3 values. These are enough for obtaining most of the effects.

10	52	63	42	74
86	24	**45**	28	82
62	91	17	24	2
49	19	18	36	75
41	15	78	17	14

0	1	0
0	0	1
0	0	0

Figure 1. Convolution matrix

The new value for 45 is calculated in the following manner:

52x0+63x1+42x0+24x0+45x0+28x1+91x0+17x0x24x0=91

The method can be used for implementing edge detection algorithms such as Sobel [6] and Frei-Chen [7]. The goal of an edge detection algorithm is to identify points of an image where the intensity changes abruptly. There are many factors that need to be taken into consideration, such as surface orientation, discontinuities, lightning changes, very similar textures, etc. The aim of edge detection is to apply edge detectors to an image and to receive a set of lines or curves that delimits objects. The result is a gray-scale image where each pixel value indicates whether it is or not on the boundary of an object. There are many algorithms and the results are very sensitive to image characteristics. Usually it is better to try multiple edge detection algorithms and choose the one that is best for the case. In every algorithm there is a need to establish the threshold. For every pixel, if the value is above the threshold it is considered part of the edge, otherwise not. After this step named binarization we can use other algorithms to discover the edges in the image.

The Sobel filter uses two 3x3 convolution matrices to detect vertical and horizontal gradients of the image.

$$G_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \qquad G_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix},$$

$$|\Delta I| = \sqrt{(G_x * I)^2 + (G_y * I)^2}$$

Figure 2. Sobel edge detection matrix.

These masks are applied to the 3x3 footprint of every RGB color in the image. The results are then used to obtain the gradient value.

Another algorithm used for edge detection is Frei-Chen edge detector. The algorithm uses nine 3x3 convolution masks. The weighted sum of all convolution results is used to determine the final value for each pixel.

$$G_1 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & \sqrt{2} & 1 \\ 0 & 0 & 0 \\ -1 & -\sqrt{2} & -1 \end{bmatrix} \quad G_2 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & 0 & -1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & 0 & -1 \end{bmatrix} \quad G_3 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 0 & -1 & \sqrt{2} \\ 1 & 0 & -1 \\ -\sqrt{2} & 1 & 0 \end{bmatrix}$$

$$G_4 = \frac{1}{2\sqrt{2}}\begin{bmatrix} \sqrt{2} & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -\sqrt{2} \end{bmatrix} \quad G_5 = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad G_6 = \frac{1}{2}\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

$$G_7 = \frac{1}{6}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \quad G_8 = \frac{1}{6}\begin{bmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{bmatrix} \quad G_9 = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Figure 3. Frei-Chen edge detection matrices. [7]

The first four kernels are used for edges, the next four are used for determining lines and the last one is used for smoothing out the result. The projection equation used by the algorithm is presented in figure 4.

$$cos\, e = \sqrt{\frac{M}{S}} \quad where \;\; M = \sum_{ke\{e\}} (G_k * I)^2 \;\; and \;\; S = \sum_{k=1}^{9} (G_k * I)^2$$

Figure 4. Frei-Chen projection equation. [7]

In practice the results of Frei-Chen algorithm are better because the algorithm is less sensitive to noise and is able to detect edges with small gradients. Also Sobel can be improved by a normalization factor which in Frei-Chen is represented by the ninth mask [7].

CANVAS IMAGE PROCESSING

Using HTML5 canvas element to render the processed video image involves two steps.

In the initialization step the canvas element is created and added to the DOM tree. The canvas has the same size as the source video. The context object is also created in this step and saved for using in the display loop.

The second step is setting the display loop. The display loop is based on the *requestAnimationFrame* method. For each animation frame the current video frame is drawn on the canvas. Before being displayed, the image is processed by applying one or more convolution operators. Processing

is performed on the *ImageData* objects obtained by using the *getImageData* method of the *canvas context* object.

GLSL PROCESSING

The proposed WebGL solution uses the OpenGL ES rendering pipeline to process the source video frames. Figure 5 shows the solution architecture.

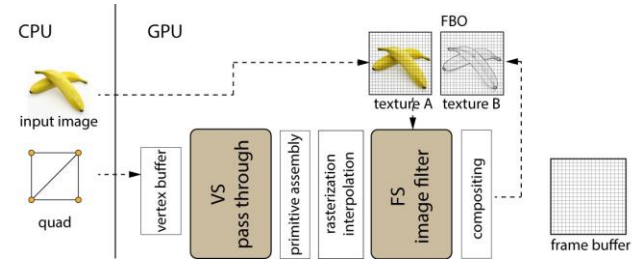

Figure 5. OpenGL pipeline for image processing [8]

In the WebGL initialization phase we create the GLSL program and the necessary attributes and uniforms (video frame texture, kernels …). The program consists of two shaders. The first one is a simple pass-through vertex shader that processes the quad used to display the processed frame. The second one is the fragment shader that applies the convolution operators for every pixel of the texture. The simplified single matrix variant:

```
precision mediump float;
uniform sampler2D u_image;
uniform vec2 u_textureSize;
uniform float u_kernel[9];
varying vec2 v_texCoord;
void main() {
vec2 onePixel = vec2(1.0, 1.0) / u_textureSize;
vec4 colorSum = texture2D(u_image, v_texCoord +
onePixel * vec2(-1, -1)) * u_kernel[0] + ...;
float kernelWeight = u_kernel[0] + u_kernel[1] +
...;
...
gl_FragColor = vec4((colorSum / kernelWeight).rgb,
1.0); } }
```

The video frame processing loop is constructed using the *requestAnimationFrame* method. For each video frame the following operations are performed: the current video frame is transferred from the video element to the GPU texture memory and the OpenGL ES pipeline is executed in order to process and display the video frame.

BENCHMARK SETUP

In order to determine the merits and performance characteristics of the two image processing techniques across devices we set up a benchmark.

The hardware used for benchmarking is representative for desktop, laptop and smartphone devices.

	Desktop	Laptop	Phone
Description	i7-920, 18GB, HD 7870 XT	i7-4500U, 8GB, HD Graphics 4400	HTC One M8 - Krait 400, 2GB, Adreno 330
CPU Cores	4	2	4
CPU Frequency	2.67 GHz	1.8 – 3 GHz	2.3 GHz
GPU Cores	1536	20	4
GPU Frequency	925 – 975 MHz	200 – 1100 MHz	550 MHz

Table 1. Hardware used for benchmarking.

For measurement we used *DOMHighResTimeStamp* objects (accurate to the thousandth of millisecond) obtained through *Performance API*. This technique was used to measure image acquisition and processing times. The actual WebGL drawing / processing time cannot be measured directly from the browser. Because of this we used an indirect method based on the fact that the browser will honor a *requestAnimationFrame* request only after the current drawing operation is completed. This technique allows us to obtain an adequate measurement for the WebGL performance for values >16ms.

PERFORMANCE ANALYSIS

Real time webcam edge detection

The first test was performed using the webcam as a video source. The video was obtained using the *getUserMedia* method. The resulting 640x480 video stream was channeled to a visible *video* element on the page and used as source for WebGL and *canvas* processing.

	Desktop	Laptop	Phone
WebGL FPS	60	60	34
Acq. / Proc.	0.4 / *	1.3 / *	5.2 / *
canvas FPS	37	32	7.2
Acq. / Proc.	0.4 / 24.6	1.1 / 26.6	4.8 / 111

Table 2. Sobel edge detection on 640x480 webcam video. Acquisition and processing times in milliseconds.

From the results presented in Table 2 we can see that even the application of a relatively simple convolution operator in real time is possible only with the more powerful x86 CPUs. Only those CPUs have the necessary speed to apply the operators in the required 30ms time frame using a single thread.

The use of the GPU shaders for applying the operator in real time is possible on all devices. Even the low power Adreno 330 is able to finish the task in under 30ms.

The image acquisition time (transfer from the video stream to canvas or GPU texture memory) is similar for *WebGL* and *canvas* across devices.

In the following two subsections we analyze the most important factors that influence the performance for this class of problems: video size (total number of pixels that need processing) and number of operations per pixel.

Sensitivity to video source size

In order to determine the sensitivity of the processing with respect to the number of processed pixels we use the same video re-encoded at 4 standard resolutions. The video was streamed from the web server and the processing was performed in real time on the client.

	Desktop	Laptop	Phone
320x240 (0.1 mp)	60 / 60	35 / 60	16 / 50
640x480 (0.3 mp)	30 / 60	19 / 60	7 / 45
1280x720 (0.9 mp)	16 / 60	10 / 58	3 / 40
1920x1080 (2.1 mp)	7 / 60	4 / 40	1.5 / 35

Table 3. Average number of frames per second using *canvas / WebGL*.

Unlike the canvas processing time which evolves linearly, the WebGL processing time remains almost constant. Almost all FPS differences in the WebGL case are caused by variations in HTTP transfer, video decoding and frame drawing time.

Sensitivity to processing complexity

All the tests performed up to now used the relatively simple Sobel operator (2 matrices). Other operators (like the Frei-Chen presented above – 9 matrices) and filter combinations can require a larger number of operations. In order to determine the canvas / WebGL sensitivity to the number of operations performed per pixel we apply the same convolution matrix multiple times per pixel.

Operators	Desktop	Laptop	Phone
2	60	60	53
10	60	60	48
20	60	60	43
50	60	60	55
100	60	60	20
500	60	51	3.5

Table 4. Average number of frames per second using *WebGL*.

The results show that the WebGL processing technique presented above can be used to apply at least 50 convolution operators even on a mobile phone. The desktop computer was able to process up to 25000 matrix operations maintaining a frame rate above 30 FPS. The canvas results are not presented in the table because the performance degraded sharply after 2 operations.

FUTURE RESEARCH

Because the results of the proposed WebGL processing technique are very promising we plan to extend our research on more complex image processing and computer vision algorithms. The next two types of problems we plan to address are motion detection (control UI using webcam) and object recognition (especially 1D and 2D barcode scanning).

Another future direction of research is combining the presented GLSL processing technique with the new Web Workers API for exploiting the multiple CPU cores available in modern devices for algorithms that cannot be parallelized efficiently on the GPU.

CONCLUSION

In this paper we presented an improved method of processing images in the context of HTML applications. Our measurements show that *WebGL* is much better suited for real time video processing even on mobile devices. The standard canvas-based processing is not fast enough for most video processing tasks. In the future the performance gap between the two techniques will probably increase even more based on the fact that the number of GPU cores available in mobile devices will grow much faster than the number of cores and the performance improvements in WebGL 2 / OpenGL ES 3.0.

REFERENCES

1. ***.WebGL. 2013, OpenGL ES 2.0 for the Web. Retrieved May 12, 2015 from https://www.khronos.org/webgl/
2. ***.2015. WebGL Earth. Retrieved May 12, 2015 from http://data.webglearth.com/doc/webgl-earthch1.html
3. ***.2015.MDN (Mozilla Developer Network), Retrieved May 12, 2015 from https://developer.mozilla.org/en-US/docs/Web/WebGL
4. ***.2013.The OpenGL ES Shading Language, The Khronos Group Inc., Retrieved May 12, 2015 from https://www.khronos.org/registry/gles/specs/2.0/GLSL_ES_Specification_1.0.17.pdf
5. ***, WebGL Stats, Retrieved July 17 2015 from http://webglstats.com/
6. Samta Gupta, Susmita Ghosh Mazumdar. 2013. *Sobel Edge Detection Algorithm*, International Journal of Computer Science and Management Research, Vol 2, Issue 2, pp. 1578-1583, February 2013
7. Daniel Rakos. 2011, Frei-Chen edge detector, Retrieved May 12, 2015 from http://rastergrid.com/blog/2011/01/frei-chen-edge-detector/
8. Gábor Sörös. 2014. *GPU-Accelerated Joint 1D and 2D Barcode Localization on Smartphones*, Proceedings of the 39th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), Florence, Italy. IEEE, pp. 5095-5099, May 2014

Diversifying Search Results Using Semantic Resources

Cristian Neamțu
"Alexandru Ioan Cuza" University,
Faculty of Computer Science
General Berthelot, No. 16
cristian.neamtu@info.uaic.ro

Adrian Iftene
"Alexandru Ioan Cuza" University,
Faculty of Computer Science
General Berthelot, No. 16
adiftene@info.uaic.ro

ABSTRACT
In the last years, multimedia content has grown increasingly over the Internet, especially in social networks (like Facebook or Flickr), where users often post images using their mobile devices. In these networks, the content is later used in search operations, when some users want to find something using a specific query. Nowadays, searching into these networks is primarily made using the title, description and the keywords associated to resources added by users that have posted the content. In this paper we address the problem of query ambiguity. The usage of semantic resources, like ConceptNet and DBpedia ontologies, has been proven to retrieve a better set of results.

Author Keywords
Image Retrieval; Search Diversification; Semantic Resources; ConceptNet; DBpedia.

ACM Classification Keywords
H5.2. Information interfaces and presentation; H3.3. Information Search and Retrieval.

INTRODUCTION

Over time, various theories involving search results diversification [11] have been developed, theories that have been taken into consideration [5]: (i) content [7], i.e. how different are the results to each other, (ii) novelty [3, 4], i.e. what does the new result offer in addition to the previous ones, and (iii) semantic coverage [16], i.e. how well covered are the different interpretations of the user query. The problem of query ambiguity was approached in [2, 13 and 14].

Capannini et al. [2] approached the problem by mining the query log for specializations of a newly submitted query. The query log is divided into sets of possible user sessions; later these sets are further refined in sessions by finding chains of linked queries. The system has been evaluated using the metrics and the datasets provided for the TREC 2009 Web Track's Diversity Task. In terms of efficiency the system performs faster than its competitors and, in terms of effectiveness, outperforms IA-Select [1] and shows comparable performance with XQuAD [15].

Navigli et al. [14] approached the problem by constructing a set of configurations as follows: for each word from the original query, the system selects a sense and then builds a semantic network for that sense and adds it to the configuration. The semantic network is built by extracting some elements (synonyms, hyponyms, homonyms, etc.) from the WordNet lexical database [6] and from finding words that co-occur by running a natural language processor on the SemCor annotated corpus [12]. At the end, the system returns the configuration with the highest score, obtained by counting the number of common nodes across the semantic networks. They devised an experiment in order to test five different sense-based expansion methods and all of them show an improvement over the plain query-methods.

Minkoo et al. [13] proposed a system that captures the user query concepts. They approach the problem by building a set of rules of related concepts, using top-down refinement strategy, where the rules are represented as an and-or tree. Next, the tree is transformed into to a neural network with the same topology and then applies the back propagation algorithm in order to adjust the weights based on the user's relevance feedback. A set of rules is determined by a fuzzy evaluation of the tree (with the updated weights) and finally these rules are used in an extended Boolean retrieval model.

In a recent paper from 2014, the authors build a novel image retrieval framework that performs a semantic interpretation of the user queries and returns a diversified and accurate result set [8].

OUR MODEL

In this section the system's workflow is described along with its main components. The system has been designed to expand ambiguous queries, in order to obtain different types of results. Thus, the initial query is expanded with ConceptNet or DBpedia ontologies, then a set of images are retrieved from Flickr, finally these are cached and are presented to the end-user.

Query processing

Given a query q from a user, the system builds a set of categories (clusters of related entities that are annotated with a common, broader concept). At the beginning, the initial query is passed to the ConceptNet sub-module; if it cannot obtain any relevant data, the initial query will be passed to the DBpedia sub-module. In order to decide whether the DBpedia sub-module should be used, the system analyzes the number of categories and the total number of elements within each cluster received from the ConceptNet sub-module.

ConceptNet

For a given query *q*, the sub-module removes the stop-words only from the beginning and the end of the query due to the way the data is processed and stored in the ConceptNet ontology. Then it retrieves from ConceptNet [10] a list of relations between different concepts, where at least one of them is related to *q*.

Next, the system extracts from each relation, the more general concept based on its type and stores it temporarily. At the next step, for each extracted relation, the system finds all it's the specializations and then removes the elements that are considered too long. The system considers a query too long if it has a large number of words (5-6) or total length of the string is greater than a threshold). These queries are removed because they may not produce any results when submitted to Flickr.

At the next step, the system calculates the similarity coefficient between two clusters, using the Jaccard Index [9], and merges the similar clusters. Lastly, the system removes the all clusters that do not provide new data (i.e. a large part of the elements can be found in larger clusters). The remaining nodes will be used in the image retrieval process.

Given the query *gnu* has at least two senses: the first one is *a genus of antelopes*, the other one being *operating system*. The system retrieves from ConceptNet the relations and the associated concepts. Next it starts building the clusters based on the relations (ignoring longer elements e.g. "gnu large African antelope have…"). For this query the system cannot find any clusters that can be merged however there are a few small clusters that will be removed, although they might be relevant (e.g. the cluster *aircraft* contains the element S*opwith Gnu*). At the end the module returns three clusters: *mammal* which contains two subspecies, *software* (contains a list of programs related to the gnu operating system) and *organization* (folk group and another element related to the Free Software Movement).

DBpedia

Using the DBpedia Lookup[1] web service the system queries for a list of entities that might be related to the initial query *q*. Each one of the returned elements contains: a label, an URI, a description and a list of categories.

Next, these elements must be filtered because it may contain concepts that are not relevant to the query. A slightly modified version of the Tf-Idf [11] score is used to filter the non-relevant elements, as follows: first, the system builds a corpus from the definitions of every element, then for each element, the system calculates the Tf-Idf weight vector for each of its categories. For words that are very common, the Tf-Idf score will return a special value instead of 0. Finally, if the number of vectors that are zero exceeds a threshold it removes the element.

[1] DBpedia Lookup: https://github.com/dbpedia/lookup

For example, for the query *android*, some of the concepts obtained from DBpedia Lookup are*: Android (operating system)*, *Android (robot)* and *Testosterone*. After filtering the irrelevant elements, the underlined concept is removed.

For the remaining elements the system will first search among its categories for one that has a label similar to the one of the element. (i) If found, system will run a query to obtain specializations for that resource. (ii) Otherwise, the system will attempt to identify a set of common concepts by running a query for each category of an element, intersecting the results and keeping those elements that appear frequently.

Given the query *microprocessor* the system attempts to build a set of clusters using the data from ConceptNet. The module will not produce any useful results because majority of the clusters are small. Next the system will obtain from the DBpedia Lookup web service the following elements: Central Processing Unit, Embedded system, Microprocessor, 32-bit, CPU Cache, Microcontroller, Instruction set, ARM architecture and Motorola 68000. In the next step, it will keep only the elements that are considered relevant to the query and will expand them. The end result produced by the module is **Central Processing Unit** (Dual-voltage CPU, Tag RAM, etc.), **Embedded system** (Logic analyzer, Debit card, etc.), **Microprocessor** (KOMDIV-64, Apple A4, etc.), **CPU Cache** (Smart Cache, Tag RAM, etc.), **Microcontroller** (Intel MCS-48, Motorola MC14500B, etc.), **Instruction set** (Berkeley RISC, MIPS-X, etc.), **ARM architecture** (Tegra, ARM Cortex-A15, etc.) and **Motorola 68000** (Motorola 68000).

Image processing

This module receives a category previously built by the query expansion modules and returns a set of relevant images for it. Depending on the size of the built category (cluster), this module will apply one of the two different methods.

The first method is applied only to small clusters (up to 6 elements). For each element of such a cluster the module will perform a search on Flickr to obtain a set of photos whose title, description or tags contain the specified text. The other method is applied to large clusters. These will be divided into smaller sets. Next, for each set, it will perform a new search in order to obtain a set of pictures whose tag list contains at least one of the elements from that set.

Next, the images will be grouped by the user's id and the system will cluster the elements inside a non-singleton group based on the title using the edit-distance and it will select a single image from each cluster. Finally the system will select the top K images from the previous step.

CASE STUDIES

Case Study 1 – Query *Formula One Racer*

Given the query *formula one racer*, the system returns a single cluster containing approximately 70 elements (see

Figure 1) where four of them are not related to the query. For instance: *Mike Park* does not belong to the cluster (he is an American musician).

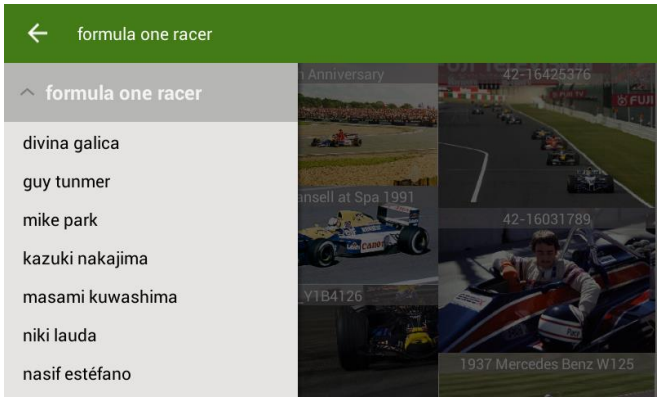

Figure 1. A few elements from the returned cluster.

In Figure 2 is presented a subset of the results returned by the system. A fraction of the images returned are not relevant to the initial query, but are related to the career of that person. For example, David Brabham was a formula one racer, however he worked in other fields of motorsports such as: Touring Car Championships and Le Mans competitions). A relevant example for this case would be the image in the lower right corner of the Figure 2.

Figure 2. A small subset of the images returned after processing the query *formula one racer*.

Case Study 2 – Query *Eiffel*

After submitting the query *Eiffel*, the system returns 7 clusters, all of them are presented in Figure 3: The finer grained clusters are: *Architectural Structure*, *Band*, *Programming Language* and *Film*. The rest of the clusters have a greater degree of generality and contain elements from the other clusters. For example: the cluster *Organization* contains some of the elements from the cluster *Band*.

The cluster *Architectural Structure* contains the following elements: *Eiffel Tower*, *Eiffel Bridge Zrenjanin*, *Eiffel Bridge Ungheni* and *Ponte Eiffel*. Below is presented a subset of the images returned for this cluster.

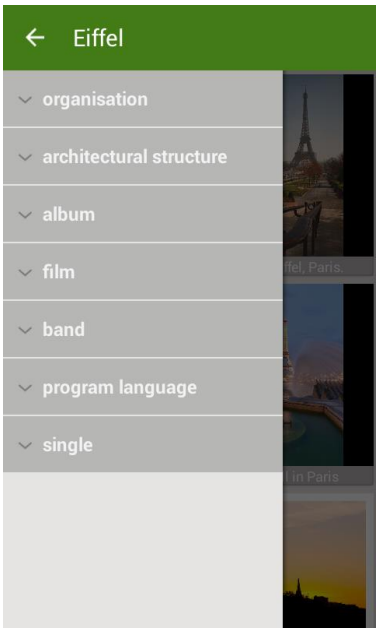

Figure 3. The clusters returned for the query Eiffel.

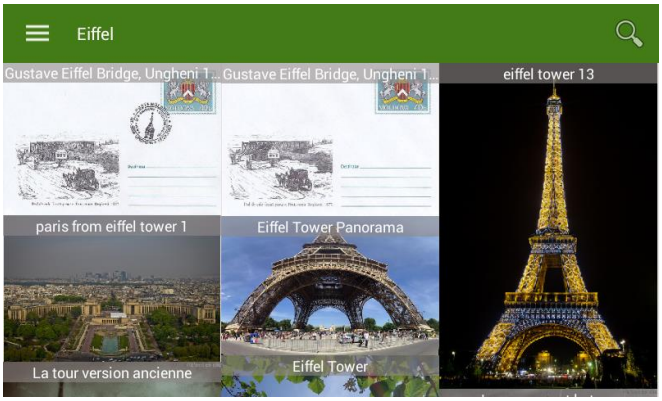

Figure 4. The results for the *Architectural Structure* category.

The system has been built to expand ambiguous queries that are generally short, for which the system returns relevant results. However, submitting more complex and elaborate queries may result in a failure to return a set of images.

We attempted to tokenize the user query, clear all the stop-words and try to find a set of linked concepts based on the processed query, but it resulted in set concepts that could not be used in the current context.

EVALUATION

In this section the system will be evaluated from three different points: (1) *the time required to process a query* (with and without caching), (2) *the relevance of the items inside a cluster* and (3) *the relevance of the images returned by the system*. In order to evaluate the system, there were considered 10 queries.

The time required to process a previously submitted query is on average 4 seconds, while the time required to process a newly submitted query may take up to 60 seconds. The execution time depends on the number of elements in a

cluster (retrieving images for smaller clusters may take up to 10 seconds, while bigger clusters take a significantly longer time).

Regarding the relevance of the elements inside a cluster, results show that approximately 70% of the clusters elements are related to the initial query. In a relevant cluster, up to 80% of the items are relevant to initial query and are relevant for the current cluster.

After analyzing the set of images returned by processing a cluster it results that on average 65% of the images are relevant to the query. In general, the relevant images are among the first results; however we found some exceptions due to the fact that the cluster contains some irrelevant items positioned at the beginning. We remark that, there are some cases in which the system cannot obtain, due to some limitations, a relevant set of images.

For instance, when "*a view from Eiffel Tower*" is used as initial query, the system identifies two meanings: (i) *the panorama of Paris from the top of the Eiffel Tower* and (ii) *the movie*. The system tries to retrieve from Flickr a set of images for (ii) the second meaning of the query, but it receives results for the first sense.

CONCLUSIONS

This article presents a method in which semantic resources like ConceptNet and DBpedia are used to improve the quality of searches performed in an image retrieval system. Until now, the results are promising and we see how 70% of the clusters elements and up to 80% of the items are relevant to initial query.

In the future, we want to improve the current system by combining semantic resources ConceptNet and DbPedia (now the system uses them separately). Also, we want to replace the Tf-Idf score with other function.

ACKNOWLEDGMENTS

The research presented in this paper was funded by the project MUCKE (Multimedia and User Credibility Knowledge Extraction), number 2, CHIST-ERA/ 01.10.2012.

REFERENCES

1. Agrawal, R., Gollapudi, S., Halverson, A., and Ieong, S. Diversifying search results. In *WSDM '09*, (2009), 5–14.
2. Capannini, G., Nardini, F. M., Perego, R. and Silvestri, F. Efficient Diversification of Web Search Results. In *Proc. VLDB Endow*., 4(7), (2011), 451-459.
3. Carbonell, J. G. and Goldstein, J. The use of mmr, diversity-based re-ranking for reordering documents and producing summaries. In *SIGIR*, (1998), 335-336.
4. Clarke, C. L. A., Kolla, M., Cormack, G. V., Vechtomova, O., Ashkan, A., Bttcher, S. and MacKinnon, I. Novelty and diversity in information retrieval evaluation. In *SIGIR (2008)*, (2008), 659-666.
5. Drosou, M. and Pitoura, E. Search Result Diversification. In *SIGMOD Record*, 39(1), (2010), 41-47.
6. Fellbaum, C. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998.
7. Gollapudi, S. and Sharma, A. An axiomatic approach for result diversification. In *WWW'2009*, (2009), 381-390.
8. Iftene, A. and Alboaie, L. Diversification in an image retrieval system. In *IMCS-50. The Third Conference of Mathematical Society of the Republic of Moldova dedicated to the 50th anniversary of the foundation of the Institute of Mathematics and Computer Science*. (2014).
9. Jaccard, P. *The distribution of the flora in the alpine zone*. New Phytologist, 1912.
10. Liu, H. and Singh, P. Focusing on ConceptNet's natural language knowledge representation. In *Commonsense Reasoning in and over Natural Language Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES'2004)*, (2004).
11. Manning, C.D., Raghavan, P. and Schütze, H. *Introduction to Information Retrieval*. Cambridge University Press New York, NY, USA, 2008.
12. Miller, G. A., Chodorow, M., Landes, S., Leacock, C. and Thomas, R. G. Using a Semantic Concordance for Sense Identification. In *Proceedings of ARPA Human Language Technology Workshop*, (1994).
13. Minkoo, K. and Vijay, R. Adaptive Concept-based Retrieval Using a Neural Network. In *Proceedings of ACM SIGIR Workshop on Mathematical/Formal Methods in Information Retrieval*, Athens, Greece, (2000).
14. Navigli, R. and Belardi, V. An Analysis of Ontology-based Query Expansion Strategies. In *Workshop on Adaptive Text Extraction and Mining (ATEM), ECML 2003*, (2003).
15. Santos, R. L. T., Macdonald, C. and Ounis, I. Exploiting query reformulations for web search result diversification. In *WWW'10*, (2010), 881–890.
16. Zheng, W., Wang, X., Fang, H. and Cheng, H. Coverage-based search result diversification. In *Journal IR* (2012), 433-457.

Motives and Characteristics of Facebook Use by Students from a Romanian University

Alexandru Balog, Costin Pribeanu
ICI Bucharest, Romania
Bd. Maresal Averescu nr.8-10, Bucuresti
alexb@ici.ro, pribeanu@ici.ro

Ion Ivan
ASE Bucharest, Romania
Calea Dorobanti, 15-17, Bucuresti
ionivan@ase.ro

ABSTRACT

The increasing popularity of Facebook among university students is raising several research questions regarding the characteristics of use. In the context of modern school, the academic use and the usage related to the university context are of special interest. This paper aims at analyzing the motives for the use of Facebook and the characteristics of Facebook use by students from a Romanian university. The results show that the motives of using Facebook are mainly related to communicating with friends, to find out what happens in their university and to keep in touch with former high school colleagues. The results revealed significant differences between male and female students in terms of motives of using, network size, frequency and duration of use. Female students have more Facebook friends and spend more time on Facebook than male students. Moreover, significant differences exist between undergraduate and master students.

Author Keywords

Facebook, social networks, educational use of Facebook, motives of using Facebook.

INTRODUCTION

The proliferation of Facebook among university students is a provocation for teachers to look for modalities of exploiting the educational potential of social networks. The shift of paradigm from student centered learning to social learning requires new activities such as meeting, chat, active participation, critical thinking, information and resources exchange, collaboration, and debate [1, 3, 14]. This change requires understanding why students use Facebook and the main characteristics of use.

Although the number of Facebook users in Romania continuously increased, very little is known about the reasons why university students spend time on Facebook, about the relationship between Facebook use and their work at university as well as about differences related to enrollment (college / master), and gender. It is worthy to mention that despite the well-known gender differences, the studies carried on in other countries rarely report the network size, frequency and duration of use by gender.

Based on two qualitative studies carried on in Lithuania and Romania, a questionnaire was developed and tested during a pilot study [13]. The questionnaire was revised and then administrated again in 2015 in Romania.

This paper presents an analysis of reasons why university students use Facebook and an analysis of differences by gender and programme of study. The sample was collected in March 2015 at the Academy of Economic Studies (ASE) in Bucharest.

The following research questions are guiding this study:

- Which are the motives of using Facebook by university students and to what extent are these related to the university context?
- Which is the average size of Facebook friends network and how many of them are students?
- How often do university students log on Facebook and how much time do they spend daily?
- Which are the variations of use by programme of study (college / master) and gender?

The rest of this paper is organized as follows. The following section briefly presents recent researches in the social networking sites with an emphasis on Facebook use. Next, the method and variables used in this study are described. Afterwards, the results are presented and analyzed. The paper ends with conclusion and future research directions.

RELATED WORK

As shown by Brown & Adler [3], the development of students' communities is as important as the access to the educational content. Social networking websites support students to develop the skills they need in a community of practice. Facebook enables users to create a profile that represents them and to present their concerns, activity and opinions in an open space.

According to Selwyn [14] Facebook is an open space for informal education, an arena where students have the possibility to present their identity and model their behavior. He argues that Facebook stimulates critical thinking since students can analyze and share opinions as regarding various situations that occur in their university.

The study of Ellison et al. [5] reports a positive association between the students' perception as regarding the social capital and the satisfaction with life in university, which suggests a positive influence of Facebook for the integration and participation in a university community.

Valenzuela et al. [15] found a positive relation between the intensity of use on one side, and the social trust, civic engagement, and political participation, on the other side.

Park et al. [11] found four primary needs for joining Facebook groups: socialization, entertainment, self-status seeking, and information. Using a sample of 1715 college students they found that these gratifications vary on gender, home-town and year in school.

Lampe et al. [10] analyzed the perceptions of users as regarding the value of Facebook as information source. The study reveals the specific way in which users are converting their social capital in information and concludes that the information usefulness and seeking behavior are positively associated with their perception regarding the social capital.

Junco [8] analyzed the relationship between the frequency of Facebook use, participation in Facebook activities, and time spent on curricular and co-curricular activities. The results show that Facebook is positively influencing the time spent on co-curricular activities.

The study of Arteaga-Sanchez et al. [1] shows that Facebook could support collaborative and cooperative learning. In this respect, teachers should analyze how students interact with the technology and the ways in which Facebook could support social learning and enrich the learning experience.

Lamanuskas et al. [9] analyzed the use of social networking websites by university students from five countries: Lithuania, Romania, Czech Republic, Turkey, and Ukraine. Most of students (74.4%) are visiting daily a social networking website. The most frequently mentioned motives are communication, learning new things and information exchange, photos and videos sharing.

Following the data provided by Facebrands.ro [6], the number of Facebook users in Romania was eight millions in March 2015 which represents a penetration rate of 39.76%. The highest adoption rate is in the range of 15-24 years (33.12%), which means that a young person out of three has a Facebook account.

In Romania, Băltăreţu & Balaban [2] used a focus-group method to investigate the use of social networking websites by young people and concluded that the main reasons are related to social aspects: need of communication, socialization, and participation in collective actions. Popa & Marhan [12] explored the creative potential of the students that are using Facebook. The results show that Facebook plays a central role in students' concerns and the exchange of information content generated by other people prevails over the generation of new content. The study of Grosseck et al. [7] focuses on how students perceive the use of Facebook for academic purposes. Their results show that the majority of students spend significant time on Facebook more for social uses and less for academic purposes.

METHOD

Sample

The questionnaire in Romanian has been administrated to university students from ASE Bucharest in March 2015. Students were asked to answer general questions (faculty, programme of study, year of study, age, and gender), questions regarding the use of Facebook (network size, frequency, and duration of use), and then to evaluate several items on 7-points Likert scale.

From the total of 451 questionnaires received a number of 37 were eliminated because of incomplete data. The working sample has 414 observations, from which 156 men and 258 women. The age of participants is varying between 18 and 37 years with a mean of 21.28 years (SD=2.78). Most of the participants (368) are students at Cybernetics, Statistics and Economic Informatics (CSIE) faculty and the rest of 46 at Marketing faculty. As regarding the study programme, 313 are college students and 101 master students.

Variables

This paper examined a part of a larger data set collected on Facebook users students. Two categories of variables are analyzed that refer to the motives for using Facebook (Table 1) and the characteristics of usage (Table 2).

Table 1. Motives for using Facebook

Item	Statement
FBU1	I use Facebook to present myself
FBU2	I use Facebook to find out what happens in my university
FBU3	I use Facebook to get in touch with new people
FBU4	I use Facebook to get advice about something I am interested into
FBU5	I use Facebook to get access at shared resources
FBU6	I use Facebook to participate in group discussions
FBU7	I use Facebook to keep in touch with former high-school colleagues
FBU8	I use Facebook to locate old friends
FBU9	I use Facebook to communicate with my friends
FBU10	I use Facebook to find out what is new and innovative

Six variables are related to the university context and information (FBU1, FBU2, FBU4, FBU5, FBU6, and FBU10). The rest of four variables refer to communication and socialization (FBU3, FBU7, FBU8, and FBU9).

Table 2. Characteristics of usage

Variable	Questions
FBF	How many FB friends do you have in your FB network ?
FBFS	How many of your FB friends are university students ?
FBFSU	How many of your FB friends are students in this university ?
DaysWeek	How many days per week do you use FB ?
LogDay	How many times per day do you log on FB ?
MinDay	On average, how many minutes per day do you actually use FB ?

The characteristics of Facebook usage are related to the network size, frequency and duration of use.

Data analysis

The collected data was analysed using SPSS 16.0. The statistical tools used in this study include descriptive statistics (frequencies, percentage, mean, and standard deviation) and one-way ANOVA.

RESULTS

Motives of using Facebook

As shown in Table 3, the motives of using Facebook are mainly related to communicate with friends (FBU9), to find out what happens in university (FBU2), and to keep

in touch with former colleagues (FBU7). Next four reasons are to get access at shared resources (FBU5), to participate in group discussions (FBU6) and to locate old friends (FBU8), and to find out what is new (FBU10). Although the order of preferences suggests that social motives prevails, it can be noticed a positive perception of Facebook as a potential support for social learning. The answers at three questions have mean values bellow 4.00, showing the university students are less interested to use Facebook to present themselves (FBU1), to get in touch with new people (FBU3), and to get advice (FBU4).

Table 3 shows that for all variables female students scored higher than male students. To test whether there were any differences in the mean ratings of motivations between male and female students, one-way ANOVA tests were performed. There were statistically significant differences between male and female in all variables, excepting for FBU1, FBU3 and FBU6. Highest differences are noticed for FBU8 and FBU7, suggesting that female students are more interested than male students to locate old friends and keep in touch with former high school mates.

Table 3. Motives for using Facebook: differences between male and females students

Variable	All	Male (156)	Female (258)	*F*	*p*
FBU1	2.80	2.72	2.86	0.742	.389
FBU2	5.31	4.82	5.61	24.470	.000
FBU3	3.57	3.56	3.58	0.013	.911
FBU4	3.94	3.61	4.14	9.248	.003
FBU5	4.90	4.45	5.18	21.835	.000
FBU6	4.81	4.60	4.94	3.851	.050
FBU7	5.02	4.50	5.34	25.689	.000
FBU8	4.58	4.01	4.92	27.295	.000
FBU9	5.76	5.54	5.88	5.378	.021
FBU10	4.51	4.05	4.79	18.522	.000

The mean and standard deviation values by study programme are presented in Table 4. The three most important reasons to use Facebook are the same for college and master student (FBU9, FBU2 and FBU7). College students have a higher perception for all variables.

Table 4. Motives for using Facebook: differences between college and master students

Variable	College (313)	Master (101)	*F*	*p*
FBU1	2.88	2.56	3.073	.080
FBU2	5.38	5.09	2.546	.111
FBU3	3.73	3.07	11.257	.001
FBU4	4.10	3.44	11.589	.001
FBU5	5.03	4.52	7.820	.005
FBU6	5.00	4.25	15.150	.000
FBU7	5.11	4.77	3.002	.084
FBU8	4.68	4.25	4.610	.032
FBU9	5.86	5.45	6.217	.013
FBU10	4.57	4.33	1.536	.216

To test whether there were any differences in the mean ratings of motivations between college students and master students, one-way ANOVA tests were performed. The analysis revealed no differences between college and master in FBU1, FBU2, FBU7, and FBU10. In contrast, is a significant difference in all other motivations. Highest differences are noticed for FBU6, FBU3, and FBU4, suggesting that college students are more interested than master students to participate in group discussions, to get in touch with new people, and to get advice.

Network size

As regarding the network size, the number of Facebook friends is large. About 32% of respondents have between 500 and 1000 friends and about 22% over 1000 friends. The number of friends that are students is about twice smaller, 33% of students having in their Facebook network between 250 and 500 students and 24% over 500 friends. The mean number of students at this university has a weight of 19.38% in the Facebook network.

The number of Facebook friends is varying between 7 and 5000, with a mean value of 840.60 (SD=825.69). The number of friends which are students is varying between 7 and 2000 (M=419.64, SD=409.69), which represents about half of the mean number of FB friends. From these, 38.8% are students at ASE (M=162.92, SD=176.12).

Female students have more Facebook friends. To test whether there were any differences in the mean ratings of the network size between male and female students, one-way ANOVA tests were performed. The results revealed a significant difference between male and female students in terms of network size for all variables (Table 5).

Table 5. Network size differences between male and females students

Variable	Male (156)	Female (258)	*F*	*p*
FBF	695.55	928.30	7.86	.005
FBFS	316.04	482.28	16.61	.000
FBFSU	127.49	184.33	10.36	.001

Furthermore, the analysis revealed a significant difference between college and master students with regard to number of friends (F=5.89, p=.016) and number of student friends (F=5.82, p=.016). Analysis revealed no difference with regard to number of friends studying at the same university (F=2.10, p=.148).

Frequency and duration of use

The frequency of use is varying between a day and seven days per week (M=6.57, SD=1.18), which shows that most of the university students are using Facebook daily. As regarding the daily use, few students are logging on Facebook once (4.6%) or twice (11.8%). Most of them are logging on three times per day or more (55.1%), while 118 students (28.5%) are continuously logged on. The average amount of time spent per day is varying between 2 and 600 minutes (M=108.04, SD=109.37).The mean number of minutes per day is pretty high: 52.7% up to one hour, 25.8% between one and two hours, and 21.5% over three hours. The intense use of Facebook should not be surprising. Similar situations have been reported in other countries [4].

Female students log on Facebook more often and spend more time on Facebook than male students. To test whether were any differences in the mean ratings of the reported usage between male and female, one-way

ANOVA tests were performed. The results revealed a significant difference between male and female students in terms of reported usage for all variables (Table 6).

Table 6. Reported usage differences between male and females students

Variable	Male (156)	Female (258)	*F*	*p*
DaysWeek	6.24	6.77	20.47	.000
LogDay	2.96	3.15	6.24	.013
MinDay	84.20	122.46	12.22	.001

Moreover, reported data showed that the college students log on more often and spend more time on Facebook than master students. A one-way ANOVA showed that differences are statistically significant for the number of days per week (F=5.14, p=.024) and the number of minutes per day (F=7.41, p=.007).

CONCLUSION

This study contributes to the understanding of the motives and characteristics of Facebook use by university students in Romania. The results show they use Facebook especially to communicate with friends, to find out what happens in their university, and to keep in touch with former high school colleagues. The gender analysis shows significant differences between female and male students. Female students are more interested than male students to locate old friends and keep in touch with former high school mates.

The analysis by programme of study shows higher mean values for college students than for master students. Although several differences exist that are statistically significant is not possible to draw a conclusion regarding the usage behavior since Facebook penetration might be different from generation to generation.

Female students log on Facebook more often and spend more time on Facebook than male students. The frequency and duration of use show an intense use that might have negative effects on the students' academic performance. The characteristics of Facebook use by university students should challenge teachers to find ways to orient the use for academic purposes. In this respect, the relatively large number of Facebook friends from the same university is an argument towards the educational use.

There are inherent limitations of the study. The sample is from one university and most students are from the same faculty. The perceptions might differ according to the university / faculty profile. Future research may study Facebook motives and usage from a cross-cultural perspective to identify patterns.

To have a wider and more complete perspective on Facebook use by university students a larger sample is needed. Next studies will collect samples for several universities and focus on the analysis of variation by profile (technical, economics, and social sciences), programme, and year of study.

ACKNOWLEDGEMENT

This work was supported from a national grant financed by ANCS under TEHSIN 0923 0207.

REFERENCES

1. Arteaga Sánchez, R., Cortijo, V., & Javed, U. (2014). Students' perceptions of Facebook for academic purposes. *Computers & Education*, *70*, 138-149.
2. Bălțărețu, C.M., Balaban, D.C. (2010). Motivation in Using Social Network Sites by Romanian Students. A Qualitative Approach. J. of Media Research, 6, 67-74.
3. Brown, J. S., & Adler, R. P. (2008). Open education, the long tail, and learning 2.0. *Educause review*, *43*(1), 16-20.
4. Chang, C. C., Hung, S. W., Cheng, M. J., & Wu, C. Y. (2014). Exploring the intention to continue using social networking sites: The case of Facebook. *Technological Forecasting and Social Change*.
5. Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12(4)*, 1143-1168.
6. Facebrands.ro (2015) http://www.facebrands.ro/demografice.html, accessed on 8 April 2015.
7. Grosseck, G., Bran, R., & Tiru, L. (2011). Dear teacher, what should I write on my wall? A case study on academic uses of Facebook. *Procedia-Social and Behavioral Sciences*, *15*, 1425-1430.
8. Junco, R. (2012). The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. *Computers & Education*, *58*(1), 162-171.
9. Lamanauskas V, Šlekiene V, Raguliene L, Iordache DD, Pribeanu C, Bilek M, Cavas B, Mazurok T. (2013) Social networking websites from the point of view of university students: a comparative analysis, *Problems of Education in the 21st Century 57*, 61-78.
10. Lampe C., Vitak, J., Gray, R., Ellison N.B. (2012) Perceptions of Facebook's value as information source. *Proceedings of CHI 2012*, 3195-3204, ACM.
11. Park, N., Kee, K. F., Valenzuela, S. (2009). Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *CyberPsyc. & Behavior*, *12*(6), 729-733.
12. Popa, C., Marhan, A.M. (2013). Profilul creativ al adolescenților utilizatori de Facebook, *Revista de Psihologie 59(2)*, 167-178.
13. Pribeanu, C., Balog, A., Lamanauskas, V. Šlekienė V. (2015) Social capital on Facebook as perceived by Lithuanian university students. *Journal of Baltic Science Education 14(1)*, 132-141.
14. Selwyn, N. (2009). Faceworking: exporing students' education-related use of Facebook. *Learning, Media, and Technology 34(2)*, 157-174.

Valenzuela, S., Park, N., & Kee, K. F. (2009). Is There Social Capital in a Social Network Site?: Facebook Use and College Students' Life Satisfaction, Trust, and Participation. *Journal of Computer-Mediated Communication*, *14*(4), 875-901.

Movies Classification

Daniel Gavril
"Alexandru Ioan Cuza" University,
Faculty of Computer Science
General Berthelot, No. 16
daniel.gavril@info.uaic.ro

Adrian Iftene
"Alexandru Ioan Cuza" University,
Faculty of Computer Science
General Berthelot, No. 16
adiftene@info.uaic.ro

ABSTRACT
Every device that is released nowadays has the capability to display videos. From small devices like smart phones to bigger ones likes desktops or smart TVs, movies are available at any time. Therefore, getting important, fast and reliable data about movies at any time and everywhere is very important when it comes to money. Computer users like to get live experiences based on their interests, searches or their personal profile and their needs include movies. When people gather to someone's place or a family unites on holidays, finding a good movie to watch for that moment becomes hard. There are different tastes in movies, persons that like drama more than science-fiction or some that will enjoy movies based on the director of that movie or important actors. Using a good search engine that might understand users desires based on recent searches or interests might come in handy. *Movies Classification* application runs on Windows operating systems and tries to understand user perspective when he tries to find a movie that might suit his tastes. The app updates live the suggested movies based on the customizable profile that every registered user must set.

Author Keywords
Movies Classification; Client-Server Architecture; User Profiling.

ACM Classification Keywords
H5.2. Information interfaces and presentation; H3.3. Information Search and Retrieval.

INTRODUCTION
In the last years, the main search engines use the history of user activities in order to provide more accurate results. Google[1], Yahoo[2], Bing[3] are ones of the greatest companies that during decades have worked on search engines that will help users find related data for a given input. There are some search engines for movies that show recommendations based on their preferences. Even if the main subject is movies, finding correlated data and categorize it for each user, it is a difficult task. Currently, there are around 80.000 TV movies, more than 1.500.000 actors and more than 389.000 directors registered to IMDb database[4] [2]. Jinni[5], Taste Kid[6] or Nanocrowd[7] are some web apps that allows you to get movie recommendations in real time. For example, *Jinni* is a recommendation engine that helps you find films based on your mood, time available, setting, or reviews. *Taste Kid* is another example of finding related data for a movie from other areas: music, images, and books. [6]

In what follows, we present a desktop application that is compatible with Windows operating system which offers for authenticated users suggested films based on their profile and searches from the "Internet movie database". The client application queries with a service hosted online the IMDb database and then downloads and stores related data for movies, directors, and actors in a local database in records related to a user profile.

SYSTEM ARCHITECTURE
The architecture of the application is based on a client-server architecture (see Figure 1). At the *server level* the main components are based on a SQL database and on a WCF [7] service. At the *client level* we used the WPF [8] and related API's used to get data from IMDb and YouTube[8].

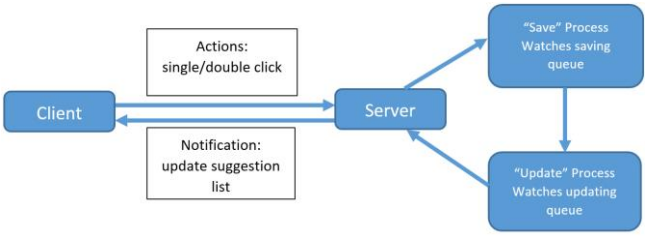

Figure 1. Application Architecture.

Both client and service components use the API's for IMDb and a particular case of the application use the YouTube Data API for trailers. The API's for IMDb are: OmdbApi[9] used to search for movies and MyApiFilms[10] used to find more details (directors and actors with associated movies,

[1] Google: https://www.google.ro/
[2] Yahoo: https://ro.search.yahoo.com/
[3] Bing: http://www.bing.com/?setlang=ro
[4] IMDb: www.imdb.com
[5] Jinni: http://www.jinni.com/
[6] Taste Kid: http://www.tastekid.com/
[7] Nanocrowd: http://www.nanocrowd.com/
[8] YouTube: https://www.youtube.com/
[9] OmdbApi: http://www.omdbapi.com/
[10] MyApiFilms: http://www.myapifilms.com/

genres, plot, votes etc.) for a movie based on a unique identifier (set by IMDb).

The client is structured following the MVVM [3] design pattern recommended by Microsoft community for this type of project. There are three main windows: "login", "find movies" and "profile". The "login" window is used to authenticate the user in the application. Therefore, the service can register the actions related to every user in the database. Also, the application has the functionality for new users to register.

The "find movies" window (Figure 2) is divided in 4 parts: the top part has a search section where the user can search for new movies by title and year (optional). The bottom part is divided in three: the left side is used to show the search results, in the middle area is presented the details for the selected movie and the right side is used for suggested movies that the service recommends. There is a secondary window that is opened on double click event on a movie from search list or suggested movies list and give more details about it including a trailer from YouTube.

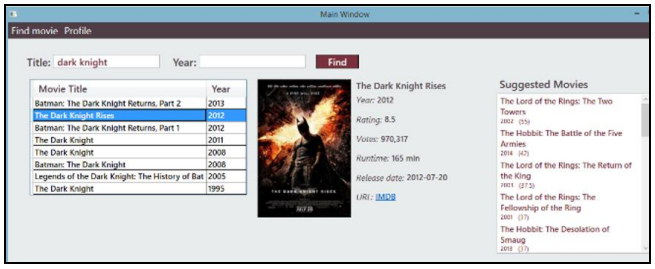

Figure 2. Find movies window.

The "profile" window (Figure 3) offers the possibility to change preferences about movies, directors or actors. Also, there is a small description about the application functionalities.

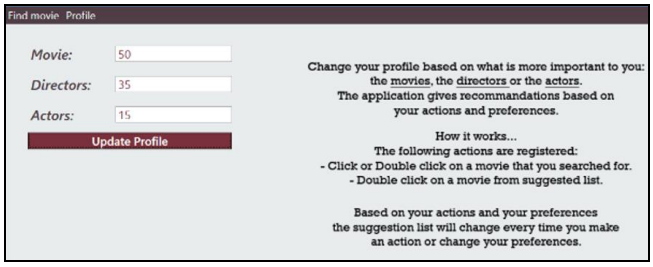

Figure 3. Profile window.

The client interacts with the service in the following cases and each one has its significance: the user clicks on a movie from search list meaning that has some interest for that movie, double click on the same movie means that user starts to get interested in that movie and double click on a suggested movie reflects user high interest on a particular movie. The service may be hosted in IIS[11] web server on a virtual machine or locally and has configured the connection string to the database (locally or hosted in Azure cloud[12]). The service starts two tasks (processes) which checks every five seconds if there is any data to process from their corresponding queue. Every user action is added in the saving queue because the service stores every related data for directors and actors. Therefore, there are some restriction based on the unique id of the movies, directors and actors and saving simultaneously might corrupt or fail the saving process. After a successfully processing of the data, the service sends the results to an updating queue where the corresponding process updates for current user, based on the action significance, the scores for each related entity.

Using a duplex connection between client and service (WCF feature), the service sends to the user a notification that the updating process has finished and he must update the suggested movies list.

CASE STUDIES

Case Study 1

In this section we will present different scenarios related to user actions and how the scoring model works.

Let assume that it is the user's first login. Therefore, there are no movies related to him. For example, let's assume he starts to search for "The Hobbit" and no given year. The OmdbApi returns 8 matches that contains the given text.

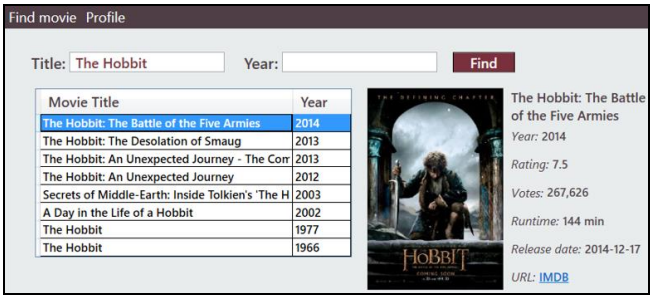

Figure 4. Single click selection.

From the list, the user single clicks "The Hobbit: The Battle of five armies" (Figure 4). The action is sent to the server and saves the movie with its details using MyApiFilms. Also, the service saves related movies from the directors and the main actors (with no details). For this pick, there are saved 74 movies in the database and 15 main actors. The service suggests only 1 movie because the single click action is not very relevant to the system. A low score is added for the movie, director and the 15 main actors.

Next, let's assume that the user double clicks on the same movie (Figure 5). The service will not download any data because it was saved previously so it skips directly to the scoring part. The double click action signifies that the user is starting to get interested, therefore he might like movies

[11] Internet Information Services: https://www.iis.net/

[12] Azure cloud: http://azure.microsoft.com/en-us/

related to the director(s) of this movie. A medium score is added to director of selected movie, to actors of this movie

and to every movie of the director. The service suggests this time 14 movies all related to *Peter Jackson.*

Figure 5. Double click window with no profile set.

Let's assume that the user double clicks on a movie from the suggestion list – "King Kong". This actions suggests that the user might like a movie and shows high interest. Therefore, a high score will be added to director, to the actors of this movie and to the director movies. The service will download any related data - in the database are now 259 saved movies. The suggestion list doesn't change by number because "King Kong" has the same director as "The Hobbit". But now there are 29 main actors associated with current user.

Every movie associated with its directors and actors will determine a score based on the Weighting and Scoring Model [5]. Every user has a profile where he must set "how much" will influence the movie itself, the directors or the actors the ranking of the suggestion list (Figure 2). The total of the percentages must be 100. The following study was structured based on the results from the previous one.

Case Study 2

With the current scores for movies, director and actors, if the user sets the percentages of movie to 100 and the rest to 0, the directors and actors will not influence the scoring (Figure 6). Therefore in the top of the list will be the movies that were associated with the most significant action (Top 3: "King Kong", "The Hobbit: The battle of five armies" and 2 other movies related to "The Hobbit" trilogy).

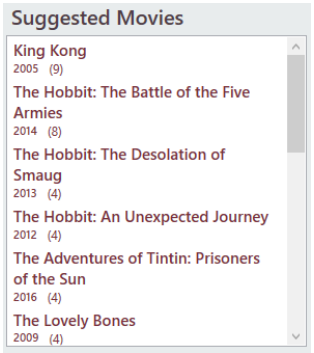

Figure 6. Suggested Movies for Profile Value for *Movie*=100.

Changing the percentages of the movie to 50, the directors to 50 and the actors to 0, the weight of the scores will be split equally between the movie and the director (Figure 7). This set up is useful when there are other directors associated with the user.

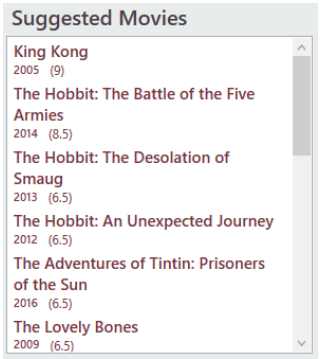

Figure 7. Suggested Movies for Profile Values for *Movie*=50 and for *Directors*=50.

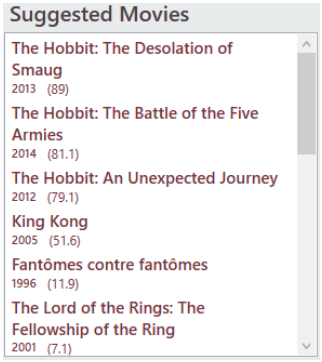

Figure 8. Suggested Movies for Profile Values for *Movie*=50 and for *Directors*=20 and for *Actors*=30.

Let's include actors in the scoring model. If the user changes the percentages of the director to 20 and the actors to 30, the suggestion list will show on the top the movies where actors appeared more in the search list (Figure 8). "King Kong" is now on rank 4, top 3 is occupied by the Hobbit trilogy because in all 3 movies appeared the same main actors. Therefore, they "weight" more in the scoring process.

EVALUATION

The process of downloading and saving a movie has the longest execution because a request to MyApiFilms includes a very large amount of data to work with. Moreover, the service must not save twice the same movie, director or actor because of the IMDb id's uniqueness constraint.

Let's assume that the database is empty and a user searches for "Dark Knight" and selects "The Dark Knight Rises" from 2012. The click action is sent to service and starts to work with the given data: the request to MyApiFilms took 10.179 seconds, saving all data to database took 14.259 seconds with a total of 24.474 seconds. In the database are saved: 230 movies, 15 actors, 1 director, with 571 relations between movies and actors and 9 relations between movies and directors. The process of setting the scores between user and movies, director and actors for single click action took 1.168 seconds, but for double click action took 1.562 seconds. Searching again for the same movie the process of saving the movie takes 0.011 seconds, updating the score costs 0.254 seconds for single click and 1.005 for double click.

Testing the application for 10 movies the average score for API to respond is 11.233 seconds, saving to database costs 11.732 seconds, updating the score for single click 1.255 seconds and for double click 1.520 seconds. Moreover, there are saved in the database 233 movies, 1.1 directors, 15 actors, 444 relations between movies and actors and 13.3 relations between movies and directors.

MyApiFilms gives for some movies actors that are not the main cast. This error can't be controlled by the application, therefore the solution might be to contact the developer of the API to look for a solution or correct the request.

Being limited by each API to 2000 request per day, saving the movies to the database save some requests. Also, accessing the data from the database with WCF is faster and easier than downloading and processing it every time.

The WCF service can be hosted in a Web App and can expose the structure of the SOAP messages. Therefore, integrating the service in a different client (Web client) is very easy with .NET technologies.

Deploying the WCF service to Azure cloud is not possible because of the Callback Contract that the service expose. The load balancer can't keep a duplex connection more than 1 minute. Therefore, a different approach for the server-client communication is needed in this case.

Using Entity Framework [1] for mapping the database increases fast and easy development if new features or models will be added. Moreover, Entity has a feature called lazy loading which means that any related data is loaded when the given query requests it.

A custom web scrapper for IMDb will improve the saving process and will offer the chance to download specific details based on user preferences.

CONCLUSION

"Movies Classification" is a desktop application easy to maintain and use, with fast results and live experiences for each user. We decided to create a desktop application, because it can be faster than the web application which depends on the browser's processing power or other elements that requires web control.

For the future we intend to reduce more the duration of the execution for the main methods from the application. Also, we intend to search another solutions to validate information provided now by IMDb, which are not always correct.

ACKNOWLEDGMENTS

The research presented in this paper was funded by the project MUCKE (Multimedia and User Credibility Knowledge Extraction), number 2, CHIST-ERA/ 01.10.2012.

REFERENCES

1. Entity Framework: https://msdn.microsoft.com/en-us/data/ef.aspx
2. Internet movie database stats: http://www.imdb.com/stats
3. Model-View-View Model: https://en.wikipedia.org/wiki/Model_View_ViewModel
4. MyApiFilms: http://www.myapifilms.com/
5. The Weighting and Scoring method: http://www.dfpni.gov.uk/eag-the-weighting-and-scoring-method
6. Top 10 movies recommendation engines: http://www.cnet.com/news/top-10-movie-recommendation-engines/
7. Windows Communication Foundation: https://msdn.microsoft.com/en-us/library/dd456779(v=vs.110).aspx
8. Windows Presentation Foundation: https://msdn.microsoft.com/en-us/library/aa970268(v=vs.110).asp

A Software Component for Polyglot Text-to-Speech Synthesis: User Interface and Beta Testing Results

Paul Fogarassy-Neszly, Zlatomir Zinveliu
BAUM Engineering
Str. Traian Moşoiu nr. 8, 310175 Arad
pf@baum.ro, zz@baum.ro

Costin Pribeanu
ICI Bucureşti
Bd. Mareşal Averescu nr.8-10, Bucureşti
pribeanu@ici.ro

ABSTRACT

Text-to-speech synthesis has many applications in the area of assistive technologies for visually impaired people. Some applications require multilingual text-to-speech synthesis. In this case, multilingual text analysis and voice switching are desirable. In this paper an improved functional version (beta) of a software component for polyglot text-to-speech synthesis is presented. Beta testing results are useful to reveal the optimal level of parameters and suggest new directions to improve the method and algorithms. Testing results show that the new version of the component is able to automatically detect the language with a reasonable accuracy from texts with a variable degree of fragmentation.

Author Keywords

Multilingual speech synthesis, language recognition, text-to-speech synthesis, assistive technologies beta testing, usability, accessibility

ACM Classification Keywords

D.2.2: Design tools and techniques. H5.2 User interfaces.

INTRODUCTION

Many assistive technologies for visually impaired people are using text-to-speech (TTS). TTS means converting a text document into speech signals, typically by using voice fragments prerecorded by a native speaking person in the desired language [8]. If the text is written in another language then the user has to manually select a corresponding voice for that language. The TTS synthesis is mainly used by people with visual or reading disabilities (dyslectic or illiterate) in order to make accessible the electronic documents. Examples of assistive technologies using TTS synthesis are: screen readers, automatic reading machines, portable computers with voice interface, and Braille display.

There is an increasing interest in applications based on assistive technologies that are able to process texts written in two or more languages. There are many application areas that need polyglot text-to-speech, such as education for all and multi-cultural contexts, to name just two [7, 12, 13, 14].

In this case both a multilingual (polyglot) text-to-speech synthesis and voice switching are needed. This requires to analyze the text piece by piece, to detect the language for each fragment and then to select the voice available for that language. Many approaches for multilingual TTS exist [1] (see [3], [6], [7], [9], [10], and [11]) that differ with regard to the solutions adopted for text analysis and speech synthesis.

Traber et al. [10] classified speech synthesis as regarding multilinguality into four categories: monolingual, simple multilingual, mixed lingual with pre-defined language, and polyglot with language detection. In the first case, foreign words are rendered with the available voice. In the second case, language switching is accompanied by voice switching. In the third case, the synthesis process detects foreign words and adapts the pronunciation and intonation. The fourth category is able to detect the current language using multilingual text analysis and use phonetic and intonation models to generate utterances.

Romsdorfer & Pfister [7] made a more clear distinction between multilingual TTS synthesis that need manually language selection and polyglot TTS synthesis that analyze parts of text in different languages. In this case, language identification of the text is indispensable.

The main objective of the research project iT2V is to develop and implement a software component for automatic language identification and voice switching. The project is carried on in a consortium of three partners: BAUM Engineering, ETA Automatizari Industriale, and National Institute for Research and Development in Informatics – ICI Bucharest.

The development follows four steps: alpha version (proof-of-concept, functional version (beta), commercial version, and implementation in several applications. The lifecycle of iT2V is illustrated in Figure 1.

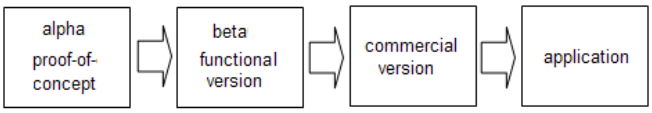

Figure 1. iT2V lifecycle

The alpha version was described in a previous work [4]. In this case, the text was written in one language and the goal was to test the language identification algorithms. In a recent paper a preliminary functional version (beta01) was presented. The evaluation results showed an acceptable accuracy when testing with four candidate languages [5]. However, switching the voice in the middle of a sentence is an important shortcoming for the user.

The objective of this paper is to present an improved functional version of the software component (beta02) and to present and discuss the evaluation results. The component has been tested with four, three and two candidate languages on texts with a variable degree of fragmentation. The results are analyzed against two

additional parameters: look-ahead (number of words considered in text analysis) and inertia at language switching.

The automatic language identification and voice switching are supported by a software component having the role of intermediate layer between the client application and the speech synthesis process. Unlike many other similar algorithms designed for language recognition, our approach is optimized for runtime; this is the main novelty of the proposed software component.

THE SOFTWARE COMPONENT

Functions

The software component supports a polyglot TTS and is able to perform multilingual text analysis, automatic language detection and automatic language switching. It plays the role of intermediate layer, voice independent, between the application and the synthesis process. Language identification is based on computing and comparing trigrams frequencies of a given text [2].

The functional version enables testing of implemented algorithms and analyzing the influence of various parameters on achieved results. It has three main modules that enable language configuration, training and dynamic recognition testing.

Configuration module

The configuration module allowed user to select the candidate language (maximum 4 and minimum 2), the desired voices and to launch the language training and dynamic recognition functions. The user interface is presented in Figure 2.

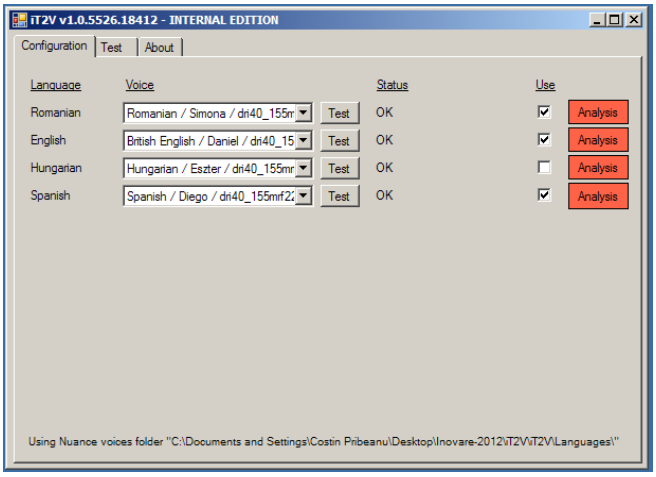

Figure 2. User interface of the configuration module

For each language a voice can be selected from a list of available voices. In the example above, the testing program is configured for 3 candidate languages.

Training module

The language recognition component requires a training phase before use. The interface of the training module enables language analysis that results in tri-gram frequencies. These frequencies can be saved in a language characteristic file in an internal format (.lang). Figure 3 presents the user interface of the training module.

The text can be loaded directly in „Analyze text" window or via a document file (.txt, .doc or .rtf). It is possible to further extend the language file with other documents, if required.

Beta02 was trained with corpora for five languages: Romanian (Ro), English (En), Hungarian (Hu), French (Fr) and German (De). The number of trigrams frequency stored for each language varies from 4.000 to 5.000.

In Figure 2 an example of language analysis for the Romanian corpus is presented.

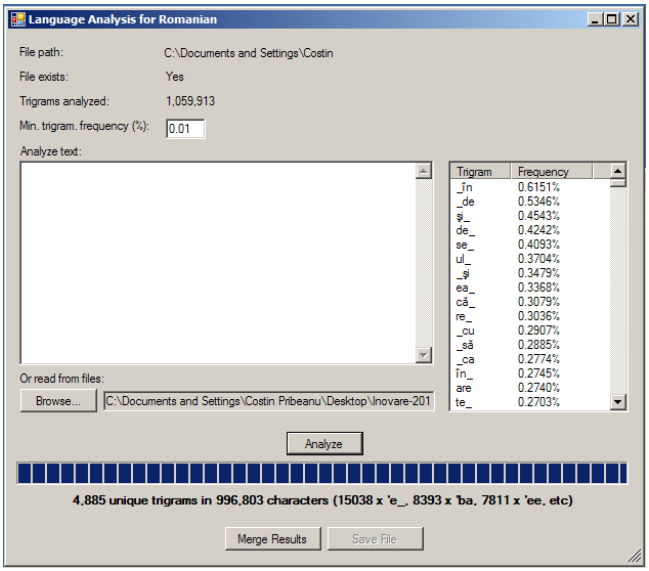

Figure 3. User interface of the training module

The analysis resulted in 4.885 unique trigrams. These could be merged ("Merge results" button) with the existing data for Romanian language and eventually saved ("Save file" button).

Dynamic language recognition module

The main module of the software component enables testing the language recognition between any selections from available languages. The main improvement in beta02 consists in controlling for language switching at the sentence level only, avoiding the disturbing voice change in the middle of the phrase.

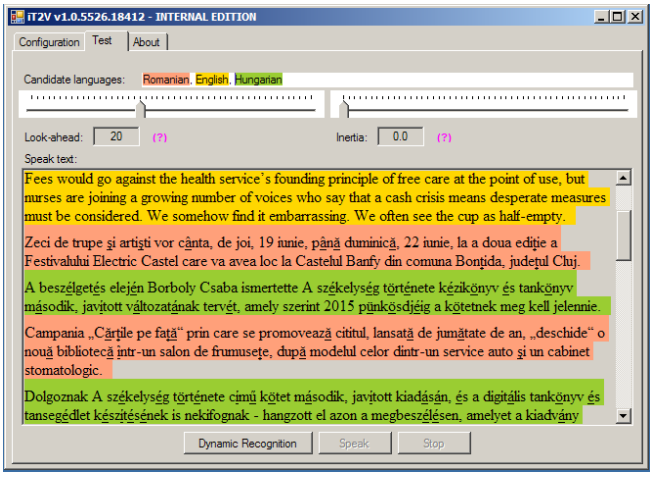

Figure 4. Dynamic recognition language user interface

The user interface is presented in Figure 4 where three candidate languages are specified (Romanian, English,

Hungarian, in this example) and marked with distinct colors.

After text analysis, the module highlights each piece of text with the corresponding language color. This makes it easier to detect errors and assess the precision.

There are two main parameters that could be varied within this module: look-ahead (LA) and inertia (I). The former is an integer specifying the number of words that are analyzed and the latter the inertia at language change, as a difference between computed criteria for candidate languages. The first parameter influences the results since the trigram method is statistical and the precision depends on the size of the analyzed sample. It also affects the response time so an optimal value is the smallest value for which the precision is satisfactory.

The second parameter represents the degree to which the program delays the language switching when the statistical criterion indicates another language for the text analyzed. This parameter affects the user experience. If there is just one word in another language (for example weekend in a Romanian text) it doesn't make sense to change the voice. Also, it is annoying for the user if the voice changes in the middle of a sentence. After testing the beta01 version, it was decided that for the target applications of iT2V is better to restrict language switching at sentence level.

EVALUATION RESULTS

Text used for testing

For testing with four candidate languages, three texts with the same content but different degree of fragmentation (number of language changes) were used: low (3), moderate (10), and high (21). The first text has a distinct paragraph for each language, hence there are only 3 language switching. For further testing with three and two languages the text content for the respective language was preserved.

The text contains sentences from newspapers (Adevărul, Times, Le Monde, Deutche Welle, and Uj Kelet). The text for each language refers to at least two different domains. The text used has 8 sentences / 173 words (Romanian), 10 sentences / 207 words (English), 9 sentences / 217 words (French), 10 sentences / 161 words (Hungarian).

An excerpt from the text with four languages (Ro, En, Fr, and De) and high degree of fragmentation is given bellow.

Les ambulanciers ont dû remettre leurs téléphones portables lorsqu'ils ont vu le visage, qui aurait beaucoup maigri durant les 170 jours d'hospitalisation à Grenoble, après son grave accident de ski, en décembre, selon le journal. Bis zu den Wahlen 2016 in Somalia müssen unbedingt wirksame politische Lösungen für das Land gefunden und moderate Kräfte unter den Islamisten mit einbezogen werden. Il n'a pas parlé, mais il a communiqué avec les ambulanciers par des hochements de tête, durant les quelque 200 km du trajet. Pentru a convinge comisiile de examen că stăpânesc bine limba română, la nivel conversaţional, elevii au trebui să cunoască stilurile limbii române. They urged politicians to say how they plan to pay for a health service that faces a £30 billion funding black hole by the end of the decade. Patient leaders attacked the plan, saying that charges would deter the poor from seeking help, push people towards A&E and require cumbersome bureaucracy to collect. Stilul beletristic este specific romanelor şi încurajează folosirea figurilor de stil, a termenilor arhaici, regionali, jargou etc. Stilul juridico-administrativ apare atunci când vorbim de documente oficiale, iar stilul publicistic este caracteristic articolelor de presă. Es fehlt an Ausrüstung wie etwa Kampfhubschraubern, um die selbsternannten Gotteskrieger in Somalia zu bekämpfen. La chambre de commerce et d'industrie de Paris-Ile-de-France, qui représente 800 000 entreprises, s'inquiète du nombre croissant de jeunes envisageant de faire leur vie professionnelle à l'étranger.

Measures

Following measures were collected: number of candidate languages, number of language switching, look-ahead parameter, inertia parameter, and number of sentences for which the language was correctly detected. The last measure enables to compute the effectiveness of language switching (rate of success) as reported to the total number of sentences in the text.

Procedure

Preliminary tests showed that the results are acceptable (effectiveness over 90%) when the parameters are varying in the range 10-40 for LA and 0.0-2.0 for I.

Testing has been carried on in six sessions. In the first three sessions the beta02 was configured for testing with four candidate languages (Ro, En, Fr, and De). LA parameter was varied with an increment of 5 and I parameter with an increment of 0.3. In the first session the text with low degree of fragmentation was used. In the second and third session the texts with moderate and high degree of fragmentation were used. Next two sessions used three candidate languages (Ro, En, and Hu) and the last session only two candidate languages (Ro and En).

Results

The testing text with no fragmentation was also used for the evaluation of beta01 thus enabling comparison. A synthesis of testing results is presented in Table 1.

NCL is the number of candidate languages, NLS is the number of language switching, LA is the look-ahead parameter, I is the inertia parameter, and EFS is the effectiveness of language switching.

The beta testing results provide with useful information by showing the optimal level of parameters. The results for beta 02 are suggesting a look-ahead parameter in the range of 25-30 with inertia between 0.00 and 0.02.

Table 1. Synthesis of results

beta	NCL	NLS	LA	I	EFS
01	4	3	10	1.9	81.25%
	3	2	10	2.0	81.25%
	2	1	10	2.0	87.50%
02	4	3	30-40	1.0	97.30%
	4	10	20-30	0.0	91.89%
	4	21	20-40	0.3	91.89%
	3	10	25-30	0.0	96.55%
	3	15	25-30	0.0-0.3	96.55%
	2	11	24-30	0.0-0.2	94.44%

As it could be observed, the testing results for beta02 are better. The effectiveness of language switching (EFS) is 97.30% for the text with low fragmentation and 91.89% for the texts with moderate and high fragmentation. Since the text content is identical, this means that the text fragmentation is influencing the results.

CONCLUSION AND FUTURE WORK

In this paper an improved functional version of a software component for polyglot text-to-speech synthesis was presented. The testing results of the beta02 functional version confirm the improvements of language detection algorithms.

Testing was done using texts with similar content but with different degree of fragmentation. The results suggest that text fragmentation is an important parameter for language detection algorithms.

In the next future the commercial version will be finalized and implemented in several applications.

Acknowledgement

This work is supported by the IT2V research project (29DPST/2013), financed by UEFISCDI under the PNCDI II Innovation Program.

REFERENCES

1. Bourlard, H., Dines, J., Magimai-Doss, M., Garner, P. N., Imseng, D., Motlicek, P. & Valente, F. (2011). Current trends in multilingual speech processing. *Sadhana*, *36*(5), 885-915.
2. Cavnar, W., and Trenkle, J. (1994). N-gram-based text categorization. Proc. 3rd Symp. on Document Analysis and Information Retrieval (SDAIR-94)
3. Chen, C. P., Huang, Y. C., Wu, C. H., Lee K. D. (2014). Polyglot speech synthesis based on cross-lingual frame selection using auditory and articulatory features. *IEEE/ACM TASLP 22 (10)*, 1558-1570.
4. Fogarassy-Neszly, P., Gherhes, V. (2014). Applications for dynamic language identification. Popovici D., M. & Iordache D. D. (Eds.) Proceedings RoCHI 2014, 51-54.
5. Pribeanu, C., Fogarassy-Neszly, P. (2014). Beta testing of a dynamic language identification software component - preliminary results. *Revista Romana de Interactiune Om-Calculator 7(3)*, 259-272, 2014.
6. Ramani, B., Actlin Jeeva, M.P., Vijayalaksmi, P., nagarajan, T. (2014). Cross-lingual voice conversion-based polyglot speech synthesizer for Indian languages. *Proceedings INTERSPEECH 2014*, 775-779.
7. Romsdorfer, H., Pfister, B. (2007). Text analysis and language identification for polyglot text-to-speech synthesis. *Speech Communication 49*, 697-724.
8. Shiga, Y. & Kawai, H. (2012). Multilingual speech synthesis system. *Journal of the National Institute of information and Communication Technology 59(3/4)*, 21-28.
9. Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufis, D., and Varga, D. (2006). *The jrc-acquis: A multilingual aligned parallel corpus with 20+ languages*. In Proceedings of LREC'2006, 2142–2147, Genoa, Italy
10. Traber, C., Huber, K., Nedir, K., Pfister, B., Keller, E., Zellner, B. (1999). From multilingual to polyglot speech sinthesis. *Proceedigs of EUROSPEECH*, Budapest, Hungary, 835–838.
11. Traber, C., (1995). *SVOX: the implementation of a text-to-speech system for German*. Ph.D. thesis, No. 11064, Computer Engineering and Networks Laboratory, ETH Zurich (TIK-Schriftenreihe Nr. 7, ISBN 3 7281 2239 4).
12. Tripathi, M., & Shukla, A. (2014). Use of assistive technologies in academic libraries: A survey. *Assistive Technology*, *26*(2), 105-118.
13. Turunen, M., & Hakulinen, J. (2000, October). Mailman-a multilingual speech-only e-mail client based on an adaptive speech application framework. In *Proceedings of Workshop on Multi-Lingual Speech Communication - MSC 2000*, 7-12.
14. Udvari-Solner, A., & Thousand, J. S. (1996). Creating a responsive curriculum for inclusive schools. *Remedial and special education*, *17*(3), 182-191.

.

Author Index