CONTENT 2011

The Third International Conference on Creative Content Technologies

ISBN: 978-1-61208-157-1

September 25-30, 2011
Rome, Italy

CONTENT 2011 Editors

Hans-Werner Sehring, T-Systems Multimedia Solutions GmbH, Germany

Wolfgang Fohl, Hamburg University of Applied Sciences, Germany

CONTENT 2011

Foreword

The Third International Conference on Creative Content Technologies [CONTENT 2011], held between September 25 and 30, 2011 in Rome, Italy, targeted advanced concepts, solutions and applications in producing, transmitting and managing various forms of content and their combination. Multi-cast and uni-cast content distribution, content localization, on-demand or following customer profiles are common challenges for content producers and distributors. Special processing challenges occur when dealing with social, graphic content, animation, speech, voice, image, audio, data, or image contents. Advanced producing and managing mechanisms and methodologies are now embedded in current and soon-to-be solutions.

We welcome technical papers presenting research and practical results, position papers addressing the pros and cons of specific proposals, such as those being discussed in the standard fora or in industry consortia, survey papers addressing the key problems and solutions on any of the above topics short papers on work in progress, and panel proposals.

We take here the opportunity to warmly thank all the members of the CONTENT 2011 Technical Program Committee, as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to CONTENT 2011. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the CONTENT 2011 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that CONTENT 2011 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of creative content technologies.

We are convinced that the participants found the event useful and communications very open. We also hope the attendees enjoyed the charm of Rome, Italy.

CONTENT 2011 Chairs:

Ajith Abraham, Machine Intelligence Research Labs (MIR Labs), USA
Jalel Ben-Othman, Université de Versailles, France
Wolfgang Fohl, Hamburg University of Applied Sciences, Germany
Raouf Hamzaoui, De Montfort University - Leicester, UK
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Zhou Su, Waseda University, Japan
Hans-Werner Sehring, T-Systems Multimedia Solutions GmbH, Germany

CONTENT 2011

Committee

CONTENT Advisory Chairs

Raouf Hamzaoui, De Montfort University - Leicester, UK
Jalel Ben-Othman, Université de Versailles, France
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Wolfgang Fohl, Hamburg University of Applied Sciences, Germany
Zhou Su, Waseda University, Japan

CONTENT 2011 Industry/Research Chairs

Ajith Abraham, Machine Intelligence Research Labs (MIR Labs), USA
Hans-Werner Sehring, T-Systems Multimedia Solutions GmbH, Germany

CONTENT 2011 Technical Program Committee

Dana Al Kukhun, IRIT - University of Toulouse III, France
Marios C. Angelides, Brunel University - Uxbridge, UK
José Enrique Armendáriz-Iñigo, Universidad Pública de Navarra, Spain
Kambiz Badie, Iran Telecom Research Center, Iran
Johannes Behr, Fraunhofer-Institut für Graphische Datenverarbeitung IGD, Germany
Kadi Bouatouch, IRISA / University of Rennes 1, France
Christos Bouras, University of Patras, Greece
Christian Bouville, IRISA – Rennes, France
Yiwei Cao, RWTH Aachen University, Germany
Eduardo Cerqueira, Federal Unversity of Para, Brazil
Raffaele De Amicis,Fondazione Graphitech - Trento Italy
Rafael del Vado Vírseda, Universidad Complutense de Madrid, Spain
Dieter W. Fellner, TU Darmstadt, Germany
Wolfgang Fohl, Hamburg University of Applied Sciences, Germany
Antonio Javier García Sánchez, University of Cartagena, Spain
Afzal Godil, National Institute of Standards and Technology, USA
Patrick Gros, INRIA Rennes - Campus de Beaulieu, France
Raouf Hamzaoui, De Montfort University - Leicester, UK
Chih-Cheng Hung, Southern Polytechnic State University, USA
Jinyuan Jia, Tongji University – Shanghai, P.R. China
Angelos Lazaris, University of Southern California (USC), USA
Maryam Tayefeh Mahmoudi, Iran Telecom Research Center, Iran
Vittorio Manetti, CRIAI - Portici & COMICS / University of Napoli "Federico II", Italy
Joan Navarro, Universidad Ramón Llull, Spain
Michael Adeyeye Oluwasegun, University of Cape Town, South Africa
Sebastian Pena Serna, Fraunhofer-Institut fuer Graphische Datenverarbeitung IGD, Germany
Sofia Pescarin, CNR ITABC - Rome, Italy
Sören Pirk, Universität Konstanz, Germany

Copyright Information

Table of Contents

Webcasting From Challenging Locations

Ronald Marsh
Computer Science Department
University of North Dakota
Grand Forks, ND, USA
rmarsh@cs.und.edu

Timothy Young
Department of Physics and Astrophysics
University of North Dakota
Grand Forks, ND, USA
tim_young@und.nodak.edu

***Abstract*—This paper will discuss the issues associated with webcasting from remote and challenging locations using open source software. Issues discussed will include the pros and cons of point-to-point versus multicasting, the use of commercial software versus open source, website design criteria, and some of the technical issues (network bandwidth requirements, chat room and webpage configuration). The authors have experience in setting up and conducting webcasts of astronomical events which have been broadcast from nine different countries as part of the Sun Earth Moon system (SEMs) project - a public outreach and informal learning project developed at the University of North Dakota (UND). The system developed is also being used to evaluate an unmanned aircraft system in support of Defense Support of Civil Authorities.**

Keywords-webcast; solar and lunar eclipses; social media.

I. Introduction

The scientific community has long been concerned with the gap between advances in natural sciences, particularly in physics and astronomy, and the level of public awareness and involvement [1-3]. Additionally, if young generations do not acquire knowledge of basic scientific concepts, the knowledge gap will continue to widen. Contrary, if young people are equipped with the underlying scientific principles, it is easier for them to stay connected with the growth of scientific knowledge throughout their lives [4]. Therefore, scientists should not only focus on research but also on bridging the gap between science and the public by means of informal science education [5-8]. Hence, the scientific community is currently looking for innovative ways of encouraging informal science learning [9].

Sun Earth Moon system (SEMs) [10] is a public outreach project conceived by the authors to bring live coverage of rare astronomical events to the public using the Internet as a way to encourage informal science learning [11]. Scientists determine beforehand from what geographical location these celestial events can be best observed; then a UND team travels to the location, sets up equipment, and shares the event with SEMs visitors in real time.

SEMs webcasts began with the June 8, 2004 Venus transit that was webcast from Delhi, India. The webcast was very successful with 37,000 visitors to the website occurring during the six hour webcast. This was encouraging as webcasting was still in its early development stage with an adoption rate of only 7% [12]. The second SEMs webcast was the October 28, 2004 lunar eclipse that was webcast from Grand Forks. Unfortunately, it was cloudy. However, the website had over 159,000 visitors in just two hours; which increased the load on the server to the point where the system administrator could not login to place a notice on the website that it was cloudy and that there would be no webcast! It was quickly realized that if the team planned to continue webcasting astronomical events that a better understanding of webcasting methodologies, the current server and software environment, and usage patterns would have to be developed.

For the first two webcasts, accounts and disk space on the UND School of Medicine's MacroMedia Breeze server were obtained. While this software is very capable and easy to use, it limits the number of concurrent viewers (as determined by the site license). Given the problems encountered with the October 28, 2004 lunar eclipse webcast, alternative webcast techniques / packages were explored.

The SEMs webcasts have some unique requirements which drove our decision making process. First and foremost is the tremendous number of visitors that can be expected over a short time period coupled with finite network bandwidth. The second is that the team must be able to broadcast from remote locations – locations were one must make do with whatever Internet connection can be obtained and with whatever equipment can be brought (due to weight and customs restrictions). Therefore, it is mandatory to be able to send a single stream from our remote site to a server located at UND and have that server provide the multitude of public streams. The third is that it is the goal of the SEMs project to give viewers the feeling of being there and to instill the excitement of being part of a global community witnessing a rare event. The fourth is to have the ability to record and store for playback (at the viewer's convenience) the video and audio. As a result, experiments with the use of the media to make the event as life-like as possible for the viewers have been continuously conducted. Unlike the other eclipse webcasts, the SEMs webcasts use streaming color video, audio and have a chat room. It was discovered that having viewers post questions

 ISBN: 978-1-61208-157-1

on the chat room and our answering via audio was very popular. However, such amenities require bandwidth and flexible software/systems.

The remainder of this paper will review the techniques and software that are available and provide a discussion of the techniques/software that have been adopted and developed.

II. BACKGROUND

Webcasting refers to the delivery of audio and video content over the web [13]. Ha and Ganahl [14] point out that "there are many different applications of webcasting in both the nonprofit and the commercial sector" where the web is used as a delivery medium, such as informational, instructional, marketing, and entertainment. Webcasting options can be broken down into two categories: the network protocol and the software used [15, 16]. The network protocol options can be further divided into two sub-categories:

1) *Point-to-point:* This is the most common protocol used by commercial webcast systems as it gives the software a mechanism to track/limit the number of viewers. Furthermore, point-to-point commonly uses HTTP port 80 which is rarely blocked by Internet Service Providers (ISPs). Unfortunately, every viewer creates/requires a separate and redundant connection to the server. Therefore, the number of viewers possible is also limited by the server site's bandwidth.

2) *Multicast:* With multicast a single stream is broadcast by the server and is replicated by all network routers encountered and therefore sent to everyone on the network (whether they want it or not). As a result, multicast consumes a massive amount of network bandwidth and therefore, many ISPs block all multicast signals. Thus, multicast is akin to AM/FM radio where anyone knowing the channel can tune in, thus multicast is not commonly supported by commercial webcast systems. Finally, to webcast over multicast, one needs a multicast address [17], and one must set the time-to-live on the packets such that the appropriate number of routers is crossed to reach the desired audience.

The software options can also be divided into two sub-categories:

1) *Commercial Software:* Commercial packages typically provide support for many cameras and microphones and provide useful features such as chat rooms, the ability to show presentation slides, have remote feed capability and have web browser interfaces. The drawbacks to using commercial packages include the price, the lack of support of multicast, license restrictions limiting the number of viewers, and the need to install web browser plug-ins.

2) *Open source:* Open source packages range widely in capability and few support chat rooms, the ability to show presentation slides, or remote feeds. We have also found camera and microphone support spotty in these packages. However, free packages have no license restrictions and more likely to support multicast.

One can also consider a push or pull strategy. The "pull" strategy requires the receiver to initiate a message transfer by explicitly contacting the sender; therefore, a website using the "pull" model only allows users to retrieve information. In the "push" model, the sender knows the identity of the receiver in advance and pushes messages in an asynchronous manner to the receiver [18, 19]. For the purposes of creating the utmost, life-like experience of being part of a global community witnessing a rare astronomical event, the "pull" strategy was found to be most effective. In addition, the nature of the webcasts and network bandwidth has limited our ability to support the "push" model.

III. UND/SEMS WEBCAST SYSTEM

Given our requirements and limited funding, we were fortunate to realize that we could develop a system that would meet our needs using open source software. Hence, the UND/SEMs webcast system relies heavily on open source software. The chat room and website are hosted on a Linux server running the High Performance Internet Relay Chat software [20] and the Apache web server [21]. The multimedia (audio and video) streams are served by three Windows XP servers each running the VLC media player [22]. All servers are located at UND. A webcam connected to a laptop computer is used to capture and forward the multimedia stream from the remote site to one of the VLC servers using a User Datagram Protocol (UDP) connection. Two copies of the VLC media player (version 8.6i) are executed on the laptop computer. The first is used to acquire the multimedia data from connected devices (webcam, microphone, etc.), to make a high quality recording on the laptop, and to stream the multimedia data to the loopback network connection (127.0.0.1). The second receives the multimedia data from the loopback network connection, transcodes the multimedia stream to reduce its bandwidth, and streams the transcoded multimedia data to UND using the UDP connection. At UND, the multimedia stream is received by the first VLC server which provides two multimedia streams, a webcast/HTTP stream and a multicast/UDP stream (the North Dakota Higher Education Computer Network possesses a block of multicast addresses). Both of these multimedia streams use the same codec as was employed by the laptop. The second VLC server receives the multicast stream and transcodes it producing a Microsoft MediaPlayer compatible webcast multimedia stream. The third VLC server receives the multicast stream and transcodes it producing a MPEG4 Real Time Streaming Protocol *(*RTSP*)* multimedia stream compatible with many mobile devices (e.g., smartphones). Experience has shown that when serving N unique (different codecs) video streams it is best to use N unique IP addresses

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-157-1

on N unique servers for stability. Hence, two video streams require two servers, each sending a uniquely encoded stream. Even a dual processor server with dual network interface cards has proven to be unstable when used for this purpose. However, multiple streams using the same codec has proven to be stable when served from a single computer with multiple network interfaces. Figure 1 depicts the webcasting system.

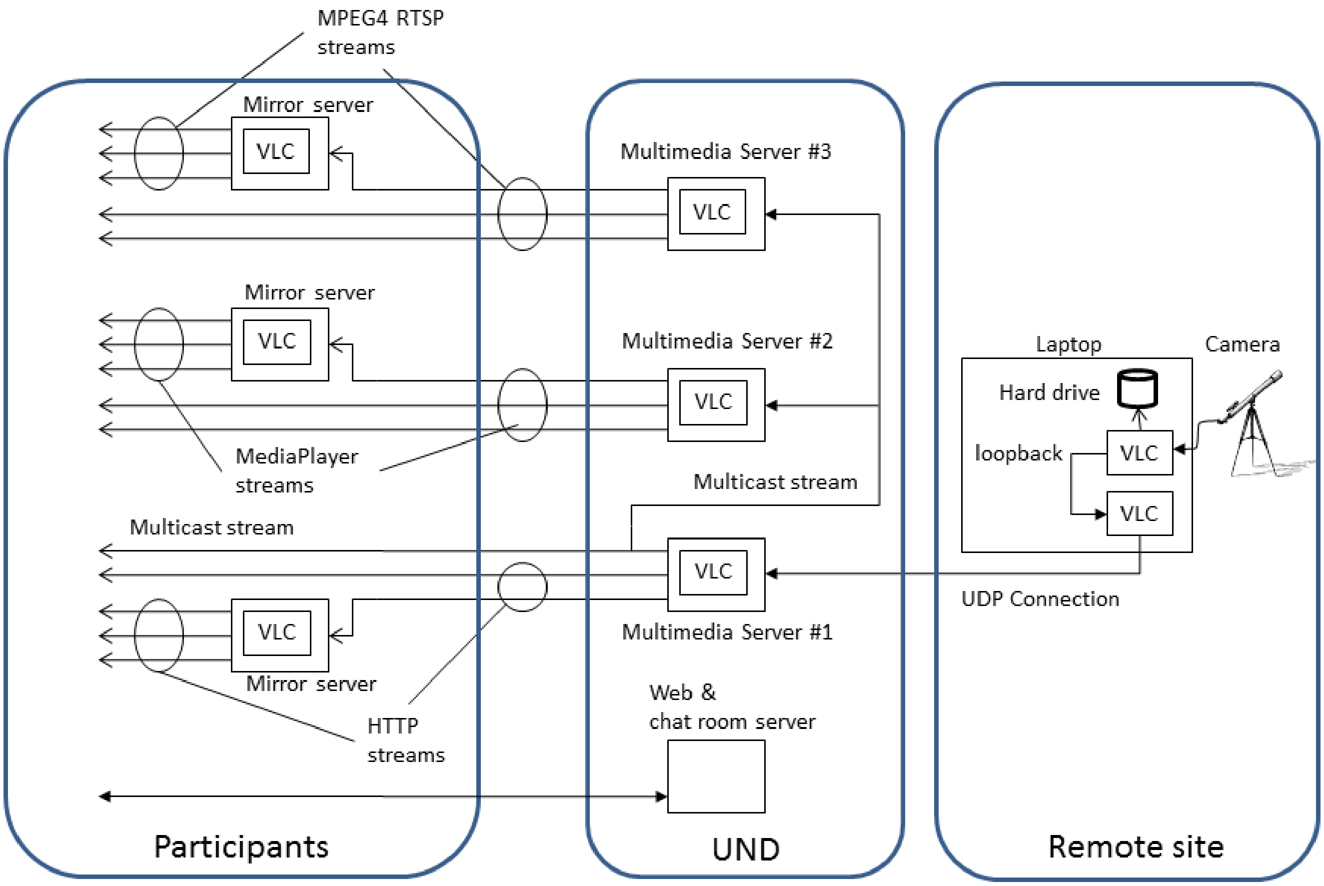


Figure 1. UND/SEMS webcast system

Fortunately, VLC can be easily configured (via command line options or a startup script) to allow mirror sites. Thus, in order to reduce bandwidth requirements at UND, a mirroring scheme has been employed. To date, the SEMs webcasts have been mirrored by Pennsylvania State University (main campus at University Park, and Wilkes-Barre campus), High Performance Computing Research Department at Lawrence Berkley National Laboratory, and Williston North Dakota School District in the United States; Universidad de Sonora in Mexico; and the University of Barcelona in Spain.

Unfortunately, VLC does not provide support for a chat room, presentation slides, has relatively poor web browser support, spotty camera and microphone support and requires unique web browser plug-ins. However, due to its popularity, a lot of people not associated with the VLC project have developed components to extend its capability. For example, JavaScript applications have been written that allow VLC streams to be viewed on a web browser (IE and FireFox).

To support social interaction [23, 24] and solicit feedback within an open environment like SEMs, the site integrates an online chat room supported by open source JavaScript code. The integration of a chat room provides visitors with an opportunity to become engaged in discussions with UND scientists and fellow viewers at a time of an event. The inclusion of a real-time discussion turned out to be very popular. Therefore, with a little development, the SEMs team was able to create a website with nearly the same look and feel as a commercial package, but that provided the non-commercial features required.

The presence of the blog is another important factor for a website's success. Originally, the SEMs project weblog consisted of postings of daily activities, trials and tribulations of fieldwork, links to local Internet-radio stations and webcams, as well as information about local cultures. The current site design allows readers to comment on the project blog. The commenting feature was implemented during the August 1, 2008 total solar eclipse webcast that was broadcasted from Xi'an, China. According to Du and Wagner [25], factors that determine the success of a blog include content, the technology used to support the blog, and its social value. The content is the information (written or media) provided. The technology used to support the weblog should be interactive and should be able to present and organize content facilitating social interaction among bloggers [26]. Research has also shown that users prefer blogs that are commented on by other visitors and support social interaction [27, 28].

Once you have arrived and successfully "setup shop" you have to determine how much bandwidth you have. Or more importantly, what is the bandwidth back to the server. While local bandwidth may be very high, there may be a low bandwidth link somewhere between you and your servers. One tool that has proven useful is Speedtest.net (www.speedtest.net) and/or other similar sites (as long as they have a server near your servers). However, the only reliable way to verify what bandwidth you have is to test the stream and to adjust the video and audio data rates (VLC provides this capability) to obtain the best quality of service possible. It helps to have someone monitoring the stream from a location at or near the server and to have them communicate to you regarding the stream quality (the chat room works nicely for this). Unfortunately, there is no obvious formula for determining the best combination of codecs, frame rates, and bit rates. Given the many variables, trial and error is the only real solution.

IV. SCRIPTS

We have developed a set of scripts that allow our system to be used by others, those scripts are listed below:

A. *Laptop VLC Script #1*

This MS Windows 7 script starts VLC instructing it to accept connect to a Vimicro USB PC Camera and the default audio device. The script then sets the video size to 320x240 with a frame rate of 6 frames per second. The script then transcodes the stream into a mpeg4 (video and audio) stream, sets the bit rates, duplicates the stream for local display, for local storage (filename: "test.mp4"), and broadcasts the stream over the loopback network connection (http port 80).

```
C:\Progra~2\VideoLAN\VLC\vlc.exe -vvv dshow://
:dshow-vdev="Vimicro USB PC Camera (ZC0301PL)"
:dshow-adev="" :dshow-size="320x240" :dshow-
caching=200  :dshow-fps=6.000000 :sout=#transcode
{vcodec=mp4v,vb=3072,scale=1,acodec=mp4a,ab=192,
channels=2}:duplicate{dst=display,dst=std
{access=file,mux=mp4,dst="C:\Users\rmarsh\Desktop\te
st.mp4"},dst=std{access=http,mux=ts,dst=127.0.0.1:80}
}
```

B. *Laptop VLC Script #2*

This MS Windows 7 script starts VLC instructing it to accept the stream on the loopback network connection and to transcode the stream into a reduced bit rate DivX3 stream. The stream is then displayed locally and forwards it via UDP to UND VLC server #1 (bbbb.cs.und.edu).

```
C:\Progra~2\VideoLAN\VLC\vlc.exe -vvv
http://127.0.0.1:80 :sout=#transcode{vcodec=DIV3,
vb=256,scale=1,acodec=mp3,ab=192,channels=2}
:duplicate{dst=display,dst=std{access=udp,mux=ts,
dst=bbbb.cs.und.edu:1235}}
```

C. *UND VLC Server #1 script*

This MS Windows XP script terminates any running version of VLC (taskkill) and restarts VLC instructing it to accept a UDP stream addressed to it on port 1235 and to rebroadcast it over multicast (UDP) on IP xxx.xxx.xxx.xxx port 1234 and over http on IP/host name bbbb.cs.und.edu port 80 with a time-to-live of 200. Note that the host name of this machine would be bbbb.cs.und.edu.

```
taskkill /f /im vlc.*

C:\Progra~1\VideoLAN\VLC\vlc.exe -vvv
udp://@:1235
:sout=#duplicate{dst=std{access=udp,mux=ts,dst=xxx.
xxx.xxx.xxx:1234},dst=std{access=http,mux=ts,dst=bbb
b.cs.und.edu:80}} --ttl 200
```

D. *UND VLC Server #2 script*

This MS Windows XP script terminates any running version of VLC (taskkill) and restarts VLC instructing it to accept the multicast (UDP) stream produced by server #1 on port 1234 and to transcode and rebroadcast it over mmsh (media player) on IP/host name cccc.cs.und.edu port 80 with a time-to-live of 200. Note that the host name of this machine would be cccc.cs.und.edu.

```
taskkill /f /im vlc.*

C:\Progra~1\VideoLAN\VLC\vlc.exe -vvv
udp://@xxx.xxx.xxx.xxx:1234
:sout=#duplicate{dst=std{access=mmsh,mux=asf,
dst=cccc.cs.und.edu:80}} --ttl 200
```

E. *UND VLC Server #3 script*

Like the previous MS Windows XP scripts, this script terminates any running version of VLC (taskkill) and restarts VLC instructing it to accept the multicast (UDP) stream produced by server #1 on port 1234. This script then transcodes the stream into mpeg4 (video and audio) and rebroadcasts it as a Real Time Streaming Protocol (RTSP) on IP/host name dddd.cs.und.edu port 554 with a time-to-live of 200. Note that the host name of this machine would be dddd.cs.und.edu. Finally, we chose not to include this

 ISBN: 978-1-61208-157-1

script at this time as we are still debugging and testing it on the different cell phone provider's systems and their smart phones available.

Note that to use any of these scripts one will need to copy and paste the text into Notepad and save it as a DOS batch file.

V. CONCLUSION

By using a combination of open source software, our own scripts, and a combination of point-to-point and multicast webcast technologies a hierarchical webcasting system that allows many more viewers than would be possible with any single approach has been achieved. A multicast stream over Internet2, primarily for university viewers, and multiple point-to-point streams, for home viewers via the SEMs site and mirror sites, is provided. A similar webcast system could easily be developed by anyone else wanting to produce such live events.

The latest development/usage is for Defense Support of Civil Authorities. For this project video acquired by a UND operated (http://www.uasresearch.com/home.aspx) Insitu ScanEagle unmanned aircraft is captured by a computer located in the mobile command center [29-32]. Two instances of VLC are again used to make a high quality local recording and a lower quality transcoded version for streaming. Using a cell phone, the transcoded video stream is then sent to the VLC servers at UND using a User Datagram Protocol (UDP) connection. From there emergency management personnel have access to the video stream via a dedicated website. We have also achieved some success in streaming that video to smart phones (iPhone, Android, etc.) using the Real Time Streaming Protocol (RTSP).

REFERENCES

[1] J. Baxter, "Children's understanding of familiar astronomical events," International Journal of Science Education, 11, 1989, pp. 502-513.

[2] J. Dunlop, "How children observe the universe," Publications of the Astronomical Society of Australia, vol. 17, no. 2,2000, pp. 194-206.

[3] P. M. Sadler, "Misconceptions in astronomy," Proceedings of the second international seminar misconceptions and educational strategies in science and mathematics (Vol. III), J. Novak, Ed. Ithaca, NY: Cornell University Press, 1987, pp. 422-425.

[4] L. D. Richardson and M. Wolfe, Principles and practice of informal education: Learning through life. New York, NY: RoutledgeFalmer. 2001.

[5] J. H. Falk, E. Donovan, and R. Woods, Free-choice science education: How we learn science outside of school, New York, NY: Teachers College Press. 2001.

[6] A. Hofstein and S. Rosenfeld, "Bridging the gap between formal and informal science learning," Studies in Science Education. vol. 28, 1996, pp. 87-112.

[7] P. D. Hurd, Inventing science education for the new millennium, New York, NY: Teachers College Press.1997.

[8] G. Sonnert and G. J. Holton, Ivory bridges: Connecting science and society. Cambridge, MA: MIT Press. 2002.

[9] L. J. Rennie and S. M. Stocklmayer, "The communication of science and technology: Past, present and future agendas," International Journal of Science Education, vol. 25, no. 6, 2003, pp. 759-773.

[10] Sun Earth Moon System, www.sems.und.edu. Last accessed June 26, 2011.

[11] R. Marsh, "The Sun Earth Moon System; Connecting Science and Informal Learning," Handbook of Research on Social Interaction Technologies and Collaboration Software: Concepts and Trends," T. Dumova and R. Fiordo (Eds), pp. 428-438, July, 2009, [ISBN: 978-1-60566-368-5].

[12] C. A. Lin, "Webcasting adoption: Technology fluidity, user innovativeness, and media substitution," Journal of Broadcasting & Electronic Media, vol. 48, no. 3, 2004, pp. 157-178.

[13] P. Miles, Internet world guide to webcasting. New York, NY: John Wiley & Sons. 1998.

[14] L. Ha and R. J. Ganahl, Webcasting worldwide: Business models of an emerging global medium, Mahwah, NJ: Lawrence Erlbaum. 2007.

[15] J. Follansbee, Hands-on guide to streaming media: An introduction to delivering on-demand media. Burlington, MA: Focal Press. 2006.

[16] J. Mann, When and how to use webcasting. Stamford, CT: Gartner. 2005.

[17] Cisco Systems, Inc, "Guideline for Enterprise IP Multicast Address Allocation," http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6552/ps6592/prod_white_paper0900aecd80310d68.pdf. Last accessed June 26, 2011.

[18] S. Mack and D. Rayburn, Hands-on guide to webcasting: Internet event and AV production. Amsterdam: Elsevier/Focal Press. 2006.

[19] B. O. Szuprowicz, Webcasting and push technology strategies: Effective communications for intranets and extranets. Charleston, SC: Computer Technology Research Corp. 1998.

[20] IRCD-Hybrid, www.ircd-hybrid.org. Last accessed June 26, 2011.

[21] The Apache Software Foundation, www.apache.org. Last accessed June 26, 2011.

[22] VideoLAN, www.videolan.org. Last accessed June 26, 2011.

[23] J. Porter, Designing for the social web. Berkeley, CA: New Riders. 2008.

[24] J. Preece, I. R. Rogers, and H. Sharp, Interaction design: Beyond human-computer interaction. Hoboken, NJ: John Wiley & Sons. 2007.

[25] H. S. Du and C. Wagner, "Weblog success: Exploring the role of technology," International Journal of Human-Computer Studies, vol. 64, no. 9, 2006, pp. 789-798.

[26] L. D. Richardson and M. Wolfe, Principles and practice of informal education: Learning through life. New York, NY: RoutledgeFalmer. 2001.

[27] C. Shirky, "Power laws, weblogs, and inequality," Clay Shirky's Writings About the Internet, 2003, unpublished Retrieved April 16, 2011, from http://www.shirky.com/writings/powerlaw_weblog.html

[28] C. Wagner and N. Bolloju, "Supporting knowledge management in organizations with conversational technologies: Discussion forums, weblogs, and wikis," Journal of Database Management, vol. 16, no. 2, 2005, pp. i–viii.

[29] R. Marsh, M. Kary, K. Ogaard, and J. Nordlie, "Development of a Mobile Ganged Phase Array Radar – Risk Mitigation System System for UAS Operations in North Dakota," Proceedings of the 2010 International Conference on Software Engineering Research & Practice, {SERP} 2010, Las Vegas, NV., 2010, pp. 403-410.

[30] R. Marsh, M. Kary, K. Ogaard, J. Nordlie, and C. Thesien, "Development of an Information Display System for UAS Operations in North Dakota," Proceedings of the IADIS International Conference Computer Graphics, Visualization, Computer Vision and Image Processing 2010 (part of MCCSIS 2010), Frieburg, Germany, 2010, pp. 193-200.

[31] H. Reza, M. Askelson, and R. Marsh, "A Fault Tolerant Architecture Using AADLs for Unmanned Aircraft Systems (UAS)," Proceedings of the 2010 International Conference on Software Engineering Research & Practice, {SERP} 2010, Las Vegas, NV., 2010, pp. 180-184.

[32] R. Marsh, M. Kary, K. Ogaard, and J. Nordlie, "Development of a Mobile Information Display System for UAS Operations in North Dakota," International Journal of Computer Information Systems and Industrial Management Applications, Volume 3, 2011, pp. 435-443.

Location-Based Mobile Collaborative Digital Narrative Platform

Sicai Lin[1], Tzu-Chieh Tsai[2], Sheng-Chih Chen[1], Shindi Lee[2]
Master Program of Digital Content and Technologies[1]
Department of Computer Science[2]
National Chengchi University
Taipei, Taiwan
sicai720@gmail.com ttsai@cs.nccu.edu.tw scchen@nccu.edu.tw shindilee@gmail.com

***Abstract*—This study is based on interactive design theory, supplemented by the concept of communication. We propose a "Mobile Collaboration Digital Narrative Platform", through which, with the aid of technologies, a user can, depending on his or her location, download one's favorite collaborative narrative content, and also have the ability to add, edit or record what is happening around; thus, linking the narrative content and the location. Through the function of "collaboration", the content can be made more diverse and rich and the reader can better immerse him or herself in it. The platform also can work in offline mode. Currently, the preliminary design of the system has been completed, and tests in different situations will be conducted and user behaviors will be recorded and then analyzed. Quantitative and qualitative evaluation and analysis of interface design, operational processes, system functions, and collaboration narrative content of the output are in progress. We believe that this study will be an important application of mobile content.**

Keywords-Location based; Digital Narrative; Collaborative Narrative; Mobile Technology.

I. INTRODUCTION

In the era of Internet and the rapidly changing technologies, the types of narrative have become diversified and rich. This also makes digital narrative have more different ways to create innovative and surprising content.

One result of the Internet boom, Tim O'Reilly emphasized that the content generated by users, through user interaction with Web 2.0, results in diverse and rich content. This also makes the form of digital narrative has had a major change. The International Telecommunication Organization (ITU) indicated that in 2010, more than 90% of the worldwide population use mobile phones, of which 9.4 million are 3G users [16]. Meanwhile, the Institute for Information Industry (III) in Taiwan said that for the 3rd quarter of 2010, 69.5 % of mobile users have subscribed to mobile Internet service [1].

The above data imply that we are gradually entering the era of mobile Internet. Our daily lives are filled with a variety of mobile devices. Thus, we can also imagine that the future media narrative will be impacted by the new technology platform and people's lifestyle changes. This will become a totally brand new outlook.

This research is focused on the new media narrative for this trend, and develops a pilot platform design. To this end, the goal of our work is to design a location-based mobile collaborative digital narrative platform. This platform must have the features of 'mobile', 'location-based', and 'collaborative'. The design is based on the user behavior and experience on using the mobile platform for digital narrative. With the designated collaboration features assisted by mobile communication technologies, users can create spark clashes with each other, and can even easily disseminate their ideas and record things with their surrounding stories.

A. *Why Mobile*

As mentioned earlier, data show that the number of mobile phones has exceeded 100% of the population, and mobile Internet subscription is also increasing at an exponential rate. Our lifestyle is thus changing: whatever we rely on computers to do before, is now being replaced by mobile devices. The other is being under the influence of globalization. The world is like a global village, and people moving for interchanging information become a trend. Therefore, high mobility device, has been become an indispensable tool for modern lives. According to the survey from a wireless service company in U.S., SinglePoint [23], the phone is the closest media to the user (Figure 1). We are in the age of speed and convenience; mobile phones are definitely becoming the best choice for users to get all the needed information.

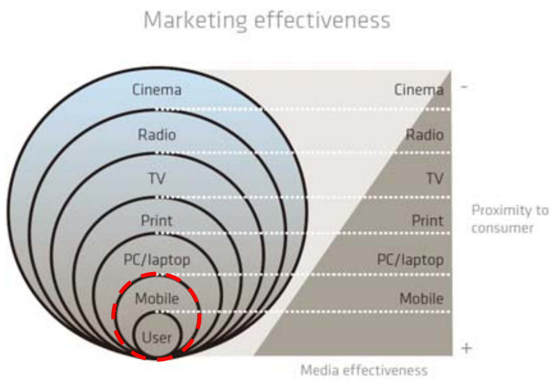


Figure 1. Relationship between media and users [23]

B. *Why Location-Based*

Nowadays, the characteristics of the smart mobile devices contain touch input, GPS, sensors, and Internet access capabilities. We can only carry one mobile phone to make records and tell a story of life without constraints of space and time. Narrative in the past has ignored the atmosphere of 'space'; however, the portability of mobile devices gives the narrative vitality for the new development opportunities to the narrative of space. Narrative scholars such as Herman said: "a storyteller prompts his or her interlocutors to relocate from the here and now of the current interaction to the alternative space-time coordinates of the storyworld." [15]. Henry Jenkins said: "In a new transmedia storytelling environment, I want to introduce an important third term into this discussion - spatiality - and argue for an understanding of game designers less as storytellers and more as narrative architects." [17]. Therefore, through the supplement of mobile devices, the narrative may be able to let users add more space into the story elements. In addition to the narrator, audiences are also assisted through mobile technology, so that the narrative content with the perception of space can have more immediate and immersive nature of the association. Just as Y.-F. Tuan said, "we can perceive the body senses the presence of space" [7].

Besides, from the users' habits and experience point of view, "location-based services" is the highest viscosity of service (Figure 2), according to market research consulting firm InsithtXplorer survey. This is because the location is a significant factor in attracting users [2]. Therefore, the location information will be built in our platform system, and is also an important basis for the user to search, edit and read.

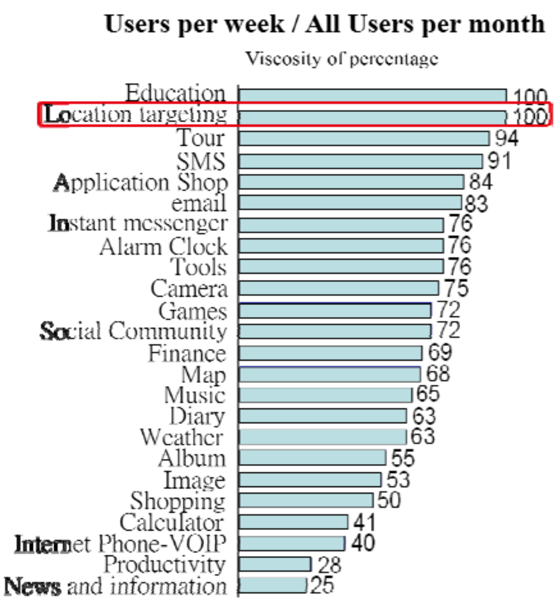


Figure 2. Using viscosity survey [2]

To this end, we also find some service and content providers trying to develop similar applications. These applications are based on the story games with location information. With these features, the games are more "stereo" and more vivid. For example, Disney's Kim Possible World Showcase Adventure [24] uses Walt Disney World theme parks as the playgrounds, and uses the Disney animated TV series Kim Possible as the story elements. In this game, everyone can join the organization using mobile devices. The system can set the roles as heroes or villains to achieve the tasks given. The Walt Disney Company hoped that the application can impress the minds of tourists, and somewhat make the park much special from others.

The Disney's application is customized and not open to the public to develop more other applications. However, it did the concept proof of the trend for location-based story games.

C. *Why Collaborative*

The earliest forms of storytelling were thought to have been primarily oral, combined with gestures and expressions. Digital storytelling means using new digital tools to help ordinary people tell stories in a compelling and emotionally engaging form, so that the story becomes more rich and diverse and full of surprises. Cao [22] had proposed a PESE (Personalized Storytelling Environment) system which used the concept of Web 2.0, namely, collaborative narrative approach to production stories. The idea for PESE is to combine both multimedia production and Web 2.0 production knowledge. Storytelling is an efficient means to fulfill learning goals. Knowledge is exchanged within communities when stories are told.

Jhao-Ling Chen [9] created collaborative narrative storylog which was integrating the social sciences with computing technology to help reveal personal brain thinking to realize physically by storytelling, immersed audio, location-specific content, and blog. Users interact with others to create the story from the formation of the partnership, and thus widen the virtual social relationship.

Collaborative narrative created the "Collective wisdom" which is an important output in the Web 2.0 era. Cooperation not only enriches the narrative content, but also broadens social relations. Through mutual cooperation with the exchange of ideas and feelings, then a story with wonderful content can be produced. Therefore, the collaboration feature will be also one of the features in our platform design.

To sum up, the contribution of this research includes the pilot study of the field of "mobile narrative". We figure out the important features that have to be added in such systems. We also base on our survey results to develop a platform to do field trial.

D. *This paper organization*

This paper is organized as follows: Section I, as aforementioned, briefed the relevant research and survey data. Then we addressed the motivation, background and goal of our work. Section II presents the case studies. We will discuss and analyze the characteristics for the existing mobile digital narrative applications. To this end, we also compare them from different usage aspects. Section III will describe our system architecture and experimental process

 ISBN: 978-1-61208-157-1

design. Section IV presents the evaluation results for the study. Finally, conclusions and future prospects will be presented in Section V.

II. CASE STUDY

A. PicPlz

Picplz (Fig. 3) is a photo story-telling application. It is built-in a variety of filter effects which allow users to take beautiful photos in different styles easily. Photos can be also tagged with the location information and messages. At the same time, they also can be synched with many social networking sites.

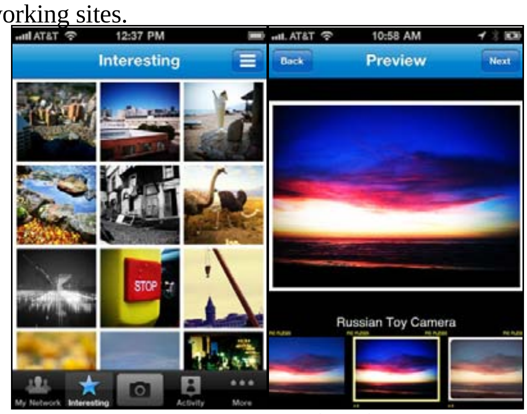


Figure 3. PicPlz interface fig [25]

B. Broadcastr

Broadcastr is a voice platform. It gives a way to record voice that you saw and heard. Anyone can upload a voice recorder to the platform with the tag of its location in the physical coordinates. The users have to be at the right location to be able to listen to the audio tagged by the same location. Thus the experience of hearing the audio will be more 'stereo'. (Figure 4).

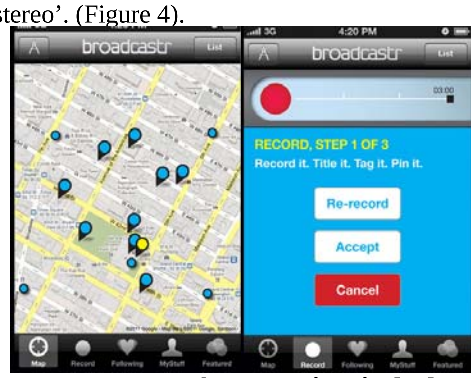


Figure 4. Broadcastr interface fig [26]

C. Instgram

Instgram is a photo story sharing application. To use it is very intuitive. Through its built-in camera effects, users do not have too much skill of photography, and are able to shoot quite photography texture photos. Briefly speaking, what you see will be what is taken. Also the recorded images can be uploaded to Facebook, Twitter and other social networking sites to share with friends (Figure 5).

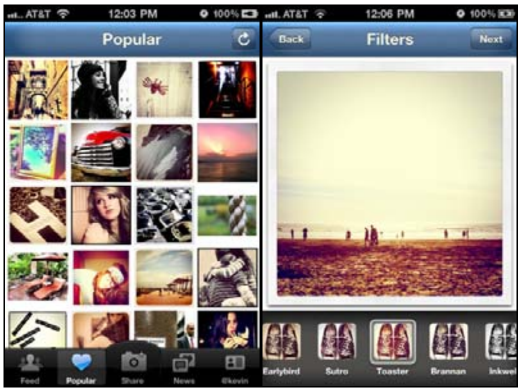


Figure 5. Instgram [27]

D. Summary

The above case studies show that using mobile devices to record our life stories or doing digital narrative creation can produce different narrative content. And allowing interaction between people will even change the content subtly. Although the above applications utilize the characteristics of mobile phones, it still did not fully use all the features of smart phones. Specifically, they are not sufficient to support the needs of the new narrative trend we mentioned in the previous section. Therefore, our system design is developed for improving all the above related applications. Table 1 summarizes the comparisons for the mobile applications including our platform, called Plastory.

TABLE 1. PLASTORY (DESIGN IN THIS RESEARCH) COMPARISON WITH OTHER APPLICATIONS (SOURCE: THIS RESEARCH)

	Plastory	Picplz	Broadcastr	Instgram
Platform	Android	Android iPhone	iPhone	iPhone
Location Information	✔	✔	✔	✔
Collaborative	✔	✖	✖	✖
Offline Edit	✔	Can take photo, Can not upload automatically	✖	Can take photo, Can not upload automatically
Media Type	Photo Sound Text	Photo	Sound	Photo

This platform will build the database to store user-edited files. Users can upload or download data, and can also query collaborative content. In order to achieve the space concept of narrative, so the contents to be queried must be restricted with the user's associated location. Through the platform, digital narrative storytelling can have a whole new content creating experience.

III. System Architecture and Experimental Procedure

A. System architecture

The system architecture is the client-server model. The server is the database for storing and processing data (story) generated by users. After editing, users can upload the stories to the database and share the content with others. In order to maintain the 'spatiality', only the content restricted to the nearby location can be queried or downloaded to the mobile device (client). The mobile app will have the features mentioned before, e.g. location service agent, information processing agent, sensing management and data analysis, etc. (Fig. 6).

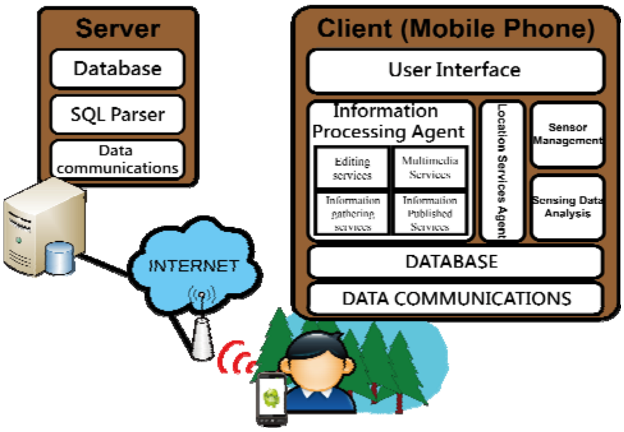


Figure 6. System architecture

B. System Design

The system is implemented on the Android phones. The features include the following:

1. Add collaborative editing.
2. Support offline editing.
3. Allows the users to use a variety of media type for digital narrative.
4. To a restricted geographical location, the user must visit the place to be able to read the existing narrative.

The followings will explain the user interface and operation procedures for our system:

First the user must apply for an account of the system to log in before using (Fig. 7). After logging, the system will remember the status automatically, unless the user clicks log off.

We design four sub-pages with the main page (Figure 8). These four sub-pages are major frequently used functions. This sub-pages design in this user interface is for easily switching pages. Fig. 8 shows the location of nearby users and currently available narrative themes.

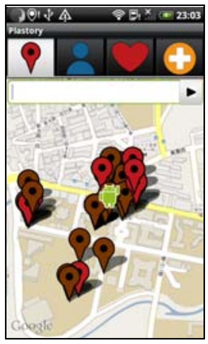


Figure 7. System Login screen, Figure 8. System main page

Click the anchor on the map (Figure 9), the system will show the list of the nearby narrative themes below. This list design is for easy distinguishing from all the themes, in the case of too many themes overlapping on the map.

Figure 9. The menu bar of clicking the anchor on map

Figure 10 is the profile page. The page will display the personal points value (the number of stars), the number of friends, the number of collaborative partners, and the number of published content. Three types of the content presentation can be chosen, namely (from left to right in Fig. 10), rendering on the map, arranged in thumbnail and list type.

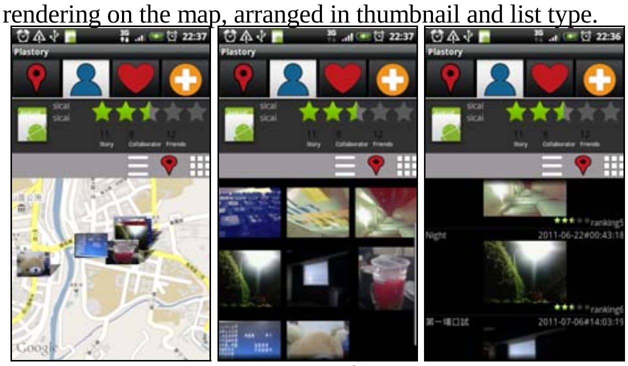


Figure 10. Profile page

Figure 11 is My Stuff page that includes my friends, my collections, and shopping cart items. "Shopping cart items" is designed to allow users to access other people's stories, so that the contents of the platform can have a high degree of interaction/collaboration in order to be able to produce more diverse and rich content.

Add new page is shown in Figure 12 The content is divided into adding new story theme, and adding collaborative story. The latter can be used by any purpose of content creation such that the collaboration will lead to many possibilities of content produced, and also change the original mode of communication.

Figure 11. My Stuff page

Figure 12. Add New page

We have introduced the system's main screen and operation procedures. After editing the content, it will be stored in of the local database and get synchronized with the Server until the networking access is available. Besides, if encountering other's mobiles in the vicinity, the newly editing content can be exchanged mutually in order to have more timely interaction and efficient collaboration with the content.

C. *Experimental procedures and evaluation methods*

As for experiment, we first design the use context by features of the platform, and then invite the users to use the platform for the designated situations. The system will record the use process which will be evaluated for the system interface and usability problems between various scenarios by the human-machine interface evaluation method. The system will also log the editing and reading situation for further understanding of user behaviors. The following will have the detailed description of the experiment:

1) Contextual Design

For evaluating the collaborative editing features for the platform, we design the following test scenario:

During the off-campus extracurricular teaching activities or collaboratively collecting data, it often takes time and is inefficient to make notes or interaction if using pens and papers by the traditional ways. However, in the mobile generation now, through the various functions on the mobile device, such as: cameras, microphones, GPS, etc., we are able to instantly record sound, image and location. Our system platform fully integrates these functions which are further coupled with the collaboration mechanism. Thus, the way of communication has changed and information dissemination becomes more real-time. A new mode of digital narrative is formed.

Besides, during the test scenario using the platform to create or read story/message, we are also interested in finding how users feel, and what users prefer in doing various narrations. Therefore, we will ask the users to use the platform in some difference venues and at all time. Then the system will have more complete log for us to analyze and evaluate.

2) Human-Computer Interface Evaluation

In this part, two evaluation methods for evaluation are adopted, namely:

- Think aloud method

  Thinking aloud method was proposed by Ericsson and Simon in 1984. This method is to allow the users to express their thinking, feeling and suggestions verbally when operating the system. What the user do and say are recorded for analysis. Nielsen [20] said that using the think aloud method to conduct usability assessments, about 80% of usability problems can be found if five persons are tested. Almost 95% of the problems can be found if ten persons are tested. Therefore, five to ten persons to participate in the test can get the best efficiency. The advantage of this method is to understand the relationship between how they use the system and what they think. Aside, the system can be directly enhanced by the suggestions that users make.

3) System log analysis

We develop the points and ranking system to motivate users to create good quality of content and to participate more frequently with the collaboration. The results will be displayed on the integral User's personal page to encourage users often use the platform. Other users can also contribute and encourage high-quality content by giving appropriate scores. The points will be used as the basis for ranking when navigating. We think that such a mechanism should improve the quality of the content, and also learn the user content preferences and usage behaviors.

The parameters used to calculate points are: using time, the number of articles published, the number of collaborative articles published, the number of new friends, viewing times of the published content, the number of content collected, the number of cited content (Add to Cart) and so on. Personal profile page will display the total score by different number of asterisk.

IV. EVALUTION RESULTS

The evaluation has two parts: interface/features design, and system performance/satisfaction. In the first part of evaluation, we found out that users did like to use mobile devices to do narrative. However, they prefer use more photos to record than audios or videos. Nevertheless, they all expressed they will consider to use audio in the future. The reason might be that the new mobile system needs to take time to influence the users' narrative behavior. Another reason could be the limitations of the inconvenient mobile device to input voice. Instead, most users prefer using short, annotated texts accompanied with the photos they take.

By the thinking aloud method and the subjects' feedback in the interface part, most subjects thought that graphical button interface design needs to be more intuitive. For instance, the text on the screen could cause misunderstanding or confusion, and the screen click feedback for user is very important, because if there is no feedback mechanism, the user will not know whether the click is successful or not. We amended the interface design by the suggestions, then implemented on the android mobile phones, and did the field trial to evaluate the performance of the system.

In the part of the overall system, users respond that too long login/upload time would miss recording important moments, and also could make users impatient not to continue operating the system. Even so, the users all gave positive appreciation to our functional design. Users agreed that our mobile collaborative digital narrative platform have reached the goal of this study. For example, the system can use different media to narrative, and narrative content can be sorted/queried. Users assents that relation of location and contents is attractive. The stories generated can incur their curiosity, and shorten time to acquaint themselves with the location. The results are consistent to the motivation and goal of the work.

V. CONCLUTION AND FUTURE ENHANCEMENT

The design of a location-based mobile collaborative narrative platform is proposed in this paper. The field experimental results show that the system achieved positive satisfaction from the users, and did fit the narrative trend in the new era.

This research can be viewed as a pilot study in the field. It is first focused on narrative behaviors. According to the interviews, most users would like their content to be shared with friends. This implies users generally prefer the social functions which can be extended to the system in the future. On the other hands, for the efficient transmission of the content in the databases, we will suggest to extend the abilities of transmission to be P2P mode or DTN (Delay Tolerant Network) in the next version. With this pilot study, we believe that the development of mobile collaborative platforms will achieve more successes.

REFERENCES

[1] Institute for Information Industry, <3rd Quarter 2010 Mobile Customers in Taiwan>, Date: November 24, 2010, Source:http://www.find.org.tw/find/home.aspx?page=many&id=275.

[2] InsightXplorer, <ARO Observation-Internet Use Observed in 2010>, Date: 2010, Source:http://www.iama.org.tw/upload/ResourceReport/20110209041212213.pdf.

[3] Liu Weigong, <Service is "Content Industry": New Trends in Service Innovation>, Date: December 2010, Source:http://itc.tier.org.tw/2010/ppt/2-3_Dr_WeiGongLiou.pdf.

[4] Chueh-Pin Chao, <Digital Learning and Experience, the Efforts of European Heritage ~ (b)>, Digital Archives and learning newsletter, VIII XII. Date: December 15, 2009, Source:http://newsletter.teldap.tw/news/InsightReportContent.php?nid=3296&lid=325.

[5] Wikipedia, 〈 Web2.0 〉 , Date: March, 2011, Source:http://zh.wikipedia.org/zh-tw/Web_2.0.

[6] Wikipedia, 〈 OpenStreetMap 〉 , Date: March, 2011, http://wiki.openstreetmap.org/wiki/Traditional_Chinese:Main_Page.

[7] Yi-Fu Tuan writes, Pan Guicheng translation. "Perspective in the Experience of Space and Place", Taipei: National Institute for Compilation, 1997.

[8] Ling-yi Huang, "Space of the Digital Age - The Time of Narrative Structure: The Facade Site as An Example", Taiwan Society of Information Society Annual Conference, pages 135-160, 2008.

[9] Jhao-Ling Chen, "The Never Ending Story: A Collaborative Narrative Storylog", Ming Chuan University, Information and Communication Engineering, Master Thesis, 2006.

[10] Chen Jianxiong translation, "Interaction Design", Taipei, Chuan-Hua Books, 2009.

[11] Xing-sheng Kuan et al., "Design Methods", Taipei, Chuan-Hua Books, 2008.

[12] Insight, National Taiwan University Research Center of Innovation and Synergy for Intelligent Home and Living Technology, <Interactive Design>, Date: December, 2010, Source: http://insight.ntu.edu.tw/zh-tw/node/871.

[13] Insight, National Taiwan University Research Center of Innovation and Synergy for Intelligent Home and Living Technology, <Human-Computer Interface Evaluation>, Date: December, 2010, Source: http://insight.ntu.edu.tw/zh-tw/node/693.

[14] S. Chatman, Story and Discourse: Narrative Structure in Fiction and Film. Ithaca,N.Y.: Cornell University Press, 1978.

[15] Herman, Story Logic: Problems and Possibilities of Narrative. Lincoln & London: University of Nebraska Press, 2002.

[16] ITU, The World in 2010, Source:http://www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf, 2010.

[17] Henry Jenkins, "Game Design as Narrative Architecture", In First Person: New Media as Story, Performance, and Game. Edited by Pat Harrington and Noah Wardrip-Fruin. Cambridge: MIT Press, 2003.

[18] R. Kristof and A. Satran, Interactivity by Design. Mountain View: Adobe Press, 1995.

[19] E. G. Mishler, Research Interviewing: Context and Narrative. MA, Cambridge, Harvard University Press, 1986.

[20] J. Nielsen, Usability Engineering, San Francisco: Academic, 1993.

[21] Nicola J Bidwell, Thomas Reitmaier, Gary Marsden, and Susan Hansen, “Designing with Mobile Digital Storytelling in Rural Africa”, CHI '10: Proceedings of the 28th International Conference on Human factors in computing systems, pages 1593-1602, 2010.

[22] Yiwei Cao, Ralf Klamma, and Andrea Martini, “Collaborative Storytelling in the Web 2.0”, STEG'08: Proceedings of the First International Workshop on Story-Telling and Educational Games at EC-TEL08, Sep. 16, 2008.

[23] SingalPoint, <White Paper on Conversational Advertising>, Date: June 2010, Source: www.dm2pro.com/downloads/20100609_3/download.

[24] <Disney's Kim Possible World Showcase Adventure>, Date: 2011, Source: http://disneyworld.disney.go.com/parks/epcot/attractions/kim-possible/.

[25] App Store, <picplz>, Date: June, 2011, Source: http://itunes.apple.com/us/app/picplz/id382339383?mt=8.

[26] App Store, <Broadcastr>, Date: May, 2011, Source: http://itunes.apple.com/us/app/broadcastr/id423169367?mt=8.

[27] App Store, <Instgram>, Date: May, 2011, Source: http://itunes.apple.com/us/app/instagram/id389801252?mt=8.

Content Management in the Context of Collaboration

Marius Ioan Podean, Ştefan Ioan Niţchi, Dan Benţa
Business Information Systems Department
Babes-Bolyai University, Faculty of Economics and Business Administration
Cluj-Napoca, Romania
Emails: {marius.podean, stefan.nitchi, dan.benta}@econ.ubbcluj.ro

***Abstract*—Collaboration is a very dynamic process that combines functionality that supports communication, management and involves content handling. During the execution of a project team members are not always collaborating and their work alternates with cooperation, when a greater emphasis is placed on a value-chain model of producing results. Focus permanently switches from the flexible content approach to the management tools according to task's specific. Content management is often regarded in term of web or enterprise content, or as document management. Documents are mere representations of content following a particular format and this research is focused on describing a method of providing flexible interaction with content objects in the context of collaboration. Collaboration requires a flexible content management solution in order to support the interaction between group members.**

***Keywords*-collaboration, flexibility, content adaptability, creative interaction, XML**

I. Introduction

Collaboration builds on communication, coordination, cooperation (the 3C model [1]) but requires some extra components also. Coordination represents the management of people and their activities and is based on altering activities for mutual benefit [2]. Cooperation goes one step further and adds resource sharing in order to achieve a shared vision [3]. Complex problems require aspects of knowledge that reside in the minds of individual stakeholders as tacit knowledge [4]. Reaching a resolution based on consensus building produces a higher-quality decision than other decision-making processes [5]. Collaboration demands a flexible content management solution that can support the managerial efforts in order to stimulate creative insight. Such a content management solution should allow groups of minds to interact with each other and allow content to be used as an externalization of their thinking [4]. Making interesting connections between content elements more evident and supporting the growth of an idea are just a few of the key area that such a system could enable. Starting from these challenges we will present a model for a content management component that handles content in a flexible manner in order to allow users to focus their action on the task at hand an not on handling technology. Our approach is based on exploiting the flexibility of XML and related tools in regard to content management. Constraints imposed by collaboration are modeled following XML principles. For this purpose, a model that describes our findings is presented together with some implementation details.

Usually content management is considered as an individual tools and not as an integrated component of collaboration. This approach shifts the focus from "intelligent content"[6] to web content or document management, leading thus to tools that are used in conjunction with a groupware solution that do not support the needs for collaboration. Our focus is on developing a content management module that could be integrated with other tools that are required by collaboration, in such a manner that is in accordance with the prerequisites of collaboration.

This paper is organized as follows. In the second section we will discuss the main characteristics that we consider that are relevant for a content management component in a collaborative system and present a model that covers this aspects. Following this discussion, in section three we will present in greater detail the main functionality of our model. The final section will present some concluding remarks and future work.

II. Content management model

A flexible content management solution that must satisfy the requirements of collaboration must take in consideration aspects like handling the entire life cycle of content [7], provide creative means of interaction with the content so that content objects can be used as externalizations of human thinking [4] and serve as a basis for negotiation and critique. The use of structured content [8] standards and open formats together with metadata [9] will allow for more flexibility to be implemented in the system enabling thus a wider range of actions that can be performed on content . The separation of content from presentation [10] and implementing a great adaptability of the content together with a single source / multiple publishing channel and formats focus enable the system to customize content to user needs. Including project and content workflow management and providing a good accessibility and adaptability will enable user to tailor the life cycle of the content according to organizational requirements.

Our model focuses on collaboration, all elements being tailored around major concerns for this process. Most systems focus on web content, document management, pub-

 ISBN: 978-1-61208-157-1

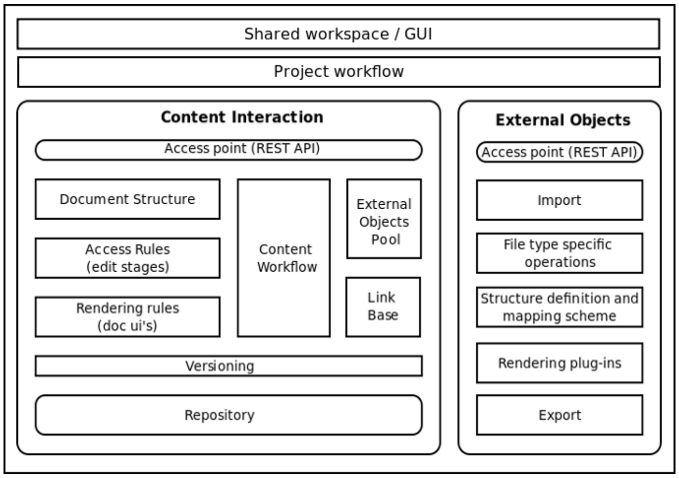


Figure 1. Collaboration enabler content management model

lishing firms and none on the needs related to collaboration. Based on the gaps identified we have defined a model for content management that focuses on four key areas: content interaction, external object management, project workflow and credential management. The principles behind the stated model will allow it to provide a great support for content management in a collaborative system.

III. Main components

The novelty of our research is that all the following components have implementations using mainly XML tools (e.g., content level workflow is handled in XProc, a pipeline-based XML processing tool). This approach enables us to provide a rich set of content interaction mechanisms required by collaboration. In the following we will stress on the design aspects of our content management model.

A. Content interaction

The main purpose of this component is to handle all interaction with the content, from it's creation to publishing. It includes the repository also and the versioning mechanism. The content interaction is separated in defining content rules (definition schemes, edit and rendering rules) and metadata, external object and link handling.

In our approach the term "document" represents a very flexible concept being regarded as a temporary view for the content. Most often refers to web representations but it can include common formats like PDF. A document is constructed from three main components: a definition, a set of editing rules and a rendering definition. A content definition consists in a vocabulary - XML Schema (schema), that defines the structure of the documents, the elements that are allowed in this structure, element data types and restriction regarding occurrences.

The access rules define sets of editing constraints that apply to roles at different editing stages. A separation in edit stages is made because during content's life cycle different sets of actions can be taken that have implications on the quality of the content. The rendering rules define visualizations for the content. These visualizations range from web user interfaces to printable documents. For each publishing format (web, print) the user must define a rendering definition. This definition must describe the relevant elements for display in the targeted media. For example an XSL-FO definition can be used for both formats but can be split and specialized for each of them. These definitions specify how content is going to look in a certain format covering aspects like page formating, styles, specific positioning in the page for elements etc. The way an element is displayed following the rendering rules is in strict relationship with the access rules. The access rules are enforced by the use of user interface elements: if an element is editable it will be displayed as an input field that allows data to be filled in, otherwise as a text element.

A request to view a document implies assembling all aforementioned elements in order to provide a representation of the content. When a representation is requested its definitions are loaded and the structure will be filtered by eliminating elements that are not accessible for the user. For this, the access rules are applied to the document definition marking as hidden elements that the user is not allowed to see. Further, the access rules are applied to the rendering definition marking elements that are editable or read only. This will result in a user interface definition that takes in consideration access rules and a schema that can be used to filter content that should not be accessible to the user.

After a document view is generated in a web format, if the user's credentials or edit stage allow it, the user can edit the content. When a document is edited individual changes are stored separately on the client (as Δ consisting in each edit step applied to content) and only these changes are sent to the server and patched on the version that they refer to (in case no other update has been made on the content). There are two ways to work on a content representation:

1) asynchronous: when the user's Internet connection does not represent a problem, each change is sent individually to the server. In the same respect, a pull operation is made at a predefined interval in order to retrieve changes made by others.
2) synchronous: multiple edit operations are stored on the client and sent to server only when the user decides to save them. This approach may lead to edit conflicts if other users have edited the same content objects.

The metadata is an important component of each content object (Figure 2). The **Repository** acts as an file system that stores all content in XML and manages content, data about content and how are they organized. Content is organized in namespaces and projects. A namespace is the higher level (e.g. it can refer to a department) that can have multiple

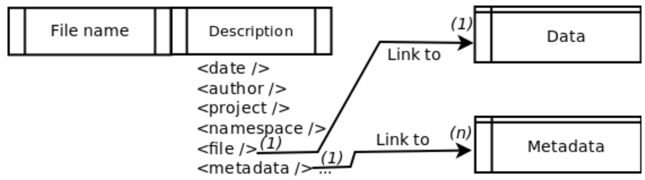


Figure 2. The structure of a content object

projects. Projects have content objects and roles associated with them. A user has access only to those projects (and inside the projects, to the files) that it has the required authorization to, according to the active role.

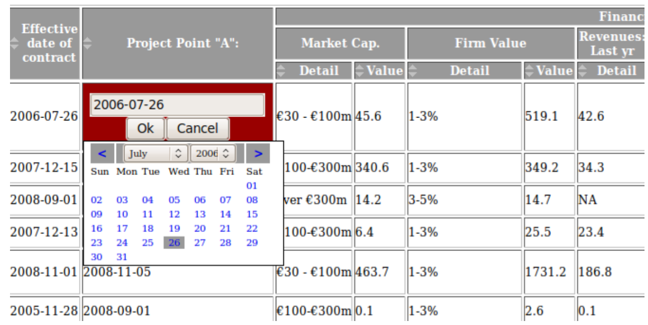
B. External objects

External objects are the gateway to introduce non standardized content in the system. This module handles *i)* the import of content from other formats, *ii)* the definition and execution of file type specific operations, *iii)* conversion to other vocabularies (where possible), *iv)* hosting and execution of special rendering plug-ins, and *v)* content export back to original format or other (closed-source) formats.

When an import request is received by the **External objects** module, based on the information regarding file type, a format plug-in is called and the import process is triggered. In the first phase the content is converted to XML following a structure defined by the format plug-in. If the conversion is successful the new content object is registered in the Object Pool from the **Content Interaction** module as raw content. The raw content will be stored in the **Repository** following the structure depicted in Figure 2.

If the format plug-in defines special operations that can be applied to the content, they will be presented to the user and then the content will be edited following the steps defined by the selected operation. For example, for a large Microsoft Excel file one useful specific operation could be a dataset cleanup that will search for similarities in string columns in order to identify typos. This operation will help increase the quality of data with minimal effort. After following the steps included in the file type specific operation, the content object is updated in the Object Pool. The next step consists in showing the user the available formats that the content can be converted to, and if the user requests the content to be converted to some format, following a mapping schema the content is transformed. The last step consists in associating the content with the active project and metadata that describes its characteristics. Applying file type specific operations and converting to other vocabularies are optional steps and can be called any time during the life cycle of the content.

All external objects plug-ins must be defined as web-services and provide an API. The module require users to provide the API and URL of these web-services in order to integrate them in the system. The plug-ins are not



Effective date of contract	Project Point "A":	Financ				
		Market Cap.		Firm Value		Revenues: Last yr
		Detail	Value	Detail	Value	Detail
2006-07-26	2006-07-26	€30 - €100m	45.6	1-3%	519.1	42.6
2007-12-15		100-€300m	340.6	1-3%	349.2	34.3
2008-09-01		ver €300m	14.2	3-5%	14.7	NA
2007-12-13		100-€300m	6.4	1-3%	25.5	23.4
2008-11-01	2008-11-05	€30 - €100m	463.7	1-3%	1731.2	186.8
2005-11-28	2008-09-01	€100-€300m	0.1	1-3%	2.6	0.1

Figure 3. External object representation: tabular data editor

necessary stored on the same server as the framework, but the application should provide a sandbox so that users can install and test their own plug-ins.

Since the content is converted to a user specified format, the possibility to transform it to a standard vocabulary should be implemented by allowing users to include transformation schemes. All schemes defined for a specific file type are managed by mapping mechanism that is responsible also with applying these transformations when requested.

The content that can be included using this mechanism is very divers, thus a mechanism to provide the proper means to display the content and interact with it is an essential requirement. Rendering mechanisms are included similar to file type specific operations, as web services. The document representation that includes an external object is similar to web mash-ups since it aggregates content and content representations from multiple sources. In the initial document only a description of the external object is present and according to the required delivery channel a proper representation is returned. If the document is delivered as a web document, then the appropriate web-service is called in order to provide a user interface that will allow users to interact with the content. For example, an Excel file that has been imported can have a user interface similar to the original editor that will allow some basic operations to be executed on the content (Figure 3). In the same manner, a definition of a genome can be provided in a 2D or 3D representation in a web document.

If the delivery channel is print media or similar, a different representation of the external object or a snapshot can be included. Delivering document representations as mash-ups enables the framework to handle a large number of content formats and providing thus a great flexibility in customizing it to user needs. External objects can be further exported back to their original format or other appropriate formats through the use of export plug-ins.

C. Project workflow

The workflows involved in the model are separated on two levels: worflows that concern content object processes and workflows that integrate project processes and content

 ISBN: 978-1-61208-157-1

objects in order to support collaborative endeavor. This module must include an workflow engine and a process map generator. The process editor is an optional component and can even be implemented separately.

After defining the project management plan, in a semi-automatic way a workflow schema must be generated by one of the subcomponents of the workflow engine. From the project management plan data regarding the tasks, their sequencing in time, dependence between them, time constraints can be extracted an imported in the workflow schema. This process will result is a partial workflow schema that must be customized by adding connections to the documents involved by each step, defining variables for decisional steps, select content representations for each edit stage of a document and attach access rules. Coming back to our article example, in the idea gathering step users are allowed to create arguments using the argument map rendering. After a certain amount of time the task will end and the next process will involve using and document like article editor that will allow users to edit the content of an argument but not allow them to remove arguments. Another use case might be a large project that includes reporting stages that require users to fill-in certain document in order to track work progress. When a report is created the owners can edit it but soon after the report has been approved they must no longer have access to this functionality.

After a workflow schema has been defined, a process map can be generated in order to provide users the information they need regarding the execution of the project. A process map is a simplified form of a workflow schema that does not contain all elements required to be executed by a worflow engine but incorporated enough details to keep users informed about the steps required in order to reach the goals and more important, the positioning in time of the project's execution and deadlines for content objects. This is a simplified form of project and risk management plan and a workflow schema.

As mentioned earlier, the workflow editor is an optional component, but the module should include a minimal customization tool that will allow users to edit the semi-automatic generated workflow schema in order to associate content objects and access rules. A workflow editor would allow user to fully customize a schema (or create one from the beginning) by adding or removing steps. Since the workflow schema must be derived from the project management plan such functionality is not compulsory.

IV. Conclusion

During the execution of a project team members are not always collaborating and their work alternates with cooperation, when a greater emphasis is placed on a value-chain model of producing results. Focus permanently switches from the flexible content approach to the management tools according to task's specific. In order to fully support the process of collaboration, the aspects that precede it or come in-between the collaboration sessions must be fully supported so that they will not represent a problem that can hinder collaboration. Both collaboration and cooperation require content support and management tools, what they differ in is their main focus.

A flexible content management solution that must satisfy the requirements of collaboration must take in consideration aspects like *i)* handling the entire life cycle of content, *ii)* provide creative means of interaction with the content so that content objects can be used as externalizations of human thinking and serve as a basis for negotiation and critique. The proposed framework takes in consideration all the aforementioned aspects and implements a very flexible approach that is targeting customization according to domain specific needs. We have designed our model starting from the idea that content needs are very divers and so rapidly changing that a content management module should not try to offer a holistic approach that will accommodate all needs, but instead provide a general framework that will allow users to define and customize both the processes involved in their projects and the content objects that are used to harness team's efforts. The actual model is the result of continuous refactoring starting from prototypes and since the implementation of the model is in an early stage we intend to finalize it in order to test the validity of the model by using it in the current projects running in our university.

Acknowledgment

This work was supported by Romanian National Authority for Scientific Research under the grant no. PN2 92-100/2008 SICOMAP.

References

[1] L. M. Camarinha-Matos and H. Afsarmanesh, *Collaborative Networks: Reference Modeling*. Springer US, 2008, ch. Collaboration forms.

[2] H. Fuks, A. Raposo, M. A. Gerosa, M. Pimental, and C. J. P. Lucena, *Encyclopedia of E-collaboration*. Hershey, PA: Information Science Reference - Imprint of: IGI Publishing, 2008, ch. The 3C Collaboration Model, pp. 637–644.

[3] T. Wolff, "Collaborative Solutions - True Collaboration as the Most Productive Form of Exchange," Collaborative Solutions Newsletter, 2005, Last Accessed: October 2010. [Online]. Available: http://www.tomwolff.com/collaborative-solutions-newsletter-summer-05.htm

[4] E. Arias, H. Eden, G. Fischer, A. Gorman, and E. Scharff, "Transcending the individual human mind - creating shared understanding through collaborative design," *ACM Trans. Comput.-Hum. Interact.*, vol. 7, no. 1, pp. 84–113, March 2000.

[5] D. Straus, *How to Make Collaboration Work: Powerful Ways to Build Consensus, Solve Problems, and Make Decisions*, ser. 107-8. San Francisco: Berrett-Koehler Publishers, 2002.

 ISBN: 978-1-61208-157-1

[6] A. Rockley, "What is Intelligent Content?" http://thecontentwrangler.com/2011/01/17/what-is-intelligent-content/, 2011, Last Accessed: February 2011. [Online]. Available: http://thecontentwrangler.com/2011/01/17/what-is-intelligent-content/

[7] U. Bartlang, *Architecture and Methods for Flexible Content Management in Peer-to-Peer Systems*, 1st ed., A. W. Ute Wrasmann, Ed. Wiesbaden, Germany: Vieweg+Teubner, 2010.

[8] A. Salminen, "Building Digital Government by XML," in *HICSS*, ser. 38th Hawaii International Conference on System Sciences (HICSS-38 2005), CD-ROM / Abstracts Proceedings, 3-6 January 2005, Big Island, HI, USA. IEEE Computer Society, 2005.

[9] T. R. Group, "The Role of Content Standards in Content Management," The Rockley Group Inc., Markham, White Paper, 2005.

[10] D. Clark, "Content Management and the Separation of Presentation and Content," *Technical Communication Quarterly*, vol. 17, no. 1, pp. 35–60, 2008.

 ISBN: 978-1-61208-157-1

Content Modeling Based on Concepts in Contexts

Hans-Werner Sehring
Content and Collaboration Solutions
T-Systems Multimedia Solutions GmbH
Dresden, Germany
Email: hans-werner.sehring@t-systems.com

***Abstract*—Content management is a generic term for different tasks dealing with content modeling, the creation and management of content instances, and the delivery of content as part of applications or documents. We studied content models for real-world entities in the area of Concept-oriented Content Management. In addition to content modeling, there is agreement that content heavily depends on context in most cases. Therefore, content management has to devise ways to consider context in its models and processes. In this paper we apply a novel generic content modeling language to address the modeling demands arising from content management that is augmented with context information.**

***Keywords*-content management; content modeling; context; contextualization; personalization.**

I. Introduction

Content management is a generic term for different tasks dealing with content modeling, the creation and management of content instances, and the delivery of content as part of applications or documents.

The definition of content is manifold. In particular, there are two basic notions of content: (1) (purely) digital content that is an entity of its own, and (2) content that is used to describe real-world entities. The latter employment of content is used for entities that cannot adequately be represented by structured data alone. One class of such content are product descriptions as found in product information management (PIM) applications. Catalogs seek to visualize products in an attractive way. Another class are complex entities, in particular ones from non-technical domains that do not rely on formal representations. In those domains content often represents (states of) a process, e.g., the inception, creation, and use of a work of art in art history.

We study content models for such real-world entities in the area of *Concept-oriented Content Management* (*CCM*) [6]. This paper presents a generic content modeling language to address the modeling demands from that area.

There seems to be agreement that content depends on *context* in most cases. This insight recently starts to have an increasing impact on content industry. Therefore, content management has to devise ways to consider context in its models, functions, and processes.

In this paper, we present a modeling language that allows combining different modeling approaches, including content description, classification, and contextualization.

The paper is organized as follows: In Section II, we formulate requirements for content models that describe entities in contexts. Section III defines a modeling language applicable for this kind of content representation. How typical modeling tasks are solved by that language is discussed in Section IV. The paper closes with conclusions in Section V.

II. Content Modeling Requirements

Content management requirements as laid out in the introduction demand for adequate content structures. To this end, content modeling is of central importance for content management applications. Typical content modeling approaches are discussed in this section.

A. Content Description

A range of modeling concepts and languages for the description of content has emerged. Figure 1 gives an overview over the most commonly used modeling techniques.

Digital content itself, from a technical point of view, simply consists of binary data representing a text, an image, a sound, etc. By adding descriptive information such data is enriched in order to be perceivable as content.

The basic level of content management is established by *meta data* and *descriptive information* that further describe the data or the entity represented by it, respectively. *Classification*, e.g. by tagging, can be seen as a specific kind of descriptive information that enables additional functionality like filtering and clustering. Classification is particularly useful if classifiers are related to each other, e.g. by narrower term and broader term relationships. Such relationships are provided by *taxonomies* and *ontologies*, e.g. defined explicitly or by description logic expressions, or freely assigned *tags* from which *folksonomies* are derived.

In CCM we also investigate *extensional descriptions* of content by providing a sample set for terms as epistemic structures. As an example three images describing the concept “strength” are shown in Figure 1 on the left-hand side. A similar approach is taken by automatic classification systems that take instances as a training set.

Many content management approaches rely on schema definitions that are based on the object-oriented paradigm. Though object-orientation offers many beneficial features

 ISBN: 978-1-61208-157-1

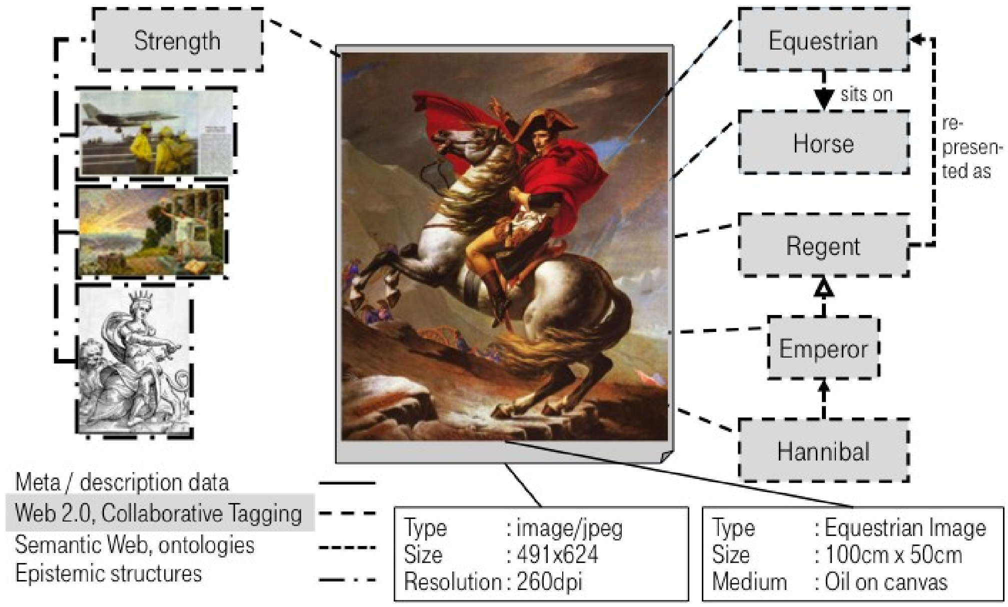


Figure 1. Content modeling approaches.

like inheritance and polymorphism, it is in several respects not well suited for those tasks.

In particular, the rigid distinction between classes and instances often is an obstacle for content modeling. Classes are supposed to provide a "blueprint" for instances. Instances are not supposed to deviate from the given form. This hinders the modeling of, e.g., variants, roles, and materializations that both represent views on one specific instance.

Furthermore, object-orientation is based on one single class hierarchy. In content management one typically has to deal with classes of different kinds. One main distinction is the one between structural definitions ("a book has a title and an author") and domain-specific classification information ("there are books on computer science and on mathematics"). Typically, the different kinds of classification exhibit different characteristics [4].

This paper presents a novel language with properties that make it suitable for content modeling. It borrows from the object-oriented paradigm, but it abstracts from certain concepts. This way it provides a basis for an integration of the above-mentioned modeling techniques.

B. Content Contextualization

In addition to content descriptions as provided by the approaches sketched in the previous section, the notion of context is of importance for content. There is agreement in the content management community that context is central to the task outlined in the introduction ("if content is king, context is the kingdom" [3]). Various studies underline the importance of context for content management [1].

The use of content is heavily influenced by the context of the user and the circumstances under which content was created. The context of content has to be considered with respect to modeling, interpretation, and delivery of content.

By introducing context into content models the definition of content variants as well as of the history of content creation and use is supported.

Context enables an extended content interpretation, with practical applications for content analysis and retrieval as well as the computation of recommendations. The value of content for a specific use can be judged if the current work context of the user is known, as well as the history of a particular piece of content by means of data provenance.

Modern content-based applications include a context-dependent delivery. Knowledge about the context of the user allows a personalized content presentation. For delivery the context often also refers to the publication channel and the device on which documents are displayed.

III. Minimalistic Meta Modeling Language

The *Minimalistic Meta Modeling Language* (*M3L*, pronounced "mel") is a modeling language that is currently under development and that has not been reported about yet. Though its further development is not particularly directed at content modeling, the rationale behind the language is based on the modeling tasks discussed in this paper.

M3L offers a rather minimalistic syntax that is completely covered by the following grammar (in BNF):

$$\begin{aligned} \textit{model} &::= \langle\textit{prop-list}\rangle \\ \textit{def} &::= \langle\textit{id}\rangle\,\text{“}\textit{is}\text{”}\,\langle\textit{id-list}\rangle \\ &\quad\;\; [\text{“\{”}\,\langle\textit{prop-list}\rangle\,\text{“\}”}\,[\langle\textit{production-rule}\rangle] \\ &\quad\;\; |\langle\textit{production-rule}\rangle\,|\,\text{“;”}] \end{aligned}$$

$$
\begin{array}{rcl}
\textit{ref} & ::= & \langle \textit{id} \rangle [\text{“}\textit{from}\text{”} \langle \textit{id} \rangle] \\
\textit{id-list} & ::= & (\text{“}\textit{a}\text{”} | \text{“}\textit{an}\text{”} | \text{“}\textit{the}\text{”}) \langle \textit{ref} \rangle [\text{“,”} \langle \textit{id-list} \rangle] \\
\textit{prop-list} & ::= & \langle \textit{def} \rangle [\langle \textit{prop-list} \rangle] \\
\textit{production-rule} & ::= & \text{“}|-\text{”} \langle \textit{ref} \rangle \text{“;”} \\
\textit{id} & ::= & \ldots (\textit{reqular expression of identifiers})
\end{array}
$$

The production for identifiers has been omitted. It is a typical lexer rule that defines identifiers as character sequences. Identifiers may—in contrast to typical formal languages—be composed of any character sequence. Quotation is used to define identifiers containing whitespace.

The semantics of M3L statements will be discussed in the subsequent sections.

The descriptive power of M3L lies in the fact that the formal semantics is rather abstract. There is no fixed domain semantics connected to M3L definitions.

A. Concept Definitions and References

A M3L definition consists of a series of *definitions* (⟨*def*⟩ in the grammar definition above). Each definition starts with a previously unused identifier that is introduced by the definition and may end with a semicolon, e.g.: `NewConcept;`

We call the entity referenced by such an identifier a *concept*.

The keyword `is` introduces the optional reference to a base concept. An inheritance relationship as known from object-oriented modeling is established between the base concept and the newly defined derived concept. This relationship leads to the visibility of the concepts defined in the context (see below) of the base concept to be visible in the derived concept. Furthermore, the refined concept can be used wherever the base concept is expected (similar to subtype polymorphism).

As can be seen in the grammar, the keyword `is` always has to be followed by either `a`, `an`, or `the`. The keywords `a` and `an` are synonyms for indicating that a classification allows multiple sub concepts of the base concept:

```
NewConcept is an ExistingConcept;
NewerConcept is an ExistingConcept;
```

There may be more than one base concept. Base concepts can be enumerated in a comma-separated list:

```
NewConcept is an ExistingConcept,
              an AnotherExistingConcept;
```

The keyword `the` indicates a closed refinement: there may be only one refinement of the base concept (the currently defined one), e.g.:

```
TheOnlySubConcept is the SingletonConcept;
```

Any further refinement of the used base concept(s) leads to an error.

Apart from the definition of new concepts the above expressions can be used to augment a concept definition if the leading identifier already has been introduced. E.g., the following expressions lead to the same definition of the concept `NewConcept` as the above variant:

```
NewConcept;
NewConcept is an ExistingConcept;
NewerConcept is an AnotherExistingConcept;
```

B. Content and Context Definitions

Concept definitions as introduced in the preceding section are valid in a *context*. Definitions like the ones seen so far add concepts the topmost of a tree of contexts. Curly brackets open a new context, e.g.:

```
Person { name is a String }
Peter is a Person {"Peter Smith" is the name}
Employee { salary is a Number }
Programmer is an Employee;
PeterTheEmployee is a Peter, a Programmer {
  30000 is the salary }
PeterTheMusician is a Peter, a Musician {
  Oboe is a playedInstrument }
```

In this example, we assume that concepts `String` and `Number` are already defined. In practice, the concept `30000` should also be given. If not, it will be introduced locally in the context of `PeterTheEmployee`, preventing reuse of the identical number.

M3L has visibility rules that correlate to contexts. Each context defines a scope in which definition identifiers are valid. Concepts from outer contexts are visible in inner scopes. E.g., in the above example the concept `String` is visible in `Person` because it is defined in the topmost scope. `salary` is visible in `PeterTheEmployee` because it is defined in `Employee` and the context is inherited. `salary` is not valid in the topmost context and in `PeterTheMusician`. Contexts with those names may be defined later on, though.

Tying a context to a concept can be interpreted in different ways. This is elaborated in Section IV.

Contexts can be referenced using the projection operator `from` in order to use concepts across contexts: `salary from Employee`.

C. Narrowing and Production Rules

M3L allows assigning one *production rule* to each concept. Production rules fire when an instance comes into existence that matches the definition of the left-hand side of the rule. They replace the new concept by the concept referenced by the right-hand part of the rule.

The following shows an example:

```
Person {
  female is the sex; married is the status
} |- Wife;
```

Whenever a `female Person` who is `married` shall be created then a `Wife` is created instead.

Production rules are usually used in conjunction with M3L's *narrowing* of concepts. Before a production rule is applied, a concept is narrowed down as much as possible. Narrowing is a kind of matchmaking process to apply the most specific definition possible.

If a base concept fulfills all definitions—base concepts and constituents of the context—of a derived concept, then the base concept is taken as an equivalent of that derived concept. If a production rule is defined for the derived concept, this rule is used in place of all production rules defined for any super concept.

The following code shows an example of combined narrowing and production rules:

```
Person { sex; status }
MarriedFemalePerson is a Person {
  female is the sex; married is the status
} |- Wife;
MarriedMalePerson is a Person {
  male is the sex; married is the status
} |- Husband;
```

There is a concept `Person`. Whenever an “instance” (a derived concept) of `Person` is created, it is checked whether it actually matches one of the more specific definitions. A married female `Person` is replaced by `Wife`, a married male `Person` by `Husband`, and every other `Person` is kept as it is:

```
Person {                           Person {
  male is the sex }          →       male is
                                       the sex }
Person {
  female is the sex;         →     Wife
  married is the status }
Person {
  male is the sex;           →     Husband
  married is the status }
```

IV. Content Modeling with the M3L

In this paper the M3L is used for concept-oriented content management as motivated in Section II. This section discusses content modeling aspects and their formulation using the M3L.

A. Content Modeling

The goal of the application of M3L for content management is to define content in a context, where both content and context consist of an identifier and nested sub contents or contexts. This leads to the following interpretations of a concept definition `A{B;}`: it may represent ...

- content `A` with partial content `B`,
- content `B` in a context `A`, or
- a context `A` with a subcontext `B`.

Object-orientation has difficulties modeling class and instance variants (Section II-A). Figure 2 illustrates the distinct layers found in object-oriented systems. Objects are instances of classes. Classes can in turn be viewed as instances of metaclasses. The layer of metaclasses is closed; metaclasses can be modeled as instances of metaclasses. In model management systems there may additionally be a meta meta layer. Furthermore, there are relationships between instances (association, aggregation) and between classes (specialization, generalization).

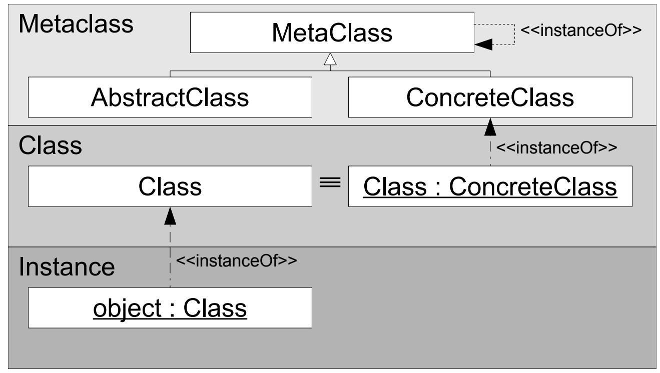


Figure 2. Levels of object-oriented models.

M3L abstracts from these distinct layers and allows interpreting concepts as classes, instances, or variants of instances. An expression `A is a B` can be interpreted in the following ways:

- `A` is a type, `B` is its super type,
- `A` is an instance of type `B`, or
- `A` is a variant, role, or similar of instance `B`.

Expressions with one of the above interpretations are used for various purposes:

- to structure content, e.g.:
  ```
  Picture is a Content {
    imgFile is a ByteArray ;
    title is a String }
  ```
  Here, `Picture` is a content class that defines two attributes `imgFile` and `title`, each of them with “type” information.
- to present content in a context, e.g.:
  ```
  PoliticalIconography is a Context {
    bonaparteCrossesTheAlps is a Picture }
  ```
  A concrete `Picture` is put in a context that helps, e.g., interpreting it.
- to define a context hierarchy, e.g.:
  ```
  PoliticalIconography is a Context {
    StrengthSymbols is a Context }
  ```
  The area of Political Iconography is defined as a context with the domain of symbols of strength as one of the specific contexts it includes.

These basic modeling means are used for various aspects of content models, some of which are discussed in the subsequent sections.

B. Domain Reuse

Contexts are often defined by a combination of other contexts. This way, existing models may (eventually partially) be reused. Reuse is achieved by putting concepts into new contexts. By the additional contextualization content can receive a new meaning or contribute to an additional use case. As part of the combination of contexts, reused concepts may be adapted for new contexts.

For the following example assume classes `Arts` and `History` to be defined that represent the respective research disciplines.

```
ArtHistory {
  Artist is the Artist from Arts;
  Painter is the Painter from Arts;
  Sculptor is the Sculptor from Arts;
  Epoch is the Epoch from History;
  Artwork {
    title is a String;
    artist is an Artist;
    epoch is an Epoch }
  Painting is an Artwork {
    artist is a Painter }
  Statue is an Artwork {
    artist is a Sculptor } }
```

The example shows the definition of a new context for art history that is composed of concepts from the domains of arts and history. The idiom *C* `is the` *C* `from` *M* makes concepts available in the newly defined context: the keyword `from` addresses a source context *M* from which to import a concept *C*. A new concept is defined in the current context as the only refinement of the original concept, but does not add any structure. This way it effectively provides a copy of the original concept from the source context.

C. Variants

Concept definitions in M3L provide a direct means to define variants of a concept in different contexts. Variants of a concept in one context are defined by means of refinement.

The following code shows a quite simple example:

```
Peter is a Person;
PeterTheEmployee is a Peter, a Programmer;
PeterTheHobbyMusician is a Peter, a Musician {
  Oboe is the playedInstrument }
```

We model a `Person` named `Peter`. Specific information on Peter is given in a context-specific way. E.g., the fact that Peter works as a programmer is stated by a concept describing Peter as an employee. A different aspect of Peter's life are his hobbies: in the example he is a `Musician` who—given that `Musician` defines a concept `playedInstrument`—is stated to play the oboe.

D. Revisions

In some applications concepts exist in different *revisions*. Revisions and their relationship have different meanings. Typical content management systems record revisions to reflect the process of content creation. This is often required for legal reasons that demand content states to be reproducible.

When modeling real-world entities through content there is an additional need for revisions. Real-world entities develop over time, so that representations of content may in fact have to cover the process of invention, creation, use, etc. This is particularly true for entities considered in history and art history. In order to reflect a process that typically manifests itself in states of the entity under consideration there have to be concepts for those states. The concepts are refinements of one common concept.

The following example shows two states of the famous painting "Napoleon Crossing the Alps":

```
NapoleonCrossingTheAlps is a Painting;
HistoricalContext {
 NapoleonCrossingTheAlpsSymbol
  is a NapoleonCrossingTheAlps, a Strength; }
Museum {
 NapoleonCrossingTheAlpsArtwork
  is a NapoleonCrossingTheAlps, a Classic; }
```

There is a state where the `Painting` is/was used as a political instrument to visualize strength. Another state represents the painting as a piece of art that is used because of its famousness (independent of the original intention).

E. Personalization

Content management users are provided with predefined concepts that they typically want to tailor to their needs by means of personalization. We investigated the use of personalization for research, teaching, and software engineering.

For example, a model definition like that for Political Iconography above is typically provided as a standardized model. Though standardization is of importance for the cooperation within a domain it is often too restrictive for individual work. Therefore, researchers want to personalize given models both in structure and in content [6].

The following shows an example for the Political Iconography model as shown above:

```
MyPI is a PoliticalIconography {
  Architect
    is the Architect from Architecture;
  Building is an Artwork {
    artist is an Architect } }
```

Here some individual researcher decided to not only consider paintings and sculptures to be pieces of art with political relevance, but to also considers buildings etc. To this end, a new context is declared and the relevant concept is imported.

Inner concepts can be personalized by recursively applying the shown refinement. This way, both sub contexts and aggregated content can be personalized.

F. Content Clusters

Classification of content is used as a parameter to many operations on content. Classification leads to (or is derived from) a clustering of the set of content objects with respect to some notion of content semantics.

Clusters define contexts in which content can be interpreted, delivered, queried, etc. To this end, the notion of context typically incorporates not just one concept at a time, but considers complex contexts for scenarios characterized by multiple aspects of content.

As an example, consider a structure like the following:

```
Tag;
someTag is a Tag;
Content { ... }
someContent is a Content, a someTag;
```

A type `Content` is used to create content objects of a certain structure. These objects can be classified by assigning tags to them. Using M3L's principles, tags are created as `Tag` objects that are used as additional base concepts for content objects.

M3L's production process allows deriving context information based on content. As an example consider the following code:

```
Location { lat is an Int; lon is an Int }
geoLocationOfHamburgPort is a Location {...}
Event {
  date is a Timestamp }
LocalEvent is an Event {
  Location location }
822ndAnniversaryOfPortOfHamburg
  is a LocalEvent
{
  geoLocationOfHamburgPort is the location;
  05/06/2011 is a date;
  05/07/2011 is a date;
  05/08/2011 is a date }
```

In this example we define basic categories for spatial and temporal classification. `LocalEvent` is defined as the base concept for content that carries characteristics from both these classification domains.

New content can be created—independent of these definitions—like this:

```
myPhoto is a Photo {
  ... is the imgData;
  geoLocationOfPortOfHamburg is the location;
  05/07/2011 is the date }
```

Since `myPhoto` fulfills all requirements of a `822ndAnniversaryOfPortOfHamburg`, myPhoto is classified accordingly.

The determination of context from content can be used for a range of operations on content. In particular, there are several uses of pattern-based analysis in community-based content collections like support for queries, tag suggestions for newly added content, and summaries generated for large collections [5].

An increasing number of applications aims to guide users by presenting recommendations for content that is supposed to be of interest to the user. Recommendations heavily depend on context, typically including the tracked users and the user for whom to compute recommendation. The model of a user context may, for instance, consist of an explicit user profile and implicit user tracking information. Recommendations can be defined as personalized content by composition of existing content [2]. In M3L such recommendations can be deduced by applying production rules.

V. Conclusion

Content modeling is of practical importance, both for digital content as well as for content-based descriptions of real-world entities. New applications demand for extended modeling support. Recently, context information was identified as a key ingredient to powerful content models.

Existing content modeling approaches can cover most content modeling requirements, but they are not well integrated. Some model aspects, in particular variants of instances with dynamic classification in changing contexts can only be expressed by auxiliary constructs in many cases.

The Minimal Meta Modeling Language as an abstract modeling language (M3L) is a proposal to an integrated content and context modeling approach. The expressive power of the M3L supports content management tasks in an integrated way, providing modeling constructs for application-specific models.

Acknowledgment

The author would like to thank Joachim W. Schmidt and Sebastian Bossung for fruitful discussions about content modeling. Numerous project partners who participated in research projects on the topic are acknowledged. The author owes a debt to his employer, the T-Systems Multimedia Solutions GmbH, for supporting the scientific work that abstracts and extends the insights gained in practical projects.

References

[1] From content to context to engagement – the future of web content management is context. Executive focus white paper, Ektron, Inc., 2010.

[2] Zeina Chedrawy and Syed Sibte Raza Abidi. An adaptive personalized recommendation strategy featuring context sensitive content adaptation. In *Proceedings Adaptive Hypermedia and Adaptive Web-Based Systems, 4th International Conference*, pages 61–70. Springer, 2006.

[3] Neil Clemmons. If content is king…context is the kingdom…. http://experiencematters.criticalmass.com/2007/12/07/if-content-is-king-context-is-the-kingdom/, December 2007. last visited May 26 2011.

[4] Joseph Goguen. An introduction to algebraic semiotics, with application to user interface design. In *Computation for Metaphors, Analogy, and Agents*, pages 242–291. Springer, 1999.

[5] Lyndon Kennedy, Mor Naaman, Shane Ahern, Rahul Nair, and Tye Rattenbury. How flickr helps us make sense of the world: Context and content in community-contributed media collections. In *Proceedings of the 15th International Conference on Multimedia (MM2007)*, pages 631–640. ACM, 2007.

[6] Hans-Werner Sehring and Joachim W. Schmidt. Beyond databases: An asset language for conceptual content management. In *Proceedings of the 8th East European Conference on Advances in Databases and Information Systems, ADBIS 2004*, pages 99–112. Springer-Verlag, 2004.

 ISBN: 978-1-61208-157-1

Magic Wako - User Interaction in a Projector-based Augmented Reality Game

Daniel Franzen, Ignacio Avellino, Fabio Mauri
RWTH Aachen University
Aachen, Germany
{*daniel.franzen, ignacio.avellino, fabio.mauri*}@*rwth-aachen.de*

Marc Jentsch, Andreas Zimmermann
Fraunhofer FIT
Sankt Augustin, Germany
{*marc.jentsch, andreas.zimmermann*}@*fit.fraunhofer.de*

***Abstract*—Augmented reality games offer a new level of player immersion into a game world. Upcoming pico projectors, which, are integrated into mobile phones, provide a way to make augmented reality games commonly available. This technologies aided by techniques such as image processing can serve as a way to creatively enhance existing games. Unless there is first work done on how projector based augmented reality applications in general can be controlled by users, investigation on augmented reality games control is little available. We present a projector based augmented reality game which, can be controlled by either real world or virtual world interaction methods. In two qualitative user explorations, one using a low-level prototype and one using a first implementation of the game, we identify user-related and technological challenges regarding interaction with the augmented reality game.**

***Keywords*-Projector phone; interaction metaphor; augmented reality; mixed reality game.**

I. INTRODUCTION

Embedding pico projectors in mobile phones provides a complete new way to display information and new interaction techniques. The first integrated projector phones are now available. Such devices overcome the inherent display limitations of mobile phones, since larger displays can now be dynamically created on rather any surface. Projector phones also open up the range of possible interactions, since the user can interact with the mobile phone while looking at a projection or even using a combination of the mobile phone's screen and the projection in parallel. For users, this methodology eliminates the necessity to switch the focus of attention between the real object and the augmented video version on the mobile device's screen, which was inherent to the See-Through-Approach [1] for creating augmented reality applications.

Recently, using the physical world as playing field is becoming popular since this increases user immersion into the game world. Examples are dance mats or Kinect [2]. But, in these cases, the actual game world is still behind the screen. Hence, using augmented reality to play games in the real world can further increase gamer immersion.

This paper researches the interaction of users with augmented reality games focussing on a congruent setup of the mobile phone's camera and projector, were the field of view and the field of projection overlap [3]. This particular spatial configuration enables the user to interact directly with the projection without any limitations, and is open for the different interaction concepts described by Rukzio and Holleis [4]. However, the congruent setup also introduces problems, since the projected image can have an immediate effect on the processing of the captured camera image. Furthermore, the physical *process* of pressing a button on the device causes that it is shaken. This affects other input modalities which, are considering position and orientation.

This paper elaborates on how users perceive the interrelation of output and input channel and how they utilize this interrelation for their interaction in a game. For being able to observe players' behavior we created an augmented reality game called *Magic Wako*. This game offers an easy to learn gameplay associated with the necessity for fast, direct and gesture-rich interactions with a mobile projector.

The remainder of this paper is organized as follows. After discussing related work in the succeeding section, we present findings of our first low-level prototype. After that, we describe experiences from the first system implementation and insights from further user tests. Finally, we conclude and give an outlook on future work.

II. RELATED WORK

The *Wear UR World* prototype [5] shows many everyday life examples how a portable projector can be used to augment everyday objects with additional information. As input modality, a wearable camera captures four-finger-gestures. For easier recognition, the fingers are equipped with colored markers. Baldauf and Froehlich [6] use the same gesture recognition approach but process the image on the mobile phone to make the setup more portable.

There have already been some games implemented which, use mobile projectors to interact with the real world. Pinhanez et al. [7] use the *Everywhere Display Projector* to let people build puzzle pictures with colored sweets. The projection shows gamers where to put a sweet on the ground. So, in this case the projector is controlling the user. In *Co-GAME*, gamers project paths onto any ground were a real-world robot toy is being guided towards a defined destination [8]. The robot is equipped with infrared LEDs that are tracked by a camera and steered by a server component. So, gamers control the game by moving themselves and the projector in their hands. *Flashlight jigsaw* is a multiplayer

 ISBN: 978-1-61208-157-1

puzzle game made for handheld projectors [9]. Players have to search on a public display for pieces of the puzzle by exploratively pointing at the screen with handheld controllers. Interaction metaphors are similar to *Magic Wako* without considering real-world objects. In *LittleProjected-Planet*, Loechtefeld et al. [10] use paintings on real world walls to let projected balls run through it. In this case, the individual manipulation of the real world creates the game world. At the same time, the information where the projector is pointing at is steering the game.

The presented game approaches do not deeply investigate user acceptance of the new input modality. However, Kawsar et al. [11] let users compare the See-Through approach against mobile projection in three different non-gaming applications. Afterwards, they conduct a qualitative study concerning user acceptance and usability issues of the interaction techniques. While preferring the larger displays of the projection approach, users figure out a higher degree of cognitive load due to more demanding hand-eye coordination. Blasko et al. [12] identify the stabilization of the projected image as the major challenge in the interaction with their wrist-worn projection display. However, we expect to find out additional challenges of projector phone interaction in the gaming context, as this requires faster and more spontaneous movements.

III. Prototype Design

We set up a first low-level prototype in order to find out about user requirements how to control the game.

A. Gameplay

MagicWako explores a new paradigm of gaming that aims at recreating the popular arcade game *Whac-A-Mole* [13]. In the original game, little moles (called Wakos) come out randomly from holes in the game board and disappear again after a short time. The aim of the game is to hit as many Wakos as possible. In our augmented reality version of that game, players have to search for the Wako with a mobile projector in their hand. The game is played on a physical playing field on which, virtual Wakos are projected by the mobile projector. As the field of projection does not cover the complete playing field, gamers have to explore the playing field to search for Wakos by moving the projector. Magic Wako differs from games presented in the previous work since firstly the projector is controlled by the user and not the other way around, and secondly only the controller and the gameboard are required to play the game; there is no need for additional world objects. Additionally, the game board can be replaces with any available object due to the initial color calibration.

B. Setup

We conducted a qualitative exploration with a low-level prototype and seven users. The aim was to find out how users would intuitively control the game with the new interaction paradigm and how they feel about it. Secondly, we wanted to know what problems might be caused by the interaction using a handheld projector for controlling an augmented reality game. The low-level prototype does not reflect the final system real conditions such as light, since the main focus is on the interaction of the control and not on the gameplay. This will be analyzed using the final implementation.

We presented the users the physical playing field, which, contained nine different colored circles made of cloth and explained the gaming rules. All users were familiar with computer games but never played an augmented reality game or used a mobile projector before. Users had to use an electric torch to simulate the projector (cf. Figure 1). Additionally, they were equipped with a mobile touchscreen phone, which, presented only one big button labeled ”Hit”. We simulated the Wakos by randomly putting and removing a card with an *X* on the circles of the playing field. Without further information on how the actual control of the game should work, for example how to hit the *X*, we asked users to play the game. We observed the players and afterwards interviewed them about their feeling of the game.

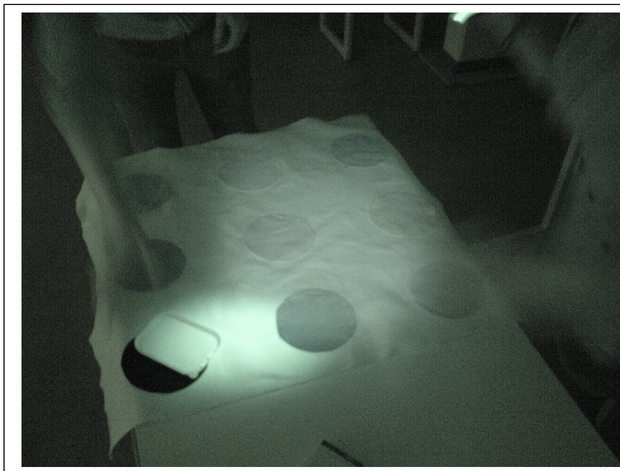


Figure 1. Torch setup for user evaluation

C. Ergonomic Findings

In the interviews, all users claimed the game to be comfortable and found the movement with the torch natural. Hence, we assume that the general approach of controlling the game is intuitive and was appropriate to be implemented in the first prototype. It seems that searching for objects with light in the real world is intuitively easy to understand even for users without experience in augmented reality games.

Three out of seven users pressed the ”Hit” button for hitting the *X*. Another three users tried to hit the physical *X* with their hand. One person wanted to use the torch itself to hit the *X*. The game setup is meant to be a hybrid between a virtual and a physical game. According to that, it seems that about half of the users favored a virtual way of hitting - pressing a button. The other half of the users intuitively preferred to act in the physical world, be it with the hand or the torch itself. Hence, we plan to offer both possibilities to hit the Wako in the prototype and compare again.

We observed that especially users who tried to hit the *X* with their hand experienced problems to keep the torchlight steady while touching a far away circle with the other hand. When users leaned towards the playing field to hit a projected *X*, they tended to move the torch thus pointing to another position where the *X* was not present. Hitting the *X* by pressing the button relieved the users from leaning towards the playing field. Nevertheless, in an integrated projector phone setup, also the physical process of pressing the button would make the device shake a bit so that the current aim of the projection would change in that moment.

Five users pointed out the problem of carrying around the playing field and asked to be able to play the game on arbitrary surfaces.

IV. System

In this section, we present our first prototype of *Magic-Wako*. It was used as to find out more user-related and technical challenges for augmented reality games control. For this, we conducted another preliminary user study.

A. Decisions

As mentioned in the previous section, two possibilities to hit the Wako were taken into account. We implemented a first version of *Magic Wako* where the user had to hit the smartphone's screen laying in front of her. Providing this modality, the input process promises to have fewer impact on the position of the projector. Furthermore it is also possible to hold the Smartphone in the other hand and do the hitting with the thumb like it would be done with a button. However, we will add a module that can determine if a Wako is hit by a hand or another physical object in future work.

To detect the currently focussed real world object, we use a color recognition engine which, can determine the currently targeted circle. To meet the users' requirement of playing the game on other surfaces as our game board, a color calibration phase was added to the game. In calibration mode, users are asked to point at a color for 3 seconds. This color is saved as one of the game colors. In this way, players can initially specify the colors of the holes were the Wakos appear. This also allows a new range of playing fields, for example on people's colored shirts in front of an unicolored background.

B. Hardware Prototype

Due to the lack of appropriate integrated projector phones on the market, we connected an Adapt ADPP-305 projector to a Samsung Galaxy S smartphone. The Galaxy S offers a TV out port and high processor power for graphic computation. The ADPP-305 is a good trade-off between mobility (battery-powered), brightness (45 lumens) and size (fits in one hand). A wireless camera is attached to the projector to guarantee the congruent setup. The wireless camera provides colored images of 640x480 pixels. This quality is high enough in order not to raise problems with color detection, given regular indoor light conditions. The prototype's weight and size allow to operate with one hand.

One of the main problems encountered with this prototype concerned the connecting cables between the projector and the smarthpone. Firstly, there were several cables which confused the users on how to handle the device comfortably. Some of the users wore the cables around their neck. Secondly, as the users moved the prototype some of the cable jacks rotated making the projected image to be unavailable for a short period of time until the projector detected the input again. In a high-level prototype were the projector is integrated into the smartphone this would not be a problem.

Figure 2 shows the system architecture: The wireless camera collects an image of the game board and sends it to a server PC. The server computes the position and size of the Wako and sends this information to the smartphone, which, creates the final image to be displayed by the mobile projector using the stored current score and time left, too.

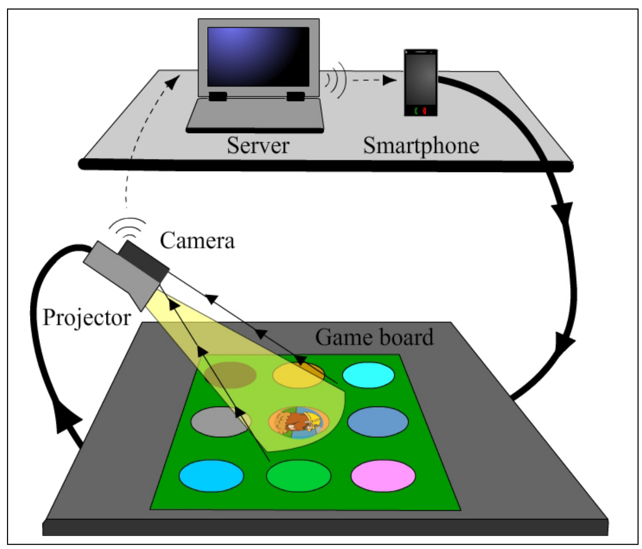


Figure 2. Game setup

Figure 3 shows a user who is playing with the first prototype.

C. Software Concepts

An image recognition module on the server detects the color of the circle where the user is pointing at by analyzing the current camera frame. The colors of a certain amount of pixels inside the shape are averaged. The actual algorithm was kept as easy as possible to execute it on a smartphone in

Figure 3. User plays with first prototype

later implementations. In the calibration phase, the images are analyzed for 3 seconds and the average value is stored for usage in the next game.

The server sends the position and size of the currently aimed spot to the game logic module on the smartphone. This module keeps track of the state of the game and passes all needed information to the image projection module, which, determines if, where and in which size the Wako shall be displayed. After this, the image is generated and displayed by the projector.

D. Lessons Learned

We conducted another qualitative exploration with the first version of *Magic Wako*. The aim was to confirm the observations of the low-level prototype testing. Additionally, we wanted to identify further user-related and technological challenges of this augmented reality game approach.

The exploration was conducted with six users which, did not take part in the first user test. They were equipped with the hardware prototype. After an introduction to the gameplay, users were asked to perform four tasks and provide feedback about the interaction: 1) Point at the top rightmost spot in the game board. 2) Point at all the spots at least once in the game. 3) Find the Wako during the game at least twice. 4) Hit the Wako during the game at least once. All the tasks were successfully performed by every user. All users managed to hit the Wako with increasing frequency towards the end of the game. It confirms that this way of handling the projector is natural and easy to learn. Additionally, every player told that the game is interesting and fun.

Three users asked for a device that is lightweighter and easier to handle. As pointed out in this section, the hardware prototype is a tradeoff because sufficient integrated hardware is currently not available. In an optimal setup, camera and projector would be integrated in the smartphone. This would reduce the size, weight and bulkiness of the hardware components. Also the server would not be necessary because all the computations could be performed on the smartphone. With regard to the shaking problem, it would have to be investigated if it is still a problem in this new setup. Alternatively a wireless button could be provided so that the user could hit it with her free hand. Although this time we explained that the Wako has to be hit by touching the mobile phone's screen, two users still tried to hit the Wako by hand. This confirms the need for a physical interaction with the game board as an alternative to the one with the smartphone.

In some cases when the Wako was projected on a circle, the color of the latter was altered by the projected light and the color detection did not work properly. This generated an annoying flickering of the projection. The color recognition algorithm was improved by increasing the similarity threshold between the detected color and the color stored during calibration. Projectors with a darker light work better because their projected image interferes less with the real world objects. At the same time, if the projection is too dark, the Wako is not properly visible. Unfortunately, background subtraction is not possible in this setup. The gameplay involves the user to navigate the field with the projector, which, constantly changes the image detected by the camera. This doesn't allow to distinguish whether the change in the detected image is due to the movement of the user or due to light changes caused by the projector.

Three users had problems to find a Wako because they were moving the camera and projector too fast thus preventing the image recognition module to work properly. The color detection process of one camera frame takes a few 100 ms. When color detection was completed, too fast users already pointed at another color. This resulted in a one-frame blinking of the Wako. That problem can be solved by optimizing the color recognition. Also a reduction of data communication steps can increase performance, for example by omitting the server between camera and smartphone.

V. Conclusion and Next Steps

Using augmented reality with a projector as a mean of game control has potential to intensify current trends in gaming. During two qualitative explorations with a low-level prototype and a first implemented version of the augmented reality game *Magic Wako*, we could identify user-related and technological challenges. We also observed that our interaction approach of controlling the game through projector movements was easy to learn and widely accepted by users.

Our explorations indicate that both modalities, hitting the Wako by button press and hitting the Wako by hand or object, are requested by a significant group of users. There is slight evidence that hitting the Wakos physically seems to be more intuitive. We will further investigate the hitting modality in the next iteration. It will have to be taken into account the movement of the persons hand holding the control when they try to hit the wako phisically with their other hand. In future work, we are going to implement the possibility to hit the Wako by hand or by using objects in order to confirm this finding with a high-level prototype. This would solve the problem of having to push a button which caused the shaking of the device. Also in the next iteration it will be investigated what kind of user group prefers what modality. This input modality will probably cause bigger usability problems because the physical input process affects the projector handling a lot. We will try to solve these problems, for example, by applying image stabilization methods. The next iteration will include a larger group of testers since the number used in the previous two user studies is not sufficient to draw solid conclusions.

The ability to play the game on arbitrary surfaces is a further matter of investigation. In this context, reachability or mobility of objects offer new problems and possibilities. Also, further investigation on this matter will possibly lead to improvements or alternative implementations on the current image processing algorithm. These might include a new approach on distinguishing the colors where the Wakos come out such as background substraction.

If more sufficient integrated projector phones will be available in the future, this will meet one of our identified user requirements. However, it will also change the nature of a button press compared to our current prototype. By now, the physical process of touching the phone's screen does not influence the position of the projector, but this is different in an integrated projector phone. We will examine how strong this limits the user acceptance. A solution could be to compensate the projector movement recalculating the position of the projected image.

Further work will concern technological issues. The color recognition module will be optimized to deal with the problem of too fast user interaction. Also, the initial color calibration allows for the game to be played in an arbitrary surface that contains a variety of nine different colors. Test will be made in another setup than the one presented with the physical board with nine dots.

VI. Acknowledgements

This research was supported by the European Commission within the BRIDGE project (project No. 261817).

References

[1] E. Bier, M. Stone, K. Pier, W. Buxton, and T. DeRose, "Toolglass and magic lenses: the see-through interface," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, 1993, p. 80.

[2] Microsoft, "Kinect, Xbox.com," 2011, available online at http://www.xbox.com/en-US/kinect; visited on June 28th 2011.

[3] J. Schoening, M. Loechtefeld, M. Rohs, and A. Krueger, "Projector Phones: A New Class of Interfaces for Augmented Reality," *International Journal of Mobile HCI*, pp. 1–14, 2010.

[4] E. Rukzio and P. Holleis, "Projector Phone Interactions: Design Space and Survey," in *Workshop on Coupled Display Visual Interfaces at AVI 2010*, 2010, pp. 408–410.

[5] P. Mistry, P. Maes, and L. Chang, "WUW-wear Ur world: a wearable gestural interface," in *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, 2009, pp. 4111–4116.

[6] M. Baldauf and P. Froehlich, "Supporting Hand Gesture Manipulation of Projected Content with Mobile Phones," in *Proceedings of the Workshop on Mobile Interaction with the Real World (MIRW)*, 2009.

[7] C. Pinhanez, F. Kjeldsen, A. Levas, G. Pingali, J. Hartman, M. Podlaseck, V. Kwatra, and P. Chou, "Transforming Surfaces into Touch-Screens," *IBM Research Report - RC22273 (W0112-0216)*, 2001.

[8] K. Hosoi, V. Dao, A. Mori, and M. Sugimoto, "CoGAME: manipulation using a handheld projector," in *ACM SIGGRAPH 2007 emerging technologies*, 2007.

[9] X. Cao, M. Massimi, and R. Balakrishnan, "Flashlight jigsaw: an exploratory study of an ad-hoc multi-player game on public displays," in *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, 2008, pp. 77–86.

[10] M. Loechtefeld, J. Schoening, M. Rohs, and A. Krueger, "LittleProjectedPlanet: An Augmented Reality Game for Camera Projector Phones," *Proc. of MRIW*, 2009.

[11] F. Kawsar, E. Rukzio, and G. Kortuem, "An explorative comparison of magic lens and personal projection for interacting with smart objects," in *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, 2010, pp. 157–160.

[12] G. Blasko, S. Feiner, and F. Coriand, "Exploring interaction with a simulated wrist-worn projection display," in *Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, 2005, pp. 2–9.

[13] B. S. Racers, "Whac-A-Mole," 2011, available online at http://whacamole.com; visited on June 28th 2011.

Real-Time Deformable Soft-Body Simulation using Distributed Mass-Spring Approximations

Ben Kenwright
School of Computing Science
Newcastle University
Newcastle, United Kingdom
b.kenwright@ncl.ac.uk

Rich Davison
School of Computing Science
Newcastle University
Newcastle, United Kingdom
richard.davison4@ncl.ac.uk

Graham Morgan
School of Computing Science
Newcastle University
Newcastle, United Kingdom
graham.morgan@ncl.ac.uk

***Abstract*—This paper investigates several methodologies for simulating soft-body objects using a mass-spring approach. The mechanisms are then expanded to include deformation information that can produce results suitable for use in real-time applications where *visual impact rather than accuracy is desired,* such as video games. Many methods use complex and esoteric methods to achieve physically accurate simulations; we target the mass-spring model because of its simplicity, using creative modifications for diverse visual outcomes.**

Keywords-soft-bodies; physics; deformation; mass-spring; games; voxelation.

I. Introduction

With the increase in computational power, movies and games have started to take advantage of simulated visual effects. Such effects include soft-body physics and deformation artifacts, which can be used to create more realistic skin/face movement in character simulations; deformable objects can bend and move when forces are applied. For example a metal bar bending when under stress, with enough stress eventually causing a permanent deformation of its shape. For games and the movie industry, visual effects are more important than accuracy, this means more novel methods come about that use eccentric models to create more visually pleasing results. One such novel method is to use springs to represent material rigidity, and use various topologies and tricks that produce close to the real thing. We explore some of these approximations to demonstrate various novel methods for simulating soft-body objects using simple mass-spring approximation.

The main contribution of this paper is the proposal of several models and strategies for use in soft-body simulations. We introduce various approximation techniques and their applicability to real-time applications such as games.

These methods are extended to demonstrate permanent deformation effects by taking advantage of the stress information stored in the mass-spring model.

Our method approaches the problem by keeping the simulations fast and uncomplicated by utilizing a mass-spring system in combination with various modeling approaches. As visual impact is the primary concern of the paper, the models are judged by whether they produce plausible results at interactive frame rates, rather than their ability to generate results that are physically accurate.

The remainder of the paper is organized as follows. Section II presents previous work done in the area, and introduces our integration scheme and mass-spring model. Section III describes our soft-body model approximations. Section IV explains our deformation method. Section V discusses preliminary results. Last, Section VI presents conclusions and future work.

II. Background and Related Work

The mass-spring system is one of the simplest physically-based models developed over the past decade [1][2], its simplicity and ease of implementation making it the most likely candidate to achieve real-time performance. Prior work has been done which models soft-body simulations using the object's pressure [3], and implicit integrators with a mass-springs model [4]. A deformable soft-body is approximated by sets of masses linked together using springs in various configurations. The mass-spring is highly parallelizable, easy to implement and involves few computations.

There are various approaches to creating soft-bodies. Elasticity and viscoelasticty models have been shown to be successful [5,6], but are not suitable for real-time applications due to their computational complexity. Using a global restraining method to create soft-bodies [7,8] allows a simulation to run at more interactive rates but demonstrates less realism.

Other methods include a finite-element approach. The approach decomposes the model into separate pieces. Work by [12,13,14,15] used tetrahedral elements, while [16] employed a hexahedral composition. Bro-Nielsen [9] takes advantage of linear elasticity to achieve real-time deformable structures. A more recent and novel method has been to use neuro-animators [10], which have a learning period, after which they can emulate coupled physical systems, including cloth.

Our method is different by using an extremely low computational model that may not produce physically accurate results, but more artistic visual outcomes.

 ISBN: 978-1-61208-157-1

A. Integration Scheme

The methods explored in this paper use a semi-explicit Euler method (also called symplectic Euler):

$$v_i^{n+1} = v_i^n + F_i^n \frac{dt}{m} \quad (1)$$

$$x_i^{n+1} = x_i^n + v_i^{n+1} dt \quad (2)$$

where v is velocity, x is position, n is the current frame, F is force, dt is the timestep, and m is mass.

This method of integration is computationally cheaper than implicit integration, but suffers from stability issues with large forces, where the time step squared must be inversely proportional to the stiffness. This is not such a problem with off-line computation, but is an issue when used in real-time applications since very small time steps are required to ensure stability. As a way to overcome this problem and ensure that our simulations remain robust and reliable, we clamp the maximum forces and allow for small visual artifacts to keep at real-time frame rates.

We also extend this checking of the forces at limits to introduce deformation artifacts.

B. Mass-Spring Model

Springs are not a perfect physical model for soft-body objects, but provide a good visual approximation.

The mass-spring system relies on the principle of Hooke's law, which states that the force applied by a spring is proportional to the load it is under.

Linear: $F_{spring} = -k_s X_{diff} + k_d v \quad (3)$

Square: $F_{spring} = -k_s X_{diff}^2 + k_d v \quad (4)$

where F is force, k_s is the spring constant, k_d is the spring damping constant, X_{diff} is the current displacement distance, and v is the velocity of the mass-spring.

Hooke's law is a useful approximation of how a real-life spring acts when compressed or expanded by external forces – it always tries to return to its basal 'restitution' length, with a greater force the further from this restitution length the spring becomes. The soft-body approximation methods described here use mass-springs between the vertices of the model, allowing flexible movement that reacts realistically to forces imparted upon it. However, it is not enough to simply recreate the model mesh using springs – the spring model is likely to collapse in on itself and become unstable under minor forces. Section III investigates methods for solving this while still allowing for real-time soft-body approximation.

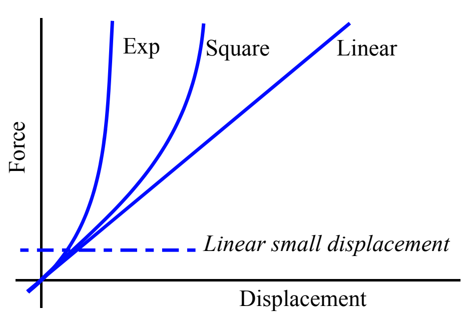


Figure 1. Ideal spring model.

Using a squared displacement-correcting spring gives us a more rigid body that still shows good flexibility for small changes, such as simulating a skin surface, or any small surface ripples while reducing the overall bendiness of the shape. Mass values chosen initially where equal and the sum of the total mass of the object. This was arbitrary, but we explored variations of the mass selection, such as making masses at the centre different, or making the mass on one edge to be larger so that object appeared to self-balance, as shown in Figure 3. We can expand on the idea of varying the mass-spring properties to create alternative effects. Other spring modifications would include different spring coefficients for expansion and contraction to produce more outlandish artistic visual effects.

III. METHODS

This section introduces various soft-body models and their applications.

A. Brute Force Method

This method is usually the first and most intuitive way of creating a soft-body object from a mesh, where spring constraints connect every vertex to every other vertex to form a rigid structure as shown in Figure 3. This produces a sufficiently stable model for soft-body simulation, but has a high cost of (n*n-1)/2 spring constraints per object.

B. Mass-Spring Voxelation Method

Splitting a model into an array of voxels is useful for the

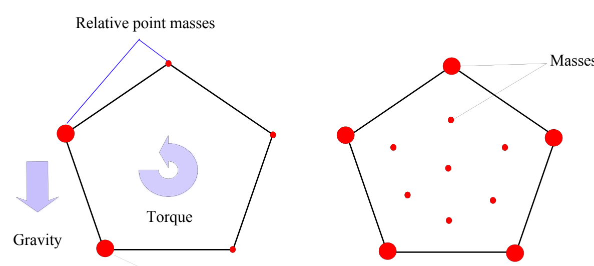


Figure 2. *Left*: Biasing the weight on one side of an object to make it balance. *Right*: Mass variation, using larger masses for the surface and smaller masses for the volume.

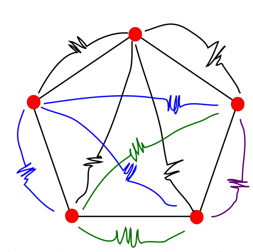


Figure 3. Brute force interconnection of surface vertices using mass-springs.

 ISBN: 978-1-61208-157-1

simulation of soft-bodies on highly tessellated shapes where using other methods would result in an inordinate number of springs. An energy efficient adaptive space deformation version was demonstrated by [11]. Our model utilizes voxelation for soft-body simulation by decomposing the shape area into voxels, then assigning each vertex to its surrounding voxels. A weighting factor determining how much each voxels change in orientation-position influences vertex movement – similar to the blend weights used in skeletal animation systems.

Figure 4 shows how these voxels are connected via mass-springs; when the voxelated approximation is deformed, the vertices of the original model are moved according to their attached voxel weightings.

This method provides advantages in performance over the brute-force method when used with high poly meshes, since the number of springs is determined by the voxel resolution and not the number of vertices. It is less suitable for low-poly models, where the underlying voxels will be able to form shapes the model vertices cannot accurately reproduced. This can be countered by tessellating the model, increasing the number of vertices and allowing greater freedom of movement.

C. *Uniform Grid Mass-Spring Distribution*

Another method of combining model rigidity with soft-body dynamics is using a uniform grid. Via this grid, mass is equally distributed across the shape, with extra vertices added to the model at points where the grid and model intersect. Neighboring vertices are then joined using mass-springs to form a rigid model.

This method offers the advantage of having a mass-spring distribution throughout the object, allowing a closer approximation to real soft-body physics. The point-mass distribution is uniformly through-out compared with the voxel method previously which uses a random scattering approach.

D. *Internal / external shape scaffold*

When creating surface springs, the model lacks rigidity to keep its original shape, due to it having multiple positions where its springs are still the same length, or lack enough strength to keep its original shape. Using a shape with a reduced number of vertices to attach the surface vertices to increase rigidity, either having the virtual rigidity model inside our outside the original.

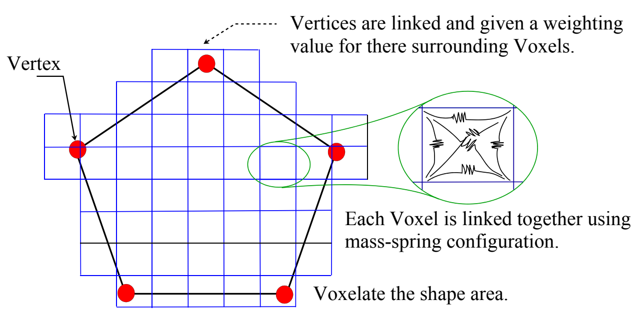


Figure 4. Mass-Spring Voxelation Method.

E. *Shell Method*

Complex and concave shapes, such as tubes give problems where additional mass-springs are added to increase rigidity. This additional rigidity causes the model to bend and twist un-naturally. Using a duplicated virtual surface shell mesh with springs, and attached to the original surface points, gives us a more rigid surface which mimics the behavior we would expect for more complex shapes. One such example of where this method gives good results is the crushing of a tube pipe.

This method extends the uniform approach, by only includes voxels intersecting the surface or within the shape.

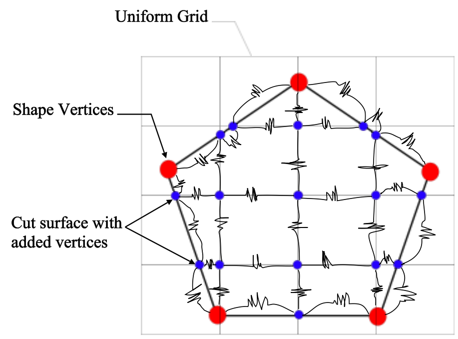


Figure 5. Uniform mass-spring grid design.

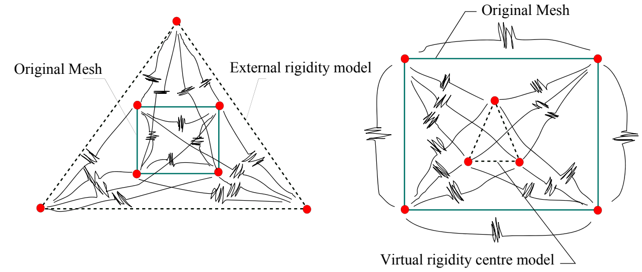


Figure 6. Using a triangle as an inner and outer skeleton for rigidity.

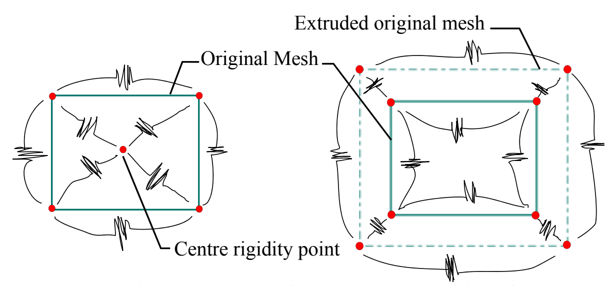


Figure 7. *Left*: Add a single reference point to attach surface points to help the soft-body hold its shape. *Right*: Using an extruded or intruded shell surface to introduce rigidity to the soft-body model

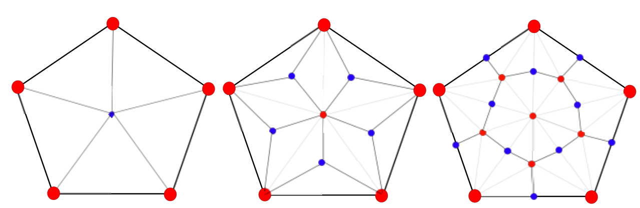


Figure 8: An example of a shape being recursively split into subsections

F. Tetrahedron Mass-Spring Distribution

This method recursively subdivides the shape into subsections - tetrahedron sections in 3D, or triangles in 2D. At each level we subdivide the largest subsections into smaller subsections until a minimum size volume is reached. Because simple shapes are used for each subsection, the mass and volume can be calculated accurately. This can then be used to scale the mass-spring constraints to give a more uniform force distribution across the body. Each subsection is then split into mass-springs along its edges.

G. Voronoi Regions and Delaunay Triangluation

The shape is partitioned using Delaunay triangulation to get a uniform random distribution of springs to help give the body rigidity.

IV. DEFORMATIONS

As materials are stretched, they reach an *elastic limit*, beyond which they will become permanently deformed, and Hooke's law will no longer accurately approximate its elastic properties.

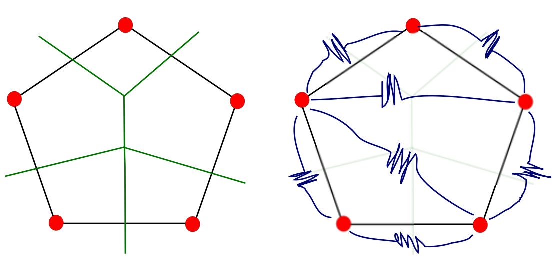
We extended our simulation to accommodate this elastic limit, so that once a mass-spring reached a specific force threshold its stiffness and restitution length is recalculated, permanently deforming the model.

V. RESULTS

The results do not physically represent how a soft-body would squish and stretch in the real world, but that was not the aim of our work. Instead we have focused on the visual results for real-time, interactive applications such as games. We have shown diverse techniques for creating approximate soft-body simulations with extension of the underlying principles to handle deformation by taking advantage of the mass-spring stress information between frames.

For comparison, Table 1 gives a brief summary outlining the different methods.

Initial methods have been simulated in 2D to test their viability and assist in quick prototyping, while the principles can easily be extended into 3D. The 2D methods gave us simple building blocks which helped us take the designs in more outlandish directions.



(a) (b)

Figure 9. (a) Voronoi Regions, (b) Delaunay Triangle are used to determine where springs are placed between masses.

VI. CONCLUSION AND FUTURE WORK

The results show flexible methods to simulate soft bodies. We used the basic Euler mass-spring model for the simulations to get early results and see how they compare. We would hope to test out more reliable integration methods such as Verlet. The Early work has been with 2D shapes but hope to extend it to 3D. We would also like to vary the mass distribution throughout the models, varying the mass properties at different points and see how the results compare to realistic models. It would be interesting to enhance the simulation to display force and energy distribution throughout the objects. Additionally, the easily parallelizable nature of mass-springs makes these methods ideal candidates for GPGPU processing.

VII. REFERENCES

[1] J.E. Chadwick, D.R. Haumann, and R.E. Parent, "Layered construction for deformable animated characters," ACM SIGGRAPH Computer Graphics, vol. 23, Jul. 1989, pp. 243-252.

[2] G. Miller, "The motion dynamics of snakes and worms," Computer Graphics, Volume 22, Number 4, August 1988, vol. 22, 1988, pp. 169-178.

[3] M. Matyka and M. Ollila, "Pressure model of soft body simulation," SIGRAD2003, November, 2003, p. 20–21.

[4] J. Mesit, R. Guha, and S. Chaudhry, "3D Soft Body Simulation Using Mass-spring System with Internal Pressure Force and Simplified Implicit Integration," Journal of Computers, vol. 2, Oct. 2007, pp. 34-43.

[5] D.T. and A. Witkin, "Physically based model with rigid and deformable components," 1988, p. IEEE Computer Graphics and; Applications; pages 41.

[6] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer, "Elastically deformable models," ACM SIGGRAPH Computer Graphics, vol. 21, Aug. 1987, pp. 205-214.

[7] A. Pentland and J. Williams, "Good vibrations: Modal dynamics for graphics and animation," Computer, vol. 23, 1989.

[8] A. Witkin, "Fast Animation and Control of Nonrigid Structures."

[9] M. Bro-Nielsen and S. Cotin, "Real-time Volumetric Deformable Models for Surgery Simulation using Finite Elements and Condensation," Computer Graphics Forum, vol. 15, Aug. 1996, pp. 57-66.

[10] R. Grzeszczuk, "NeuroAnimator : Fast Neural Network Emulation and Control of Physics-Based Models NeuroAnimator : Fast Neural Network Emulation and Control of Physics-Based Models," Science, 1998.

[11] M. Botsch, M. Pauly, M. Wicke, and M. Gross, "Adaptive Space Deformations Based on Rigid Cells," Computer Graphics Forum, vol. 26, Sep. 2007, pp. 339-347

[12] E.G. Parker and J.F. O'Brien, "Real-time deformation and fracture in a game environment," Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation - SCA '09, 2009, p. 165-175.

[13] G. Irving, J. Teran, and R. Fedkiw, "Invertible finite elements for robust simulation of large deformation," Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation - SCA '04, 2004, p. 131-140.

[14] M. Müller, J. Dorsey, L. McMillan, R. Jagnow, and B. Cutler, "Stable real-time deformations," Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation - SCA '02, 2002, p. 49-54.

[15] A.W. Bargteil, C. Wojtan, J.K. Hodgins, and G. Turk, "A Finite Element Method for Animating Large Viscoplastic Flow," vol. 26, 2007, pp. 291-294.

[16] V. Baudet, M. Beuve, F. Jaillet, B. Shariat, and F. Zara, "Integrating Tensile Parameters in Hexahedral Mass-Spring System for Simulation," WSCG'2009, 2009

A. Brute Force	Every vertex (point-mass), link all-to-all
B. Voxelation	Voxels (or bricks) form a weighted skeleton
C. Uniform Grid	Uniform grid interior, surface sliced and joined
D. Internal/External Scaffolding	Skin linked to exterior/interior reduced poly shape
E. Shell Method	Shells (or layers), extruded surface normal, linked together to form a rigid surface structure
F. Tetrahedron	Partition into Tetrahedrons
G. Voronoi Regions/ Delaunay Triangulation	Partition using Voronoi Regions and Delaunay Triangulation

Table 1. Comparing the various methods.

Classification of Pathologic and Innocent Heart Murmur Based on Multimedia Presentations of Acoustic Heart Signals

Božo Tomas
University of Mostar, Faculty of Mechanical Engineering and Computing, Mostar
University "Herzegovina", Faculty of Social Sciences Dr. Milenko Brkić, Bijakovići
Bosnia and Herzegovina
bozo.tomas@hteronet.ba

Darko Zelenika
University of Mostar, Faculty of Mechanical Engineering and Computing, Mostar
University "Herzegovina", Faculty of Social Sciences Dr. Milenko Brkić, Bijakovići
Bosnia and Herzegovina
darko_koga@hotmail.com

Željko Rončević
University Clinical Hospital Mostar, Clinic for Children's Diseases Mostar,
Cardiology Department, Bosnia and Herzegovina
zeljko.roncevic@mocable.ba

Antonija Krtalić
University "Herzegovina", Faculty of Social Sciences Dr. Milenko Brkić
Bijakovići, Bosnia and Herzegovina
antonija.krtalic@hteronet.ba

***Abstract*— Heart is a source of sound, a "live instrument" of very unstable structures. A sound produced by acoustic activity of the heart contains information about health state of the heart. However, very few physicians can establish medical diagnosis just by using the technique of listening of heart sounds called auscultation. Spectral composition of acoustic heart beat signals in children without heart murmur (Normal), and with pathologic heart murmur of Ventricular Septal Defect (VSD) were compared with innocent Still's murmur (Still) as to determine the parameters of the Still's murmur. Goertzel algorithm implementation can provide a basic understanding of the spectral compositions of the heart sounds and murmurs. Still's murmur is analyzed in more detail because it is incorrectly diagnosed by many doctors. The graphic illustrations in this paper include 3D graphic presentation of heart sound signals as the main achievement of the paper.**

Keywords-3D graphic presentation of a heart sound; Goertzel algorithm; Still's heart murmur; time-frequency composition.

I. INTRODUCTION

It is well known that there are correlations between the health state of the heart and the sound produced by its beating. The quality of finding of the cardiac auscultation is influenced by previous knowledge, experience and even musical talent [1]. Heart auscultation, defined as the process of interpreting acoustic waves produced by the mechanical action of heart is a noninvasive, low-cost screening method and is used as fundamental tool in the diagnosis of cardiac diseases [2]. Unfortunately, heart sound interpretation by auscultations is very limited to human ear competence and depends highly on the skills and experience of the listener. At heart health diagnosis, the most important point is the selection between a healthy heart (there are no deformations on heart) and sick heart (there are deformations on heart). Deformations of the heart are manifested through pathologic murmurs, either there are no murmurs or there are innocent murmurs on a healthy heart [3].

The sound emitted by a normal cardiac cycle is composed of the first heart sound (S1), a time period between the first and second heart sound called systole, a second heart sound (S2) and a time period between the second (S2) and first heart sound (S1) called diastole. Heart murmurs are extra sounds (noise) heard through systole or diastole. Systolic heart murmurs can be physiologic (innocent) without disease and pathologic caused by heart diseases. Innocent vibratory murmur was described for the first time in 1909 by an English paediatrician George F. Still (1868-1941) [4]. A good example of pathologic murmur is harsh systolic murmur of ventricular septal defect (VSD) caused by turbulent blood flow through a defect („a hole") in ventricular septum. Cardiac auscultation, especially cardiac auscultation in children is a skill difficult to master. The human ears and stethoscope are not perfectly adapted to the heart sounds. Heart sound and murmurs are nonstationary in time and are arranged in the low frequency range. However, acoustic stethoscopes will be used for a long time but will gradually be repressed by electronic stethoscopes [5].

Spectral analysis of the heart sound data in most of published studies was obtained by applying a Fast Fourier Transform (FFT) and/or Wavelet Transform (WT) [2][6][7][8][9]. In this paper, the Goertzel algorithm was used and obtained results could be compared with FFT and WT. The algorithm was introduced by Gerald Goertzel (1920-2002) in 1958 [10].

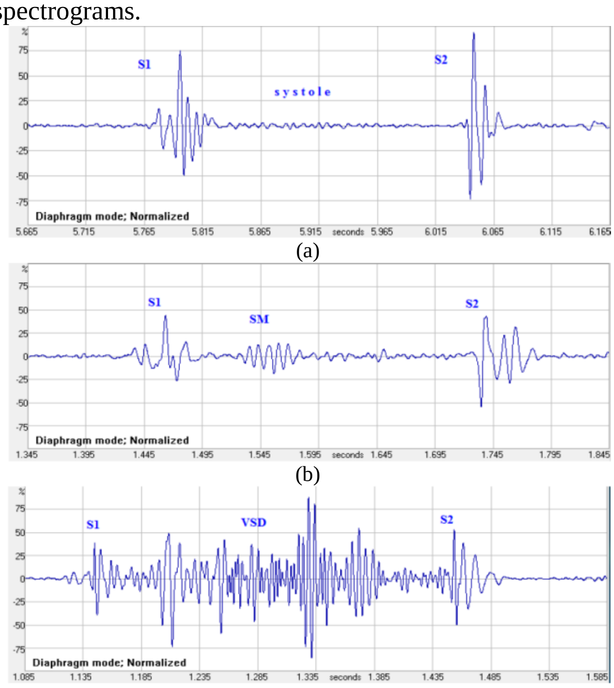
The paper is organized as follows. In Section II, recording and display of heart sound signals are shortly presented. The results of tone analysis in Still's murmur are in Section III. 3D graphics presentations of a heart sounds and Still's murmur are presented in Section IV. Section V concludes the paper with final remarks.

II. ANALYSIS PROCEDURE

During the children examination in outpatient clinic by the paediatric the cardiologists, their heart sounds were recorded with an electronic stethoscope. Each soundtrack of acoustic activity of the heart was taken for a period of 8 seconds for each patient. Children were classified into

 ISBN: 978-1-61208-157-1

three groups: children without heart murmur - Normal, children with physiological Still´s innocent murmur - Still and children with pathological murmur associated with congenital heart disease - VSD. All children were examined with ultrasound for accurate diagnosis of congenital heart disease. Fig. 1 shows one Normal, one Still and one VSD in time and Fig. 2 shows theirs spectrograms.

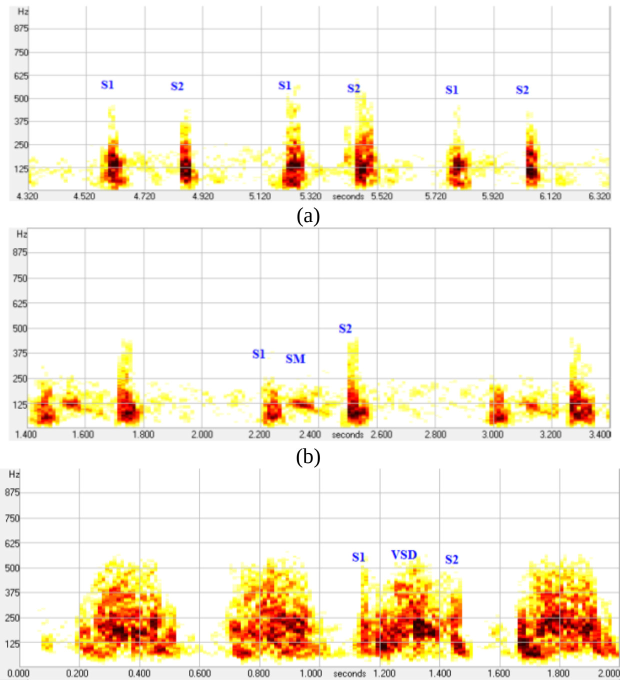


(a)

(b)

(c)

Figure 1. Heart sounds in time. (a) Normal in time. (b) Still in time. (c) VSD in time



(a)

(b)

(c)

Figure 2 Spectrogram of heart sounds. (a) Spectrogram of Normal. (b) Spectrogram of Still. (c) Spectrogram of VSD

Heart sound signals were recorded with electronic stethoscope **3M**™ Littmann® Model 4000 in e4k format. After infrared transfer to a personal computer, signals were transformed to wav format with 3M™ Littmann® Sound Analysis Software. Spectrogram (based on FFT) and time performance of heart signals can be shown by this program.

It can be seen from figures that Still begins after the end of the first heart sound (S1) and stops before the beginning of the second heart sound (S2). A Still’s murmur is audible from the beginning of the systole till after the mid systole. Twenty (20) records of Still’s heart murmur, 20 of VSD and 10 records without murmur were used for time and spectral energy analyzing.

Heart sounds were being recorded with sampling frequency of f_u=8000 Hz and 16 bits of mono-configuration and resolution of quantization. Further, the whole systolic duration was isolated (by hand) from the heart wav files, and recorded by name Still, VSD or Normal, depending on whether it was recording of innocent Still’s murmur, pathologic murmur of VSD or signal without murmur. This optimal sound sample was then processed using digital signal analysis.

III. SOUND ANALYSIS OF STILL’S MURMUR

Pre-recorded cardiac signals of 20 children with Still’s murmur are deduced for frequency of maximal sound energy and frequency bandwidth of Still’s murmur i.e.**,** Still’s tone. For every cardiac signal, elapsed time of every systole is determined, and for that time interval is secluded (manually) for spectral analysis. Still begins to vibrate at the beginning of systole instantly after the end of first heart sound (S1) and stops before beginning of second heart sound (S2). Spectral analysis needs to be done in that interval. In that time interval (around 100 ms) in time frames of 20 ms, i.e., N=160 samples in discrete time spectral energy is calculated.

Tuning coefficients (fixings) of frequency are set in that order that you can get frequency resolution of 5 Hz. Likewise spectral energy on selected time frame of systole is done within lag of 5 ms or 40 samples on discreet time axis. In Fig. 3, the spectral energy is shown in 5 consecutive time frames (scanned of 40 ms of systole). Final position of Still’s time interval of 20 ms (N=160) is chosen on maximal point of its spectral energy. This Still’s murmur was recorded with amplification of 2x (in stethoscope) for better illustration.

In Fig. 3, the spectral energy of third time frame is marked in red which represents maximal energy amount in scanned time. That time frame represents final time position of Stills’s murmur and in that frame frequency and bandwidth of Still’s murmur is calculated. For this Still’s murmur frequency is 130 Hz and bandwidth is around 50 Hz. In final time position Stills’s murmur is musical tone and it has maximal energy.

Statistic analysis of 20 Still’s murmurs resulted that tones of Still’s murmur are situated on bandwidth span of 90-170 Hz, while the average tone frequency is 118,75 Hz and the average bandwidth of Still’s murmur is 40,75 Hz. Frequency of tone is the frequency on which the murmur has maximal energy (top of curve in Fig. 3) and in most cases it is between 110 and 130 Hz. Frequency bandwidth

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-157-1

(B =Fmax-Fmin) is determined in a way that frequency of tone falls in half of strength Fmin (left of the top of curve) and Fmax (right of the top of curve).

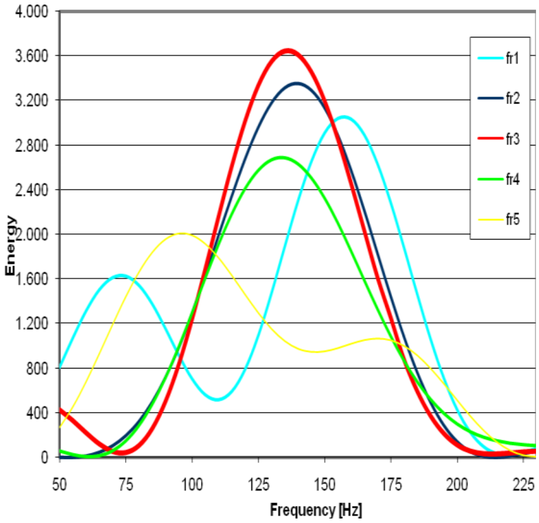


Figure 3. Final position of Still's murmur

If we do not relocate a bit from the first tone S1 or if we analyze time frame at the end of systole we will not get Still's musical tone because of the ending of the first tone S1 i.e., the beginning of the second tone S2.

Likewise Fig. 3 clearly shows that Still's murmur is mainly unsteady signal. Mainly all cardiac signals are unsteady in time span. Likewise spectral energy is mostly concentrated in narrow frequency bandwidth. The conclusion from represented material is that Still's murmurs are real musical tones and Still's murmur is a narrowband tone. That tone oscillates in frequency and amplitude. Still's murmur begins in higher frequency and that frequency deteriorates in time. That change (deteriorating) of frequency is similar to glissando tone played on guitar.

Likewise Still's murmur oscillates on its amplitude during its brief lifespan (crescendo-decrescendo). Therefore Still's murmur begins to play silently and is enhanced as it progresses and it is the loudest in the middle and after that gradually begins to decline and stops. All this occurs in systole. Still's murmur begins at the start of systole, "plays" a bit longer than half of the systole and disappears by the next systole.

IV. 3D GRAPHIC PRESENTATION OF STILL'S MURMUR S1 AND S2

Sound, which transfers heart sound signal, carries many information about heart murmurs. In this study, systolic heart signals were analyzed. Spectral energy of VSD is chaotically distributed in a broad frequency spectrum bandwidth between 80-500 Hz and frequency of the biggest VSDs peak is bigger than 200 Hz. Spectral energy of Still and Normal is regularly distributed in narrow frequency bandwidth under 200 Hz. Still has resonance in frequency bandwidth between 90-170 Hz while Normal has no resonance. Spectral energy below 80 Hz was mainly caused by ambient noise and by late oscillations of first heart sound. At early systole there are two sounds at the same time: the late oscillations of the first heart sound in low frequency about 70 Hz and the beginning of resonance of Still's murmur at higher frequency about 150 Hz [11]. Human ear can't distinct it.

Fig. 4 illustrates the distribution of spectral energies of the first heart sound S1, the second heart sound S2 and Still's murmur in final time positions. Spectral energy was calculated on time frames of 25 ms, i.e., N=200 samples in discrete time. Spectral energy was calculated with distance points of 4 Hz. This signal is amplified by 2x.

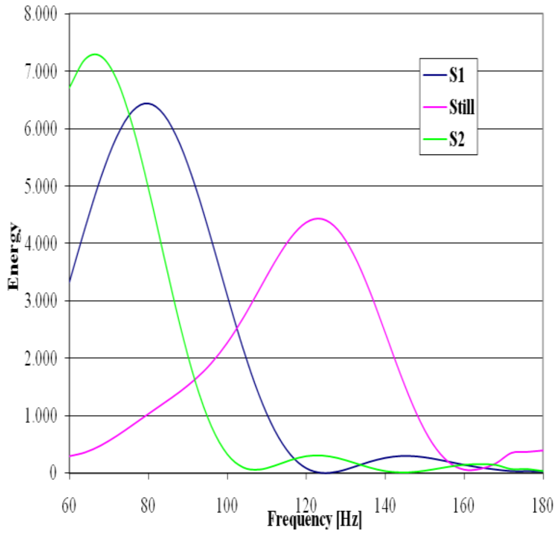


Figure 4. Comparison of the spectrum energy of Still S1 and S2

Fig. 5 and Fig. 6 illustrate 3D graphic presentations of spectral composition of one typical cardiac signal with Still's murmur.

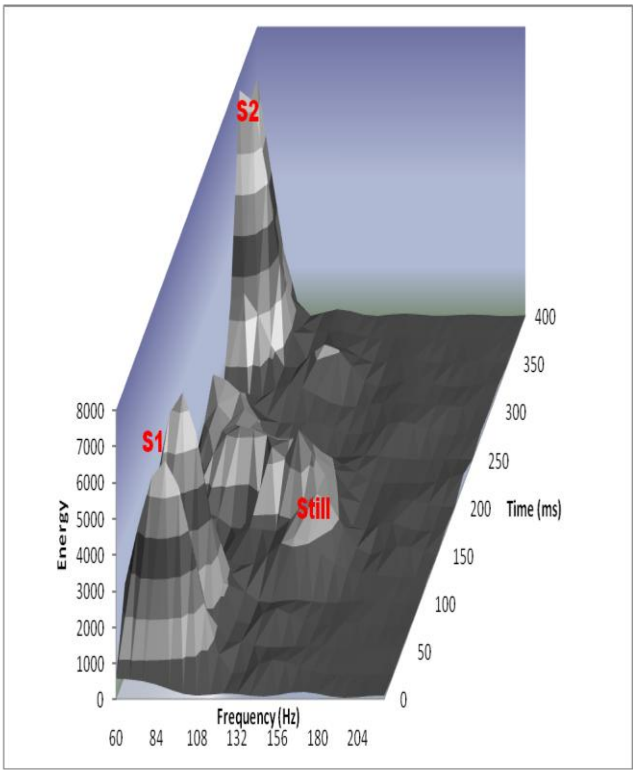


Figure 5. Time – frequency composition of Still's heart murmur

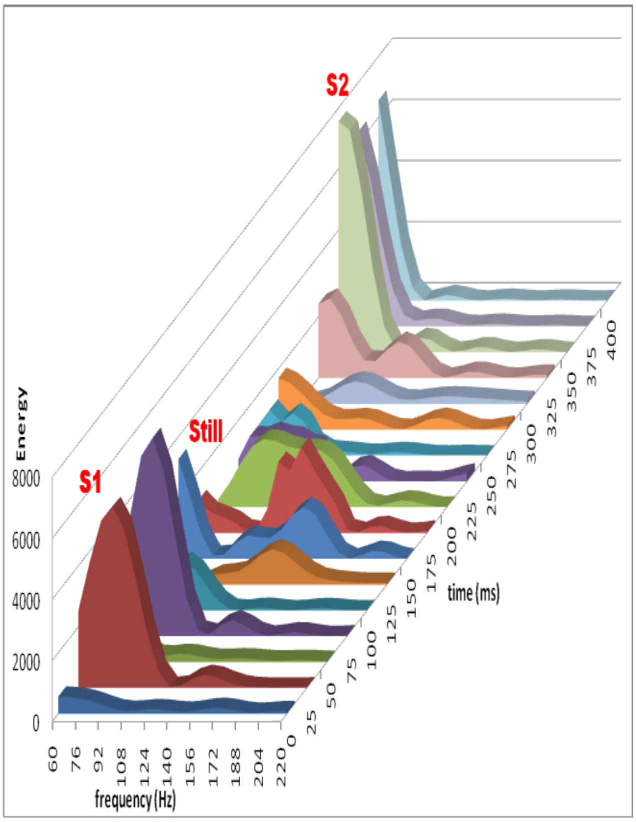


Figure 6. Time – frequency distribution of the spectrum energy of Still's heart murmur

These time-varying frequency graphic presentations show the distribution of spectral energy of heart sounds S1 and S2 as well as Still's murmur. Time frames are 25 ms, i.e., N=200 samples in discrete time and spectral energy was calculated with distance points of 8 Hz. The first heart sound S1 as well as the second heart sound S2 have bigger energy and lower frequency than Still's murmur.

V. CONCLUSION AND FUTURE WORK

Heart sounds are a good example of periodic, yet variable physiological signals [9]. Since heart sounds exhibit marked changes with time and frequency, they are therefore classified as nonstationary signals. To understand the exact feature of such signals, it is thus important, to study their time–frequency characteristics [6].

The differences between the frequency parameters and the spectral compositions of heart signals with innocent and with pathologic murmurs are determined and shown in this paper. For spectral analysis of the cardio signals (recorded by the stethoscope) the Goertzel algorithm was used. It was shown that the frequency parameters as well as the spectral compositions of innocent Still's murmur and pathologic VSD murmur clearly differed. The spectral display of the cardio signals could be diagnostic assistance system for the classification of these murmurs in other words for the health diagnosis of a patient (auscultation-visual diagnosis).

Doctors using auscultation technique for VSD murmur mostly give right diagnosis, but in case of Still's murmur incorrect diagnosis are very often. In order to help doctors in recognition of innocent Still's murmur and of course and other murmurs, it is necessary to create software solutions with graphical displays of the cardio signals spectral energy in real time. Such software would be good tool to doctors at cardiac diagnosis. This paper is one of the first steps in creating such a software package.

ACKNOWLEDGMENT

This paper is funded by the Bosnian-Herzegovinian Federal Ministry of Education and Sciences within the scientific research project in 2010. The title of this project is: Classification of Pathologic and Innocent Heart Murmur Based on Multimedia Presentations of Acoustic Heart Signals.

REFERENCES

[1] B. Tomas and Ž. Rončević, "Spectral Analysis of Heart Murmurs in Children by Goertzel Algorithm," The First International Conference on Creative Content Technologies CONTENT 2009, November 15-20, Athens/Glyfada, Greece, pp. 612-615

[2] F. Javed, P.A. Venkatachalam, and M.H. Ahmed Fadzil, "A Signal Processing Module for the Analysis of Heart Sound and Heart Murmurs," Journal of Physics, Conference Series 34 (2006), pp. 1098-1105

[3] B. Tomas, Ž. Rončević, D. Zelenika, and A. Krtalić, "Pathologic and Inocent Heart Murmurs Classification by Spectral Analysis of Acoustic Cardiosignals," 18th Telecommunications Forum TELFOR 2010, 23-25 November, Belegrade – Serbia (in Serbian), pp. 575-578

[4] G.F. Still, "Common disorders and diseases of chilhood," 1st ed. London: Frowde, Hodder & Stoughton, 1909.

[5] Ž. Rončević, "Is there future for cardiac auscultation in children?," Paediatr Croat 2009; 53 (Supl 1), pp. 27-29 (in Croatian) http://hpps.kbsplit.hr/hpps-2009/pdf/dok06.pdf

[6] S.M. Amin Debbal and Bereksi-Reguig Fethi, "Features for Heartbeat Sound Signal Normal and Pathological," Recent Patents on Computer Science, 2008, vol. 1, no. 1, pp. 1-8

[7] D.A. Balster, D.P. Chan, D.G. Rowland, and H.D. Allen, „Digital Acoustic Analysis of Precordial Innocent Versus Ventricular Septal Defect Murmurs in Children," The American Journal of Cardiology, 1997 Jun 1, vol. 79, no. 11, pp. 1552-1555.

[8] M. Unser and A. Aldroubi, "A Review of Wavalets in Biomedical Applications," Proceedings of the IEEE, vol. 84, no. 4, April 1996, pp. 626-638, doi: 10.1109/5.488704.

[9] G. Amit, N. Gavriely, and N. Intrator, "Cluster Analysis and Classification of Heart Sounds," Biomedical Signal Processing and Control 4 (2009), pp. 26-36

[10] G. Goertzel, "An algorithm for the evaluation of finite trigonometric series," American Mathematics Monthly, vol. 65, (January 1958), pp. 34-35

[11] M. El-Segaier, O. Lilja, S. Lukkarinen, L. Sornmo, R. Sepponen, and E. Pesonen, "Computer-Based Detection and Analysis of Heart Sound and Murmur," Annals of Biomedical Engineering, vol. 33, no. 7, July 2005, pp. 937-942 http://circ.ahajournals.org/cgi/content/full/103/22/2711

Speaker Labeling Using Closed-Captioning

Keita Yamamuro
Hosei University
Tokyo, Japan
Email: keita.yamamuro.xn@stu.hosei.ac.jp

Katunobu Itou
Hosei University
Tokyo, Japan
Email: it@fw.ipsj.or.jp

***Abstract*—There has recently been much research on annotation systems for television broadcasting because of interest in retrieving highlights from television programs. However, most of the methods developed have specialized in only one genre. Therefore, in this study we targeted three genres drama, animation, and variety and developed a system of annotating indexical information through metadata obtained from television captions. Specifically, the information from the captions is used to create a phoneme HMM that is then used for speaker identification. The proposed system selects the most appropriate phonemic model from several candidate models based on the Bayesian information criterion (BIC) of likelihood and data. Characters in 70 television programs were identified with a recognition accuracy of 39.6%. Television captioning can already identify about 50.0% of the characters in a show, and when we combined captioning with the proposed system, 70.0-80.0% of the utterances in one program were correctly identified.**

***Keywords*-Speaker Identification; Model Selection; HMM; Television Broadcasting; Speaker Diarization.**

I. Introduction

Recently, an incredible amount of video content is being broadcast by multi-channel services, and images that viewers demand cannot immediately be obtained. There are a few different kinds of server-type broadcasting services. Most use home servers that can retrieve and store television programs and metadata [9]. Such services access the metadata and retrieve highlights. Scene information in the metadata is crucial for retrieval and editing, but it is extremely time consuming to manually extract it. In response to this problem, several systems to efficiently process metadata have been proposed. These automated systems combine image-recognition, speech-recognition, and natural-language processing technologies [1], [2], [3], [4]. In this study, we propose a system of annotating indexical information through metadata obtained from television captions.

The remainder of this paper will proceed as follows. First, we introduce the background research in Section Ⅱ. Subsequently, we detail methods of speaker identification by model selection in Section Ⅲ. We present an experimental setup and results in Section Ⅳ. We present conclusions and future work remarks in Section Ⅴ.

II. Background research

Many methods of annotating information from television broadcasts have previously been proposed. These methods use various technologies, including image recognition, speech recognition, and natural-language processing, to retrieve television programs, edits content, and enhances speech.

For example, information retrieval interface that uses image processing [5]. The video has been annotated by image recognition. And that information will be searched by keyword.

Methods of analyzing speech in news broadcasts have been particularly researched [6], [7] because news programs contain quite varied information: speaker information is often crucial when trying to pinpoint the relevant information in a large amount of content. Annotation of scene information from sports programs has also been extensively researched [8], and program highlights are selected and edited based on the relevant scene information.

Much of the research described above is targeted at either news or sports programs. Speech from news programs does not contain a great deal of noise, as it is clearly uttered, and speech in sports programs can be easily processed with specific key words like “ goal” or “ shot.” In other words, analyzing these types of shows is not terribly difficult. However, many other types of television programs are more complex and require painstaking annotation. In this study, we targeted drama, animation, and variety shows and developed a method that can identify speaker information within them.

III. Speaker identification

The purpose of this research was to come up with a way to annotate indexical information in television programs. The proposed method can do this by using not only a program’s audio track but also its closed television captions. The annotation is performed by processing metadata obtained from television captions and speaker identification. In general, 50.0% of the utterances in a typical television program are annotated by television captions because this is the percentage that contains implicit speaker information. The proposed method identifies who utters the remaining 50.0%. Title information is used to determine the speaker’s identification. The proposed method constructs HMMs of all phonemes on the basis of the content of the utterance in the caption information. An HMM identifies the speaker in each phoneme of one utterance if the appropriate model has

 ISBN: 978-1-61208-157-1

been selected. Obviously, the identification rate is improved if the best model is selected. We perform this selection using decentralization, likelihood, and the amount of data in the models.

A. Metadata

Metadata contain the title, category and genre of the television program being broadcast. For example, they contain plot outlines, the names of performers and producers, and broadcasting times. Metadata has the advantage of being easily searchable, so there has been much research on the different types of metadata, e.g., that of a soccer program [9]. In this example, the metadata includes the content of the game based on the classification of actual keywords (e.g., “goal” and “kickoff”). There has also been research on classification according to topics in news programming [2]. Such topics are classified depending on the image, speech, and natural-language processing metadata. This research has made it possible for users to retrieve the scenes they want from the metadata. The purpose of that study was to analyze information from utterances in television programs: for example, “who is speaking?”, “when are they speaking?”, and “what are they saying?”

B. Television captions

The Ministry of Internal Affairs and Communications is working to spread the use of captioning in television broadcasting throughout Japan [10]. The present percentage of broadcast captioning is 40.1% of the total broadcasting time. The aim is to boost this percentage up to 69.9%, thus providing wider access to the information it contains.

This captioning information includes the time and content of the captioning displays as well as the kinds and colors of fonts used [11]. Television captions for main characters are in color. This means that some speakers can be identified by the font information. However, at present this type of identification works only about 40-60% of the time because often there is television captioning with no speaker information and no utterances. Although its use is limited, speaker information can be effective at extracting scenes and identifying people. In this study, we classified speech in television broadcasting by using information from the television captioning.

Although the speech of all characters contains speaker information, this is only applies to some speeches. We developed a method to identify speakers using these data.

In the proposed method, information from television captioning is used to create a model for speaker identification. Captions have information about the beginning and end of a speech, and speech that identifies the speaker is extracted from this information. When a speaker is successfully identified, the extracted speech is used for training data, and if it is uncertain who the speaker is, it is used as test data. Several speakers can also be identified by the color of the font and the content of the television captioning. Speech by main characters is in colored font, and because a speaker’s speech is all in the same color, all of his/her speech can be identified. Information on the speaker might be included in the content of television captioning where the name of the speaker is in parentheses. Because other speech by the speaker cannot be identified, this speech is used as training data.

The content of television captioning is useful to construct a model that identifies the speaker. Because the content of the speech is understood, the speech can be analyzed phoneme by phoneme.

C. Phoneme alignment

The proposed method uses phoneme alignment to identify speech from phonemes. The speaker model in the present study was constructed with this phoneme information. We used Julian, which is speech recognition decoder software, for the alignment. Julian is open-source and high-performance software for large vocabulary continuous speech recognition decoders used by speech-related researchers and developers [12]. Julian aligns phoneme units to obtain speech-recognition results. The frames of the phoneme boundaries and the average acoustic scores per frame are calculated. The phoneme alignment in this study gives the verbal information to be identified beforehand according to the television captions. The analytical precision of the phoneme alignment is improved by giving prior information. The results of this analysis were used in this study.

D. Speaker identification model

In this study, we used a hidden Markov model (HMM) as the speaker model. HMM is a probabilistic model. It is used to represent the voice feature amount. Speaker identification is determined by the likelihood of the HMM and the input speech. The speech data from television broadcasting was used to train the HMM. The speaker model of each speaker was trained by using phoneme sections analyzed by alignment. The model in three states handled the HMM of phoneme units. There were 35 kinds of phonemes. The procedure to train the model was as follows.

First, voice activity detection was executed by using television captioning to determine the different sections of phoneme alignment. This detection uses information from the beginning and end of a television caption.

Next, phoneme alignment was used to extract phoneme information to construct the phonemic model. The preprocessing during the phoneme alignment converted the content of television captions into phonemes through morphological analysis. Sections of voice corresponding to phonemes were distinguished by phoneme alignment based on these sounds. Information on each phoneme was used to extract the mel-frequency cepstral coefficient (MFCC) features for either training or test data. MFCCs from known speakers were used to train the HMM while those for uncertain speakers were used as test data.

Finally, speaker features were trained to the speaker model with the training data obtained from the television captions. All characters' names appear at some point in the captions, and these names plus utterance information are used for the initial data training. The amount of training data differs depending on the program, which means that there is always the possibility of insufficient data. We successfully identified various speakers by using this model.

However, alignment could not be used to analyze all speech. For instance, sometimes noise overlapped with voices, or there was a discrepancy between the captions and the speech. Because the alignment failed with this analysis, the phoneme unit model could not be used to accurately identify all speakers. In cases where the alignment failed, we were able to identify speech with a Gaussian mixture model (GMM) of one state and 32 mixtures. This model was trained with the average features of all phonemes.

We used the hidden Markov model toolkit (HTK) for feature extraction and for the HMM and GMM training [13]. HTK, a software toolkit for building speech recognition systems using continuous density hidden Markov models, is what the proposed method uses to analyze speech data. We also used it to construct speaker models. The constructed model identifies the speaker by HTK. The identified speech data is television captions that cannot specify the speaker. This data makes up 40.0-60.0% of all utterances. The proposed method identifies this data with HTK and then labels it.

E. Mel-frequency cepstral coefficient (MFCC)

There are some kinds of speech feature. In the present study, the MFCC is used as a feature for the speaker identification. The MFCC is analyzed by a filter bank on the Mel frequency axis. And, spectral analysis of the output from multiple filters arranged along a frequency axis is carried out. The MFCC is provided by discrete cosine transform (DCT) of the power in each band obtained in the result. The human ear is fine-tuned for hearing low-frequency sounds but not high-frequency sounds. The feature of MFCC is the same as the feature. Even if the same phoneme is analyzed, MFCC shows the feature that varies from person to person.

F. Effective model selection

Although the speaker model could identify speakers from the test data MFCCs, the amount of phoneme data used for training varied depending on the phoneme. This means there might not be enough data to learn all the phonemes in the trained speaker model.

Recognition model selection [14] and noise model selection [15] have been proposed as model selection methods. In this study, we used BIC, which is an information criterion for the performance of a probabilistic model corresponding to the amount of training data. There are other information criteria similar to BIC that have previously been proposed [16]. In many cases, the performance of a model is measured with these standards and then the most appropriate model is selected. In this study, we used 35 phoneme units as candidates and the top five were used for speaker identification. Even if there was not much training data, the proposed system could identify the speaker according to the model that was selected.

G. Assessment of identification results

We used the proposed method to select five models for the speech identification. However, the identification results were not all the same: different identification results were output by all models in cases where the system could not specify who the speaker was. In this study, we required just one identification result, which we obtained by comparing the likelihoods of all speakers being identified.

First, the identification results of one phoneme were output corresponding to the number of candidates. All likelihoods were output at the same time. Similar processing was executed for one speech. Next, the output likelihoods of all candidates were totaled. The candidate with the highest total was assumed to be a correct identification result.

IV. Experimental setup

A. Description of experiment

We performed experiments on the speaker identification system to evaluate the accuracy of the annotation used for scene retrieval. We prepared assessment data and performed speaker identification with the constructed model. In this study, we annotated information from recorded television broadcasts. We did not have much data to begin with, so we gradually added more as we went through other broadcasts. The proposed method was evaluated by comparing its performance with that of a model already known to be effective. We deemed models effective if they could identify all the speech in a particular television program. The evaluation data were extracted from drama, animation, and variety programs broadcast in Japan. In total, we used 70 programs: 20 dramas, 30 animations, and 20 varieties. Shows that were 30 minutes long averaged about 400-600 sentences, and shows that were 60 minutes long averaged about 900-1100. About 40-60 % of the speech was identified by television captioning and the remainder by the proposed method. We stipulated that speech used for training must consist of at least one sentence. Table I summarizes the analytical conditions for the speech. Speaker identification was performed by the selected model and the method using all the phonemic models was evaluated at the same time for the sake of comparison. BIC identified the trained model as the best. If alignment failed, speech was identified by the GMM described earlier.

We used different episodes from the same television programs to obtain additional character data. For example, if characters from a drama in a second episode appeared

 ISBN: 978-1-61208-157-1

Table I
ANALYSIS CONDITIONS

Number of television programs	70
Number of identified people (Average from one television program)	10
Time television program lasted	30 min, 1 or 2 hr
Sampling frequency	16 kHz
Quantization bit rate	16 bits
Frame period	10 ms
Frame length	25 ms
Feature amount	MFCC (1-12) logarithm power (1) +Δ (total 26 dimensions)

in the first episode, data from the first one were added to the second one as training data. The character data therefore increases if characters appear several times. This technique allowed us to double our training data.

B. Experimental results

Figure 1 shows the experimental results obtained by adding data. Additional data improved the average identification accuracy by 2.84% for dramas, 3.41% for animation, and 1.16% for variety. This clearly demonstrates that using speech data from another episode of a program as training data improves the identification accuracy.

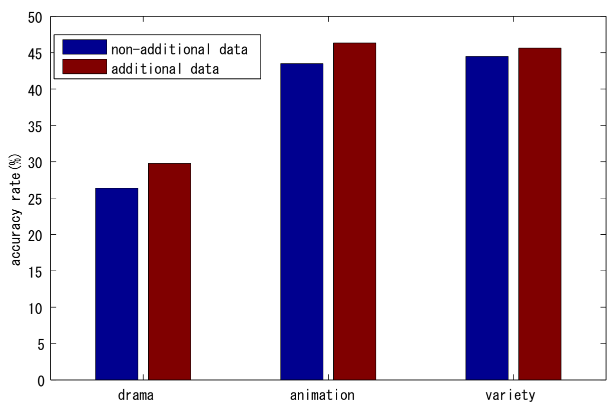


Figure 1. Results from additional data

Next, Fig. 2 shows the experimental results for model selection. Model selection improved the average identification accuracy by 11.75% for dramas, 15.15% for animation, and 15.85% for variety. 44.48% of characters were accurately identified in variety shows, which was the strongest result in this experiment. However, other studies have shown that 80.0-90.0% of speakers in news programs can be identified [17]. It is more difficult to identify speakers in dramas, animation, and variety shows than in news programs because in news programs there is less background music (BGM), fewer speakers, and clearer speech. In our experiment, we could not identify speech very well when there was an overlap of BGM and sound effects (SEs). We need to adapt the proposed method so that it can deal with such problems.

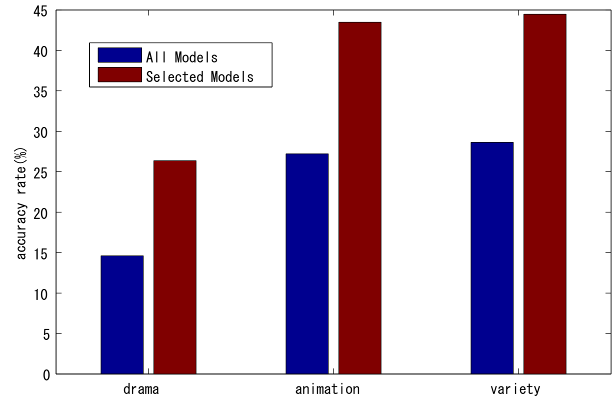


Figure 2. Results from model selection

C. Discussion

The identification accuracy improved with additional data. The identification models had different performances before and after the data were added. For example, some of them could not identify speaker features at the beginning, when there was not sufficient data, but then dramatically improved when the amount of data was increased. Take the case of a model that could accurately identify vowels such as /a/ and /i/. In this case, the model was well selected because these vowels can be used to identify speaker features when there is not much data. However, when the amount of data increased, the identification model had more freedom to use data other than vowels (e.g., consonants like /k/ and /t/). The additional data gave the model more choices, which led to improved identification accuracy.

Accurate model selection also increased the percentage of speakers who were identified. As stated previously, the proposed method selects the most effective model from among several candidate models. However, sometimes selection was difficult because each phonemic model's performance changed depending on the speaker if there was not much training data. It is therefore important that the model used for speaker identification considers all the phonemic models.

The proposed method did not deal well with overlapping BGM and SE, probably because speaker features might have been cancelled due to BGM. The identification accuracy might improve if speech is enhanced or noise is rejected.

The proposed method also selects a second-best model, so if the first model fails to identify a speech, this second one might be able to. Comparisons between the first and second models showed that there was usually not much difference in terms of identification likelihood. Therefore, identification results were requested from both models when there was

 ISBN: 978-1-61208-157-1

little difference. As a result, the recognition accuracy in all television programs was improved, in some cases by as much as about 10.0%. This demonstrates the importance of considering the difference in likelihoods.

V. Conclusions and Future Work

The speakers in 70 television programs were identified through television captioning information and model selection in the present study. The 44.48% correct identification rate for variety shows was the strongest result. This identification improved by 17.10% due to model selection. Overall, the proposed method was effective, but it was deficient when compared with the conventional one in terms of identifying speech in news programs.

We need to re-think how the model selection is performed. In this study, we selected models by using BIC, but because the best model changes depending on the speaker, the speaker identification should really use all the models. Ideally, the model selection should analyzes the model performances individually for each speaker and then attach the appropriate weight to the result.

Additionally, in the future we intend to combine the proposed method with others. We need to further examine speech enhancement and noise rejection and determine a way to remove overlapping noise. We also intend to study image recognition. Many researchers have added metadata to television broadcasting and combined two or more processes [2], [18]: for example, image enhancement combined with natural language processing. Previous studies have shown that image recognition used on characters' faces is quite effective. When television characters speak, their image is usually shown on the screen. Therefore, if characters in a scene can be identified by image recognition, it would make it easier to narrow down the target person in speaker identification.

References

[1] Photina Jaeyun Jang and A. G. Hauptmann, "Improving acoustic models with captioned multimedia speech", ICMCS-99 International Conference on Multimedia Computer Systems, pp. 767-771, 1999.

[2] H. Kuwano, Y. Matsuo, and K. Kawazoe, "Effects of Task-Cost Reduction on Metadata Generation Using Audio/Visual Indexing and Natural-Language Processing Techniques", The Institute of Image Information and Television Engineers, Vol. 61, pp. 842-852, 2007, (in Japanese).

[3] S. Meignier, D. Moraru, C. Fredouille, J. Bonastre, and L. Besacier, "Step-bystep and integrated approaches in broadcast news speaker diarization", Computer Speech and Language, Vol. 20, pp. 303-330, 2006.

[4] I. Yamada, M. Sano, H. Sumiyoshi, M. Shibata, and N. Yagi "Automatic generation of metadata for football games using announcer's commentary", IEIC Technical Report, VOL. 104, pp. 37-42, 2005, (in Japanese).

[5] A. Gordon, "Using annotated video as an information retrieval interface", International conference on intelligent user interface, New Orleans, LA, pp. 133-140, 2000.

[6] A. Messina, R. Borgotallo, G. Dimino, D. Airola Gnota, and L. Boch, "A Complete System for Automatic News Programme Annotation Based on Multimodal Analysis", Image Analysis for Multimedia Interactive Services, pp. 219-222, 2008.

[7] Chuck Wooters, "The ICSI RT07s Speaker Diarization System", Proceedings of the Second International Workshop on Classification of Events, Activities, and Relationships, Vol. 46, pp. 509-519, 2007.

[8] M. H. Kolekar, K. Palaniappan, and S. Sengupta, "A Novel Framework for Semantic Annotation of Soccer Sports Video Sequences", IET 5th European Conference on Visual Media Production, pp. 1-9, 2008.

[9] A. Baba, Y. Nishimoto, K. Ishikawa, H. Nakamura, T. Yoshimura, and T. Kurioka, "A Study on Metadata Technology for Broadcasting System based on Home Servers", IEIC Technical Report, VOL. 104, pp. 11-16, 2004, (in Japanese).

[10] "Results of broadcast captioning" Ministry of Internal Affairs and Communications, pp. 1-3, 2009, (in Japanese).

[11] R. Turetsky and N. Dimitrova, "Screenplay alignment for closed-system speaker identification and analysis of feature films", Proc. ICME, Vol. 3, pp. 1659-1662, 2004.

[12] T. Kawahara, A. Lee, K. Takeda, K. Itou, and K. Shikano, "Recent Progress of Open-Source LVCSR Engine Julius and Japanese Model Repository", Proc. ICSLP, Vol. 4, pp. 3069-3072, 2004.

[13] P. C. Woodland, J. J. Odell, V. Valtchev, and S. J. Young, "Large vocabulary continuous speech recognition using HTK", Proc. ICASSP, Vol. 2, pp. 125-128, 1994.

[14] M. Nishida and T. Kawahara, "Unsupervised Speaker Indexing Using Speaker Model Selection Based on Bayesian Information Criterion", IEIC Technical Report, vol. J87-D-II(2), pp. 504-512, 2004, (in Japanese).

[15] Z. Zhipeng and S. Furui, "Noisy Speech Recognition Based on Robust End-point Detection and Model Adaptation", Proc. ICASSP, Vol. 1, pp. 441-444, 2005.

[16] Shiro Ikeda, "Construction of Phone HMM Using Model Search Method", IEIC Technical Report, Vol. J78-D-2(1), pp. 10-18, 1995, (in Japanese).

[17] D. A. Reynolds and P. Torres-Carrasquillo, "Approaches and Applications of Audio Diarization", Proc. ICASSP, Vol. 5, pp. 18-23, 2005.

[18] K. Sekiguchi, M. Kosugi, and N. Mukai, "Automatic Indexing of the Baseball Game by Using Video and Audio Information", The Institute of Image Information and Television Engineers, Vol. 106, pp. 41-46, 2006, (in Japanese).

[19] T. Kosaka, T. Akatsu, and M. Katoh, "Speaker Vector-Based Speaker Identification with Phonetic Modeling", IEIC Technical Report, Vol. J90-D No. 12, pp. 3201-3209, 2007, (in Japanese).

 ISBN: 978-1-61208-157-1

Effect of Contrast on the Quality of 3D Visual Perception

Mahsa T. Pourazad
TELUS Communications Company, Canada
University of British Columbia, Canada
pourazad@ece.ubc.ca

Zicong Mai, Panos Nasiopoulos
University of British Columbia, Canada
{zicongm, panos}@ece.ubc.ca

***Abstract*—There are several different factors that affect the perceived quality of 3D content. Our objective in this paper is to study how the change of contrast between the objects of interest and the background in a scene will affect the overall 3D visual perception. For this study, we captured outdoor and indoor scenes in 3D format. For each scene, the brightness of the background was consistent and the object's brightness was changing using an external light source and/or a reflector disk. Subjective evaluations were performed, with the subjects being asked to rate the 3D perceptual quality of each sequence. The results showed that the Weber contrast between objects of interest and background should be within the range of -0.35 to 0.55 to provide viewers with high quality of 3D experience.**

Keywords-3D TV; quality of experience; 3D perception; contrast

I. INTRODUCTION

Recently, 3D video has received increased attention among investors, researchers and technology developers. The introduction of 3D TV can only be a lasting success if the perceived image quality provides a significant step up from conventional 2D television, while maintaining the same viewing comfort. The availability of high quality 3D content will also be a key factor to this success. Recoding 3D content – let alone high quality - is much more demanding and challenging than that of its 2D counterpart in terms of camera setup and configuration, requiring both the director and camera operators to be experts in stereoscopic geometry and camera calibration [1]. In general, 3D content production needs different considerations and provisions beside the ones found in the conventional 2D video production. There are many factors and parameters that could affect the perceptual quality of 3D media. While the effects of different acquisition parameters on the 3D perception have been studied before, their influence on the perceived quality has not been assessed quantitatively. More research and studies are required in order to improve our understanding of the different factors that affect a viewer's perception of 3D video content. This knowledge will allow us to capture high quality 3D content that may help reduce or even eliminate the visual discomfort of the viewers and thus improve the overall quality of experience. To this end, the study by Goldmann et al. has addressed the effect that the distance between stereo cameras has on the perceptual quality of the captured videos [2]. The work presented by Xu et al. [3] investigated the relation between the distance of the object(s) of interest from the camera and the quality of the perceived images when watched on different size displays. Another important factor that affects the visual quality of 3D content is brightness [4]. The study by Pourazad et al. [5] investigates the effect of the scene's brightness on perceptual 3D quality.

Inspired by the above work, in this paper we study the effect of the contrast between the object(s) of interest and the background on the visual quality of 3D content. In our study, we capture outdoor and indoor scenes with different contrast levels between the main object(s) and the background. Then we perform extensive subjective quality assessment experiments to quantify the perceived quality of the 3D experience at different levels of contrast. The objective is to identify if there is a contrast range that will lead to good 3D representation.

The rest of the paper is structured as follows. Section II elaborates on our experimental setup. Subjective evaluations are presented in Section III. Section IV elaborates on our experimental results. Conclusions are drawn in Section V.

II. EXPERIMENTAL SETUP

In our experiment, we aim at investigating the effect of changing the object's contrast with respect to the background on the perceived quality of captured 3D videos. For this comparison, 3D videos of indoor and outdoor scenes are captured using stereo cameras. For each scene the brightness of the object(s) changes from an under-exposed to an over-exposed level, while the brightness of the background is adjusted to a normally exposed level (not over/under exposed) and is kept relatively unchanged for all the different recordings of the same scene. As a result, the recorded videos have different contrast between the object(s) of interest and the background. To capture such test video sequences we use two identical full HD cameras (Sony HDR-XR500V 1080 60i NTSC) with baseline distance of 9cm. We used the same settings on both cameras, which were aligned in parallel and attached to a bar that was custom-made for this purpose. Subsequently, the bar was secured to a tripod. Since zoom lenses may differ [6], only the extreme ends of the zoom range were used to prevent unsynchronized zooming. For the indoor scenes, the brightness of the object(s) was changed by using a dimmable 1000W fluorescent video light source (FloLight FL-220AW). For the outdoor scenes, since the emitted light from the light source was insufficient for changing the brightness of the object(s) (due to the presence of sunlight), we used a collapsible circular reflector disc with multiple impacts to reflect different levels of sunlight on the object(s).

 ISBN: 978-1-61208-157-1

Fig. 1 shows our camera setup, the light source and the reflector used in our experiments. In general, capturing outdoor scenes was much more challenging compared to indoor scenes, due to the presence of sunlight and the change of weather conditions which kept altering the background brightness.

Figure 1. Stereo camera setup used in the experiments.

In order to calculate the contrast between the object(s) and the background first we measured the luminance of the object(s) and background using a multifunction light meter (Sekonic L-758Cine). Then, we employed Weber contrast definition [7] as:

$$\text{Contrast} = \frac{L_o - L_b}{L_b} \qquad (1)$$

where L_o is the luminance of an object and L_b is the average luminance of the background. We measure luminance since it indicates how much luminous power is perceived by the human eye when viewing the surface from a particular angle.

III. Subjective Evaluation

For this experiment, six stereoscopic test sequences (two outdoor and four indoor) and one demo video were captured using the stereo camera setup described in Section II. Fig. 2 shows a snapshot of our test sequences. For each scene the camera exposure is adjusted such that the background area is neither overexposed nor underexposed. Then, the brightness of the object(s) is changed from an underexposed level to an overexposed level within multiple steps, with the brightness-change remaining visually differentiable (see Fig. 3). In both cases, outdoor and indoor, each sequence is approximately 10 seconds long. For each scene recording, we ensure that while the object's brightness changes, the content of the scene and the background luminance remain unchanged.

To quantify the perceived quality of the 3D experience at different levels of contrast, we performed subjective quality assessment tests. The viewing conditions of our subjective test were set according to the ITU-R Recommendation BT.500-11 [8]. Eighteen observers participated in our subjective tests: seven females and eleven males, ranging from 23 to 60 years old. All subjects had none to marginal 3D image and video viewing experience. A 65" Full HD 3D display (©Panasonic, Plasma, TC-P65VT25) was used in our experiment. Based on our own subjective tests of many 3DTV sets, the above display offers the best crosstalk reduction performance and that is the reason it was chosen for our tests.

At the beginning of the experiment, a demo was played starting from a very dark object-exposure to a very bright one to help viewers become familiar with the test process and show them the quality-change range expected. After the demo, the viewers were shown each stereoscopic test sequence in random order of object-exposure levels. Between stereo videos of different object-exposures, there were three-second gray intervals that allowed the viewers to grade the perceptual quality of 3D content from 1 to 10 (continuous scale) and relax their eyes before watching the next video. The perceptual quality reflects whether the displayed scene looks pleasant in general. In particular, subjects were asked to rate a combination of "naturalness", "depth impression" and "comfort" as suggested by Hyunh-Thu et al. [9].

IV. Results and Analysis

After collecting the experimental results, we checked for the outliers based on the TU-R Recommendation BT.500-11 [8] and then the mean opinion scores (MOS) from viewers were calculated. Fig. 4 shows the average perceptual 3D quality (MOS) versus brightness of the object(s) for all six stereo sequences. As it can be observed, the acceptable brightness level for objects in outdoor scenes is much higher than those in indoor scenes, due to the presence of sunlight. Here, the numerical value of object(s) brightness could not be used as a guideline for capturing high quality 3D content, i.e., we can not conclude if the brightness level of the object(s) falls in a certain range (where MOS is greater than 6) then the subjective quality of 3D picture will be acceptable. The reason is that what viewers see and evaluate (perceived 3D visual quality) is not really measured in the real world but rather what has been captured by the cameras and displayed on the 3DTV. In other words, the final brightness level of the objects has been influenced by the exposure settings of the cameras as well as the setting parameters (and limitations) of the 3D display system itself. A way of quantifying how the final brightness level affects 3D perception would take into consideration the luminance

Figure 2. Snap shot of captured indoor and outdoor test sequences.

in a scene as well as the camera and display models. This study is part of our future work.

In order to investigate the effect of the contrast on 3D perception, the contrast between the object(s) and the background is calculated based on equation (1) as the difference between the average luminance of the background and that of the object of interest. Fig. 5 shows the average subjective scores for quality of 3D content versus contrast for all six sequences.

A general observation that applies to both outdoor and indoor scenes is that the stereo video sequences with Weber contrast levels of -0.35 to 0.55 between the object and background are more appealing to the viewers (these correspond to rating scores above 6, which may be regarded as acceptable quality). Note that although the visually acceptable range of object’s brightness (MOS over 6) is higher for the outdoor scenes compared to that of the indoor scenes, as shown in Fig. 4, the range of contrast that ensures high 3D quality is similar for both cases. It is also observed that low scores are associated with high contrast scenes, which in Fig. 5 appear at both ends of the horizontal axis, as contrast here is the difference between the objects’ brightness and that of the background. It is well known that crosstalk artifacts in 3D displays become severe when the contrast is high.

Figure 3. Same scene with different brightness levels of the object.

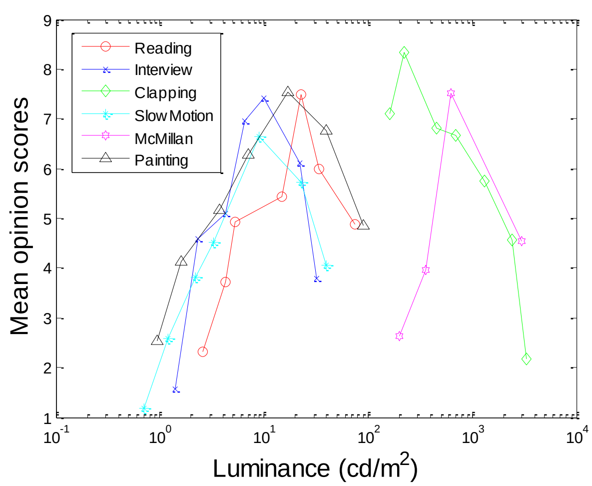


Figure 4. Snap shot of captured indoor and outdoor test sequences.

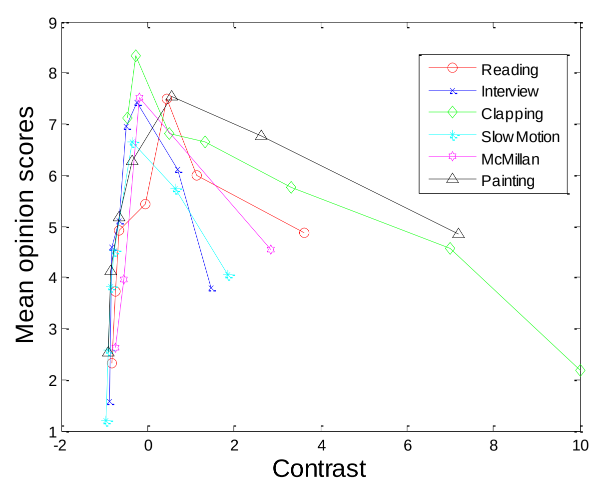


Figure 5. Perceptual 3D quality score versus the contrast between object and background.

 ISBN: 978-1-61208-157-1

In summary, this study indicates that content producers may improve the overall 3D quality of experience by adjusting the brightness of the objects to ensure that the Weber contrast between the background and the objects of interest falls between -0.35 to o.55.

V. CONCLUSION

The era of user-centric multimedia has already begun, and quality plays a central role in it. Attention to the quality of 3D content is even more important since low-quality 3D videos can produce eyestrain, headache, and generally unpleasant viewing experience for the viewers. Contrast is one of the important factors that affect the visual comfort and quality of 3D videos. In this study we addressed the problem of understanding the effect of contrast on the 3D quality by performing extensive subjective quality assessment experiments to quantify the perceived quality of the 3D experience at different levels of contrast.

According to our results, a general observation that applies to outdoor and indoor scenes is that the stereo video sequences with Weber contrast levels of -0.35 to 0.55 between the object(s) and the background are more appealing to the viewers. In summary, content producers may improve the overall 3D quality of experience by adjusting the brightness of the objects of interest in a scene to ensure that the Weber contrast the objects and background falls within the suggested range levels.

ACKNOWLEDGMENT

The authors thank the eighteen anonymous subjects that participated in our test.

REFERENCES

[1] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntiëns, “A survey of perceptual evaluations and requirements of three-dimensional TV,” IEEE Trans. Circuits & System for video technology, vol. 14, no. 3, pp. 381-391, 2004.

[2] L. Goldmann, F. D. Simone, and T. Ebrahimi, "Impact of acquisition distortion on the quality of stereoscopic images," in International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, Arizona, USA, January 13-15, 2010.

[3] D. Xu, L. Coria, and P. Nasiopoulos "Guidelines for capturing high quality stereoscopic content based on a systematic subjective evaluation," in IEEE International Conference on Electronics, Circuits, and Systems, ICECS 2010, pp. 166-169, Athens, Greece, 2010.

[4] B. Mendiburu, “3D Movie Making – Stereoscopic Digital Cinema from Script to Screen.” Elsevier, 2008.

[5] M. T. Pourazad, Z. Mai, P. Nasiopoulos, K. Plataniotis, and R.K. Ward, “Effect of brightness on the quality of visual 3D perception,” ICIP 2011, in press.

[6] Adobe. (2010, July) 3D stereoscopic editing with premiere pro CS5 Win/OSX. [Online]. http://tv.adobe.com/watch/davtechtable/3d-stereoscopic-editing-with-premierepro-cs5-winosx/. Access date: July 2011.

[7] R. Shapley, and C. Enroth-Cugell, “Visual adaptation and retinal gain controls,” Progress in Retinal Research, Vol. 3, pp. 263-346, 1984.

[8] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU-R, Tech. Rep. BT.500-11, 2002.

[9] Q. Hyunh-Thu, P. L. Callet, and M. Barkowsky, "Video quality assessment: from 2D to 3D challenges and future trends," Proc. of 2010 IEEE 17th International Conference on Image Processing, (ICIP), pp. 4025-4028, 2010.

SPS: A Web Content Search System Utilizing Semantic Processing

Joseph Leone
Dept. of Computer Science and Engineering
University of Connecticut
Storrs, CT 06269-3155 USA
Joseph.2.Leone@uconn.edu

Dong-Guk Shin
Dept. of Computer Science and Engineering
University of Connecticut
Storrs, CT 06269-3155 USA
shin@engr.uconn.edu

Abstract—**This paper describes a Web content search system that employs semantic processing. The system, called SPS (semantic processing system), consists of a crowd-sourced ontology, a component for updating and extending the ontology, an NL parser, a semantic matcher, and a content representation formalism called semantic processing (SP) logical form. A typical web search results in hundreds of pages. The user then carries out the tedious and daunting task of sifting through each page to find the relevant/interesting information. SPS aims to improve the relevance by building a layer of automated filtering on top of conventional search engines. SPS takes a user's natural language query, composes it into a keyword query, augments the keyword query with additional keywords, and presents it to the search engine. The query when augmented with additional keywords produces a richer search result set. SPS sifts through each search result page extracting grammatical and semantic information to compute page relevance. We present the architectural framework for SPS and also illustrate how it uses semantic processing to improve the quality of search results.**

Keywords-Web content mining; semantic processing; dynamic ontology development; collaboration system; information retrieval; biomedical literature mining

I. INTRODUCTION

One of the major problems with the Internet is its inability to find quality information with a higher precision. Web search engines produce either a list of too many items or a list of too few, with most of the items not relevant to users' interest/query. In many cases, the search outcome does include relevant items but currently, web surfers have the tedious and daunting task of sifting through web search result pages to find the relevant/interesting information.

The current manner of web searching can be divided into two phases: the "look" phase and the "find" phase. In the "look" phase a user presents keywords to the search engine and the search engine returns a set of pages the engine considers relevant to the user. In the "find" phase the user sifts through the search engine results to find the *actual* relevant/interesting information. We say *actual* because search engines either may not return any relevant information at all or the relevant information is buried somewhere in the collection of returned pages.

An examination of how look phase functions reveals why relevant information is usually buried. Look phase proceeds as follows. Using the user supplied keywords, a search engine retrieves pages that contain those keywords. The search engine then applies ad-hoc heuristics and machine learning techniques to the retrieved pages to compute their relevance. The heuristics that a search engine employs are word position, and utilization of HTML markup; Google uses the PageRank [1] algorithm. In Google's PageRank algorithm, linking determines relevance. The more links point to a particular page the higher Google believes the page to be relevant. If too few links point to a page, it will check whether the links are from pages that are deemed high quality (e.g., from universities, government offices, hospitals, etc.). The typical machine learning techniques search engines employ are word frequency, information gain, odds ratio, and Bayesian analysis on the text words.

Using heuristics and machine learning is helpful in computing page relevance. However, the relevance that search engines compute is generally not accurate because both heuristics and statistics-based machine learning techniques are unable to deal with

1. polysemy at the phrase level (e.g. "juvenile victims of crime" vs. "victims of juvenile crime")
2. synonymy at the word level, i.e. different words having almost the same meaning ("throttle" & "accelerator"; "road" & "street")
3. same words having different meanings ("soap bar" & "singles' bar")

In addition, heuristics and machine learning sometimes produce results for purely statistical reasons with no real "semantic" relevance to the user's query.

Given the innate ambiguity of expressing a query via keywords, the look phase provides less improvement potential. In contrast, the find phase has much greater potential for improvement/automation because it is presently expected to be carried out manually by a person. We argue that if Web search incorporates even partial natural language capabilities that could extract grammatical and semantic information from both the user's query and from visited pages, relevance could be computed more precisely and the quality of the search results would be greatly improved.

This paper proposes a semantic processing system (SPS) that improves Web search by increasing keyword quality during the look phase and automating the find phase. The intent of our approach is not to replace current search engines (e.g. Google), but to work in conjunction with them. The SPS is layered between the search engine and the human user.

 ISBN: 978-1-61208-157-1

Section II discusses related work. Section III details SPS, its components, and SPS logical form (or simply, *SP form)* which is the internal knowledge representation formalism used by SPS. Section IV shows an example of SPS processing a query against the biological literature. In the biological literature genomic structures, proteins, and other phenonena are generally described using natural language. We illustrate using examples why SPS outperforms traditional keyword search.

II. Related Work

The Web search community has been exploring use of semantic processing to improve the relevance of query results. Sieg et al. [16] proposed a semantic approach utilizing ontology-based user profiles to personalize the Web search. In their work, each user's search interests and preferences are modeled into ontological profiles. Unfortunately, this group's proposal to use the ontology is limited to organizing the user's context rather than modeling the general world knowledge. In utilizing the ontology to compute the relevance, this group uses conventional statistics methods. Another proposal for the use of semantic processing is SPARK by Zhou et al. [17]. Given a keyword query, SPARK outputs a ranked list of expanded queries which are obtained in three steps, term mapping, query graph construction and query ranking. The authors emphasize the novelty of query ranking, but this group's ontology construction is rather ad hoc and is far from the general knowledge representation frameworks originating from natural language process. Most recently, Shabanzadeh et al. [18] proposed expanding query using semantic relations. In this work, semantic relations are extracted from a lexical database called WordNet where semantic relations are basically limited to hypernymy/hyponymy (is-a relation)

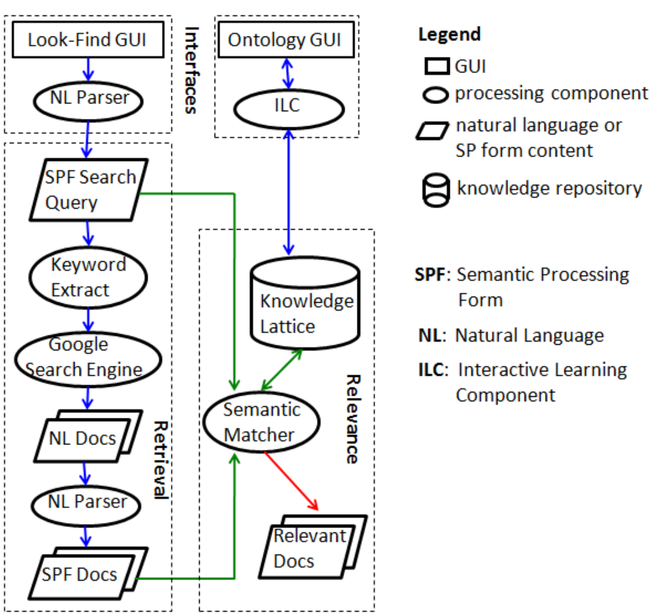


Figure 1. Semantic Processing System (SPS) architecture

and meronymy/holonym (part-of relation). As such, this group does not exploit the wealth of techniques available from the natural language processing either.

III. System Architecture

In comparison with previous attempts, our proposed approach is more comprehensive by encompassing phrase parsing, knowledge-representation, query formation, ontology construction, and semantic matching. Figure 1 shows the three main parts (interface, retrieval subsystem, relevance computation subsystem) of the SPS architecture and their internal components:

SP form: A knowledge representation formalism used by SPS.

Canonical forms. Internal (not shown in Figure 1) lexical templates that specify every possible mapping from natural language to SP forms. Each canonical form represents a semantic pattern that occurs in English.

Parser for converting natural language to SP form.

Keyword Extract for creating a keyword query. The number of keywords in the search engine query may be greater than the number of words in the natural language query. The extra keywords account for synonymy and subtype/supertype.

Knowledge Lattice. A data structure for representing words, their subtype / supertype relationships, and their synonyms. Included in the data structure is a set of operations for reasoning about the relations between words.

Interactive Learning Component for updating and extending the Knowledge Lattice.

Semantic Matcher for computing retrieved pages' relevance.

SPS assists in two ways. During the look phase SPS accepts a natural language query, extracts from the query significant concepts, composes the concepts into a keyword search query that accounts for polysemy at the phrase level and synonymy at the word level, and presents the search query to a global search engine (e.g. Google). During the find phase, for each search engine result page, SPS semantically computes (using grammatical and semantic information extracted from retrieved pages and search query) the relevance of the result page to the user's natural language input query. This section is devoted to discussing five key computational aspects of SPS architecture.

A. Computable Representation

Every SPS architecture component, except NL Parser, is implemented in Lisp; and the core technology that underpins SPS, called *SP form*, is a computable internal knowledge representation expressed in Lisp notation. SP form is inspired by the semantics and logical foundation of Conceptual Graph (CG) [2,3]. However, an SP form looks very different from *conceptual graph interchange form* (CGIF) and only superficially resembles CG *linear form*. Moreover, many elements from CG's logical foundation (e.g., schematic cluster, prototypes, context, type definitions, aggregation, etc.) are not used. Nonetheless, some CG nomenclature (e.g., Knowledge Lattice, graph, subgraph,

 ISBN: 978-1-61208-157-1

projection) is retained where there is great similarity at the abstract level between CG elements and SPS elements.

1) Canonical forms (i.e. semantic patterns) In their analysis of the English language Quirk et al. [4] determined that sentences are composed of clauses that in turn are made up of syntactic and semantic elements. Syntactic elements, i.e., subject, verb, object, complement, adverb, etc. are participants in the meaning of a clause. Semantic elements, i.e., agent, instrument, affected, etc. are the roles participants play [3,5]. Quirk identified thirty-three clause patterns, which account for all active English sentences.

For each of Quirk's clause patterns Leone [6,7] developed a *canonical form*. A canonical form is a conceptual graph lexical structure capable of representing the semantics of an active English sentence. Canonical forms are derived by mapping clause participants to concepts, and clause semantic roles to conceptual relations. A canonical form conveys information about the possible semantic roles of the participating syntactic constituents, provides predicates for representing each semantic feature, functions as a semantic pattern or template for capturing a particular class of meaning, and serves as a guide for mapping language (i.e. parser output) to logic (i.e. SP form). For each clause pattern, there is one and only one canonical form.

Below are some sample English language phrases, the phrase clause pattern, a canonical form corresponding to the clause pattern, and the phrase expressed in conceptual graph linear form. A complete set of clause patterns and canonical forms can be found in [6,7].

I bought her a gift.	S V O_i O_d
[c1] ← (agent) ← [c2] → (obj) → [c3] → (rec) → [c4]	
[I] ← (agent) ← [buy] → (obj) → [gift] → (rec) → [her]	

He put it on the shelf.	S V O_d A_{place}
[c1] ← (agent) ← [c2] → (obj) → [c3] → (loc) → [c4]	
[He] ← (agent) ← [put] → (obj) → [it] → (loc) → [shelf]	

He caught the ball.	S V O_d
[c1] ← (agent) ← [c2] → (obj) → [c3]	
[He] ← (agent) ← [catch] → (obj) → [ball]	

2) SP Form A sentence le*xical structure* consists of multiple phrases and each phrase is composed of a triple comprising a role and two participants. In SP form, each phrase is expressed as a role and two participants.

(<role> (<direction1> <participant1>) (<direction2> <participant2>))

The collection of such phrases (i.e., sp forms) constitutes a sentence.

Figure 2 shows the SP form syntax. The direction symbol → that points away from the role is read as "*is*", and the direction symbol ← that points to the role is read as "*of*". For example, *(manner (← eat) (→ fast))* is read as "manner of eat is fast".

<sentence>	:	(<phrase>+)
<phrase>	:	(<role> <participant> <participant> <participant>*)
<role>	:	function word (e.g., determiners, adverbs and prepositions) that clarifies relationships between concepts e.g., color, agent, location, obj, etc., i.e., conceptual relation
<participant>	:	(<direction><kernel>)
<direction>	:	← | →
<kernel>	:	<content> | (<content> <referent>)
<content>	:	content word (e.g., noun, adjective and verb) from catalog of conceptual types, e.g., dog, train, etc., i.e., concept
<referent>	:	<empty> | <individual> | <set> | <reference> | <measure> | <quantifier>
<individual>	:	a proper noun e.g., Snoppy, Clifford, Emma, etc.
<set>	:	(<individual>*)
<reference>	:	$
<measure>	:	@<number>
<number>	:	integer or floating point number
<quantifier>	:	@every

Figure 2. SP form Syntax

Participants could have a referent field. Figure 3 shows examples of participant referent field types, referent values, and their representation syntax in both CG and SP form. Figure 4 shows a natural language sentence expressed in SP form. Note that the sentence has three phrases.

Type	**CG**	**SP**
generic	[dog]	dog
generic set	[dog: {*}]	(dog (*))
individual	[dog: Snoopy]	(dog Snoopy)
set referent	[dog: {Snoopy, Lassie}]	(dog (Snoopy Lassie))
definite reference	[dog: #]	(dog $)
measure	[speed: @55]	(speed @55)
universal quantifier	[man: ∀]	(man ∀) (man @every)

Figure 3. Participant Referent Field Types

B. NL Parser: Stanford typed dependencies

The Stanford typed dependency (SD) [8] parser represents sentence grammatical relationships as typed dependency relations, i.e., triples of a relation between pairs of words, such as "the subject of *going* is *John*" in the sentence "John is going to Boston by bus". Each sentence word (except head of sentence) is the dependent of one other word. The dependencies are represented as *relation_name (<governor>, <dependent>).* All are binary relations: grammatical relation holds between a governor and a dependent.

This representation, as triples of a relation between pairs of words, is well suited for mapping SD parser output to SP forms.

 ISBN: 978-1-61208-157-1

Figure 4 shows an SD parse of the sentence “John is going to Boston by bus”. The parse output, which is the syntax tree and the SD dependencies, is mapped to SP forms via canonical forms. Note that for very complex sentences an additional parser that outputs parts of speech tags, for example [9,19], may be needed to determine the appropriate canonical form.

```
Parsing [sent. 1 len. 8]: [John, is, going, to, Boston, by, bus,.]
 (ROOT
   (S
     (NP (NNP John))
     (VP (VBZ is)
       (VP (VBG going)
         (PP (TO to)
           (NP (NNP Boston)))
         (PP (IN by)
           (NP (NN bus)))))
     (..)))
```

nsubj(going-3, John-1) aux(going-3, is-2) prep_to(going-3, Boston-5) prep_by(going-3, bus-7)	(agent (← go) (→ (person John))) (dest (← go) (→ (city Boston)) (inst (←go) (→ bus))

Figure 4. Stanford dependency parser output and sp form

C. *Crowd-Sourced Knowledge Lattice & Interactive Learning Component*

Systems built using a knowledge-engineered approach suffer a disadvantage: they require much labor-intensive, difficult, manual work from a knowledge engineer in creating the ontology (i.e., explicitly defining every concept to be represented). The few general-purpose content languages that have been developed [10, 11] are usually bloated and unlikely to capture the intricacies of every possible domain.

This shortcoming is addressed by having SPS end users construct the knowledge lattice collaboratively. Will users provide extensive feedback? The Oxford English Dictionary, the world's greatest dictionary, was built in the 19th-century by a network of far-flung etymologists using postal mail. Today we have the Internet and the ideals of open source: share the goal, share the work, and share the results. Existing collaboration environments (e.g., Usenet News, eBay, Amazon reader surveys, epinions, tor, Foldit, Phylo, EteRNA) and community projects (e.g., Linux, GNU, Emacs, standards working groups, the Human Genome Project, raising a barn, etc.) demonstrate that users do provide feedback because of shared interest and perceived benefit.

The knowledge lattice resides on a central server and is accessible and modifiable by the Interactive Learning Component (ILC) of every SPS instance. The ILC is a teachable system [12,13] that acquires knowledge through dialog, and is responsible for maintaining and extending the knowledge lattice.

When SPS encounters a word not present in the Knowledge Lattice, SPS’ Interactive Learning Component asks the user the position of the word in relation to other words in the lattice, and the word’s synonyms if these are not available from a digital dictionary. Users are expected to indicate the position of the word in relation to other words. Updating the Knowledge Lattice may trigger other recursive Knowledge Lattice updates. The small incremental contributions from SPS’ global population of users allow the Knowledge Lattice to be built quickly and with little effort from any one individual user. Figure 5 shows a Knowledge Lattice fragment and Figure 6 the lattice’s computational representation.

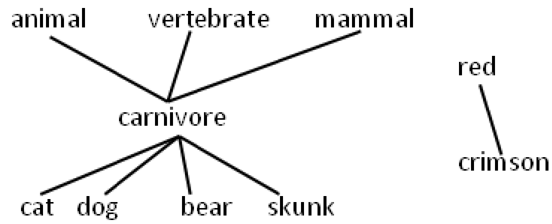


Figure 5. Knowledge Lattice Fragment

Word	Supertype	Subtype	Synonyms
carnivore	(animal vertebrate mammal)	(cat dog bear skunk)	(meat-eating flesh-eating predatory raptorial)
red	( )	(crimson)	(scarlet ruby cardinal flushed rosy wine sanguine)
crimson	(red)	( )	( )

Figure 6. Knowledge Lattice Internal Representation

The Knowledge Lattice stores no word definitions but only the subtype / super-type relations of a word and the word’s synonyms. One key architectural decision of SPS is that computing relevance would not require use of “electronic” word definitions. In SPS, the relation of a word to other words in a phrase (e.g., role of agent, obj, etc.), the synonyms of a word (e.g., scarlet, red), and the subtype / super-type relation of a query word to a target word (e.g., person, girl) are used to determine if a target phrase matches (i.e., is relevant) a query phrase.

Now we introduce Knowledge Lattice operations. A lattice [2] is a structure consisting of a set L (of type labels), a partial ordering $\leq$, and two dyadic operators $\cup$ and $\cap$. If a and b are elements of L, $a \cap b$ is the *maximal common subtype* of a and b, and $a \cup b$ is the *minimal common supertype* of a and b. For any a, b, and c in L, these operators satisfy the following axioms:

- $a \cap b \leq a$ and $a \cap b \leq b$.
- If c is any element of L for which $c \leq a$ and $c \leq b$, then $c \leq a \cap b$.
- $a \leq a \cup b$ and $b \leq a \cup b$.
- If c is any element of L for which $a \leq c$ and $b \leq c$, then $a \cup b \leq c$.

Below are examples of Knowledge Lattice operations applied on the lattice fragment shown in Figure 5.

- $\leq$ subtype
  e.g., [crimson] $\leq$ [red]
- $\cup$ *minimal common supertype* i.e. least upper bound.
  e.g. [cat] $\cup$ [dog] = [carnivore]

[cat] and [dog] have many common supertypes including [animal], [vertebrate], and [mammal]. However, [carnivore] is the minimal common supertype.

- ∩ *maximal common subtype* i.e. greatest lower bound.

e.g. [vertebrate] ∩ [mammal] = [carnivore]

[vertebrate] and [mammal] have many common subtypes including [cat], [dog], [bear], etc. However, [carnivore] is the greatest common subtype.

D. Keyword Extract

Keyword Extract creates a keyword query for a traditional search engine (e.g., Google). Each keyword of the natural language query is augmented with additional keywords that account for synonymy and with its maximal common subtypes and minimal common supertypes. For example, if the natural language query is "*Carpets in Toyota cars cause accelerator to stick*." and the Knowledge Lattice is as shown in Figure 7 then the generated keyword query contains all the synonyms, supertypes, and subtypes of each word in the natural language query. E.g., (*or* "accelerator throttle pedal choke car vehicle …").

Word	Supertype	Subtype	Synonym
accelerator	( )	( )	(throttle choke pedal)
car	(vehicle)	( )	(auto automobile motor-vehicle wheels clunker rustbucket)
carpet	( )	( )	(rug mat floor-covering blanket cover cloak)
cause	( )	( )	(root origin mainspring basis trigger foster make-happen create produce generate induce beget provoke)
stick	(wood)	( )	(cane pole club baton attach affix fasten paste pin tack stay last persist hold)
Toyota	(car automobile)	(Lexus Camry Corolla)	( )

Figure 7. Knowledge Lattice – car, carpet, cause

E. Semantic Matcher

Semantic Matcher (SM) determines which retrieved pages are relevant to the user's query. The inputs to SM are the user's query and the retrieved pages. SM carries out the relevance computation by applying the *restriction*, *projection*, *maximal-common-subgraph*, and *match* operations against the retrieved pages, each of which is explained below briefly. We note that these operations consult the Knowledge Lattice.

Restriction This operation transforms a concept into a more specific type. It replaces

- a more general concept with a more specific one, e.g. (animal) & (dog) => dog
- a generic referent with an individual referent e.g. (dog) & (dog Rufus) => (dog Rufus)
- individual/set-referent and set-referent with their union e.g. (dog Rufus) & (dog (Snoopy Lassie)) => (dog (Rufus Snoopy Lassie))
- two individuals with their union e.g. (dog Lassie) & (dog Rufus)) => (dog (Lassie Rufus))
- numerous other combinations for each participant type (see Figure 3), not listed due to space limitation

Maximal-Common-Subgraph This operation finds the largest subgraph that two graphs *u* and *v* have in common. For example,

u	*v*
(color (← (dog Rufus)) (→ brown)) (location (← (dog Rufus)) (→ porch))	(agent (→ dog) (← eat)) (obj (← eat) (→ bone)) (color (← dog) (→ brown))

The common subgraph is (color (← (dog Rufus)) (→ brown)) the generic referent (dog) is restricted to the individual referent (dog Rufus).

Match Two graphs *u* and *v* match if there is a subgraph *u*' of *u* such that

==> roles are the same in *v* and *u*'

==> pairs of corresponding concepts in *v* and *u*' have a *maximal common subtype*

For example,

u : "John likes white elephants."

v : "The boy likes animals."

u parser output:	*u* SP form:
nsubj(likes-2, John-1) amod(elephants-4, white-3) dobj(likes-2, elephants-4)	(agent (← like) (→ (person John))) (color (← elephant) (→ white)) (obj (← like) (→ elephant))

v parser output:	*v* SP form:
nsubj(likes-2, Boy-1) dobj(likes-2, animals-3)	(agent (← like) (→ boy)) (obj (← like) (→ animal))

Knowledge lattice operations indicate that (person) is a supertype of (boy), and (animal) is a supertype of (elephant); therefore, the graphs *u* and *v* match.

Projection This operation maps a general graph *v* to a more specialized graph *u*. The mapping is a subgraph of *u*, called a *projection* of *v* in *u*. A projection determines if a graph is a subgraph of another graph. E.g. adapted from [14].

u	*v*
(child (← (man John)) (→ (girl Mary))) (child (← (man John)) (→ (boy Bob))) (agent (← love) (→ (boy Bob))) (obj (← love) (→ (girl Mary)))	(child (← person) (→ person))

projection of *v* in *u*:
(child (← (man John)) (→ (girl Mary))) (child (← (man John)) (→ (boy Bob)))

Knowledge Lattice indicates that (person) is a super-type of (man), (boy), and (girl).

The difference between projection and match is that in match either graph can be specific or general; in projection, a more general graph is used to find a more specific one.

 ISBN: 978-1-61208-157-1

IV. Keyword Search vs Phrase Search

One specific application of SPS has been developing a phrase search mechanism for text mining of biomedical literature. Text mining of biomedical literature is a trendy topic in bioinformatics and we illustrate how SPS can improve relevancy in this application. Given below are two sentences from two different web pages [15]. These two pages are polysemous at the phrase level.

Page 1	We then present evidence that Sip1, Sip2, and Gal83 each interact independently with both Snf1 and Snf4 via distinct domains.
Page 2	The catalytic subunits of Arabidopsis SnRKs, AKIN10 and AKIN11, interact with Snf4 and suppress the snf1 and snf4 mutations in yeast.

A Google search against these pages with the query “interact with snf1” returns both pages 1 and 2 because all the search query keywords appear in both pages. But a SPS search returns only page 1 because the search query phrase occurs in page 1, but not page 2. This difference is the result of using the Semantic Matcher which utilizes the Knowledge Lattice and the Semantic Matcher operations to compute the relevance of page 2 to the user’s query.

Figure 8 shows the SP form of each sentence. The search query in SP form is *(obj (← interact) (→ (protein Snf1)))*. This SP form search query matches the page 1 phrase *(obj (← interact) (→ (protein (Snf1 Snf4))))* but does not match any SP forms in page 2. There is a page 2 phrase, *(obj (← interact) (→(protein Snf4)))*, similar to the search query; but the page 2 phrase contains a different protein than the search query protein.

Page 1 sp form:	**Page 2 sp form:**
(agent (← present) (→ (person We))) (obj (← present) (→ $evidence)) $evidence: (agent (← interact) (→ (protein (Sip1 Sip2 Gal83)))) (obj (← interact) (→ (protein (Snf1 Snf4)))) (manner (← interact) (→ independent)) (instr (← interact) (→ domain)) (type (← domain) (→ distinct)) [1]SnRKs [Snf1 (sucrose non-fermenting-1)-related protein kinases	(type (← subunits) (→ catalytic)) (agent (← interact) (→ subunits)) (kind (← subunits) (→ SnRKs[1])) (type (← SnRKs) (→ (plant Arabidopsis))) (equi (← SnRKs) (→ (protein (AKIN10 AKIN11)))) (obj (← interact) (→ (protein Snf4))) (agent (← suppress) (→ subunits)) (obj (← suppress) (→ mutation)) (type (← mutation) (→ (protein (snf1 snf4)))) (loc (← suppress) (→ yeast))

Figure 8. Page 1 and 2 SP forms

V. Conclusion and Future Work

This example, albeit simple, demonstrates that phrase search produces results with higher relevance. Phrase search is superior because the atomic unit for matching is not a keyword but an inter-related collection of keywords, i.e. a phrase, that form meaning. The inter-relation expresses grammatical and semantic information, and is captured in SP forms. In contrast, in keyword search, the interrelation of keywords is only approximated by statistical quantities (e.g., word frequency, information gain, odds ratio, etc.) of the page containing the keywords.

We are currently examining the scalability of our method by applying the phrase search to flexibly finding gene regulatory relationships reported in the biomedical literature.

References

[1] S. Brin and L. Page, “The Anatomy of a Large-Scale Hypertextual Web Search Engine,” www.db.stanford.edu/~backrub/google.html

[2] John F. Sowa, Conceptual Structures: Information Processing in Mind and Machine, Addison-Wesley, 1984.

[3] John F. Sowa and David Dietz, Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks/Cole Pub Co, 1999.

[4] Randolph Quirk, et al. A Grammar of Contemporary English, Longman Group UK Limited, 1987.

[5] Fillmore, Charles J. (1968) "The case for case", in E. Bach & R. T. Harms, eds. Universals in Linguistic Theory, Holt, Rinehart and Winston, New York, pp. 1-88.

[6] J. Leone, "Synergistic Model for Memory Recall", Master’s thesis, University of Connecticut, Storrs, CT, Aug. 1989.

[7] D. G. Shin and J. Leone, “AM/AG Model: A Hierarchical Social System Metaphor for Distributed Problem Solving,” International Journal of Pattern Recognition and Artificial Intelligence, Vol. 4, No. 3, September 1990.

[8] http://www-nlp.stanford.edu/software/stanford-dependencies.shtml (last accessed 6/22/2011).

[9] http://www.connexor.eu/technology/machinese/demo/syntax/index.html (last accessed 6/22/2011).

[10] M. R. Genesereth and S. P. Ketchpel, “Software Agents”, Communications of the ACM 37 (7), 1994, pp. 48-53.

[11] D. B. Lenat, “CYC: A large-scale investment in knowledge infrastructure”, CACM 38(11), 1995, pp. 33–38.

[12] N. Haas and G. G. Hendrix, "An approach to acquiring and applying knowledge," Proc. of AAAI, 1980, pp. 235-239.

[13] N. Haas and G. G. Hendrix, "Learning by Being Told: Acquiring Knowledge for Information Management". In Michalski, R.S., Carbonell, J.G., and Mitchell, T.M. (Eds.), Machine Learning: An Artificial Intelligence Approach. Morgan Kaufmann Publishers, Inc., Vol. I, 1983, Chapter 13.

[14] Jean Fargues, et al., “Conceptual graphs for semantics and knowledge processing”, IBM Journal of Research and Development, Vol. 30, No. 1, January 1986.

[15] http://www.ihop-net.org/UniPub/iHOP/gs/32484.html (last accessed 6/22/2011).

[16] A. Sieg, B. Mobasher, and R. Burke, “Learning Ontology-Based User Profiles: A Semantic Approach to Personalized Web Search”, IEEE Intelligent Informatics Bulletin, Vol. 8, No.1. Nov. 2007.

[17] Q. Zhou, C. Wang, M. Xiong, H. Wang and Y. Yu, “SPARK: Adapting Keyword Query to Semantic Search”, The Semantic Web (2007), Vol: 4825, Publisher: Springer, Pages: 694-707

[18] M. Shabanzadeh, MA Nematbakhsh, N. Nematbakhsh, “A Semantic based query expansion to search”, in International Conference on Intelligent Control and Information Processing (ICICIP) 2010, IEEE, Pages: 523–528, (2010).

[19] http://www-nlp.stanford.edu/software/tagger.shtml (last accessed 6/22/2011).

 ISBN: 978-1-61208-157-1

A Framework for Creativity in Search Results

James Sawle, Fania Raczinski, and Hongji Yang
Institute of Creative Technologies
De Montfort University
Leicester, UK
{jsawle, fania, hyang}@dmu.ac.uk

***Abstract*—Although trying to define creativity has been a hot area of research in many fields, the field of information retrieval has remained under developed. Over the report we attempt to define a structural definition of creativity which could be applied to search results in order to aid users in their creative endeavours. After defining creativity for search, we have then devised a simple metric based upon it, to show that there is a need for this research. The results, whilst positive, could be interpreted as a poor definition of creativity, and as such this is a sounding paper for future work.**

***Index Terms*—information retrieval; computational creativity**

I. Introduction

Over the past decade search has been focused on returning the smallest number of results which correlate to the user's information need. This has been a logical trend to pursue, as 92% of people use the internet as their first port of call when looking for everyday information [1].

However, this has meant that the creativity inspired by 'surfing the web' has over time slowly diminished. This research is not advocating the end of document based search; however, we propose that a new search engine architecture, which aims to inspire the creativity of its users, can only be beneficial to the landscape of the world wide web.

Over the course of the paper, we define what we mean by the creativity of a search result, with respect to a single result as well an entire set. The concepts presented in this paper, are inspired by 'Pataphysics, a pseudo-philosophy defined as "the science of imaginary solutions, which symbolically attributes the properties of objects, described by their virtuality, to their lineaments" [2].

The rest of the paper is organised as followed. Section 2 explores definitions of creativity from both computer science and psychology. In Section 3, we outline a general definition of creativity in search, which can be used to create a metric. Section 4 will see a simplistic metric to be used for the purpose of evaluating the concept as well as some experimental data.

II. Definition of Creative Search

A. A Framework to Base Creativity Upon

Creativity is a subjective topic, with different people defining the creative worth of a piece of information differently; however, Newell, Shaw and Simon [3] devised a definition based upon four criteria to categorise the creativity of a given solution or answer.

1) The answer is novel and useful (either for the individual or for society)
2) The answer demand that we reject ideas we had previously accepted
3) The answer results from intense motivation and persistence
4) The answer comes from clarifying a problem that was originally vague

Each of these criterion for creativity approach a definition from a different perspective. Whilst trying to relate this to information retrieval, it should be simple to see that criterion 1 relates to the goal of the search, whilst criterion 4 relates to the information need, or starting point. What may be less obvious however, is that criterion 3 relates to the scale of the search and hence the number of dead ends that may be encountered and that criterion 2 suggests which search paths should be avoided whilst looking for creative results.

Whilst this framework gives us a very high level definition of creativity, it is hard to apply it in its current form. Through applying some of the more prevalent techniques used in the field of computational creativity, we can attempt to reduce this down into a more precise definition.

B. P-Creativity and H-Creativity

Boden [4] defines that there are two forms of creativity, P-creativity and H-creativity. P-creativity or *'psychological'* creativity, is an idea or solution that is new to the person who came up with it. An idea that represents *'historical'* creativity, H-creativity, on the other hand, is one which has not been thought of by anybody before and can therefore be deemed a historical-sociological category [5]. H-creativity is subsequently a special case of P-creativity which many people consider to be the more important of the two, as this is what drives forward human knowledge.

When we relate these concepts to search results we end up with some interesting outcomes.

1) Single Search Result: A single search result is most likely to be P-creative or neither. This is because, for it to be H-creative, there must be some logic in the document that nobody else has noticed, or drawn the same conclusions from different information. For the single result to be neither P-creative or H-creative, the user must have a thorough understanding of the topic, and the result must add no new information.

2) Set of search results: A set of search results is most likely to be P-creative. It is highly unlikely that a user would

 ISBN: 978-1-61208-157-1

have explored every possible creative avenue over a set of results, unless the set is not of a trivial size. But by the same logic, if a large range of ideas are contained, it is unlikely that the set will be H-creative, as somebody is likely to have linked them together.

The question becomes, is there a link between maximising the chance of something being P-creative and H-creative or is the link more subtle. Or is it enough for a search engine to try and improve the chances of P-creativity for a user.

C. Exploratory and Transformational Creativity

Boden [4] goes on to define the concepts of exploratory and transformational creativity. She defines exploratory creativity to be the exploration of a space of partial and complete possibilities. This therefore suggests that there are rules that confine this space. If we were therefore to alter the rules that define the space, and subsequently alter the space that we are exploring, this is defined as transformational creativity [6].

Whilst this does give us a nice slant to look at creativity, comparing the trade-off of traditional problem spaces compared to augmented ones, this is very difficult to model, combined with the fact that the solutions found by tweaking the rules that confine the space can easily rule out the solution in the traditional space [7].

D. Bisociation

Bisociation makes a distinction between the routine skills of thinking on a single 'plane', and the creative act, which operates on more than one plane [8]. This means, with Koestler's definition, that we must define creativity as a set of results such that they are simultaneously associated with two habitually incomparable contexts.

It is clear to see how this model extends from that of Boden's theory of exploratory and transformational creativity. The fact that more than one 'plane' must be considered will force a transformational process to occur. However, unlike transformational creativity, both processes must be considered, the exploratory and transformational. Subsequently, we should not end up with a solution that can't exist within the rules defined by the original problem, even if we transcend into transformational creativity, as long as we finish the process in the plane that we started in.

E. Conceptual Blending

The idea of combining different thought processes, whilst more elegant than transformational creativity, does not give us a nice definition that applies to search results as well as tying in with our underlying philosophy. Conceptual blending is a step closer. This general theory of cognition, formally called Conceptual Integration Networks [9], allows us to look at a number of different dataspaces, and attempt to 'blend'/merge them in such a way that the new dataspace tries to simulate how we use large amounts of information and bring it together to form new ideas.

F. Combinatorial Creativity

Both of the above concepts fall into the general category of combinatorial creativity. This is a logical assumption of modelling creativity, as people tend to come up with solutions by first looking at new combinations of currently existing ideas. This therefore allows us to consider the idea of creativity as a search process through the space of all possible combinations, therefore this fits into the idea of search engines.

Whilst conceptual blending explores the idea of combining different thought processes and bisociation, looking at different planes of creative thought; let us consider the idea of placing the data itself into different concepts, enabling us to get the following areas of combinatorial creativity to explore with respect to creative search based upon philosophical.

- Placing a familiar object in an unfamiliar setting or placing an unfamiliar object into a familiar setting.
- Blending two superficially different objects or concepts
- Comparing a familiar object to a superficially unrelated and semantically distant concept
- Searching through a number of different concepts that are related to each other but could be considered as swerving away from the original concept. This is based upon Epicurus's theory of clinamen from his doctorine of atomism [10].

III. Definition

The above definitions, allow us to define creativity in search results with pre-existing concepts agreed by the academic community.

It is clear, that in the case of search results, we still have the issue of a group of results providing greater creative inspiration to one user than another. This tends to be a problem with most metrics, the problem of objectiveness vs subjectiveness. With subjectiveness being a quality that is important, it means that we have a problem getting repeatable results. We therefore need to build a definition that is as objective as possible, whilst not overlooking some of the dynamic properties that it may be possible to model.

At this stage it is important to stress that this is not an attempt to model the creative process, but to give a model for how useful a set of results might be in inspiring creativity.

A. A Single Result

It is intuitive for us to start with a single result. Whilst maximising the possibility for a single result being H-creative, it is very unlikely that this will be the case with a full set of results. The issue becomes, measuring how P-creative an individual result is to a search result.

It seems sensible to assume, that if a result has no relevance to the search request, then the result will have no chance of inspiring P-creativity. The more information about the search request a single result has, increases the chance of a result inspiring P-creativity, therefore using relevance metrics.

 ISBN: 978-1-61208-157-1

B. Set of Results

To improve the chances of inspiring creativity, a group of related results which discuss a number of different areas of the topic would logically improve the quality of the results. As stated, if we maximise the breadth of information of a single result it would improve creativity, we should therefore attempt to do the same across the entire set.

The issue however, is that the majority of users do not look past the top 10 search results [11]. Whilst this is unlikely to be the case for people using a search engine targeted at inspiring creativity, it does make sense to try to reduce the overall amount of data provided. We must therefore penalise repetition in the results provided, forcing a more diverse set of results.

This can be taken a step further by only considering a certain number of results and ignoring the ordering, because there is no simple way to define how ordering affects the creative process. With a lack of defined ordering, it means that having endless results would be tedious and counter-productive. Whilst we have no strong view on the exact number of results that should be considered, we believe that it should not be substantially greater than 10, for the reason discussed above.

The way that each result is provided to the user will affect how the user perceives the results. A diversity of different document types, e.g., text, images, sound, we believe would improve the quality of creativity inspiration.

C. Results as a Set of Sets

We could extend this concept to the next logical step of returning results as a set containing multiple related sets of results. In this analogy, each of the inner sets could relate to an individual concept related to the information need, and a clearer relationship between concepts, how they relate to each other and how the results represent the concept they are contained within would exist.

The question becomes how we measure the creative quality of this type of result. Due to the structure of the results, we can attempt to model the creativity in different levels allowing us to try and abstract the problem as much as possible.

Due to the fact that this is not a method that is currently used to return search results, we shall not explore it further at this point in time. However, we believe that this would be a logical way to return results in the future.

IV. Example Metric

As the above definition is meant as a guideline for defining creativity, this section attempts to give a real world example. The metric defined below is a contrived example to show how it could be applied with current search results.

A. Algebraic Definition

Taking the definition defined in Section III-B, we have derived the following abstract metric.

Let us define a query as q, a set of results as r and an individual result as d. As such $r = \{d_1, d_2, d_3, ..., d_i\}$ where i is the number of results examined.

For the quality of a single result, we shall define $P(q, d)$ as a measure between 0 and 1, where 1 is the optimal value.

To reduce the amount of data duplication in the returned results, we shall define $D(r)$ which has to return a value between 0 and 1, where 0 means that no data is duplicated.

Let us define $T(r)$ as a way to weigh the final outcome of the metric to ensure that a diverse set of document types are returned. This metric will return 1 if a satisfactory balance is returned, and 0 if only a single document type is returned.

We can therefore compile these measures into a single metric, the Search Creativity Metric or SCM:

$$SCM = T(r) \cdot \frac{1}{i} \sum_{j=1} P(q, d_i) \cdot (1 - D(r))$$

As such, this metric will always return a value between 0 and 1, with 1 being the optimal value.

B. Fleshing Out the Metric

To enable us to apply any experimental data to the metric, we must first give definitive definitions to each of the functions provided above, $P(q, d)$, $D(r)$ and $T(r)$.

1) D(r): As this measures the number of duplicate results in a return set, we can easily define it as the number of results that have a majority of information that is contained within another article. This allows the following definition

$$D(r) = \frac{Number of results with data in previous results}{Number of results}$$

As we relate each result to the previous results in the list, the results must always be $0 < D(r) \leq 1$. This makes sense, as even if all of the results are identical, there may still be some creative inspiration contained in the first result. This also allows us to penalise results heavily for leaning too much on one area of information.

2) T(r): As with $D(r)$, we need to define this measure so that we penalise for a lack of diversity, but do not eradicate all results, as this would not reflect the possible creative quality of the information returned.

For this definition, we will need to leverage on the definitions provided earlier. Let i is the number of results within the result set r. We can therefore define n to be the number of different result types that are returned, and σ to be the standard deviation of the number of results for each media type. It is interesting to note that $0 \leq \sigma < \frac{i}{2}$, such that $\sigma = \frac{i}{2}$ means that the results are biased to only one result.

$$T(r) = \begin{cases} 1 - \frac{2 \cdot \sigma}{i} & : n > 2 \\ 0.1 & : n \leq 2 \end{cases}$$

For the case of this sample measure, we have defined that for a result set to be considered to be broad enough, that it must contain at least 3 different media types. This measure has no empirical backing.

TABLE I: Example Results

URL	$P(q, d)$	Reason
www.unicorn-darts.com	0.3	Company called Unicorn due to the single point on a dart
en.wikipedia.org/wiki/Unicorn	1.0	Contains mythology as well as related animals
www.unicorn-grocery.co.uk	0.0	No relation to unicorns
www.unicorn theatre.com	0.25	Uses the mythology of unicorns to draw children into theatre
http://katemckinnon .files.wordpress. com/2008/07	0.7	Image of a unicorn but purely as a distraction from the rest of the article
http://www.unicorn centre.co.uk/	0.9	Unicorn mythology about the soul applied to a spiritual ideal including image
http://31st-and-chi.blogspot.com/ 2010/07/bunch-of-pictures-of-unicorns.html	0.9	Large array of unicorn pictures. One is identical to result 5.
http://disgrasian .com/2010/09/ unicorns-really-do-exist-and-theyre-asian/	0.9	Picture of unicorn and asian 2 horn unicorn.
http://www.youtube .com/watch?v= Q5im0Ssyyus	0.4	Comedy cartoon video about unicorns.
http://www.youtube .com/watch?v= v25MaXwopNI	0.8	Music for a cartoon character.

3) P(q, d): With respect to the relevance of an individual result compared to the information need, there are a number of different methods that could be used. For example, keyword analysis in text documents and image recognition in images, it is clear that a separate method would be needed for each media type that is returned.

With this in mind, for the example below, the individual relevance of a given result will be manually determined and a brief explanation given. The focus will be more on the relevance of the result to the information need, with some weighting given if there is a creative link.

C. Experimental Data

To show this metric in practice we will need to get real world data about a topic. We have used Google to search for results on the following creative need - unicorns from Greek mythology.

The search term input into Google on Thursday 12th May was 'unicorn'. Below is a table of a url to each result, their assigned P rating and a brief description of the reason why. We have taken the top 10 results including the first 4 images and videos.

Due to the repeated result in result 4 and 6, $D(r) = 0.1$ and $T(r) = 1 - \frac{2 \cdot 0.94}{10} = 0.812$. If we then feed these results into the SCM metric we get.

$$SCM = 0.812 \cdot \frac{1}{10} \cdot 6.15 \cdot (1 - 0.1) = 0.449$$

A 0.449 result for us represents a set of results that contain some creative merit, but which also could be improved. This result could be enhanced, based on this metric, if four of the results were to have been replaced with more relevant results.

We still need to understand whether the low result is due to the fact that the results are not inspiring creativity as we presume, or that the definition that we have provided is not complete and that we need to extend it further. It is planned, that we take this research further to answer the question using in-depth empirical analysis.

V. Conclusion and Future Work

Over the course of the report, we have attempted to define what we mean by creativity with respect to search engine results using the concepts from computational creativity. The definition is focused more on the structure and relationship between the results returned than the content of the results themselves. This will allow us to define this separately after carrying out further experiments.

This is evident from the metric that we generated to show how the definition could be used. We believe that the low result shows that the return set does not have a high creative merit; however, more testing will be needed to check whether this is the case, or whether the definition needs to be redefined.

We believe, that whilst this paper has little empirical backing, it has highlighted a short fall in the information retrieval domain, namely that of creative search. Even from the simple test that was conducted, it is apparent, that even when we reach a metric for measuring the creative quality of results, a new form of search engine will be required to achieve top quality results consistently.

The next stage of the research will focus on applying what we have learnt and combine quantitive and qualitative analysis to try and develop a new metric with a strong empirical backing. This means that our definition of creativity will likely need to be adapted over time; however, this could allow us to develop a metric that evolves over time to adapt to what the users consider to be creative search.

References

[1] D. Fallows, "The Internet and daily life," *Pew Internet and American Life Project*, 2004.

[2] A. Jarry, *Selected works of Alfred Jarry*, 1st ed., R. Shattuck and S. W. Taylor, Eds. Grove Press, 1965.

[3] A. Newell, J. G. Shaw, and H. A. Simon, *The process of creative thinking*. New York: Atherton, 1963, pp. 63 – 119.

[4] M. Boden, *The Creative Mind: Myths and Mechanisms*, 2nd ed. Weidenfeld & Nicolson, 1992.

[5] A. Brannigan, *The Social Basis of Scientific Discoveries*. Cambridge University Press, 1981.

[6] G. A. Wiggins, "A preliminary framework for description, analysis and comparison of creative systems," *Knowledge-Based Systems*, vol. 19, pp. 449–458, 2006.

[7] M. A. Boden, *Computer Models of Creativity*. Cambridge University Press, 1999, pp. 351–372.

[8] A. Koestler, *The Act of Creation*. London: Hutchinson & Co, 1964.

[9] G. Fauconnier and M. Turner, "Conceptual integration networks," *Cognitive Science*, vol. 22, no. 2, pp. 133–187, 1998.

[10] Epicurus, *The Epicurus Reader*, B. Inwood and L. P. Gerson, Eds. Hackett, 1994.

[11] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a very large web search engine query log," *SIGIR Forum*, vol. 33, no. 1, pp. 6–12, 1999.

Realtime Computation of a VST Audio Effect Plugin on the Graphics Processor

Wolfgang Fohl
HAW Hamburg
University of Applied Sciences
Hamburg, Germany
Email: fohl@informatik.haw-hamburg.de

Julian Dessecker
Steinberg Media Technologies GmbH
Hamburg, Germany
Email: J.Dessecker@steinberg.de

***Abstract*—A plugin system for GPGPU real time audio effect calculation on the graphics processing unit of the computer system is presented. The prototype application is the rendering of mono audio material with head-related transfer functions (HRTFs) to create the impression of a sound source located in a certain direction relative to the listener's head. The virtual source location can be controlled in realtime. Since HRTFs are measured only for certain incident angles, a interpolation for intermediate angles has to be performed in realtime. Plugins are implemented using the VST software development kit offered by Steinberg Media Technologies. Two GPU processing frameworks for a NVIDIA graphics processor were evaluated: CUDA and OpenCL. The overall processing speed can be increased by the factor 2.2 with the GPGPU modules. When calculating the FIR filter outputs by fast convolution on the GPU, the processing speed can even be increased by the factor ten.**

***Keywords*-GPGPU computing; VST plugin; Spatial audio; Head-related transfer functions;**

I. Introduction

This article presents a case study for the application of GPGPU techniques to the realtime audio signal processing. GPGPU computing (i.e., execution of general purpose computation on the graphics processor) has raised considerable interest with the availability of software development kits (SDKs) that offer an access to the massive parallel computing capabilities of modern graphics processors. The two most common frameworks are OpenCL, which provides an vendor-independent access to GPGPU computing, and CUDA-C, which is an extension to the C language for GPGPU on NVIDIA graphics processors [1] [2]. While CUDA is a proprietary framework restricted to NVIDIA GPUs, OpenCL is an open standard maintained and published by the Khronos Group [3]. The actual OpenCL development framework is provided by the GPU manufacturers.

The initial motivation for this work was the promise of GPGPU to drastically accelerate tasks that can be parallelised. This is the case for realtime audio processing. Input and output data of audio processing units are buffered, which offers the opportunity to calculate all the output samples of a buffer simultaneously from the input samples. Furthermore, the calculation of a single output sample offers opportunities for parallel execution: Most audio processing tasks consist in evaluating a finite difference equation for the N output buffer elements y_{out}:

$$\{y_{out}\} = \left\{ y_i | y_i = \sum_{k=0}^{N} f_k(x_{i-k}) - \sum_{k=1}^{N} g_k(y_{i-k}) \right\}_{(i=0\ldots N-1)} \tag{1}$$

Here, $\{y_{out}\}$ is the array of output samples, i is the sample index, x is the input signal, and k is a summation index.

In the case of a linear and time-invariant audio processor, the functions of eq. 1 are merely multiplications with constant values $\{a_k\}$ and $\{b_k\}$:

$$\{y_{out}\} = \left\{ y_i | y_i = \sum_{k=0}^{N} b_k \cdot x_{i-k} - \sum_{k=1}^{N} a_k \cdot y_{i-k} \right\}_{(i=0\ldots N-1)} \tag{2}$$

Very often in audio processing, filters with a finite impulse response (FIR filters) are employed , for which the difference equation simplifies to

$$\{y_{out}\} = \left\{ y_i | y_i = \sum_{k=0}^{N} b_k \cdot x_{i-k} \right\}_{(i=0\ldots N-1)} \tag{3}$$

which is the *convolution* of the impulse response $\{b_i\}$ of the filter with the input signal $\{x_i\}$. So for each value of the output buffer, the summation of eq. 3 has to be performed, where the computation for each output value as well as the computation of each summand can be calculated simultaneously. The convolution operation of eq. 3 can be considerably sped up by using the *fast convolution* algorithm. The basic idea is given by:

$$\{b_i\} \quad \overset{\text{FFT}}{\Rightarrow} \quad \{B_j\} \tag{4}$$

$$\{x_i\} \quad \overset{\text{FFT}}{\Rightarrow} \quad \{X_j\} \tag{5}$$

$$\{y_{out}\} \quad \overset{\text{IFFT}}{\Leftarrow} \quad \{B_j \cdot X_j\} \tag{6}$$

A FFT is applied to the filter coefficients and the input signal block, the resulting arrays are multiplied element-wise, and the product is transformed back. The two arrays $\{b_i\}$ and $\{x_i\}$ have to be brought to twice the buffer length by zero-padding, which in turn will give an output signal of twice the buffer length. The first half is added to the second half of the previous processing step and transferred to the output buffer.

The audio processing task that shall be executed as prototype application is the spatial rendering of mono audio signals by head-related transfer functions (HRTFs) according to the procedure described in [4].

In the following sections, the algorithm for spatial rendering is described, followed by an overview of the GPGPU and VST plugin architectures. Finally, our solution is described together with some implementation details, and the results of performance measurements are presented and discussed.

 ISBN: 978-1-61208-157-1

II. Overview of Technical Concepts

A. Spatial Audio Rendering

In order to make a mono audio signal appear to be coming from a certain direction of incidence, the signal is filtered by head-related transfer functions (HRTFs), that mimic the inter-aural time delays, level differences and differences in frequency response. If the resulting stereo signal is replayed via headphones, the perceived signals are similar to real signals emanating from a source at the corresponding location.

Since the coefficients of the filters could only be measured at certain discrete angles as indicated in Fig. 1, an interpolation has to be performed for intermediate angles. The input signal is routed to four filters, each representing one corner in the interpolation grid of Fig. 1, then a linear interpolation of the output signals is performed. Details are given in [4].

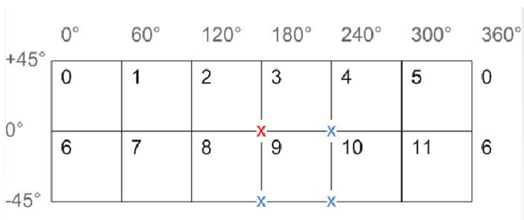


Figure 1. Grid of measured HRTF data

B. GPGPU Computing

The parallel processing architecture of the graphical processing unit (GPU) found on modern hardware suits in an optimal way the needs of many numerical processing tasks. The release of the CUDA SDK by NVIDIA in 2007 and the approval of the OpenCL standard in 2008 opened the field of GPGPU computing to the community of developers and researchers.

GPGPU computing offers a performance boost for algorithms that can be parallelised, as is shown in the code snippets in figure 2. Both programs perform an array addition. While in conventional CPU processing the program iterates over all array elements, the GPU program starts one thread for each array element, provided there are enough processing units. The threads on the GPU are programmed as so-called "kernels".

```
// Kernel definition
__global__ void VecAdd(float* A, float* B, float* C)
{
  int i = threadIdx.x;
  C[i] = A[i] + B[i];
}

int main(void)
{
  // CPU operation
   for (int i = 0; i < N, i++)
      C[i] = A[i] + B[i];
  ...
  // Kernel invocation with N threads
  VecAdd<<<1, N>>>(A, B, C);
  ...
}
```

Figure 2. Array addition on the GPU with the CUDA SDK

One drawback of GPGPU computing on consumer-grade GPUs has to be mentioned. Since these GPUs are designed for optimum video game performance, there is no need for checking the integrity of the GPU memory, since memory errors would only affect the currently displayed video frame. High reliability can be either attained by software means as described in [5], or by using GPU hardware dedicated to GPGPU computing, which are equipped with ECC-protected memory [6].

C. VST Plugin Architecture

Steinberg Media Technologies developed a plugin system for the extension of audio workstation software with external effects and with external virtual instruments. Developers can obtain a SDK after registering on the Steinberg website [7].

A plugin consists of two parts, the *processor* and the *edit controller*. The processor does the audio signal processing, the edit controller provides the GUI for parameter visualisation and modification.

Data transfer between host and plugin is performed by means of the VST plugin interface methods: One block of audio data is provided by the host in the input buffer of the plugin. Then the host calls the `process` method of the plugin, the plugin then has to compute the audio samples and transfer them to the output buffer, where the host will fetch it.

III. System Implementation

A. Technical Details

The software was implemented on a PC with an Intel Core2Quad Q9400 processor operated at 2.66 GHz, a NVIDIA GeForce 9600 graphics processor with 650 MHz core clock, and a Creative ES 1371 sound card. The operating system was Windows XP SP 3, the VST host application was Steinberg Cubase 5. The VST SDK was version 3.1.0.

B. Filter Module Architecture

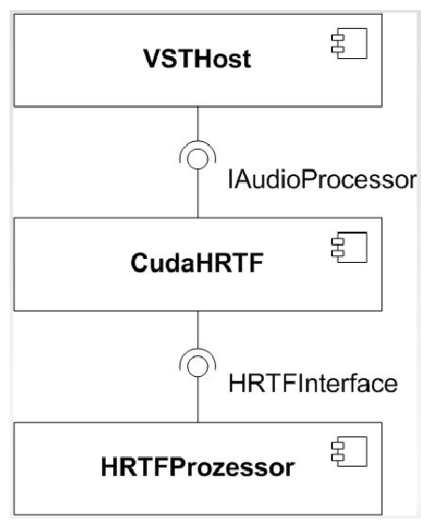


Figure 3. System architecture with VST Host (Cubase), VST plugin, and HRTF processor

In this project, the `CudaHRTF` VST plugin was developed as a wrapper for various filter modules with various implementation of the HRTF rendering algorithm. The filters modules were created as independent dll files. Four filter dlls were developed:

filt_cuda.dll
: Filter implementation using the CUDA-C SDK and fast convolution in the frequency domain

filt_ocl.dll
: Filter implementation using the OpenCL SDK and time-domain convolution

filt_sse.dll
: Filter implementation on the CPU using the SSE2 (Streaming Single-Instruction-Multiple-Data) extension, time-domain convolution

filt_fpu.dll
: Filter implementation using standard FPU code and time-domain convolution

Fig. 3 shows the component design, showing the VST host, the VST plugin, and one filter module.

Note on Time-Domain Convolution: The computation steps for time-domain convolution of the input signal x_i with the FIR filter coefficients $\{b_i\}_{i=0...L-1}$, where L is the filter length according to

$$y_i = \sum_{k=0}^{L} x_{i-k} \cdot b_k \tag{7}$$

can only be parallelised for the *products* in the sum. The summation itself has to be either performed in a sequential loop, or by successively cutting the summand array into halves and adding these halves in parallel operations. With the abbreviation $s_{i,k} = x_{i-k} \cdot b_k$ the sum of eq. 7 can be rewritten as

$$y_i = \sum_{k=0}^{L/2} s_{i,k} + s_{i,k+L/2+1} \tag{8}$$

This procedure of cutting the summand array into halves can be repeated $\log_2 L$ times.

The plugin accepts mono audio data from the VST host, and provides stereo audio data.

The class *HRTFModule* connects the VST plugin and the filter component. The class diagram is given in Fig. 4.

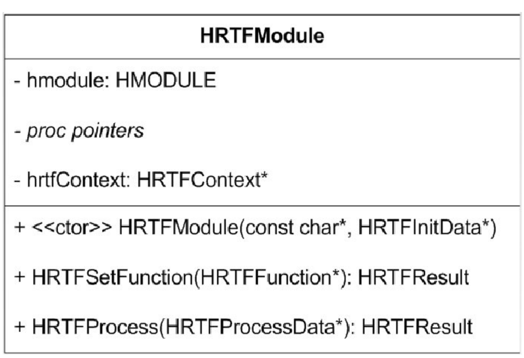


Figure 4. HRTF filter component class diagram

The data structure *HRTFProcessData* contains the necessary data for the *HRTFModule* to perform the HRTF rendering.

Four filters are referenced by their ID, containing the coefficients for the interpolation limits for azimuth and elevation angles. Two

```
typedef struct HRTFProcessData {
   int      numSamples;      // I/O buffer size
   int      filterID[4];     // Filters for interpolation
   float    weightStart[4];  // Weights at block start
   float    weightEnd[4];    // Weights at block end
   float    *input;          // Mono input signal
   float    *leftOutput;     // Stereo output: left
   float    *rightOutput;    // Stereo output: right
} HRTFProcessData;
```

Figure 5. HRTFProcessData data structure

sets of weight factors contain the interpolation weights at block start and block end, so that a smooth movement of the source can be rendered. A problem occurs when the source position crosses the limit of the current interpolation cell. Possible solutions were either to extrapolate beyond the limits or to limit the source movement to the end of the interpolation interval and continue with a new interpolation interval in the next blocks (see Fig. 6). Prototypes for both approaches were implemented in Matlab. It turned out, that the extrapolation approach resulted in audible clicks at block limits, whereas there were no audible clicks in the second approach.

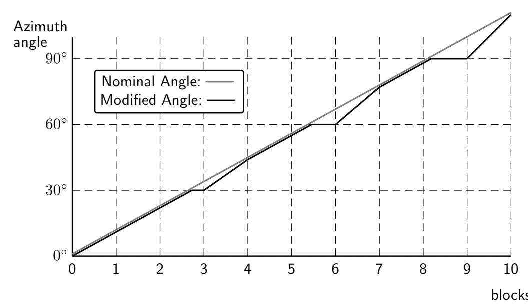


Figure 6. Dynamic angle interpolation at interval limits. The gray line shows the desired values of the azimuth angle of the virtual source, the black line shows the actual rendered angles due to block-wise data processing.

C. Filter Design

HRTF filters were designed in Matlab as FIR filters using previously measured data of a dummy-head system. Frequency and phase responses were measured at azimuth angles from 0° to 180° in steps of 30°. The data for the interval from 180° to 360° were obtained by interchanging the left and right channels. Elevation angles were -45°(below), 0°(plane), and 45°(above). In order to easily configure the plugin filter components, a XML DTD was defined, and the Matlab filter design program produced a corresponding XML file as output, which in turn could be read by the filter component. For XML parsing the TinyXML library is used [8].

D. Code Instrumentation

An audio plugin is a realtime application and as such very sensitive to modifications of the runtime environment of the plugin. This makes it difficult or impossible to use standard tools like debuggers and profilers. The approach used in this project is code instrumentation. Each call to the *process* methode of the filter module is framed with calls to the start and stop methods of the timer object which in turn calls the Windows API function `QueryPerformanceCounter`, to get high resolution timing information. The stop method calculates the elapsed time and updates

```
HRTFResult HRTFProcess(HRTFProcessData* data)
{
  REQUIRE(hrtfProcessPtr);
  timer.start();
  HRTFResult result = hrtfProcessPtr(hrtfContext, data);
  timer.stop();
  return result;
}

// .. The timer methods
inline void start()
{
  QueryPerformanceCounter(&t0);
}

inline double stop()
{
  QueryPerformanceCounter(&t1);
  double cur = (t1.QuadPart - t0.QuadPart) * factor;
  avg = (avg + cur) * 0.5;
  max = max > cur ? max : cur;
  min = min < cur ? min : cur;

  return cur;
}
```

Figure 7. Code Instrumentation for Performance Measurement

the log. This approach implements a performance measurement with minimum interference with the normal plugin operation. It has to be noted tough, that the measured time is “wall-clock time”, not CPU time, so to a small extent the measured results depend on the scheduling of the plugin by the operating system.

IV. Results and Discussion

A. Listening Impression

All filter components were tested with pieces of solo vocal music. The filters produced naturally sounding position rendering of the input audio material without clicks and other artefacts when moving the controls to change the virtual source location.

B. Performance Measurements

All performance measurements were executed with 20 seconds of audio playback. During this time the azimuth and elevation angles of the virtual source are varied according to a path that has been recorded once and was replayed by the automation functionality of the VST host.

Module	t_{min}/ms	t_{max}/ms	t_{avg}/ms
CUDA-C *Fast Convolution*	2	56	2
OpenCL *Time Domain Convolution*	8	18	10
CPU-SSE2 *Time Domain Convolution*	12	36	17
CPU.dll *Time Domain Convolution*	12	31	22

It can be seen, that the GPU algorithms perform significantly faster than the FPU algorithms, which was to be expected. An irritating observation is the large maximum value of the execution time for the CUDA filter module, while the average execution time is equal to the minimum execution time. This indicates, that this long time has occurred very few times, probably only once. Unfortunately the way of code instrumentation does not give any information, when and how often such a large execution time occurs. It can be assumed, that this large time occurs during the setup phase of the FFT algorithm (during setup a *plan* is created). If this assumption is confirmed by further experiments, the creation of the FFT plan can be moved to the plugin constructor, so it would not cause audio dropouts.

V. Conclusion

In this article, a GPGPU based realtime audio effects processor was presented. In particular, spatial audio rendering by filtering the mono input signal with the head related transfer functions for the corresponding angle of sound incidence has been performed. The FIR filtering algorithm has been moderately customised to exploit the benefits of GPGPU computing, leading to an increase in computation speed by a factor of 2.2.

During the listening and performance measurement test no observable memory errors occurred. A systematic test for memory error problems has to be conducted. For high reliability requirements a graphics card dedicated to GPGPU computing must be employed. These cards have ECC protected memory, which consumer level cards do not have.

The reason for the large maximum execution time for the CUDA fast convolution implementation has to be identified and removed.

VI. Acknowledgements

This article is the result of a project course on professional audio application development held at the Hamburg University of Applied Sciences (HAW) together with Steinberg Media Technologies. Thanks to the Steinberg developers, especially Ralf Kürschner, Dr. Nico Becherer and Yvan Grabit for their support and thanks to the students Leonhard Dahl, Tobias Hassenklöver, Ines Ouanes, Marcus Rohwer, Areg Siradeghyan, Alexander Vette, and Özhan Yavuz for their contributions to the software.

References

[1] *OpenCL Programming Guide for the CUDA Architecture*, 3rd ed., NVIDIA Corp., 2010.

[2] *NVIDIA CUDA C Programming Guide*, 3rd ed., NVIDIA Corp., 2010.

[3] “Khronos opencl api registry,” Khronos Group, Accessed 2011-05-05. [Online]. Available: http://www.khronos.org/opencl/

[4] W. Fohl, J. Reichardt, and J. Kuhr, “A System-On-Chip Platform for HRTF-Based Realtime Spatial Audio Rendering,” in *Proc. of the Second International Conference on Creative Content Technologies (Content10)*, Lisbon, Portugal, 2010.

[5] M. Dimitrov, M. Mantor, and H. Zhou, “Understanding software approaches for GPGPU reliability,” in *Proceedings of 2nd Workshop on General Purpose Processing on Graphics Processing Units*, ser. GPGPU-2. New York, NY, USA: ACM, 2009, pp. 94 –104, ACM ID: 1513907.

[6] “Why choose tesla,” NVIDIA Corp., Accessed 2011-05-05. [Online]. Available: http://www.nvidia.com/object/why-choose-tesla.html

[7] "Vst plugin sdk," Accessed 2011-05-02. [Online]. Available: http://www.steinberg.net/en/company/developer.html

[8] L. Thomason, Y. Berquin, and A. Ellerton, "TinyXML Project Page," Accessed 2011-04-28. [Online]. Available: http://www.grinninglizard.com/tinyxml/index.html

 ISBN: 978-1-61208-157-1

A Collaborative Content Publisher

Wan-Ning Wang
Department of Information Management
Yuan Ze University
Taoyuan, Taiwan
s961733@mail.yzu.edu.tw

Hsiang-An Wang
Research Center for Information Technology Innovation
Academia Sinica
Taipei, Taiwan
sawang@iis.sinica.edu.tw

***Abstract*—We propose a web application called Collaborative Content Publisher (CCPublisher), which enables users to publish content and post their work on blogs or elsewhere. Users can participate in collaborative writing of the content posted on different web pages and synchronize each content version. A web annotation function is also provided for users to annotate multimedia content on the web pages. Users can integrate the annotated multimedia content with their collaborative work and update the annotation after the original annotation has been modified. Annotations of multimedia content in different versions of the collaborative work will be consistent if the same multimedia is used. With the CCPublisher, users can provide feedback on any web page, exchange ideas, and enrich the shared collaborative content as well as the web content. It therefore facilitates collaborative learning and the efficient accumulation of knowledge.**

Keywords-Collaborative learning; Web annotation; Multimedia annotation; sharing annotation; Publisher

I. INTRODUCTION

The concept of collaborative learning, a popular topic in E-learning 2.0 [1], refers to a group of learners working together to achieve a common goal [2]. In the collaborative learning process, learners share ideas or modify each other's work. Thus, learners are no longer just information receivers; instead, through contributing, sharing and collaborating, they can also become information providers. As a result, collaborative learning increases a learner's sense of achievement and motivation to learn, and also facilitates the development of critical thinking [3].

An increasing number of educationalists are exploring the possibility that online services could be used to enhance collaborative learning. For example, the State University of New York used Flicker's tagging feature [4] in an online history course to enable students to annotate and discuss a series of paintings [5]; while SUNY Geneseo launched a collaborative writing project using Mediawiki [6]. The above services may not have been designed for educational purposes initially, but they have accelerated online collaboration among learners and created new educational opportunities.

Existing online collaborative learning applications, such as Wikipedia [7] and Flicker, have certain limitations. For example, most of them require users to work collaboratively on a centralized platform. Although users can share their collaborative content or post it on web pages, they cannot provide feedback or exchange ideas on the shared content. This limited one-way sharing also prevents synchronizing changes with the original content. If the original content has been modified, the authors have to re-share it, which may take a great deal of time; otherwise, the shared content will be inconsistent with the latest version.

To resolve the above problems, we have developed a web application called Collaborative Content Publisher (CCPublisher), which enables users to publish content that can be embedded in web pages. Editing tools are provided so that users can edit and give feedback directly on the published content. New feedback is transferred to copies of the published collaborative content automatically so that all versions are synchronized. We also provide a web annotation function for users to annotate multimedia in web pages and transform it with the annotations into collaborative content. The annotations of the multimedia in the collaborative content will be consistent with the original annotations. Through the content embedded in blogs or elsewhere, users can easily exchange comments, co-edit the content and expand the current version. CCPublisher increases the opportunities for interaction between users and facilitates the efficient accumulation of information.

The remainder of this paper is organized as follows. Section 2 examines existing online services related to collaborative learning; Section 3 presents the features of CCPublisher; and Section 4 describes the synchronization mechanism. Section 5 contains some concluding remarks.

II. RELATED WORK

In this section, we discuss existing online services for collaborative learning, with the focus on those that provide collaborative writing platforms and web annotation tools. We also consider online services for sharing collaborative content.

Collaborative writing platforms, such as Wikipedia, Moodle wiki [8] and Google doc [9], can be used for collaborative learning. Wikipedia allows users to contribute content and collaborate with other users to modify the content; thus, most articles have multiple authors. An article will change dynamically after modification, and anyone can track who made a particular change. Moodle wiki, a component of the Moodle system, is also a collaborative writing platform. It enables students to participate in collaborative writing with peers and instructors. Teachers can see when each student writes a new article and view the records of the student's previous online work. Meanwhile, students can view and modify each other's work. Moodle wiki supports group projects and can be used to develop group reports and maintain class records of activities on given topics. Google also provides a writing platform called Google Docs, which allows users to edit a document online and collaborate with other users to edit their contributions.

Collaborative learning can also be achieved by using a web annotation feature, such as Tagtoo [10], Diigo [11], or Reframe It [12], to construct a collaborative environment. Tagtoo allows the user to tag images on any web pages. The tag, which can be a Facebook account or a link, will be shown on the image and shared via Facebook. Diigo enables users to annotate, archive and organize web pages. It is easy for students to participate in collaborative research projects, and ask questions while reading web pages. Meanwhile, Reframe It puts users' comments in a sidebar that changes for any page. Users can comment on any page, even if it does not have a comment function.

However, none of the above collaborative learning services provide a convenient sharing function, such as the capability to update shared collaborative content. Moreover, users cannot participate in collaborative writing on the shared collaborative content. For example, after sharing collaborative content via the embedded functionality of Google Docs, users cannot update the new version of embedded content as the original content changes, or provide feedback, such as comments or annotations. Tagtoo allows users to embed tags in web pages, but a tagged image cannot be shared with other web pages.

Using other services to share collaborative content also has some limitations. For example, Calameo [13] enables users to upload PDF or PowerPoint files and convert them into embeddable Flash files, which can then be used to embed collaborative content in web pages. Although users can comment on shared content, they cannot transfer their comments to the original collaborative work or show the comments on the same content embedded in other web pages. Another platform called Apture [14] allows publishers and bloggers to link and integrate multimedia or web pages in a dynamic layer above their own pages. Users can view the latest version of the content, but they cannot give feedback on it.

To update web annotations, we exploit the annotation functions presented in our previous work [15]. The functions described in this paper allow users to add annotations to multimedia on web pages. They also enable users to interact with one another and maintain the consistency of annotations of the same multimedia on different web pages. Furthermore, we introduce a function that synchronizes the annotations of multimedia in the collaborative content with the annotations of the original web multimedia.

III. THE FEATURES OF CCPUBLISHER

CCPublisher is comprised of three components: the Collaborative Multimedia Web Annotator (called the "Web Annotator" hereafter), the Collaborative Content Editor (CCEditor), and the Collaborative Content Reader (CCReader).

Fig. 1 shows the process for publishing content with the CCPublisher. First, the web page loaded by a user is embedded by the Web Annotator (step 1). The Web Annotator enables users to annotate web multimedia content that lacks an annotation capability and transfers the annotated multimedia content to the CCEditor for editing (step 2). On completion of the editing, the user can publish his/her collaborative content via the CCReader (step 3), which can be embedded in the web pages with an embedded code (step 4). We describe each component in detail below.

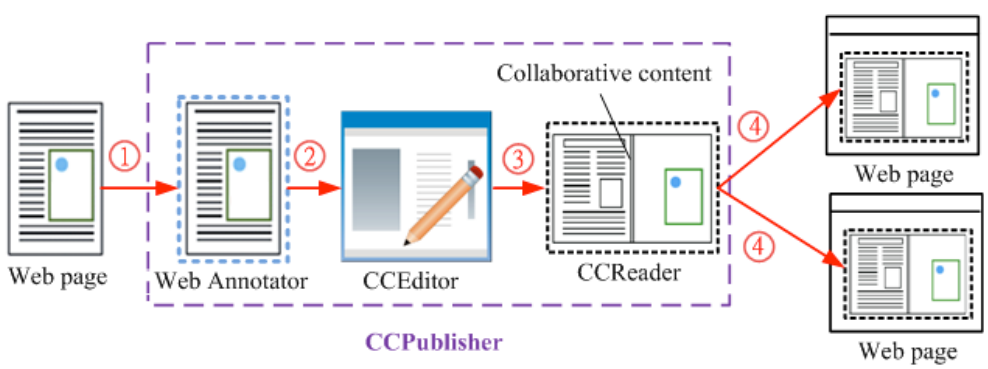


Figure 1. The Publishing Process of CCPublisher.

A. Collaborative Multimedia Web Annotator

To annotate multimedia content on web pages that lack an annotation capability, the Web Annotator uses a parser to embed the annotation function in the web multimedia content. Fig. 2 shows an example of an image on a web page embedded with the annotation function. The annotation tool allows users to draw lines or type text on existing web content. After completing the annotation process the author can select the public mode for other users to view or re-edit the annotations, or select the private mode so that only he/she can change the annotation. Any annotations are displayed with the multimedia and recorded on the Multimedia Annotation Record so that users can track who made a particular change. The annotations made by a user will be loaded into the CCEditor for further editing.

B. Collaborative Content Editor

The CCEditor is used to publish collaborative content with the CCReader. The interface of the CCEditor is shown in Fig. 3. The editing tools on the left-hand side of the interface are: add lines/figures and text, or adjust the thickness, transparency, and color. The author can import materials from other web pages or local computer. The multimedia that was annotated by the Web Annotator will also be loaded into the CCEditor.

If the author wants to import the annotations added to the multimedia by the Web Annotator, CCEditor or CCReader, he/she can click on the multimedia. The CCEditor will then display a pop-up window that shows the Multimedia Annotation Record (Fig. 4). The author can import and edit the annotations.

Figure 2. An image embedded by the Web Annotator.

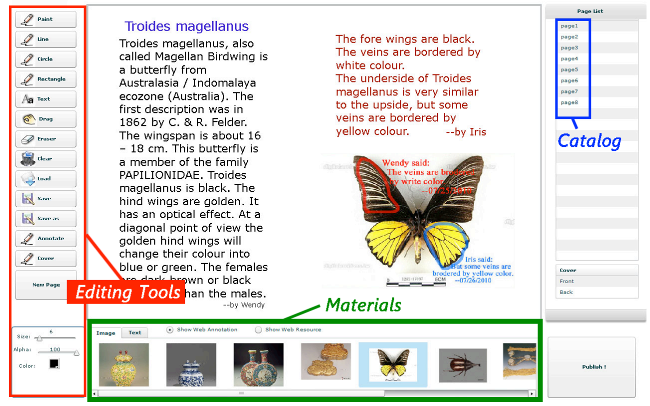


Figure 3. The Interface of the Collaborative Content Editor.

Figure 4. An example of a pop-up window that shows the Multimedia Annotation Record.

The CCEditor allows the author to store changes as he/she complete each page, as well as sort the pages in the catalog and review the work. On completion of the editing, the CCEditor generates the collaborative content with the CCReader, which then creates an embedded code and a URL for the author to post the collaborative content on blogs or elsewhere. The CCPublisher will also store the collaborative content so that the author can go back to the CCEditor to re-edit it.

The CCEditor provides three types of sharing mechanism. First, the author can select the public mode, which allows anyone on the Internet to modify and republish any part of the content. Second, the author can set content as "read-only" so that others can only view the material. Third, the author can restrict access so that only nominated members of a group can modify and republish the content. In addition, to prevent other users making inappropriate modifications of the collaborative content, the author can vet changes and decide which ones to accept.

C. Collaborative Content Reader

The CCReader enables users to view and make feedback on the collaborative content that the CCEditor generates

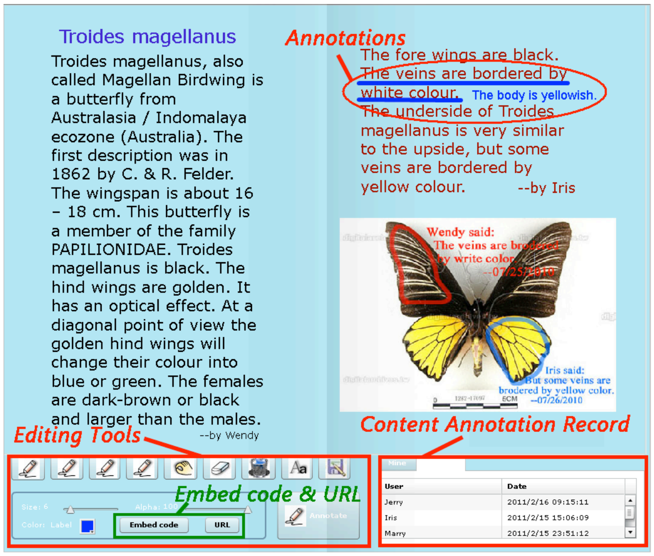


Figure 5. The Interface of Collaborative Content Reader.

and embeds on web pages. Fig. 5 shows an example of the CCReader interface that authors can use to add annotations with the editing tools. If users want to share the collaborative content via the CCReader, the embedded code and URL are provided.

Any part of the collaborative content can be modified and republished if the author has given permission. The changes will be recorded on the Content Annotation Record so that everyone can track who made a particular change, and the author can decide which version to accept.

In addition, users can import annotations added to the multimedia by the Web Annotator, CCEditor or CCReader. When a user clicks on the multimedia, the CCReader will display a pop-up window that shows the Multimedia Annotation Record for users to import annotations. The interface of the pop-up window is the same as that in Fig. 4. It also provides editing tools for users to edit existing annotations or add new ones. The changes will be displayed on the multimedia in the collaborative content and transferred to the original multimedia resource embedded by the Web Annotator. (We describe this mechanism in Section 4).

CCPublisher is implemented with PHP and Adobe FLEX. PHP is used to implement the web applications, transfer the annotation record and access the database; and Adobe FLEX is used to implement the interfaces that communicate between the web browser and the web server. We use the Apache HTTP Server 2.0 as the web server and MySQL as the database.

IV. The Synchronization Mechanism of CCPublisher

The synchronization mechanism ensures that all versions of shared collaborative content are consistent. As collaborative content may be posted on two or more web pages for users to access, CCPublisher uses a synchronization mechanism to maintain the consistency of all shared content. Users can choose to synchronize all copies of the collaborative content or synchronize the multimedia annotations in the collaborative content and the Web Annotator.

A. Synchronizing copies of collaborative content

Fig. 6 shows an example of synchronizing copies of collaborative content. The content and copies are posted on web pages A, B and C. After one of the pages has been modified, the change is sent to CCPublisher (step 1), which stores the modification and transfers it to copies of the collaborative content posted on other web pages when users access the CCReader (step 2). After users update their copies, the modification will appear in the collaborative content (step 3).

B. Synchronizing multimedia annotations in collaborative content and the Web Annotator

Annotations of multimedia in different versions of the collaborative content and the Web Annotator will be consistent if the authors utilize the same multimedia resource. Fig. 7 shows an example of synchronizing multimedia annotations in collaborative content and the Web Annotator.

Content A and content B contain the same image obtained from the same web page. When reading content A, the user annotates the image, and the new annotation is transferred to the CCPublisher (step 1). When other users access content B, the annotation will be transferred to content B by the CCPublisher (step 2). After users update this annotation, it will be displayed in the image (step 3). The annotation will also be transferred to the original image resource embedded by the Web Annotator (step 4), and appear on the Web Annotator after users update the annotation (step 5). Similarly, when authors use the Web Annotator to annotate multimedia on web pages, the annotations are loaded to the multimedia in the embedded collaborative content by the CCPublisher.

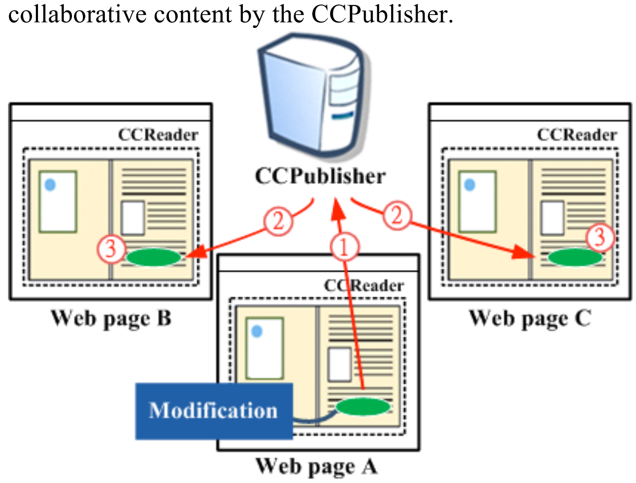


Figure 6. Synchronizing copies of collaborative content

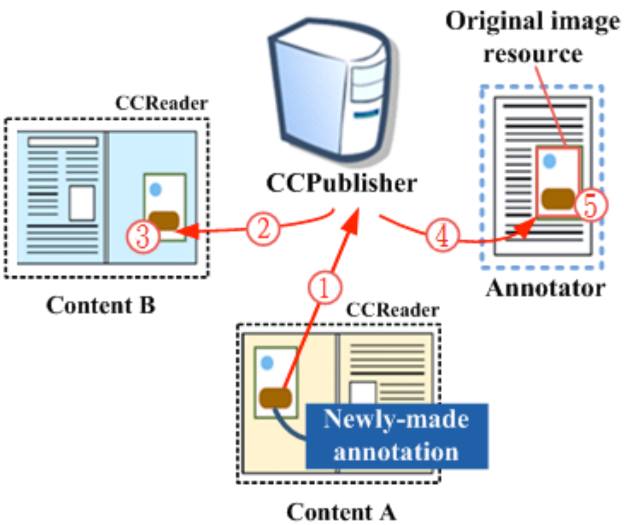


Figure 7. Synchronizing multimedia annotations in collaborative content and the Web Annotator.

In summary, the CCPublisher synchronizes all shared versions of the collaborative content. It also synchronizes the multimedia annotations in the collaborative content and the Web Annotator. The synchronization mechanism allows users to exchange ideas and participate in collaborative writing more efficiently by providing immediate feedback and more communication opportunities.

V. CONCLUSION

We have presented a web application called CCPublisher, which enables users to participate in collaborative writing of the content posted on web pages and helps authors synchronize different versions of the content. CCPublisher also solves the problem of one-way knowledge sharing and the inconvenience of updating different versions of shared content. In addition, we provide a web annotation function for users to collaborate with each other on web content that lacks an annotation capability. The annotations added to the web content can be integrated into the collaborative content. They will also be consistent with their original multimedia annotations. CCPublisher allows users to provide feedback, exchange ideas on any web page, and enrich the shared collaborative content as well as the web content. It therefore facilitates the accumulation of knowledge and provides more opportunities for collaboration.

In our future work, we will expand the functionality of collaborative learning, e.g., by developing the annotation capability to include video annotation. We also hope to implement CCPublisher on portable readers, thereby providing users with more mobility. Moreover, we hope that, in the near future, all users will be able to access the source code to expand the functionality of collaborative learning or implement our application on other e-learning platforms. Adding these functions would increase the convenience and utility of CCPublisher and enhance the concept of collaborative learning.

ACKNOWLEDGMENT

This research was partially supported by the Taiwan E-learning and Digital Archives Program (TELDAP), which is sponsored by the National Science Council of Taiwan under NSC Grants: NSC100-2631-H-001-020, NSC100-2631-H-001-012

REFERENCES

[1] S. Downes, “E-Learning 2.0,” ACM eLearn Magazine: Education and Technology in perspective, Oct. 2005. http://elearnmag.acm.org/featured.cfm?aid=1104968.

[2] P. Dillenbourg, “What do you mean by 'collaborative learning'?,” In Collaborative-learning: Cognitive and Computational Approaches, P. Dillenbourg (Ed), Oxford: Elsevier, 1999, pp.1-19.

[3] A. A. Gokhale. “Collaborative Learning Enhances Critical Thinking,” Journal of Technology Education, vol. 7, 1995, pp.22–30.

[4] R. Giles, How to Use Flickr: The Digital Photography, 1rd ed., Course Technology PTR, 2006.

[5] S. O'Hear, “e-learning 2.0 - how Web technologies are shaping education,” ReadWriteWeb, Aug. 8, 2006, http://www.readwriteweb.com/archives/e-learning_20.php.

[6] P. Schacht, “The Collaborative Writing Project,” in Using Wiki in Education, sec. 2, S. Mader Eds. 2007, pp.12-23.

[7] P. Ayers, C. Matthews and B. Yates, How Wikipedia Works: And How You Can Be a Part of It, 1rd ed., No Starch Press, 2008.

[8] K. Brandle, “Are you ready to ‘MOODLE’?,” In Proc. of Language Learning & Technology, vol.9(2), 2005, pp. 16-23.

[9] Google Doc, http://www.google.com/google-d-s/b1.html.

[10] Tagtoo, http://www.tagtoo.org, 07.18.2011.

[11] B. Yuan, and B. Tan. “Explore Diigo for the First Time Applying to Learning Hitting the Target,” In Proc. of the Second International Symposium on Networking and Network Security, 2010, pp. 133-135.

[12] Reframe It, http://reframeit.com, 07.18.2011.

[13] Calameo, Publish and share documents, http://www.calameo.com, 07.18.2011.

[14] Apture, http://www.apture.com, 07.18.2011.

[15] C. Lin, F. P. Chu, H. A. Wang, and J. C. Ko, “An Interactive Multimedia Web Editor,” In Proc. of 8th IEEE Fourth International Conference, IEEE Press, 2008, pp. 341–342.

 ISBN: 978-1-61208-157-1

Using Frame-based Lexical Chains for Extracting Key Points from Texts

Sudabeh Mohamadi
University of Tehran
Tehran, Iran
E-mail: sumohamadi@ut.ac.ir

Kambiz Badie
Iran Telecom Research Center
Tehran, Iran
E-mail: k_badie@itrc.ac.ir

Ali Moeini
University of Tehran
Tehran, Iran
E-mail: moeini@ut.ac.ir

***Abstract*—In last years, many automatic systems have been designed for text summarization and extracting key phrases, but still no system has been suggested for extracting key points. According to our definition, key points are high-level concepts extractable from a text that consist of words that may not necessarily exist in the original text. In this paper, we try to design an automatic system for extracting key points by using lexical chains. In this system, we use FrameNet for shallow semantic parsing of texts. As the final attempt, we present the set of tuples that contain important concepts of an original text with the related semantic roles. With use of generalization from parts onto whole, we can then have the claim of extracting a higher-level concept, which stands for a key point. Comparing the output of this system with human abstract, we perceived that 42 percent of cases generated by this system are similar to those generated by human.**

Keywords-automatic summarization; keyphrase extraction; abstract; lexical chain; generalization.

I. INTRODUCTION

Recently, the documents data are remarkably increasing and we need to have access to these data easily and rapidly. These data may belong to video, sound, text or image format. Text data may exist in web pages, books, articles, emails, documents of organizations, etc. Using these data leads to consuming much time to the extent that finding needed information becomes very hard and sometimes impossible. One of the ways for fast and suitable access to text information is automatic summarization of text [1]. The goal of automatic summarization is to take an information source, extract content from it, and present the most important content to the user in a condensed form and in a manner sensitive to the user's or application's need. In fact, summary is a brief statement that presents the main points in a concise form. There are two types of summary: extract and abstract.

In the former, summary is formed by reusing the portions of the main text like words or sentences. Words sequences that come into summary are the same as that of the main text. The words sequence can be used as phrases, sentences or paragraphs. In this type of summary, the most important information of the main text is copied to the final summary.

In an abstract, the content is an interpretation of an original text. In fact, abstracts consist of new phrases that describe the content of the original text. In this type of summary, words sequences that come into abstract are not the same as words sequence of the original text. Producing an abstract contains topic fusion and text generation stages. The main problem with text generation is that the new text should contain several sentences that must be coherent. For some specific applications, summary and abstract may be not suitable, as their structure of sentences may be complicated. For example, in search engines, we need to match key phrases or key points of texts with these of in the query [2]. In these cases, we can use key phrases or key points instead of summary. Key point and key phrase extraction is performed at word level while text summarization is performed at sentence level. Key phrases and key points can be considered as sets of words or concepts that present a brief representation of the original text. Key phrase extraction is highly related to automated text summarization where the most indicative words in a document are selected as key phrases.

In contrast with text summarization, key point extraction does not require coherence between sentences. In our definition, key points represent important concepts in the text that have the semantic relation with central topic of the original text. They consist of words that may not be necessarily in the original text. We can consider key points as a set of phrases that are semantically related to most of the portions of the text. In this paper, we try to extract key points.

Although information about text obtained from abstract are more than key points and key points cannot be considered as the alternatives for abstract but we can use them in specific applications such as indexing in search engines or text categorization. In addition, the key points

 ISBN: 978-1-61208-157-1

assigned to the text can help the reader distinguish whether a document is relevant or not.

As mentioned before, key points must have the most relevance with the concepts in the original text. The best way for presenting relevance between words and senses is using lexical chains. Lexical chains contain a set of words or senses related to each other with semantic relations. The words and senses in the same lexical chain are related to each other from the viewpoint of semantic relation. Portions of the text that are covered by each lexical chain are different from other lexical chains. Furthermore, the number of words and the type of relations between words would be different for each lexical chain [3]. Different criteria exist for measuring the strength of lexical chain. The number of words and the type of relations between them are particularly important here. We try extracting stronger chains because the strength of a chain indicates its importance [4].

There are several lexical resources for computing semantic distance between two nodes of lexical chains. For instance: Dictionaries, Thesauri, semantic nets, WordNet and FrameNet [5]. Most of the text summarization or key phrase extraction systems make use of WordNet ontology.

In this paper, for the first time, lexical chains are built by FrameNet ontology [6]. After the strength of chains was extracted, the extracted frames are generalized from parts to whole to obtain a high-level of concept. In the key phrase extraction or summarization systems, this stage does not exist. It is exclusive for our system. Two methods exist for this generalization. In one method, we generalize two sub-frames to a super-frame when both have the same super-frame. In second method, where the intermediate frame is super for the first frame and sub for other, we can generalize the first frame to the other frame. Finally, we obtain the list of the tuples that present the important concept of the original text with the related semantic roles. Output of this system can be used in clustering and classification properly.

The paper is organized as follows. In Section 2, we present the related work. The suggested approach includes five stages that would be explained in Section 3. We present the experimental results and the evaluation in Section 4. Finally, we conclude and suggest possible future improvements in Section 5.

II. OVERVIEW ON EXISTING APPROACHES

Currently, there is no system for key point extraction. However, many other technologies, such as Automatic Summarization [1], Information Retrieval [7] and keyword and key phrase Extraction [2] can be mentioned. In this section, we present a brief review on these technologies. The focus here is specifically on review of Automatic Key Phrase Extraction systems.

In 1999, Witten et al. [8] presented KEA algorithm for automatically extracting key phrases from text. KEA identifies candidate key phrases using lexical methods, calculates feature values for each candidate, and uses a machine-learning algorithm to predict which candidates can be suitable as key phrases. KEA's extraction algorithm has two stages: (1) Training that creates a model for identifying key phrases, using training documents where the author's key phrases are known. (2) Extraction that chooses key phrases from a new document. KEA finds less than half of the author's key phrases.

In 2000, Turney [2][9] used an approach for automatically extracting key phrases from texts as a supervised learning task. He performed two types of experiments to test his approach. His first set of experiments applied the C4.5 decision tree induction algorithm to this learning task and the second set of experiments applied the GenEx algorithm to the task. The experimental results showed that GenEx algorithm could generate better key phrases than C4.5 algorithm.

Avanzo and Magnini [10] presented the LAKE System (Learning Algorithm for Key phrase Extraction) that first considered a number of linguistic features to extract a list of candidate key phrases, then used a machine learning framework to select significant key phrases for a document. The two features that they used are reasonably effective but they did not consider any semantic features of key phrases. This system utilized key phrases extraction for summarization.

Turney and KEA algorithm used first occurrence position in text and frequency based features. Later, Hult [11] extended their systems by integrating more linguistic features like part of speech tags. He used four features: term frequency, collection frequency, relative position of the first occurrence and the POS tag(S) assigned to the term.

Hulth improved automatic keyword extraction, using more linguistic knowledge. He used supervised machine learning algorithm by adding linguistic knowledge to the representation such as syntactic features, rather than relying only on statistics such as term frequency and n-grams. He showed that keyword extraction from abstracts can be achieved by using simple statistical measures as well as syntactic information from the documents. He used approaches such as n-gram; chunking and pattern, then computed recall, precision and f-score for these approaches and then compared them. Extracting with chunking approach gives a better precision, while extracting all words or sequences of words matching any of a set of POS tag patterns gives a higher recall. The highest f-score is obtained by n-gram approach [11].

Ercan and Cicekli [12][13] are the first to use the lexical chains in keywords extraction. They proceeded to automatic keyword extraction of texts by supervised learning algorithm. They used lexical chains for this task and built them using the WordNet ontology. Ercan and Cicekli extracted keywords instead of key phrases because most of the phrases did not exist in WordNet data source. Thus,

 ISBN: 978-1-61208-157-1

lexical chains were constructed just for words. They used seven features. Four of which are lexical chain's features. Then evaluated different combination of the seven features and concluded that lexical chain's features improves keyword extraction task. Their lexical chain based features focus on members of the lexical chains rather than the whole lexical chain.

In 2010, Sarkar et al. [14] presented a neural network based approach to key phrase extraction from scientific articles. For predicting whether a phrase is a key phrase or not, they used the estimated class probabilities as the confidence scores which are used in re-ranking the phrases belonging to a class: positive or negative and they finally compared their system with KEA and concluded that their proposed system performs better than KEA.

III. THE PROPOSED APPROACH

The suggested approach includes five stages that would be explained below.

A. Segmentation

In the first stage, after acquiring the input text, it must be segmented by a segmenter. The main reason of segmentation is to prevent from construction of huge chains. If lexical chains are constructed in the whole text, the size of chains becomes very large and consequently a large space is needed for their storage. On the other side, construction of chains in the entire text is very time consuming, because we must check the relation between each frame with the others for the whole text. Therefore, we divided the original text into smaller segments and then constructed chains in these segments.

One of the ways of text segmentation is to use text segmenters. The duty of the text segmenter is to divide the original text into segments that represent the same topic. It tries to break the text into thematically meaningful segments. There are several applications for text segmentation as text segmenter. One of these applications is Marphadorner. Marphadorner implements two linear segmentation methods, which use measures of lexical cohesion to produce segments: Marti Hearst's TextTiler [15] and Freddy Choi's [16]. Both try to find those portions of a text in which the vocabulary changes from one subtopic to another.

B. Shallow Semantic Parsing of Input Text

After original text was segmented, it must be parsed semantically. For this task, we use FrameNet dataset. In fact, in this stage, syntax and semantic structure of original text are identified. One of the applications for this goal is SHALMANESER. It is a SHALlow seMANtic parSER used to assign semantic classes –frames– and semantic roles –frame elements (FEs) – to original text automatically. To do so, it performs two stages. Firstly, disambiguates word senses that correspond to semantic classes with FRED and then assigns semantic roles by ROSY. The dataset for SHALMANESER is the FrameNet dataset [17].

C. Constructing Lexical Chains

Lexical chain construction is performed in three stages as follow:

1) Select Candidate Terms

Our goal is to extract the key points or the key concepts of the text, hence we consider frames as candidate terms that present concepts. The frame that is assigned to the lexical unit, expresses the concept of that lexical unit in the special position. So, the frame can be considered as the concept of its lexical unit, because when the word evokes a frame, it means that the frame is one of the word's concepts.

2) Select Appropriate Chain

We use FrameNet for recognition of the relation between frames and computing the semantic distance of frames as a lexical resource. In this algorithm, three types of relations are defined:

a) the extra-strong relation type: between a frame and its repetition occur.

b) the strong relation type: between two frames connected with one of frame-to-frame relation like these:

Inheritance- perspective on- sub frame- precedes- inchoative of- causative of- using [6]. You can also see details of these relations in FrameNet project. In the strong relation, two frames are connected directly.

c) the medium-strong relation type: between two frames connected to each other using another frame that is called intermediate.

In this algorithm, we just consider one frame as intermediate but to improve an extended algorithm, we can use relation with two or more intermediate frames.

3) Insert the Frame in the Chain

To select an appropriate chain, we added frames in order to place in the paragraph. Suppose n chains were constructed and now we want to find an appropriate chain for frame a. At first, we investigate the relation between a and each frame in chain j of n. If frame a has the extra-strong relation with one of the frames of j, a belongs to chain j. otherwise we must check strong relation like extra-strong relation for a. if strong relation was not found for it, we investigate medium-strong the same as other two relations.

According to this priority, we find the appropriate chains for the candidate frame. Three types of state can occur. If no appropriate chain is found, then a new chain is created and the candidate frame is inserted into it. Whenever, one appropriate chain was found, the frame is inserted into it. If two appropriate chains were defined, they are joined to each other. When the candidate frame is inserted in the chain, the chain will be updated. The new frame is then connected to the other frames in chain according to their types of relation with them.

Algorithm 1 is the pseudo-code describing lexical chains construction.

 ISBN: 978-1-61208-157-1

Algorithm 1: Lexical Chains Constructor (Frames)

```
Start
for each Frame a from (1…m) do
   for each Chain j from (1…n) do
      for each Frame(1…k) in Chain(1…n) do
            if Frame_a has Extra-strong relation with Frame_k then
                  Add Frame_a to Chain_j
                  Update Chain_j
                  break
      else if Frame_a has strong relation with Frame_k then
                  Add Frame_a to Chain_j
                  Update Chain_j
                  break
      else if Frame_a has Medium-strong relation with Frame_k then
                  Add Frame_a to Chain_j
                  Update Chain_j
         end if
         else ConstructNewChain(Frame_a)
      end for
   end for
end for
end
```

D. Semantic Distance Between Frames

The semantic distance between frames depends on the relation type between them. We described three types of relations and according to them we must define three values for semantic distance.

Barzilay and Elhadad experimented different states and concluded that the following weights can be the best. So we use those in this approach [4].

In the first type, 10 should be added to the distance for each repetition. For example, if repetition of the frame is two in one paragraph , the distance becomes 10 and if it repeats n times in the paragraph, we must add " (n-1)*10" to the distance of chain. In the second type, where two frames are connected directly, the distance would be equal to 7. In the third type, when two frames are connected with other frame as intermediate, the distance would be 4.

E. Scoring of Chains

After the original text is converted to several chains by the presented algorithm, in this stage, we must identify the strongest chains for extraction. There is no formal way to evaluate chain strength, as there is no formal method to evaluate the quality of key points. Hence, we rely on an empirical methodology.

In our approach, we select the number of texts. The text has been parsed by using FrameNet dataset. In the beginning, we construct chains for those described above, and then we score chains according to different criterion. There are several formal measures on the chain for scoring as follows: chain length, number of chain's member and weight of relations between members of chains. Elhadad and Barzilay have presented other criteria like: distribution in the text, text span covered by the chain, density, graph topology and number of repetition. They concluded that only the length is a good predictor of the strength of a chain. They supposed that the length is the number of occurrences of members of the chain that we call number of chain's member [3][4].

In our algorithm, we use four different features for scoring the chains

Feature 1: the number of chain members

In this feature, we compute the number of chain's member where score of each chain is equal to this.

Feature 2: sum of the weight of frame-to-frame relations

In this feature, score of chain is equal to sum of the weight of relations. The way of scoring was described previously. Notice that whenever there is more than one type of relation between two frames, only the relation with more weight is considered.

Feature 3: Feature 1+Feature 2

This feature is created by sum of two former features. Sometimes, the number of frames is high, while the semantic distance between them is weak. This feature balances them.

Feature 4: score (chain) = length * homogeneity

Barzilay and Elhadad experimented different features and concluded that this feature is the best for extracting keywords. In this formula

(1) Homogeneity =1- (number of distinct occurrence / length).

IV. Experimental Results and Evaluation

It should be mentioned in the beginning of this section that since no system still exists for extracting key points in the way we have elaborated in this paper, our basis for comparison is just human being who is asked to extract the key points from texts in an intuitive manner.

A. Extracting the Strong Chains

In this stage, we must extract chains with maximum score. For this goal, we need to have one threshold. A selected threshold for this algorithm is: average (scores) +2 * standard deviation (scores) the same as Brazilay and Elhadad criterion. So we recognize chains with scores higher than threshold as strong chains and extract them for use in key points. In fact, the chain would be extracted if

(2) Score (chain)>average (scores) + 2*standard deviation.

B. Generalization From Parts to Whole

In this stage, we perform generalization from parts to whole. If two frames have the medium strong relation, that means they are connected to each other with an intermediate frame, and we can generalize them in two ways. In this state, we achieve a higher level of concepts. For example

(1) Alice writes a note with pen.

(2) Alice draws a plan with pencil.

In (1), author writes with pen tool and in (2) creator draws with pencil tool. Write evokes Text_creation frame and draw evokes Create_physical_artwork. Relation of them is the same as Figure 1.

The Intentionally_create frame is the intermediate frame for Text_creation and Create_physical_artwork frames. This frame is not seen in the original text but it is a super-frame for the other two and both of them can be generalized to Intentionally_create. This frame has two core frame elements: creator and created_entity. In these sentences, creator is assigned to Alice and creator_entity is assigned to note and plan. "instrument" is one of the non_core elements for the Intentionally_create that has been evoked by pen and pencil. As a consequence, after generalization, the following tuples are created.

(Intentionally_create, Creator, Creator_entity, instrument)
(Intentionally_create, Alice, Note/plan, pen/pencil).

In this example, the intermediate frame is super-frame for both frames, so both frames are generalized to this frame. If intermediate frame is super-frame for one of the frames and sub-frame for the other, we can generalize sub-frame of intermediate to super-frame of it. In Figure 2, frame 2 is the intermediate for 3 and 1. Therefore, frame 1 is generalized to 3.

This stage is the final stage for key point extraction systems. After this, we achieve the list of tuples that contains a frame as the first member and frame elements as the other members. These tuples indicate the main concepts in original texts.

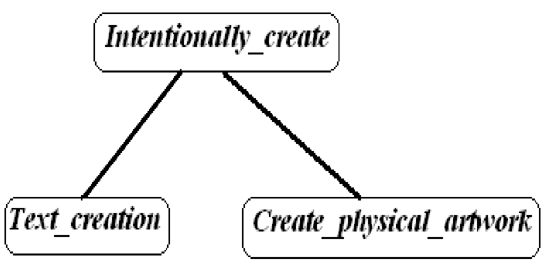


Figure 1. Example of relation between two frame

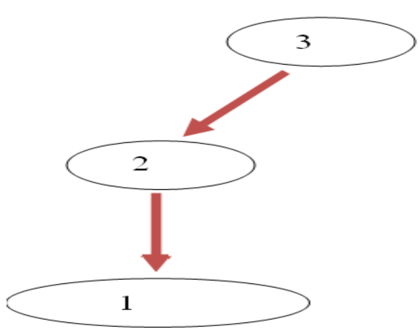


Figure 2. Generalization from parts to whole

C. Evaluation

For evaluation of this system, we use the texts that exist in FrameNet project of Berkeley [6]. FrameNet contains a number of full texts with their annotation and FE and frames would be included in them. We choose five of them to experiment our algorithm. Our system extracts key points from these, and then compares the output of automatic system with human extraction key points. Five students are chosen for extracting key points from the full texts. Two texts are given to each of them so each text is investigated by two students. We can use three performance criteria to evaluate this system [2].

(3) Recall = correct / (correct+missed)

(4) Precision = correct / (correct+wrong)

(5) F-measure = (2*recall*precision) / (recall+precision)

As it is seen in Table 1 to Table 3, Feature 1, Feature 2 and Feature 3 are very similar to each other and Feature 4 gives the worst result. The recall of Feature1 is better than the two other features but with regard to precision and f-measure criteria, Feature 2 is better than other features and Feature 3 is better than Feature1.

Feature 4, which has its most emphasis on number of iteration of frames, gives poor results. This feature focuses on the concept frequency and ignores relations between the frames. Since our goal is extracting the key points, the relations between frames are very important. Therefore, the Feature 4 is not suitable.

In Table 4, the average of recall and precision is shown. Comparing these with f-measure, we conclude that the recall and precision are balanced because the average of them is very similar to f-measure.

Table 1. THE RECALL CRITERIA

	Feature 1	Feature 2	Feature 3	Feature 4
Text 1(Madonna)	52%	52%	50%	36%
Text 2(Stephanopoulos Crimes)	31%	28%	28%	25%
Text 3(Bell Ringing)	42%	42%	42%	42%
Text 4(Loma Prieta)	34%	34%	34%	16%
Text 5(Dublin)	62%	62%	60%	61%
Average	44%	43%	42%	36%

Table 2. THE PRECISION CRITERIA

	Feature 1	Feature 2	Feature 3	Feature 4
Text 1(Madonna)	44%	44%	44%	25%
Text 2(Stephanopoulos Crimes)	11%	14%	14%	6%
Text 3(Bell Ringing)	30%	30%	30%	30%
Text 4(Loma Prieta)	60%	60%	60%	58%
Text 5(Dublin)	67%	75%	73%	73%
Average	42%	44.6%	44.2%	38%

 ISBN: 978-1-61208-157-1

Table 3. THE F-MEASURE CRITERIA

	Feature 1	Feature 2	Feature 3	Feature 4
Text 1(Madonna)	47%	47%	47%	29%
Text 2(Stephanopoulos Crimes)	16%	18%	18%	35%
Text 3(Bell Ringing)	35%	35%	35%	35%
Text 4(Loma Prieta)	43%	43%	43%	22%
Text 5(Dublin)	64%	67%	65%	68%
Average	41%	42%	41%	37%

Table 4. THE AVERAGE OF RECALL AND PRECISION

	Feature 1	Feature 2	Feature 3	Feature 4
Text 1(Madonna)	48%	48%	47%	30%
Text 2(Stephanopoulos Crimes)	21%	21%	21%	15%
Text 3(Bell Ringing)	36%	36%	36%	36%
Text 4(Loma Prieta)	47%	47%	47%	37%
Text 5(Dublin)	63%	69%	66%	67%

In these formulae, the variable *correct* represents the number of times that the human-generated key phrase matches the machine-generated key phrase. The *wrong* variable represents the number of concepts that the machine extracts but the human does not. The *missed* variable represents the number of concepts, which are extracted by human but not by machine. These performance criteria have been brought in Table 1, 2 and 3.

V. CONCLUSION AND FUTURE WORK

There are some systems for summarization and extracting key phrases from text, but there is no system for key point extraction. In this paper, we presented an approach to the task of extracting key points and concepts from English texts. For this goal, we have used lexical chains that are constructed based on FrameNet ontology. We then experimented four features based on lexical chains and achieved the expected results. This system extracts key points as high-level concepts from the original text. In 42 percent of the cases, the concept which generated by this system is equal to the concept generated by human. Although the output is complicated and difficult to understand for usual users but this approach is very useful in classifying and clustering systems.

The suggested system depends much on semantic parsing systems. Therefore, the extension of our system would call for improvement of semantic parsing systems. One of the shortcomings of this work is that we only consider one intermediate frame for the third type of relation. In future, relation with more intermediate frames can also be considered. In addition, in this work, just the medium-strong relation generalizes from parts to whole. In the future, strong relation can be considered too. Also, we can investigate more features which are based on concepts in the chains instead of the whole chains.

REFERENCES

[1] V. Gupta and G.S. Lehal, “A Survey of Text Summarization Extractive Techniques” Journal of Emerging Technologies in Web Intelligence, Vol 2, pp. 258-268, Aug 2010.

[2] P.D. Turney, “Learning to extract keyphrases from text”, National Research Council, Institute for Information Technology, Technical Report, 2002.

[3] C. Santos, “Alexia - acquisition of lexical chains for text summarization”, Master’s thesis, University of Beira Interior, 2006.

[4] R. Barzilay and M. Elhadad, “Using lexical chains for text summarization”, Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization, pp. 111-121, 1997.

[5] G. Hirst and A. Budanitsky, “Lexical chains and semantic distance”, Proc. Eurolan, 2001.

[6] J. Ruppenhofer, M. Ellsworth, M. Petruck, and C. Johnson, “FrameNet II: Extended theory and practice”, International Computer Science Institute,University of California at Berkeley, 2006.

[7] E. Greengrass, “Information Retrieval: A Survey”, Citeseer, DOD Technical Report TR-R52-008-001, November 2000.

[8] I.H. Witten, G.W. Paynter., E. Frank, and C. Gutwin, “KEA: Practical automatic keyphrase extraction”, In Proc. of the 4th ACM Conference on Digital Libraries, pp. 254-256, 1999.

[9] P.D. Turney, “Learning algorithm for keyphrase extraction”, Journal of Information Retrieval, pp. 303-336, 2000.

[10] E.D. Avanzo and B. Magnini , “Keyphrase extraction for summarization purposes: The LAKE system at DUC-2004”, Proceedings of the Document Understanding Conference, Boston, USA., 2004.

[11] A. Hulth, “Improved automatic keyword extraction given more linguistic knowledge”, In Proceedings of the 2003 Conference on Empirical Methods in Natural Language processing Japan, August 2003.

[12] G. Ercan and L. Cicekli, “Using lexical chains for keyword extraction”, Information Processing & Management, pp. 1705-1714, 2007.

[13] G. Ercan, “Automated text summarization and keyphrase extraction”, Master thesis, 2006.

[14] K. Sarkar, M. Nasipuri, and S. Ghose, “A new approach to keyphrase extraction using neural networks”, Vol. 7, Issue 2, No 3, pp. 16-25, Arxiv preprint arXiv:1004.3274, 2010.

[15] M.A. Hearst, “TextTiling: Segmenting text into multi-paragraph subtopic passages”, Computational linguistics, pp. 33-64, 1997.

[16] F.Y.Y. Choi, “Advances in domain independent linear text segmentation”, Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, pp. 26-33, 2000.

[17] K. Erk and S. Pado, “SHALMANESER: a toolchain for Shallow Semantic Parsing”, Proceedings of LREC, Genoa, Italy, 2006.

PPETP: A Peer-to-Peer Overlay Multicast Protocol for Multimedia Streaming

Riccardo Bernardini, Roberto Cesco Fabbro, Roberto Rinaldo
DIEGM – Università di Udine, Via delle Scienze 208, Udine, Italy
{riccardo.bernardini,roberto.cesco,roberto.rinaldo}@uniud.it

***Abstract*—One major issue in multimedia streaming over the Internet is the large bandwidth that is required to serve good quality content to a large audience. In this paper we describe a protocol especially designed for peer-to-peer data distribution to a large number of users. The protocol is suited for the efficient distribution of live multimedia and it can exploit even the limited resources of residential users. Special care was paid to make the protocol *back-compatible* with existent multimedia tools and protocols, so that software and protocols already multicast-enabled require only minor changes to be adapted to the new protocol. The flexibility, the openess and the features of the proposed protocol makes it an interesting solution for streaming content to large audiences.**

***Keywords*-Data transmission; multimedia streaming; overlay multicast; peer-to-peer network; push networks**

I. Introduction

A problem that is currently attracting attention in the research community is the problem of streaming live content to a large number of nodes. The main issue to be solved is due to the amount of upload bandwidth required to the server that, unless multicast is used, is equal to the bandwidth required by a single viewer (some Mb/s for DVD quality) multiplied by the number of viewers (that can be very large, for example, it is reported that in 2009 the average number of viewers per F1 race was approximately $6 \cdot 10^8$).

The upload bandwidth problem is not limited to the "large audience" scenario, but it can also be found at smaller scales. Consider, for example, the case of a medium-size community with 100–1000 members (e.g., a political party or a fan club) that wants its own IPTV channel to stream events to its members or, maybe, organize virtual meetings. If the association aims to "YouTube quality" video (hundreds of Kbit/s), the overall bandwidth is in the order of hundreds of Mbit/s. Although this is within current technology capabilities, the implementation of such services could prove too expensive for the association.

Multicast is of course a possible solution, but it has its drawbacks too. Maybe the most difficult issue in using multicast, in applications like these, is that the audience is expected to be spread among several different Internet Service Providers (ISPs) and multicast across different Autonomous System (AS) is not trivial, both on a technical and on a administrative side.

An approach that recently attracted interest in the research community is the use of peer-to-peer (P2P) solutions [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14]. With the P2P approach each viewer re-sends the received data to other users, implementing what could be roughly defined as an overlay multicast protocol where each user is also a router. Ideally, if each user retransmitted the video to another user, the server would just need to "feed" a handful of nodes and the network would take care of itself. This could be beneficial to both the "large audience" and the "fan club" scenarios.

Unfortunately, the application of the P2P paradigm to multimedia streaming has some difficulties such as

- **Asymmetric bandwidth** Depending on the media type and quality, the typical residential users, connected to the Internet via an ADSL, could have enough download bandwidth to receive the stream, but not enough upload bandwidth to retransmit it. Therefore, the solution of having the user retransmit the content to another user is not applicable and more sophisticated solutions are needed.
- **Heterogeneous nodes** The network can include nodes with different upload capabilities, from residential users with few hundreds kbit/s up to nodes with upload bandwidth of several Mbit/s. A good P2P structure should be able to exploit the bandwidth of each peer as much as possible, both for low-bandwidth and high-bandwidth nodes.
- **Sudden departures** A node can leave the network at any time, possibly leaving other nodes without data for a long time.
- **Security** P2P networks have several security issues [15]. Here we simply cite the *stream poisoning attack* where a node sends incorrect packets which cause an incorrect decoding and are propagated to the whole network by the P2P1 mechanism.
- **Network Address Translators (NAT)** Several residential users are behind at least the NAT built-in in their modem and this is a problem for P2P solutions since the NAT makes the user PC unreachable by outside peers.

This article describes the *Peer-to-Peer Epi-Transport Protocol* (PPETP), a peer-to-peer protocol developed as part of the project *Corallo* hosted on *SourceForge*. This paper is organized as follows: Section II describes the design goals that guided the development of PPETP, Section III gives an overview of PPETP and introduces some "PPETP jargon," Section IV describes the idea of "reduction procedure" that is at the core of PPETP, Section V shows the similarities between PPETP and IP multicast and how these similarities allow one to reuse with PPETP all the tools (protocols, formats, and so

 ISBN: 978-1-61208-157-1

on) developed for multicast with a minimal change, Section VI illustrates some practical examples of use, Section VII shows some results about some performance aspects of PPETP, Section VIII presents the conclusions.

II. Design goals

Our objective was to design a protocol that would solve the problems enumerated above, that could be used in several applicative contexts (not just video streaming) and that would look, from the application level, like another transport protocol. More precisely, our design goals were

- **Multicast-like structure** From the application level the protocol must look like a multicast protocol, with an API (Application Programming Interface) similar to the well-known BSD sockets. This would simplify the integration of the new protocol in existing protocols and applications.
- **Usability with heterogeneous networks** The system must be able to exploit efficiently the bandwidth of each user, both for high- and low-upload bandwidth users.
- **Robustness with respect to data losses** In particular, the video must not stop even if one or more peer suddenly leave the network.
- **Security** The protocol must counteract possible attacks that a malicious user could try. In particular, it must be difficult to poison the data stream and, in case of an attack, it should be possible to find out the culprit.
- **Usability with NAT** The protocol must take care by itself of the possible presence of NATs. The application programmer should not worry about the presence of NATs.
- **Usability with any data type** Like a true transport protocol, the developed protocol must be able to carry any type of data.
- **Flexible topology** The protocol must not be tied to a specific network topology, nor to a specific peer-discovery technique, but it must be possible to use it with different peer discovery procedures and topologies.

Remark II.1 (What PPETP is not)

A P2P streaming system is a complex piece of software that must take care of several things: transferring data, finding new peers, tracking content and so on. We would like to emphasize here that PPETP is designed to take care only of the efficient data distribution; other important aspects of the P2P streaming application (e.g., building the network) are demanded to extra-PPETP means. This is similar to what happens with TCP: the standard specifies how data is carried from a host to another, but does not specify, for example, how one host finds the other, this being handled by protocols such as DNS.

III. Overview of PPETP

The goal of this section is to give a brief overview of the structure of PPETP and to introduce some PPETP jargon that will be used in the following. For the sake of brevity, many details will be omitted. A more detailed description can be found in the Internet Draft [16].

A PPETP network is made of several nodes that exchange data and control information over a non necessarily reliable protocol (currently PPETP is built on UDP, but other protocols, such as the Data Congestion Control Protocol [17], can be added in the future). Since each node streams autonomously to a (fairly stable) set of nodes, a PPETP network can be considered a *push* network. If node A receives data from node B, we will say that A is a *lower peer* of B and that B is an *upper peer* of A. (This nomenclature is inspired to the typical picture of a tree structured network with data flowing from top to bottom).

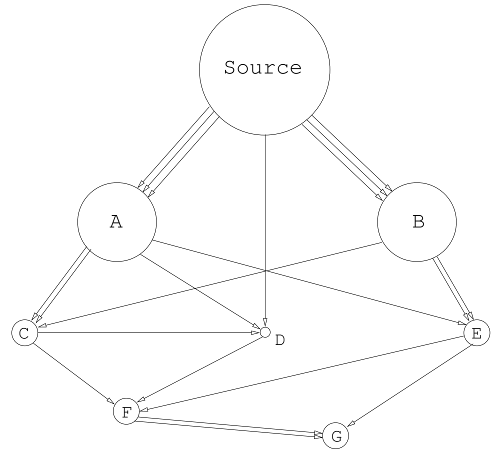


Figure 1. Example of a PPETP network for multimedia streaming.

A key characteristic of PPETP is that every node does not upload to other nodes the whole content stream, but a reduced version of it that requires less bandwidth. The details of how the reduced streams are produced are described in the following. Here it suffices to say that the original content can be recovered as soon as the node receives a minimum number $N_{\min}$ (typically chosen off-line) of reduced data. It follows that in a typical PPETP network a node has many upper peers and, possibly, many lower peers.

Fig. 1 shows an example of a possible PPETP network for multimedia streaming with $N_{\min} = 3$. Each arrow represents a reduced stream, each circle represents a node and the node upload bandwidth is represented by the circle size. For example in Fig. 1, node A (an upper peer of C, D and E) sends to C *two* different reduced streams. Note also that the source "feeds" directly nodes A and B by sending them three different reduced streams. Finally, note that the network of Fig. 1 is an irregular mesh, showing that not only tree-structured networks are possible with PPETP. As already said, by design PPETP does not mandate any particular network topology nor any specific way to find the peers. The specific application can construct the network as it sees fit.

IV. Data reduction procedures

As said above, the cornerstone of PPETP is a special type of network coding called *reduction procedure*. The idea is that, in order to take an advantage of even small upload bandwidths, a node does not propagate the multimedia packets, but a *reduced version* of them obtained by processing each packet with a *reduction function*. The result of the reduction function is a *reduced packet* whose size is (typically) a fraction of the size of the original packet. Thus, the reduced packets

 ISBN: 978-1-61208-157-1

transmission can fit the limited upload bandwidth of each peer. The *reduction factor*, that is the ratio between the size of the content packet and the size of its reduced version, will be denoted with R and supposed approximately constant (we say *approximately* since we admit slightly variations due, for example, to padding).

The reduction function is parametrized by a *reduction parameter* so that different reduced versions of the packet can be obtained by processing the same packet using different reduction parameters. The reduction procedure has the property that a node can reconstruct the original content packet as soon as it knows a sufficient number of reduced versions of the packet itself.

An overview of the typical behavior of a PPETP node is the following: the node, after receiving at least $N_{\min}$ reduced versions of the same content packet, recovers the packet itself and moves it toward the application level. Moreover, if the node has some lower peers, it reduces the recovered packet and sends the reduced versions to its lower peers. Nodes with larger upload bandwidth can serve several peers and also send to the same peer several different reduced streams, as exemplified in Fig. 1.

Example IV.1 (An example of reduction procedure)

The description of reduction function given above is very general and abstract. This follows the specification of PPETP that, for the sake of future extensions, does not impose a specific reduction procedure, but demands its description to documents called *reduction profiles*.

Since the abstract idea of a reduction procedure could be difficult to grasp on an intuitive level, we cite as an example of reduction procedure the algorithm described in [18] (used in the *Vandermonde* reduction profile).

To reduce the size of a content packet by a factor R, the packet is interpreted as a vector with entries b_0, b_1, ... in $\mathrm{GF}(2^d)$ (the finite field with 2^d elements) and every R-tuple of values b_0, b_1, ..., $b_{R-1} \in \mathrm{GF}(2^d)$ is replaced by

$$c_r = b_0 + rb_1 + \cdots + r^{R-1} b_{R-1}$$

where r is an element of $\mathrm{GF}(2^d)$ randomly chosen by (or assigned to) the node at start-up. The reduced packet is obtained by concatenating the values c_r obtained as above. Note that since this procedure replaces a sequence of R elements with a single element, the required upload bandwidth will be R times smaller than the bandwidth of the multimedia content.

In order to recover the original content packet a node contacts at least R peers and receives from them their reduced packets. It is easy to show that if each peer chooses a different value for r, then the node can recover the original values b_0, b_1, ..., by solving a linear system associated to a Vandermonde matrix [18].

A. Reduction profiles

The reduction procedure described above is not the only possible approach for data reduction. For example, other network coding procedures (e.g., digital fountains) could be used. In order to allow for future adoptions of different techniques, PPETP does not define a specific reduction procedure, but demands such a definition to side documents called *reduction profiles*. This makes possible to extend PPETP with new reduction procedures without changing its core definition.

Two profiles currently are defined: the Vandermonde profile (described above) and the Basic profile that does no reduction at all and it is thought for streams with very low bandwidth (e.g., RTCP streams) where the bandwidth saving would not be worth the additional complexity of a "true" reduction profile.

Although PPETP does not mandate any special characteristic to a reduction profile, it is expected that future reduction profiles will share with the Vandermonde profile the following important characteristics.

- **Size reduction** The size of the reduced packet is a fraction ($\approx 1/R$) of the size of the original content packet.
- **Parametrization** The reduction procedure depends on a set of parameters. Using different parameters gives rise to different reduced versions of the content packet. In the Vandermonde profile the reduction parameter is the value r.
- **Reconstruction** The content packet can be recovered from the knowledge of a suitable number $N_{\min}$ of different reduced versions (intuitively, $N_{\min} \geq R$). In some cases, such as in the Vandermonde profile describe above, $N_{\min} = R$, but this can be different in other profiles. For example, in an hypothetical reduction profile based on digital fountains, $N_{\min}$ would be a random variable with average slightly larger than R. In the following, for the sake of simplicity, we will suppose $N_{\min}$ deterministic.

B. Consequences of the reduction procedure

The reduction procedure in PPETP allows us to meet some of the previously stated design goals.

a) Exploitation of low-bandwidth nodes: Since the size of a reduced packet is a fraction of the size of the original content packet, the corresponding upload bandwidth is a fraction of the bandwidth of the content stream.

For very small upload bandwidths (that would required too large reduction factors) PPETP allows to introduce a *puncturing* that can be *random* (the packet is sent with a given probability) or *deterministic* (packets are sent according to a pattern). A careful use of puncturing allows for a finer control of the upload bandwidth.

b) Usage with heterogeneous networks: It is easy to manage the case when nodes have different bandwidths. For example, nodes with large upload bandwidth can serve several peers, Moreover, nodes with a large upload bandwidth can produce different reduced streams (by using different reduction parameters with the same content packet) and send more than one stream to the same lower peer (see Fig. 1).

c) Robustness to data loss: To counteract the risk of packet losses (due, e.g., to network congestion or peer leaving) the node requests data to $N > N_{\min}$ peers and recovers the content as soon as it receives $N_{\min}$ packets.

d) Security: To prevent stream poisoning, the node requests data from $N > N_{\min}$ peers, recovers the packet using $N_{\min}$ reduced packets and checks that the remaining packets are coherent with the reconstructed packet. This procedure can counteract a coordinated attack from $N - N_{\min}$ peers and, with a slightly variation, it allows to find (and punish) the node(s) that tried the attack.

If we punish who tries a poisoning attack, a malicious user could try a *defamatory attack* by sending corrupt packets while pretending to be another peer. In order to avoid this type of attack, PPETP allows a node to sign the packets that transmits.

e) Distributed parameter assignment: The reduction parameter used by the node can be assigned by an external entity or it can be chosen autonomously by the node (maybe at random). The latter is especially interesting, since it does not require a centralized actor for parameter assignment. The probability that two peers choose the same parameter can be made negligible by using a large enough parameter space.

f) Independence on the data type: Since the reduction procedure handles the content packets just as "sequences of bits," PPETP can be used with any type of data (e.g., audio/video, encoded with scalable or multiple description encoders, even encrypted data).

V. PPETP AND MULTICAST

As said above, one of our objectives was to design a protocol that looked, from the application point of view, like a multicast protocol. A first step toward this objective was the development of a protocol such that all the "P2P-related matters" (e.g., data reduction and reconstruction, handshaking with new peers and so on) could be handled inside the library implementing PPETP. In this way, the PPETP API could be made similar to the BSD socket API.

Moreover, in order to make the integration of PPETP with existing protocols (e.g., SDP [19], RTSP [20], SIP [21]) simpler, we decided to introduce the concept of *address* of a PPETP session in the form of an (host, port) pair. The problem is that a PPETP network, being distributed, has not a "natural" address. However, since a PPETP session needs to be configured (e.g., to set the reduction profile, the reduction parameters or any cryptographic credential used to communicate with other peers), the host part was chosen to be the address of a *configuration server* used to get the configuration data. (Note that every P2P network needs at least a "starting point" used by users to join the network; the starting point for a PPETP network is the configuration server.) The role of the port is played by the *session number*, a 16-bit integer that, together with the host address, uniquely identifies the PPETP session.

The configuration server is queried via a special protocol designed to be light-weight and stateless, so that it is less prone to Denial-of-Service (DoS) attacks and it can handle also a large number of connections. If needed, the configuration server can redirect the users, after authentication, to a more powerful protocol (e.g., an HTTP-based one) or maybe a distributed one (where the configuration data are obtained from others peers).

VI. EXAMPLES OF USE

In order to make clearer the just given overview of PPETP, this section describes two possible typical uses of PPETP: a live streaming application and a conference application.

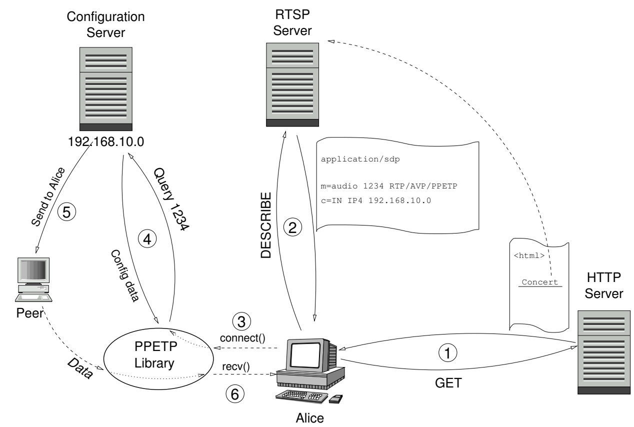


Figure 2. Example of establishment of a PPETP session

A. Live streaming

Suppose Alice wants to watch a concert streamed over PPETP. A possible sequence of actions is the following (see also Fig. 2)

1) Alice goes to the web page of the streamer, finds a link related to the concert and clicks on it.
2) The link points to an RTSP server. The browser launches a "viewer" that queries the RTSP server to get a program description (in SDP format [19]) that says that the program is streamed over PPETP.
3) The viewer opens a PPETP socket and connects it to the session address found in the SDP description by using a function similar to the BSD `connect()`.
4) The connection function queries the configuration server that replies with the configuration data.
5) Now Alice's upper peers must be notified to send data to Alice. This can be done in several ways, for example
   a) The PPETP network is fully managed by the video provider. In this case, the configuration server chooses the upper peers and asks them (via suitable control packets) to send data to Alice. If an upper peer is behind a NAT, the control packet will also cause the initiation of a suitable NAT traversal procedure. This is the case shown in Fig. 2.
   Although this centralized solution could seem to introduce a "single point of failure," it must be said that in this case there is an actor (the streaming provider) that is interested in doing the streaming. If the provider's host fails, the whole system makes no sense. Moreover, this centralized solution allows for a finer control of some PPETP network characteristics such as the quality of service assigned to Alice and the locality of the network.
   b) The server chooses the upper peers, sends the list to Alice and let her contact the peers.
   c) The server sends to Alice a list with some possible peers. Alice contacts few peers, asking for data; if a peer has no more bandwidth available, it refuses the request and Alice tries another peer until she gets enough peers. Note that, with this setup, it is difficult to make sure that Alice gets only its fair

 ISBN: 978-1-61208-157-1

share of resources.

d) A possible "strongly distributed" solution is the following: the nodes are organized in a Distributed Hash Table (DHT) where a set of "keys" (e.g., b-bit integers) is assigned to each node. The address of the DHT "entry points" is included in the configuration data. Alice randomly draws few keys, searches for the corresponding nodes and contacts them. Nodes that run out of bandwidth, refuse the request.

6) Alice receives reduced data. As soon as enough data are available, content packets are recovered and moved to the application level. The viewer reads the recovered data by means of a function similar to the BSD `recv()`, gives the data to the decoder and the result is shown to the user.
7) Suppose now that Bob joins the network and that the server assigns him Alice as an upper peer. Alice's host will receive a control packet that asks to send data to Bob.
8) In response to the received request, Alice's host applies the reduction procedure to the recovered packets and sends the result to Bob.
9) When Alice wants to stop to watch the concert, sends a TEARDOWN request to the RTSP server that in turn sends suitable control packets to Alice's upper peers, asking them to stop the transmission toward Alice and maybe redirecting them to the lower peers of Alice. Alternatively, Alice herself can redirect her upper peers to her lower peers.
10) If Alice suddenly leaves (maybe because of a bug), her lower peers notice her absence because they stop receiving data from her. As a consequence of this, they search for new peers. Note that if the network was built with redundancy, the *users* associated to Alice's lower peers *would not notice* the sudden departure since they will keep receiving enough data to recover the content packets.

B. Conferencing

The multicast-like nature of PPETP makes it an interesting solution for conferences. Conference management can be done via SIP [21], including in the session description the address of the PPETP session. In this case, every node "injects" its data on the network via a function similar to the BSD `send()` and reads from the PPETP socket the packets produced by the other nodes. The problem of separating the packets according to their source is outside the scope of PPETP and it pertains to the application. For example, if RTP is used, packets can be partitioned according to their SSRC [22].

C. Comments to the examples

It is worth to emphasize that most of the P2P management (e.g., NAT traversal, handshaking with the new peer) is handled by the PPETP library and it does not arrive at the application level. It should be clear from the examples above that the application just needs to (i) open a PPETP socket and connect it to the PPETP address, (ii) read/write data from/to it and (iii) close it when done.

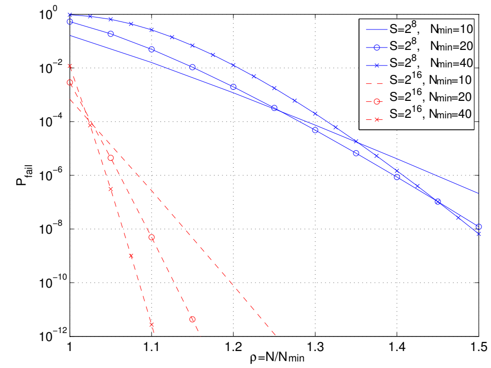


Figure 3. Failure probability P_{failure} as function of the redundancy ρ and the parameter space size S.

Note also that since for PPETP a packet is just "a collection of bytes," any type of data can be transmitted over PPETP. This means that all the currently available streaming tools (e.g., RTP, RTCP, audio/video coders, scalable or multiple description coding, encryption procedures) can be transparently used with PPETP.

VII. PPETP Performance

In this section we give some figures about the performance of PPETP. For the sake of simplicity we will suppose that $N_{\min} = R$ (as it happens with the *Vandermonde* profile).

g) Failure probability: If each peer chooses at random its reduction parameters, it could happen that the set of reduction parameters associated to the N upper peers of a node has less than $N_{\min}$ different parameters (*failure event*) and the node will never be able to recover the content packets. Intuitively, the probability P_{failure} of such event gets lower when the redundancy ratio $\rho = N/N_{\min}$ or the parameter space size get larger. This intuition is confirmed by Fig. 3 that shows P_{failure} as a function of the redundancy ρ and of the parameter space size S.

h) Robustness to sudden departures: In order to give a feeling of the robustness to departures offered by PPETP, assume that a node has $N > N_{\min}$ upper peers and as soon as a peer leaves the node searches for a new one. If too many peers leave in a short time, the node could remain with less than $N_{\min}$ upper peers (*underflow event*).

The probability P_{under} of the underflow event can be computed by modeling the set of upper peers as a queue with N servers, system size N, mean inter-arrival time (i.e., the time required to find a new peer) equal to T_{find}, and mean service time (i.e., the mean time a node remains connected) equal to T_{leave}. With this model, P_{under} is the probability of having less than $N_{\min}$ peers in the system.

A plot of P_{under} as function of ρ and $N_{\min}$ when $T_{\text{leave}}/T_{\text{find}} = 70$ and with the hypothesis of Poisson arrival times can be seen in Fig. 4 [18]. Probability P_{under} gets smaller when ratio $T_{\text{leave}}/T_{\text{find}}$ gets larger.

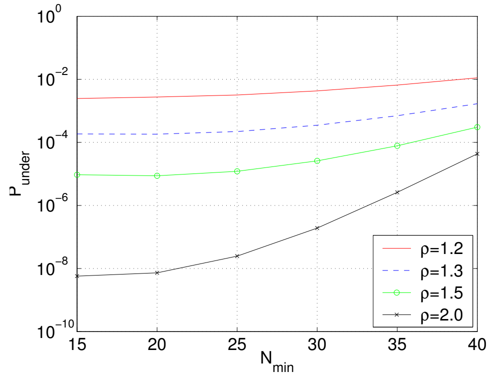


Figure 4. Underflow probability P_{under} as function of N_{min} for $T_{\text{leave}}/T_{\text{find}} = 70$.

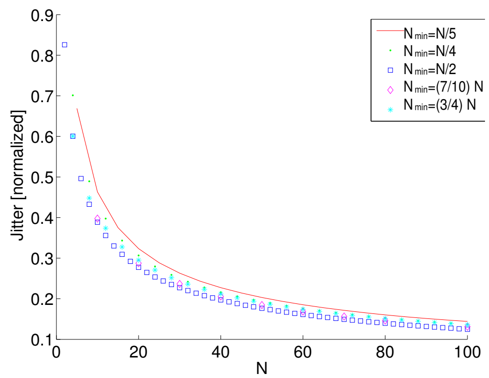


Figure 5. Jitter as a function of N (peer delays distributed as $\mathscr{N}(0,1)$)

i) Jitter reduction: A nice side effect of the use of network coding in PPETP is the reduction of the jitter observed by the node. Intuitively, this happens because the time a content packet is recovered is the time necessary for the arrival of the N_{min} fastest packets out of N. Fig. 5 shows the theoretical prediction of the jitter (i.e., the standard deviation of the reconstruction time), as a function of N_{min} and N, when the delays are Gaussian with mean m and variance σ^2. The values on the vertical axis are measured in units of σ. Note that the jitter decays as $1/\sqrt{N}$ [23]. This behavior was also verified experimentally [24].

VIII. Conclusions

This article has described PPETP, an overlay multicast protocol that allows for efficient data propagation even when some nodes have limited resources. The protocol is designed to appear at the application level as a multicast protocol, allowing for its easy inclusion in existing protocols and software. PPETP is robust against packet losses and it has tools that help counteracting possible attacks such as stream poisoning or DoS tentatives. PPETP is currently hosted by *SourceForge* as part of the open source project *Corallo*.

A. Acknowledgments

PPETP is partially funded by Italian Ministry PRIN project *Arachne*.

References

[1] E. Adar and B. A. Huberman, “Free riding on Gnutella,” *First Monday*, vol. 5, October 2000.

[2] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker, “Making gnutella-like p2p systems scalable,” in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM ’03, (New York, NY, USA), pp. 407–418, ACM, 2003.

[3] V. Fodor and G. Dán, “Resilience in live peer-to-peer streaming,” *IEEE Communications Magazine*, vol. 45, pp. 116–123, June 2007.

[4] V. Padmanabhan, H. Wang, P. Chou, and K. Sripanidkulchai, “Distributing streaming media content using cooperative networking,” in *Proc. of NOSSDAV 2002*, (Miami, Florida, USA), ACM, May 2002.

[5] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani, “Do incentives build robustness in BitTorrent?,” in *Proceedings of 4th USENIX Symposium on Networked Systems Design & Implementation (NSDI 2007)*, (Cambridge, MA), USENIX, April 2007.

[6] D. Stutzbach and R. Rejaie, “Understanding churn in peer-to-peer networks,” in *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, (Rio de Janeriro, Brazil), pp. 189–202, SIGCOMM, 2006.

[7] M. Wang and B. Li, “R2: Random push with random network coding in live peer-to-peer streaming,” *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 1655–1666, December 2007.

[8] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, “A scalable content-addressable network,” in *IN PROC. ACM SIGCOMM 2001*, pp. 161–172, 2001.

[9] M. Castro, P. Druschel, A. Kermarrec, A. Nandi, A. Rowstron, and A. Singh, “Splitstream: High-bandwidth multicast in cooperative environments,” in *19th ACM Symposium on Operating Systems Principles, 2003*, 2003.

[10] S. Marti and H. Garcia-molina, “Taxonomy of trust: Categorizing p2p reputation systems,” *Computer Networks*, vol. 50, pp. 472–484, 2006.

[11] S. Iyer, A. Rowstron, and P. Druschel, “Squirrel: A decentralized peer-to-peer web cache,” in *12th ACM Symposium on Principles of Distributed Computing (PODC 2002)*, pp. 1–10, July 2002.

[12] Y. Yue, C. Lin, and Z. Tan, “Analyzing the performance and fairness of bittorrent-like networks using a general fluid model,” *Computer Communications*, vol. 29, no. 18, pp. 3946 – 3956, 2006.

[13] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, “Chord: A scalable peer-to-peer lookup service for internet applications,” *SIGCOMM Comput. Commun. Rev.*, vol. 31, pp. 149–160, August 2001.

[14] A. Rowstron and P. Druschel, “Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems,” in *IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*, pp. 329–350, Nov. 2001.

[15] X. Hei, Y. Liu, and K. W. Ross, “IPTV over P2P streaming networks: The mesh-pull approach,” *IEEE Communications Magazine*, vol. 46, pp. 86–92, Feb. 2008.

[16] R. Bernardini, R. C. Fabbro, and R. Rinaldo, “Peer-to-peer epi-transport protocol.” http://tools.ietf.org/html/draft-bernardini-ppetp, Jan. 2011. Internet Draft, work in progress.

[17] E. Kohler, M. Handley, and S. Floyd, “Datagram Congestion Control Protocol (DCCP).” RFC 4340 (Proposed Standard), Mar. 2006.

[18] R. Bernardini, R. Rinaldo, and A. Vitali, “A reliable chunkless peer-to-peer architecture for multimedia streaming,” in *Proc. Data Compr. Conf.*, (Snowbird, Utah), pp. 242–251, Brandeis University, IEEE Computer Society, Mar. 2008.

[19] M. Handley, V. Jacobson, and C. Perkins, “SDP: Session Description Protocol.” RFC 4566 (Proposed Standard), July 2006.

[20] H. Schulzrinne, A. Rao, and R. Lanphier, “Real Time Streaming Protocol (RTSP).” RFC 2326 (Proposed Standard), Apr. 1998.

[21] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, “SIP: Session Initiation Protocol.” RFC 3261 (Proposed Standard), June 2002. Updated by RFCs 3265, 3853, 4320, 4916.

[22] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications.” RFC 3550 (Standard), July 2003.

[23] H. A. David, *Order Statistics 2nd edition*. Wiley-Interscience, 1981.

[24] R. Bernardini, R. C. Fabbro, and R. Rinaldo, “Peer-to-peer streaming based on network coding improves packet jitter,” in *Proc. of ACM Multimedia 2010*, (Florence, Italy), Oct. 2010.

A Quality Evaluation Framework Based on Distribution Measurement in Service Computing Environment

Zhenyu Liu
Shanghai Key Laboratory of Computer Software Evaluating and Testing
Shanghai, China
e-mail: lzy@ssc.stn.sh.cn

Abstract— **The importance of software quality evaluation has being gradually more and more focused. This paper studies service-based software quality evaluation in service computing environment. As is known to all, the existing evaluation technology adopts concentrated strategy during run-time. The paper compares the difference between the offline and online strategy for distributed system, and analyzing the features in service computing environment furthermore. The paper puts forward a novel quality evaluation measurement model. The attribute factoring technology and corresponding data collection strategy are described. A new evaluation method of distributed quality data acquisition is put forward in the paper, which is based on software testing technology. So, during service runtime operation, the obtained results by test method can be accurate and credible. Finally, this paper puts forward quality framework and related steps which is evaluated by distributed in service computing environment. Through example of manufacturing industry, it shows that the quality evaluation in distribution measurement framework is a effective and trustworthy.**

Keywords-Quality Evaluation; Software Testing; Distributed Measurement; Service Computing

I. INTRODUCTION

With the development of network, more and more software based on distributed architecture, such as Web Service, Service-oriented, Grid Computing SOA and cloud computing. The rapid growth of computer hardware and network, distributed system architecture make software more complicated, especially service-based computing convert the traditional computer and centralized storage approach into distributed architecture according to the demand of end user.

Service-based computing is service computing, which accomplished combining distributed processing, parallel processing and grid computing. The further research is realization of the commercial on view of user's point. Generally, basic principles of service computing are considered using the distribution by many computing resources, rather than the local computer or remote server.

This allows companies to adopt the appropriate resources of applications to access computing and storage resource, according user-demand. [1] As for service, life-cycle consisting of service development, service registration and service delivery. The three stages divided into offline testing and online testing according to the service releases correspond to the quality of service testing [2]. Offline test means the software testing without real environment. The test is executed in development environment. Before validation, this stage mainly verifies service function on the unit testing, system testing and other testing techniques related to code and service interface. Online testing is the design for the service lifetime, which focuses on non-function quality related with testing activities during execution of service.

In service environment, services register and publish using uniform computing platform. These services publish on the distributed environment. The function of service is same in different deploy environment. However, the actual quality of service should be considered when the service is being carried out. Service should be published and deployed in the middleware server and registered in specific server before use. Therefore, quality of service should be evaluated delay to the runtime phase. And continuous test is key approach to acquire original quality data.

In Figure 1, we compared different software phase in various distributed environments. The typical software phase is software development, software testing and system runtime. And the usual environment consists of network-based environment, grid environment and cloud computing environment. From the comparison, some distributed system required to postpone the online tests into run-time. The cloud environment is kind of these distributed systems. The results of software testing are obtained through the evaluation of the quality of service during the online testing.

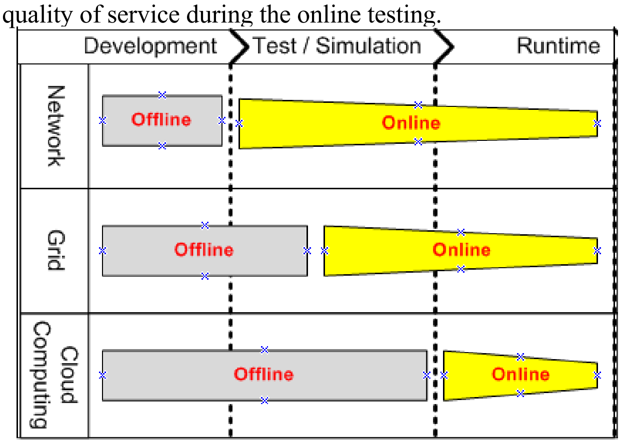


Figure 1. Comparsion of phase in distrubted environment

 ISBN: 978-1-61208-157-1

With the development of software techniques, there is change from the stand-alone application to network-based applications gradually. However, service-based computing software changes current operation and computer method. Therefore software quality is more and more to pay sufficient attention [3]. Also the users of the software quality are constantly changing from the initial needs that meet the functional requirements to non-function quality, such as run-time efficient and convenient.

As the service computing environment is the loose network, the test during the phase of running service in scenario. The scenario is closer to the genuine quality of the software runtime environment. Therefore, there is inconsistency and uncertainty for online quality of service in contrast to development period. Table 1 gives the different development period under the various architectures. In contrast to other architecture comparison, it is evidence that cloud computing environment own run loose-couple management, which mean the online quality is not analyzed and evaluated in stage of development period effectively.

TABLE I. COMPARISION WITH DIFFERENT ARCHITECTURE

Architecture	**Environment and related Stage**		
	Development	***Testing***	***Runtime***
Stand-alone	Stand-alone	Stand-alone	Stand-alone
Browser / Server	Single/Multi	Single/Multi	Network
Application / Midderleware	Single/Multi	Multi	Integrated Management
Cloud Computing	Stand-alone or Single/Multi	Single/Multi	Loose-copuled

Through the above analysis, it is necessary to transfer the validation and testing of service from the testing stage delay to the runtime stage. For the loose network applications, various users on different network nodes access the same services. But the results obtained varied greatly, which is different from the traditional stand-alone architecture. From the results of a single user to determine the software quality is not sufficient. It is also not enough to determine the quality in the same network. It is necessary that the distribution of methods used to measure this evaluation [4].

The main structure of this paper is as follows: The evaluation of quality decompose model in section 2. Section 3 introduces data collection method through testing method. The next part presents the distributed online quality evaluation framework. Finally, there are conclusions and future works.

II. EVALAUTION MODEL

Software quality evaluation model is the approach that evaluate and quality of the quantitative evaluation according to software-related needs. Boehm et al proposed a hierarchical model of software measurement [5]. McCall proposed three-level model to measure, which divided software quality into elements, standards [6]. At present, the evaluation model is main method of measure software features. ISO/IEC TR9126 standard gives a typical measurement model for software product. The proposed model consists of an external measurement metrics, internal measurement metrics and user measurement metric [7].

A. Generic Model

The quality requirements of the software have functional requirements and non-functional requirements. As for quality requirements of system design, the traditional software quality evaluation model is mainly based on quality index in evaluation of quality.

For different scenarios, quality requirement are different according to request user. Especially for cloud computing environments, quality demand of different user is also changing.

However, the existing evaluation model gives the design requirements for the proposed related measurement methods. And measurement methods and their model include some measurement items. Under many circumstance, the data from items adopt the offline process, that is to say, system is not still provided service in actual environment.

Therefore, the quality of evaluation process also meets the unique context during evaluation. A single context is not applicable for various scenarios which involve in the application of measurement models. In practice, the different scenarios will eventually lead to different measurement results.

The typical quality model is proposed by ISO/IEC TR9126 international standards. This standard consists of four parts: a generic measurement model and three measurement models from a different perspective, which is internal metrics and external metrics, and user metric measurement methods, respectively.

The generic model has six main attributes: functionality, usability, availability, performance, maintainability, portability. Each attribute also contains a number of sub-attributes. The sub-attribute measure value will determine from the quality of their respective attributes directly.

The overall quality could be calculated through the available quality data of sub-attribute.

For the given external and internal metrics, each of its sub-attributes consists of a collection of measurement items. For example, efficiency of time attributes includes sub-attributes: response time, throughput and turnaround time. The response time metrics also constituted with two metric indexes, which is average response time and response time in the worst case. From related response time, we can found that the test is effective method to obtain the metric index value.

Here, we study the properties of all of the sub-attribute of metrics. There are three approach to get index value, which consists of obtain values from testing directly, statistics and obtain by checklist or interview.

B. Attribute Factoring

As for a given measure model, the metric attribute value need to acquire after analysis sub-attribute based on specific index value. Usually attribute computation should consider the sub-attribute and its relation between sub-attributes. In order to facilitate the calculation of index value, sub-attribute is propose to bridge the corresponding attribute measurement

 ISBN: 978-1-61208-157-1

model with metric index. Therefore, a number of sub-attributes will be measured through the composition of metric index. In some cases, metric index should be computed by several values. In these conditions, an exceed three level model is need for measurement model and its relationship between the attributes is relatively cumbersome.

If this model is analyzed with top-down method, we possible constitute measure architecture with multi-level. So the relation is relatively redundancy and increase the analyzed complicated.

This goal of research is to simplify the existing multi-level model. For reason, we adopt the idea of attributes based on the direct decomposition of GQM model [8]. The idea of GQM is contain target problem, question and measurement items.

The goal of GQM model is address the object via the question, which the question is the need to raise relevant issues for this object. Finally, measure item is the access method for sources of data directly.

Definition 1: The metric index is the smallest of factors in quality measurement model, including metric index information, metric objects, collection of metric index values, and extended attributes.

For the quality assessment model, the Φ is one attribute includes a number of sub-attributes: φ1, φ2 ... φn, which each sub-attributes of φ includes metric index ξ1, ξ2 ... ξn. The ξ is the appropriate factor for acquire data.

$$\Phi = f(\varphi_1, \varphi_2, \dots, \varphi_n) \qquad (1)$$

$$\varphi_1 = g(\xi_{11}, \xi_{12}, \dots, \xi_{1n}) \qquad (2)$$

where f, g are calculation functions corresponding to measurement model. Here index value ξ is collected from the different value. The index value can be a function point, software requirement, or even the data collection composed by above single value. The collection of single value is composite value of the point values.

Definition 2: Measure item is the smallest unit in the quality model. For each index ξ, there could be several measure item β.

For the collected data sets, usually metric index data have single or multiple item composition. The item data can be characterized by general information of specific data. The multiple item value also is vector value. Vector value cannot become calculate direct in this model and need process into scalar value by the way of average or the weighted average method. Common calculate method is the average point of the calculation, the composite weighted method. The average is calculated for all metric data collected with average method:

$$\xi = \frac{1}{j}\sum_j \beta_j \qquad (3)$$

Composite weighted is a measure of all the points in accordance with the weigh distribution. Generally, assuming that the weigh value is w, and then calculated as follows:

$$\sum w_j = 1 \qquad (4)$$

$$\xi = \frac{1}{j}\sum_j w_j \beta_j \qquad (5)$$

The decomposition attributes optimized multi-metric evaluation system in measurement model. The decomposition attributes provided by defining a set sub-attribute to obtain relation between quality attributes and index. As a result, in order to get the ultimate quality attributes, we must decompose the quality attribute into final quantifiable evaluation values in accordance with the index of the quality attribute value through three-level model.

Three-level model consists of attribute, decompose attributes and their index, just as three level trees which shown in Figure 2. The root of tree represents the attributes, while the leaf nodes in tree are corresponding to different index. In decomposed model, two characters are differed from other generic quality model. The one is that same sub-attributes could have the several indexes. The other is, if there have two sub-attributes of the same index, the same index can be applied to multiple two sub-attributes. In Figure 2, the Metric Index 3 is applied not only for Sub-attribute 1.2 also for Sub-attribute 1.3. Therefore model can reduce the metric index use frequency and improve efficiency.

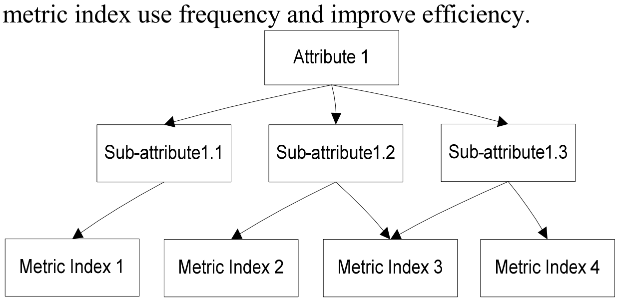


Figure 2. Direct decomposed model

III. MEASUREMENT INDEX

A. Data Collect

According to the direct decomposition model, evaluation activities need to obtain various measure items according to metric index. The item value is acquired to get the quality of data related technologies. The common technology is offline and online. Here we focus on online technology, generally real-time technology. After collecting the results of further calculations include three ways: metric index directly, function and statistics. Statistics refers to the sample on the basis of certain statistical sample. Function is the method to get metric index value through the complex calculation.

Generally, the need to obtain various types of measurement elements, namely, the direct decomposition of the metric index. The activities come mainly from the software engineering such as testing, validation and verification. The data acquisition is carried out in software engineering-related behavior of statistical analysis furthermore. Software product data adopt testing and auditing to achieve in quality model. Besides the testing, the interview is other way to obtain metric value through conducting interview.

In software quality assurance technology, testing is key approach, which can complete most acquisition of metric index value in decompose model. As far as distributed services are considered, distributed characteristics should take into account in order to obtain more accurate item data. Here the data collect is need to consider extended factors that affect the quality.

Definition 3: Extended factor θ is the quality of factors can influence the specific measure of certain external factors. For the different factors θ, its value will lead to measure the various different observations value y, assuming a factor of two values for x1 and x2, if $x1 \neq x2$, then $y1 \neq y2$, then the attribute can be considered extended factor.

For the typical extended factor, network or resource constraints is the main factor. These factors does not affect the value themselves or directly. They impact data collection of information significant in different access patterns, such as time, location, etc. Assuming the time factor t1 in day and t2 in night, obviously $t1 \neq t2$, then the time duration is $y1 \neq y2$ for a specific service. Therefore the time factor can be considered extended factor.

For the same measure item of quality can be used in the testing process in software product.

B. Test Collect Method

Software testing is the primary approach to obtain the testing results, which help to calculate quality value. The test approach can direct access to collected results. Here we consider the software testing, which is main approach of acquiring quality data in this study.

During software testing, test case is the basic elements during testing execution. In test collect method, a standard test case template is provided to facilitate the test execution and metric index collect. The standard test case template can assist comparative analysis in different extended factor. As for testing activities is high cost work. The test case reuse technique can improve the test efficiency and reduce cost of test.

Definition 4: Typical test case consists of test scripts, test resources, test context, test items and test.

```
<test case>::= <TCID> <Case Info>{<test item>}
<TCID>::= /*Test Case Unique ID */
<Case Info>::= <test context><test oracle><measure><test goal><test type><test method><version>
<test goal>::= product | project | technique
<test type>::= function | performance | security | others
<test method>::= manual | automated tool
<test context>::= /*test scenario for current case*/
<measure>::= <state><granularity >< reuse frequency >
< state>::= initial | modify | use | expired
```

Definition 5: Test items is single test step in one test case, including the item input, expected output, output and hint.

```
<test item>::= <TIID><item description>
<TIID>::= /*Test Item Unique ID */
< item description>::= <item input><expected output><output and hint>
<item input>::= /*test operation procedure and input information*/
<expected output>::= /*expected the result based on the input*/
<output and hint>::= /*output information and software hint information*/
```

Metric index data obtained from the testing results. In actual testing execution, the data can be get from the test results directly which defined in test results in test case.

Definition 6: Quality Data is collected through the execution of test cases.

$$d = f(tc\ o, \theta) \tag{6}$$

where tc is the collection of test cases, o is the measured object, θ is extended factor vector in specific test scenario according to test scenario.

The quality of each element can be tested to obtain the data, each element of the quality of the corresponding mapping relations between the testing requirements, so that the required index value for each measure can be tested through a series of test cases to be completed. With the metric system, the corresponding index value, by testing their statistical ways to get the corresponding index value.

Definition 7: If the same quality of data collection, its test cases is similar.

That is to satisfy for test case tc1 and tc2, if tc1 and tc2 contained same collecting metrics ξ, then tc1 and tc2 have the same measure item. If ξ (tc1)$\in\xi$ (tc2), then tc1$\in$tc2, i.e., we can adopted tc1 are tc2 to complete task of data collection.

In software testing, test case design account for a large proportion. Because the test engineers need more time design test cases, reuse test case can reduce the design costs if use existing test case which have been used. It is necessary that find a common test cases in order to reduce the cost of the actual test case design after test design and test execution according to Definition 7. [10][11].

IV. DISTRIBUTED EVALUATION FRAMEWORK

The model is presented above. Here service computing has the characteristics that end user access service through any node in the network. Consider metric index value is various from the various node, which access the service due to network or time. The factors are influenced on the basis of obtain quality data using the same method from all nodes. The Session 1 shows that novel software systems are distributed mode, which different from the previous stand-alone mode. Therefore, distribution evaluation model is necessary to replace the original method that a single node, centralized evaluation of the quality metrics.

As far as service computing environment is considered, users scattered in different network node access services in the network environment. Therefore test request nodes scattered in network to replace the centralized access mode.

The traditional single-node test, this may be better for a simple without network, which results of function test are inaccuracy. But for non-functional requirements testing, such as performance, availability and reliability, the test results is affected by computer performance and configuration of computer which running the software. With the growth of network technology, the current software systems require the support of the network environment, for instance, Browser / Server or Client / Server architecture for these can be run on the network. It is worth studying in network runtime environment, especially for non-functional metric index.

Therefore, distributed test strategy can be used for accessing quality. However, many industry performance testing tools support built-in concurrent test engine to carry out the efficiency testing, which these engines often require the deployment of network-specific location. These concurrent engines do not reflect the heterogeneity of the distribution of characteristics.

Relative to the single-node test, the distributed testing can be close to the actual results relatively. However the distribution is relatively complex, the accuracy of data is usually due to a variety of factors, such as network structure, bandwidth, user habits and physical environment and different. In some practical applications, it can be distributed using the grid on the existence of the physical distribution. The more satisfactory distribute data could be acquired when the test case is executed in the agent on physical grid node.

A. Evaluation System

According to the multi-attribute decompose model and related testing techniques, we presents an evaluation system. The Figure 3 shows the architecture, as well as the main model layer structure. Quality assessment method is usually based on a large number of quality data, thus the underlying layer is responsible for access to quality data. So underlying layer is metric level to adopt testing and validate assessed object.

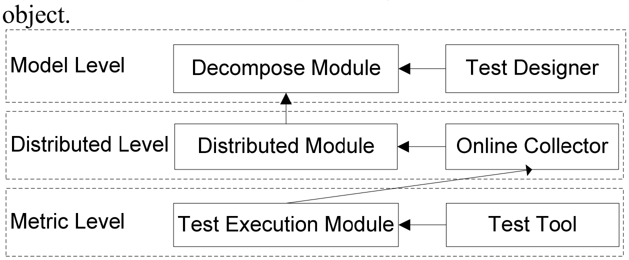


Figure 3. Evaluation System

The establishment followed the entire process from the decompose quality to generate test cases. The main steps in the process are as follows:

1) *Analyzis:* analyze user requirements for quality assessment and exemplify measure properties, required attribute of evaluated system.

2) *Find minimum set:* find the minimum measurement items with decomposed sub-attribute according to involved the metric index.

3) *Test design:* according to minimum measurement items, generate the test case ot reuse corresponding test case.

4) *Publish:* dispatch the test to various distributed access node, the node under the scenario requirements.

5) *Execute:* execute the test case automatically in distribute nodes, collect quality data and return to the central controller.

6) *Collect data:* central controller calculates the metric index value according to return quality of data.

Through the above six steps to obtain the corresponding metric index data, an online quality is available in distributed environment.

A measurement system is developed in practice. The platform concerns the design of the measurement model, in addition, the middleware in system integrated test case reuse library which support the reuse test case. Through analyzing the user's quality requirement, the system analyzes their requirement for test reuse test cases to obtain the required quality metric index data.

B. Case Study

Manufacture is the economy basic industries, which output required for a large number of social economic. Manufacturing information technology enhances the manufacturing and the development of integrated design, production, circulation and management efficiency. In the road of development in manufacturing industries, from the capital-intensive, technology-intensive development of the information-intensive, the computer introduces database technology, network technology, and distributed computing technology to the grid system integration technology.

Service computing environment is distributed computing. The dynamically change needs that require access to resources and services. Application and service request resources from the service rather than the traditional sense of physical entity. Service is that somehow self-maintenance and management of virtual resources. For manufacturing enterprises, the dynamics of self-maintenance and management of virtual resources in the intensive resources, utility computing, and information technology play a role in the progressive development for the enterprises are more concerned about the production, business and other technological innovations.

Figure 4 shows application integration environment. The entire application architecture is constituted with a virtual environment to achieve configuration, deployment, service. Integrated environment including enterprise, information systems and external information based on the software as a service (SaaS) model system, enterprise information system through the service encapsulation. The original model provides computing resources and storage resources, for integrating enterprise legacy systems. The enterprise information support upgrade and configure, some systems or new systems can use in the distributed environment.

The entire application environment in a typical manufacturing information applications such as CAD, CAM, CAE, PDM, ERP, CRM, SCM, MES, BI, and BPM, etc., design and manufacture of these platforms from the bottom and resource management to manufacturing execution, and then to business process reengineering and business intelligence, service environment provides the existence of information resources sharing and mutual relationship. In this environment, applications and system architecture meet the functional requirements. However the production scale, continuity, and other demand information in manufacturing environment is differ from the traditional systems. The requirement of customers with real-time, operational complexity and inclusiveness, flexibility and robust features is very high, so the overall quality of the model is concerned.

According to manufacturing information property, quality modeled at first. Not only function for a range of business system covering the production organization, production management and production optimization, but also the continuity of production to meet the requirements of non-functional metric index. And then establish manufacturing information quality model with general purpose based on the use of reusable test case in distribute environment.

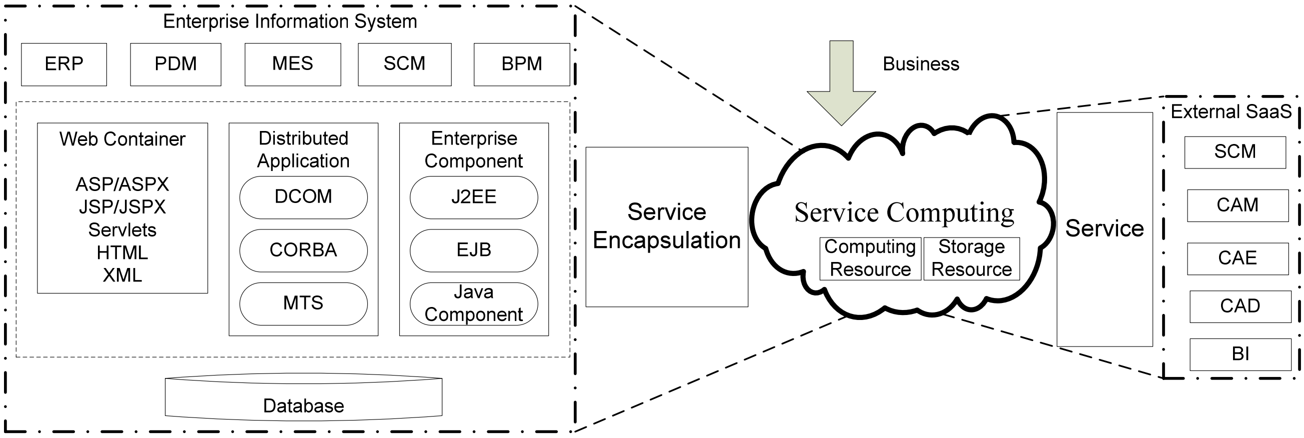


Figure 4. Typical Case with Service Computing

V. CONCLUSION AND FUTURE WORKS

In this paper, distributed quality measurement model is proposed suitable for the actual data quality during practical scenarios. The decompose quality models is key technology which goal is distribution of online access the value of the collection by running the stage. Based on distributed agents under different networks, the results of the test obtain qualitative assessment of metric values can be a more accurate in analysis system quality and runtime scenario.

Evaluation framework adopt the distributed measurement services, online execution can be more closer to the actual scenario, which not only can serve as a software engineering activities in the software online test and evaluation before online, but also provide the maintenance phase of the software to conduct regular assessment to obtain the software lifecycle quality information. The study proposes the distribution of online quality evaluation in the distributed environment for the deployment of a large user, many applications manufacturing information system and has good prospects.

The measurement model could be analyzed factors further which affecting the difference results of the final measure. On the one hand, access location of the network bandwidth will influence the evaluation results. On the other hand, the time factor is also necessary for distributed applications. With distribution of the international business applications, the peak time will occur varied in number of users and business regions.

Distributed quality evaluation is the direction of research with software technology. Future works can study related algorithms furthermore, especially evolution quality model of functions and non-functional in the system runtime and maintenance. And the benchmark library is needed to form based on numerous testing results which conform to international quality standards. In addition, the playback technology of same test cases in software regression testing could be studied for improving the authenticity of the results.

ACKNOWLEDGMENT

The work is supported by National Torch Program under Grant No. 2009GH510068 and Shanghai STCSM Program under Grant No.10DZ2291800.

The authors would like to thank without knowing the name for their helpful suggestions.

REFERENCES

[1] Chen K. and Zheng WM. Cloud computing: System instances and current research. Journal of Software, 2009, 20(5):pp.1337−1348.

[2] Bertolina A.,Angelis, De G.,Frantzen, L., Polini, A., “The PLASTIC Framework and Tools for Testing Service-Oriented Applications” Software Engineering, International Summer Schools, ISSSE 2006-2008, Salerno, Italy, pp.106-139.

[3] Offutt, J. “Quality Attributes of Web Software Applications”, IEEE Software, 2002, Vol.19(2), pp.25-32.

[4] Stephen J.H. Yang, James S.F. Hsieh, Blue C.W. Lan, Jen-Yao Chung, “Composition and Evaluation of Trustworthy Web Services”, International Journal of Web and Grid Services, 2006, 2(1), pp5-24.

[5] Boehm, B. W., Brown, J.R., Caspar, H., Lipow, M., Macleod, E. J., and Merritt, M.J. Charecteristics of Software Quality. Amsterdam: North-Holland, 1978

[6] Cavcno, J.P. and McCall, J.A. “A Framework for the measurement of software quality”. In proceedings of the Software Quality and Assurance Workshop, 1978, pp.133-140.

[7] W. Vambenepe, C. Thompson, V. Talwar, S. Rafaeli, B. Murray, D. Milojicic, S. Iyer, S, K. Farkas, M. Arlitt, “Dealing with scale and Adaptation of Global Web Services Management”, Proceedings of the IEEE International Conference on Web Services(ICWS’05) Orlando, Florida, USA, July.11-15, 2005, pp.339-346.

[8] ISO/IEC TR9126-2006, Software engineering-Product quality Part 4: Quality in use metrics

[9] Van Solingen, Rini; Egon Berghout, The Goal /Question/ Metric Method, 1999

[10] Wee Kheng Leow, Siau Cheng Khoo, Yi Sun, “Automated generation of test programs from closed specifications of classes and test cases”, Proceedings of the International Conference on Software Engineering, 2004, pp.96-105.

[11] S. Jones, “Toward an Acceptable Definition of Service”, IEEE Software, May, 2005, 22(3), pp. 87-93

Scientific Gateway: Grid and Cloud-based Visualization

Eva Pajorová
Institute of Informatics
Slovak Academy of Sciences
Bratislava, Slovakia
utrrepaj@savba.sk

Ladislav Hluchý
Institute of Informatics
Slovak Academy of Sciences
Bratislava, Slovakia
l.hluchy@savba.sk

***Abstract* - The science gateway is important component of many large-scale Earth, astronomical, environmental and natural disasters science projects. Developing the sciences portals and the science gateways looks for us coverage of requirements of large scale sciences such as, Earth science, astronomy and all sciences which are using grid, cloud or cluster computing and high-performance computing infrastructure. We participate on designing and developing an e-Science Environment for Astronomy and Astrophysics. The paper describes the main visualization facilities for Visualization Tool (VT) to qualify them for Scientific Gateway, which have been apply for our design of astronomy Visualization Tool.**

Keywords-visualization; grig; cloud; gateway.

I. INTRODUCTION

Through user-friendly web interfaces such as e-Science gateway integrated into the same environment, researchers and scientists can securely and transparently access computational and data sources, services, tools, sensors, etc. Science gateway is a computational web portal that includes a community-developed set of tools, applications, and data customized to meet the needs of a targeted community. It can hide the complexity of accessing heterogeneous Grid computing resources from scientists and enable them to run scientific simulations, data analysis and *visualization* through their web browsers [5]. Scientific gateways are able to provide a community-centric view, workflow/dataflow services and a strong support in accessing to the cyber infrastructure including grid and cloud based resources. In each of science contexts, scientific gateways play a key role since they allow scientists to transparently access to distributed data repositories (across several domains and institutions) and metadata sources to carry out search & discovery activities, as well as *visualization* and analysis ones, etc. Finally, scientific gateways can play an important role in training students (at the academic level) in the different scientific disciplinas, attract new users and representing a relevant centralized knowledge repository in the sciences context. Our paper deals with the position of *visualization* as one of the main components of scientific gateway. The scientific web portal - gateway cumulate all types of *visualization*.

Since 2004 numerous scientific gateways have been developed. Lot of scientific gateways were funded by the TeraGrid Science Gateways program [1]. The gateway paradigm requires gateway developers to compile and install scientific applications on a variety of HPC clusters available from the resource providers in TeraGrid, to build service middleware for the management of the applications, and to develop web interfaces for delivering the applications to a user's web browser. Consequently many web-service frameworks [2][3] have been designed and applied in building domain-specific science gateways. Some of them enable workflow based on the web services [4], but they commonly don't provide solutions to support web interface generation. Developers was usualy hindered. Usualy they need to spend a lot of time learning web programming, especially JavaScript and AJAX Technologies to implement a user-friendly and interactive web interface to these services.

Developed visualization tools by us take acces on properties to include them to the Scientific gateway. For example our design propose a new web based application framework for astronomy and asttrophysics environment. We start from rich experimences in lot of grid and cloud based project in e-Sciences environment. We start with proposing new framework enables astronomy and astrophysic science gateway dewelopers based on last one web resources. Visualization tool is part of gateway and proposes a new based application framework for astronomy and astrophysics environment. The framework including the can import the astronomy specific workflow scripts easily can generate web appliance for running astronomical applicationworkflows and visualization the outputs results directly from workflow execution, online visualization through their web browsers.

.

II. VISUAL REPRESENTATIONS OF DATASETS WHICH ENABLED SCIENCE GATEWAY

There are some reasons why scientists, including astrophysics, are using visual representations of datasets

- for a visual control of the execution proces
- for know-how discovery and for presentations the academics research results
- for formal publication of research results
- for a directly visual education form

A. VT architecture, and visual control of the execution process

Simulation and execution with a huge date usually spend long execution time. Good solution for execution is represented by grid and actually on cloud computing. In both infrastructures get *visualization* has the main position as a way to control the execution proces**s.** Visual control has in all infrastructure very useful position. The modal parametric studies applications include, for example, astronomical simulations on which we tested our submission tool as on-line visualization tool. The simulation was realized as a sequence of parameter studies, where each sub-simulation was submitted to the grid as a separate parameter study. The job management was rather time consuming due to the analysis of failed jobs and to their re-submission.

Visualization is included as a visual control process. The visualization tool is designed as a plug in the module. Client asks for visualization is as a “visualization client”. Output data on the storage element are the inputs data for visualization jobs. Visualization workers are to modify data to the formats, which can be visualized but also to prepare the typical visualization scenes. Client can render such scenes on the browser, can make the visual control and modify executions. For example, to immediately understand immediately the evolution of the investigated proto-planetary disc we have developed a Visualization Tool (VT). The VT is composed of several modules, which are responsible for creating scenes and converting data to the the “visualize” format. The VT is designed as a plug-in module. The components generating rendering scenes are easy to exchange, according to the requirements of the given application. In case of our “gridified” application the output data of the simulation located on the SE can be used directly as the input for the VT. The final product of the VT includes a set of files containing data in the VRML (Virtual Reality Modeling Language) format. These output files can be rendered by many available VRML web-browsers. The whole visualization process is maintained through a visualization script, whose basic function is invoking the individual VT components in successive steps, transferring data, and handling error events. The script is written using the Bourne shell scripts and all VT modules are implemented in the C++ language. The VT can be embedded into the framework described above, or can be used separately as a stand-alone program. By using the on-line VT the client can stop the execution process, change the input parameters and restart the execution process ones again. In grid environment. such architecture can be used for all applications from different sciences spheres which have the character of a parametric study.

Actually the research community needs not only “traditional” batch computations of huge bunches of data but also the ability to perform complex data processing and this requires capabilities like on-line access to databases, interactivity, fine real-time job control, sophisticated *visualization* and data management tools (also in real-time), remote control and monitoring. The user can completely control the job during execution and change the input parameters, while the execution is still running. Both tools, the tool for submission designed before and continued sequential visualization tool - provided complete solution of the specific main problem in Grid environment. The position of the visualization tool as a visual control process is shown in Figure 1. Astrophysics scientists are able to run scientific simulations, data analysis, and visualization through the web browsers. Through astronomical science gateway can scientists to import they sofisticated scripts by which the VT he can activated as well, as the output from workflow executions without writing any web related code [6].

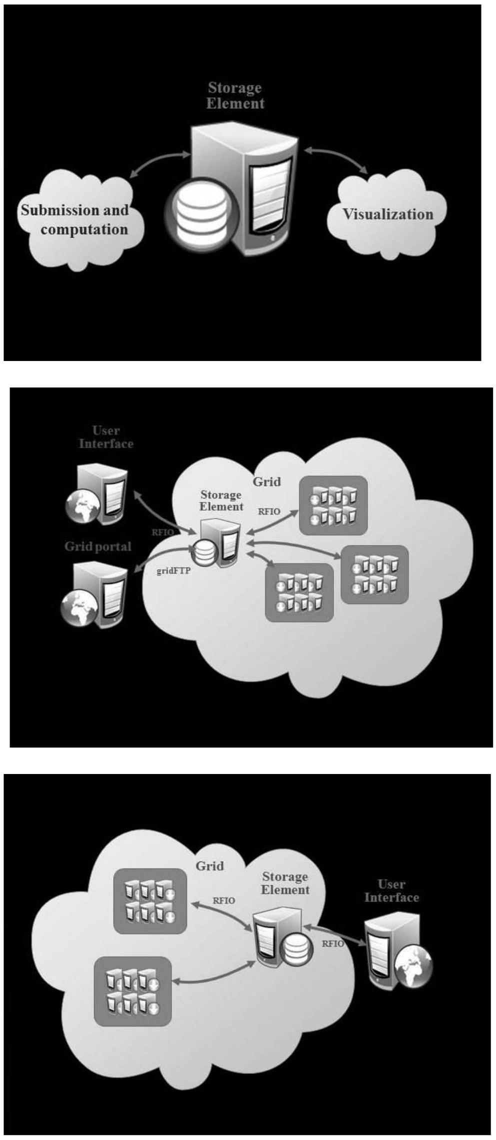


Figure 1. Example On – line visualization – the main position in grid based applications. Example when is visualization using as a control of execution process

B. VT as a new discovery for presenting academic research results

In order to demonstrate the practicalities of interchanging multi-dimensional data, we consider the case of cosmological visualization: representation of the three dimensional spatial structure of the Universe, including both observational and simulation datasets. Where such information exists, we extend this definition to include time-evolving datasets (e.g., evolution of structure formation or the hierarchical merging of galaxies), and derived data products such as catalogues and merger trees. A modern, fully-digital cosmological visualization allows the user to rotate, zoom, pan and even interactively select from datasets.

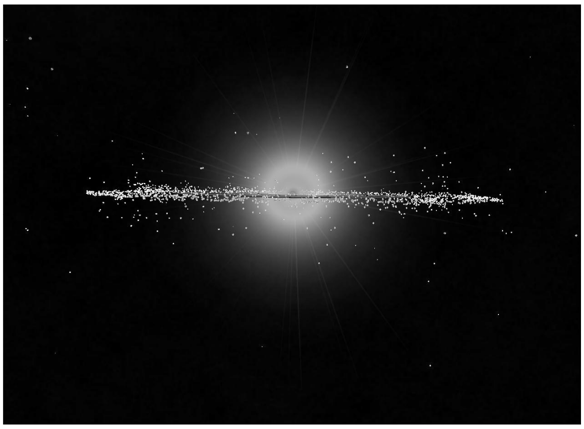
Advance in sciences and engineering results in high demand of tools for high-performance large-scale visual data exploration and analysis. For example, astronomical scientists can now study evolution of all Solar system on numerous astronomical simulations. These simulations can generate large amount of data, possibly with high resolution (in three dimensional space) and long time series. Single-system visualization software running on commodity machines cannot scale up to the large amount of data generated by these simulations. To address this problem, a lot of different developed grid-based visualization frameworks have been developed for time-critical, interactively controlled file-set transfer for visual browsing of spatially and temporally large datasets in a grid environment. To address the problem, many frameworks for grid and cloud based visualization are used.. We can go through evolution of sophisticated grid based visualization frameworks with actualized functionality. For example: Reality Grid, “UniGrid” and “TerraGrid”.

All of the frameworks have been included in the *visualization*. Frameworks were created during grid based projects and create new features for presentations the academic research academic results in visualization. Visualization resources enabled by a astronomical science gateway the top of researches experiences.

C. VT and its formal research results

Multiple visualizations generated from a common model will improve the process of creation, reviewing and understanding of requirements. Visual representations, when effective, provide cognitive support by highlighting the most relevant interactions and aspects of a specification for a particular use. The goal of scientific visualization is to help scientists view and better understand their data. This data can come from experiments or from numerical simulations. Often the size and complexity of the data makes them difficult to understand by direct inspection. Also, the data may be generated at several times during an experiment or simulation and understanding how the data varies with time may be difficult. Scientific visualization can help with these difficulties by representing the data so that it may be viewed in its entirety. In the case of data, varying in time animations can be created that show this variation in a natural way. Using virtual reality techniques, the data can be viewed and handled naturally in a true three dimensional environment (e.g. depth is explicitly perceived and not just implied). All these techniques can allow scientists to better understand their data. Viewing the data in this way can quickly draw the scientist's attention to interesting and/or anomalous portions of the data. Because of this, we encourage scientists to use scientific visualization from the beginning of their experiments and simulations and not just when they think they have everything operating correctly. This also allows the scientists to develop a set of visualization tools and techniques that will help them understand their data as their research matures. For example, depending on of our astronomical example in order to understand immediately the evolution of the investigated proto-planetary disc we have developed a Visualization Tool (VT) for astronomers.

VT for the astronomical application provides pictures from simulation of the evolution of proto-planetary disc from 1Myr to 1000 Myr. Specifically, Figure 2. shows the evolution of proto-planetary disc in the time of 1 Myr. We can see that during the 1000 Myr time that the particles were replaced from inside to outside of the spheres. Figure 2 show the result of dynamical evolution of Oort- Cloud as a part of Proto-planetary disk after its evolutionary stage which was the first Gyr (giga year) [7].



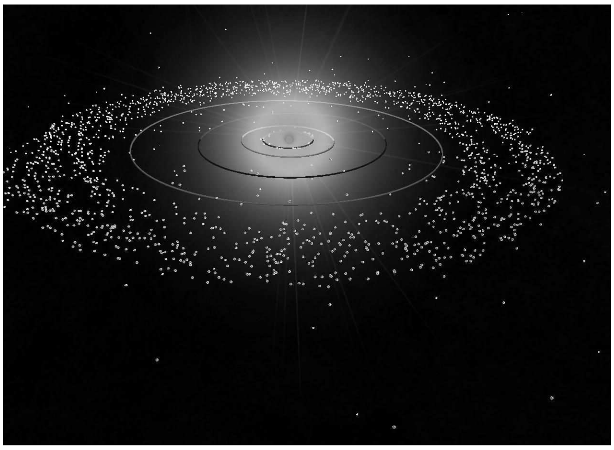


Figure 2. Example On – line visualization – the main position in grid based applications. Example when is visualization using as a control of execution process

D. *Directly visual education form*

Educational visualization uses a simulation normally created on a computer to develop an image of something so it can be taught about. This is very useful when teaching about a topic which is difficult to see otherwise see, for example, .protoplanetary disk., its evolution or evolution in Solar system. It can also be used to view past events, such as looking at the Solar system during its evolution stage, or look at things that are difficult.For astronomers, the VT has in education roles well.

III. CONCLUSION

The goal of the paper was to describe the VT architecture and to support the *visualization* as essential component in new portals - gateways technologies and to show some examples. For the future we want to extend the use of the VT for other scientific disciplines in addition to astronomy, but also for Earth Sciences with all *visualization* aspects. For the future, we plan to participate in a project in which the main activity will be to create and operating a pan-European e-Science Support Centre as a global astronomical environment in which portals such as gateways with *visualization* included will be as part of essential requirements. In the future we want instead of grid infrastructure to use the cloud resources.

ACKNOWLEDGMENT

This work was supported by Slovak Research and Development Agency under RPEU-0024-06 project, and by VEGA project No. 2/0211/09, as well as by EGEE III EU FP7 RI project: Enabling Grids for E-science III (2008-2010) FP7-222667 and by the project RECLER ITMS: 26240220029 and SMART II ITMS: 26240120029.

.

REFERENCES

[1] Wilkins-Diehr, N.,Gannon, D., Klimeck, G., Oster, S., and Pamidighantam, S., 2008. TeraGrid Science Gateways and Their Impact on Science, *IEEE Computer* 41(11):32-41, Nov 2008.

[2] Kandaswamy, G., Fang, L., Huang, Y., Shirasuna, S., Marru, S., and Gannon,D. 2006. Building Web Services for Scientific Grid Applications. *IBM Journal of Research and Development* 50(2-3), 2006.

[3] Krishnan, L. et al. 2006. Opal: Simple Web Services Wrappers for Scientific Applications. *IEEE International Conference on Web Services (ICWS 2006)*, Sep 18-22, Chicago.

[4] Oinn,T. et al. 2004. Taverna: A tool for the composition and enactment of bioinformatics workflows. *Bioinformatics Journal*, 20(17), 3045-3054.

[5] Wenjun Wu, Thomas Uram, Michael Wilde, Mark Hereld, Michael E. Papka. Accelerating Science Gateway Development with Web 2.0 and Swift. http://www.mcs.anl.gov/uploads/cels/papers/P1765.pdf.

[6] Paulech T., Jakubík M. Neslušan L., P. A. Dobczynski, and G. Leto Extended modeling of the Oort cloud formation from the initial protoplanetary disc. On 4th International Workshopon Grid Computing for Complex Problems. Oct. 27-29 2008 pp 142-150.

[7] Astalos Jan. Experiences from porting the astrophysical simulation “The unified theory of Kuiper-belt and Oort-cloud formationi” to EGEE grid. The 3rd EGEE UF .Egee user forum Viena 2009.