International Journal on Advances in Internet Technology

2014 vol. 7 nr. 3&4

The *International Journal on Advances in Internet Technology* is published by IARIA.
ISSN: 1942-2652
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Internet Technology
Volume 7, Number 3 & 4, 2014

Thierry Brouard, Université François Rabelais de Tours, France
Dumitru Dan Burdescu, University of Craiova, Romania
Carlos T. Calafate, Universitat Politècnica de València, Spain
Christian Callegari, University of Pisa, Italy
Juan-Vicente Capella-Hernández, Universitat Politècnica de València, Spain
Miriam A. M. Capretz, The University of Western Ontario, Canada
Ajay Chakravarthy, University of Southampton IT Innovation Centre, UK
Chin-Chen Chang, Feng Chia University, Taiwan
Ruay-Shiung Chang, National Dong Hwa University, Taiwan
Tzung-Shi Chen, National University of Tainan, Taiwan
Xi Chen, University of Washington, USA
Dickson Chiu, Dickson Computer Systems, Hong Kong
IlKwon Cho, National Information Society Agency, South Korea
Andrzej Chydzinski, Silesian University of Technology, Poland
Noël Crespi, Telecom SudParis, France
Antonio Cuadra-Sanchez, Indra, Spain
Javier Cubo, University of Malaga, Spain
Alfredo Cuzzocrea, University of Calabria, Italy
Jan de Meer, smartspace®lab.eu GmbH, Germany
Sagarmay Deb, Central Queensland University, Australia
Javier Del Ser, Tecnalia Research & Innovation, Spain
Philipe Devienne, LIFL - Université Lille 1 - CNRS, France
Kamil Dimililer, Near East Universiy, Cyprus
Martin Dobler, Vorarlberg University of Applied Sciences, Austria
Jean-Michel Dricot, Université Libre de Bruxelles, Belgium
Matthias Ehmann, Universität Bayreuth, Germany
Tarek El-Bawab, Jackson State University, USA
Nashwa Mamdouh El-Bendary, Arab Academy for Science, Technology, and Maritime Transport, Egypt
Mohamed Dafir El Kettani, ENSIAS - Université Mohammed V-Souissi, Morocco
Marc Fabri, Leeds Metropolitan University, UK
Armando Ferro, University of the Basque Country (UPV/EHU), Spain
Anders Fongen, Norwegian Defence Research Establishment, Norway
Giancarlo Fortino, University of Calabria, Italy
Kary Främling, Aalto University, Finland
Steffen Fries, Siemens AG, Corporate Technology - Munich, Germany
Ivan Ganchev, University of Limerick, Ireland
Shang Gao, Zhongnan University of Economics and Law, China
Kamini Garg, University of Applied Sciences Southern Switzerland, Lugano, Switzerland
Rosario Giuseppe Garroppo, Dipartimento Ingegneria dell'informazione - Università di Pisa, Italy
Thierry Gayraud, LAAS-CNRS / Université de Toulouse / Université Paul Sabatier, France
Christos K. Georgiadis, University of Macedonia, Greece
Katja Gilly, Universidad Miguel Hernandez, Spain
Feliz Gouveia, Universidade Fernando Pessoa - Porto, Portugal
Kannan Govindan, Crash Avoidance Metrics Partnership (CAMP), USA
Bill Grosky, University of Michigan-Dearborn, USA
Vic Grout, Glyndŵr University, UK

Jason Gu, Singapore University of Technology and Design, Singapore
Christophe Guéret, Vrije Universiteit Amsterdam, Nederlands
Frederic Guidec, IRISA-UBS, Université de Bretagne-Sud, France
Bin Guo, Northwestern Polytechnical University, China
Gerhard Hancke, Royal Holloway / University of London, UK
Arthur Herzog, Technische Universität Darmstadt, Germany
Rattikorn Hewett, Whitacre College of Engineering, Texas Tech University, USA
Quang Hieu Vu, EBTIC, Khalifa University, Arab Emirates
Hiroaki Higaki, Tokyo Denki University, Japan
Eva Hladká, Masaryk University, Czech Republic
Dong Ho Cho, Korea Advanced Institute of Science and Technology (KAIST), Korea
Anna Hristoskova, Ghent University - IBBT, Belgium
Ching-Hsien (Robert) Hsu, Chung Hua University, Taiwan
Chi Hung, Tsinghua University, China
Edward Hung, Hong Kong Polytechnic University, Hong Kong
Raj Jain, Washington University in St. Louis , USA
Edward Jaser, Princess Sumaya University for Technology - Amman, Jordan
Terje Jensen, Telenor Group Industrial Development / Norwegian University of Science and Technology, Norway
Yasushi Kambayashi, Nippon Institute of Technology, Japan
Georgios Kambourakis, University of the Aegean, Greece
Atsushi Kanai, Hosei University, Japan
Henrik Karstoft , Aarhus University, Denmark
Dimitrios Katsaros, University of Thessaly, Greece
Ayad ali Keshlaf, Newcastle University, UK
Reinhard Klemm, Avaya Labs Research, USA
Samad Kolahi, Unitec Institute Of Technology, New Zealand
Dmitry Korzun, Petrozavodsk State University, Russia / Aalto University, Finland
Evangelos Kranakis, Carleton University - Ottawa, Canada
Slawomir Kuklinski, Warsaw University of Technology, Poland
Andrew Kusiak, The University of Iowa, USA
Mikel Larrea, University of the Basque Country UPV/EHU, Spain
Frédéric Le Mouël, University of Lyon, INSA Lyon / INRIA, France
Nicolas Le Sommer, Université Européenne de Bretagne, France
Juong-Sik Lee, Nokia Research Center, USA
Wolfgang Leister, Norsk Regnesentral (Norwegian Computing Center), Norway
Clement Leung, Hong Kong Baptist University, Hong Kong
Man-Sze Li , IC Focus, UK
Longzhuang Li, Texas A&M University-Corpus Christi, USA
Yaohang Li, Old Dominion University, USA
Jong Chern Lim, University College Dublin, Ireland
Lu Liu, University of Derby, UK
Damon Shing-Min Liu, National Chung Cheng University, Taiwan
Michael D. Logothetis, University of Patras, Greece
Malamati Louta, University of Western Macedonia, Greece
Maode Ma, Nanyang Technological University, Singapore
Elsa María Macías López, University of Las Palmas de Gran Canaria, Spain

Olaf Maennel, Loughborough University, UK
Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France
Yong Man, KAIST (Korea advanced Institute of Science and Technology), South Korea
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Chengying Mao, Jiangxi University of Finance and Economics, China
Brandeis H. Marshall, Purdue University, USA
Sergio Martín Gutiérrez, UNED-Spanish University for Distance Education, Spain
Constandinos Mavromoustakis, University of Nicosia, Cyprus
Hamid Mcheick, Université du Québec à Chicoutimi, Canada
Shawn McKee, University of Michigan, USA
Stephanie Meerkamm, Siemens AG in Erlangen, Germany
Kalogiannakis Michail, University of Crete, Greece
Peter Mikulecky, University of Hradec Kralove, Czech Republic
Moeiz Miraoui, Université du Québec/École de Technologie Supérieure - Montréal, Canada
Shahab Mokarizadeh, Royal Institute of Technology (KTH) - Stockholm, Sweden
Mario Montagud Climent, Polytechnic University of Valencia (UPV), Spain
Stefano Montanelli, Università degli Studi di Milano, Italy
Julius Müller, TU- Berlin, Germany
Juan Pedro Muñoz-Gea, Universidad Politécnica de Cartagena, Spain
Krishna Murthy, Global IT Solutions at Quintiles - Raleigh, USA
Alex Ng, University of Ballarat, Australia
Christopher Nguyen, Intel Corp, USA
Vlad Nicolicin Georgescu, SP2 Solutions, France
Petros Nicopolitidis, Aristotle University of Thessaloniki, Greece
Carlo Nocentini, Università degli Studi di Firenze, Italy
Federica Paganelli, CNIT - Unit of Research at the University of Florence, Italy
Carlos E. Palau, Universidad Politecnica de Valencia, Spain
Matteo Palmonari, University of Milan-Bicocca, Italy
Ignazio Passero, University of Salerno, Italy
Serena Pastore, INAF - Astronomical Observatory of Padova, Italy
Fredrik Paulsson, Umeå University, Sweden
Rubem Pereira, Liverpool John Moores University, UK
Mark Perry, University of Western Ontario/Faculty of Law/ Faculty of Science - London, Canada
Yulia Ponomarchuk, Far Eastern State Transport University, Russia
Jari Porras, Lappeenranta University of Technology, Finland
Neeli R. Prasad, Aalborg University, Denmark
Drogkaris Prokopios, University of the Aegean, Greece
Emanuel Puschita, Technical University of Cluj-Napoca, Romania
Lucia Rapanotti, The Open University, UK
Gianluca Reali, Università degli Studi di Perugia, Italy
Jelena Revzina, Transport and Telecommunication Institute, Latvia
Karim Mohammed Rezaul, Glyndwr University, UK
Leon Reznik, Rochester Institute of Technology, USA
Joel Rodrigues, Instituto de Telecomunicações / University of Beira Interior, Portugal
Simon Pietro Romano, University of Napoli Federico II, Italy
Michele Ruta, Politecnico di Bari, Italy

Jorge Sá Silva, University of Coimbra, Portugal
Sébastien Salva, University of Auvergne, France
Ahmad Tajuddin Samsudin, Telekom Malaysia Research & Development, Malaysia
Josemaria Malgosa Sanahuja, Polytechnic University of Cartagena, Spain
Luis Enrique Sánchez Crespo, Sicaman Nuevas Tecnologías / University of Castilla-La Mancha, Spain
Paul Sant, University of Bedfordshire, UK
Brahmananda Sapkota, University of Twente, The Netherlands
Alberto Schaeffer-Filho, Lancaster University, UK
Peter Schartner, Klagenfurt University, System Security Group, Austria
Rainer Schmidt, Aalen University, Germany
Thomas C. Schmidt, HAW Hamburg, Germany
Didier Sebastien, University of Reunion Island, France
Zary Segall, Chair Professor, Royal Institute of Technology, Sweden
Dimitrios Serpanos, University of Patras and ISI/RC ATHENA, Greece
Jawwad A. Shamsi, FAST-National University of Computer and Emerging Sciences, Karachi, Pakistan
Michael Sheng, The University of Adelaide, Australia
Kazuhiko Shibuya, The Institute of Statistical Mathematics, Japan
Roman Y. Shtykh, Rakuten, Inc., Japan
Patrick Siarry, Université Paris 12 (LiSSi), France
Jose-Luis Sierra-Rodriguez, Complutense University of Madrid, Spain
Simone Silvestri, Sapienza University of Rome, Italy
Åsa Smedberg, Stockholm University, Sweden
Vasco N. G. J. Soares, Instituto de Telecomunicações / University of Beira Interior / Polytechnic Institute of Castelo Branco, Portugal
Radosveta Sokullu, Ege University, Turkey
José Soler, Technical University of Denmark, Denmark
Victor J. Sosa-Sosa, CINVESTAV-Tamaulipas, Mexico
Dora Souliou, National Technical University of Athens, Greece
João Paulo Sousa, Instituto Politécnico de Bragança, Portugal
Kostas Stamos, Computer Technology Institute & Press "Diophantus" / Technological Educational Institute of Patras, Greece
Vladimir Stantchev, SRH University Berlin, Germany
Tim Strayer, Raytheon BBN Technologies, USA
Masashi Sugano, School of Knowledge and Information Systems, Osaka Prefecture University, Japan
Tae-Eung Sung, Korea Institute of Science and Technology Information (KISTI), Korea
Sayed Gholam Hassan Tabatabaei, Isfahan University of Technology, Iran
Yutaka Takahashi, Kyoto University, Japan
Yoshiaki Taniguchi, Osaka University, Japan
Nazif Cihan Tas, Siemens Corporation, Corporate Research and Technology, USA
Alessandro Testa, University of Naples "Federico II" / Institute of High Performance Computing and Networking (ICAR) of National Research Council (CNR), Italy
Stephanie Teufel, University of Fribourg, Switzerland
Parimala Thulasiraman, University of Manitoba, Canada
Pierre Tiako, Langston University, USA
Ioan Toma, STI Innsbruck/University Innsbruck, Austria
Orazio Tomarchio, Universita' di Catania, Italy

International Journal on Advances in Internet Technology
Volume 7, Numbers 3 & 4, 2014

CONTENTS

pages: 161 - 171
A Study of Unsounded Code Strings at the End of Online Messages of a Q&A Site and a Micro Blog
Kunihiro Nakajima, Ryukoku University, Japan
Yasuhiko Watanabe, Ryukoku University, Japan
Subaru Nakayama, Ryukoku University, Japan
Kenji Umemoto, Ryukoku University, Japan
Ryo Nishimura, Ryukoku University, Japan
Yoshihiro Okada, Ryukoku University, Japan

pages: 172 - 182
Focusing on Older Web Users: An Experience in Patagonia Argentina
Viviana Saldaño, University of Patagonia Austral (UNPA-UACO), Argentina
Adriana Martin, University of Patagonia Austral (UNPA-UACO), Argentina
Gabriela Gaetán, University of Patagonia Austral (UNPA-UACO), Argentina
Diego Vilte, University of Patagonia Austral (UNPA-UACO), Argentina

pages: 183 - 192
Acceptance and Value of Mobile Payment Service Designs in Complex Ecosystems
Elisabeth Pergler, CAMPUS 02 University of Applied Sciences, Information Technologies & Business Informatics, Austria
Daniela Glatz, evolaris next level GmbH, Austria
Christian Adelsberger, evolaris next level GmbH, Austria
Rainer Schamberger, PSA Payment Services Austria GmbH, Austria

pages: 193 - 204
Dependency Parsing and Its Application using Hierarchical Structure in Japanese Language
Kazuki Ono, KLab Inc., Japan
Kenji Hatano, Doshisha University, Japan

pages: 205 - 217
Development, Testing, and End-User Evaluation of Pervasive Community Signatures and Micro-Agreements Infrastructure - Architecture, Android Implementation, Performance Tests, Usage Examples, and User Evaluation
Mitja Vardjan, SETCCE, Slovenia
Helena Halas, SETCCE, Slovenia
Simon Jureša, SETCCE, Slovenia
Jan Porekar, SETCCE, Slovenia

pages: 218 - 231
A Survey on Modelling Historical Administrative Information on the Semantic Web
Jesper Zedlitz, Christian-Albrechts-Universität Kiel, Germany
Norbert Luttenberger, Christian-Albrechts-Universität Kiel, Germany

pages: 232 - 242
Efficient and Adaptive Web-native Live Video Streaming

Luigi Lo Iacono, Cologne University of Applied Sciences, Germany
Silvia Santano Guillén, G&L Geißendörfer & Leschinsky GmbH, Germany

pages: 243 - 261
Improving Web Accessibility: Computing New Web Page Design with NSGA-II for People with Low Vision
Yoann Bonavero, LIRMM, CNRS and Université de Montpellier, France
Marianne Huchard, LIRMM, CNRS and Université de Montpellier, France
Michel Meynard, LIRMM, CNRS and Université de Montpellier, France

pages: 262 - 271
A Non Intrusive Method for Measuring Visual Attention Designed for the Study and Characterization of Users' Behavior in Serious Games
Zahen Malla Osman, Conservatoire National des Arts et Métiers, France
Jérôme Dupire, Conservatoire National des Arts et Métiers, France
Alexandre Topol, Conservatoire National des Arts et Métiers, France
Pierre Cubaud, Conservatoire National des Arts et Métiers, France

A Study of Unsounded Code Strings at the End of Online Messages of a Q&A Site and a Micro Blog

Kunihiro Nakajima, Yasuhiko Watanabe, Subaru Nakayama, Kenji Umemoto, Ryo Nishimura, and Yoshihiro Okada
Ryukoku University
Seta, Otsu, Shiga, Japan
Email: t13m071@mail.ryukoku.ac.jp, watanabe@rins.ryukoku.ac.jp, t090433@mail.ryukoku.ac.jp,
t11m074@mail.ryukoku.ac.jp, r_nishimura@afc.ryukoku.ac.jp, okada@rins.ryukoku.ac.jp

***Abstract*—In this study, we compare answers in a Q&A site with messages in a micro blog and discuss how we use unsounded code strings at the end of online messages. We first show that unsounded code strings at the end of answers in a Q&A site are used for not only smooth communication but an other purpose, minimum length limit avoidance. Next, we show that the length of unsounded code strings at the end of answers in a Q&A site, which are used for smooth communication, have a similar distribution pattern to those at the end of messages in a micro blog. On the other hand, the length of unsounded code strings used for minimum length limit avoidance have a different distribution pattern. Furthermore, we compare frequently used unsounded code strings at the end of answers in a Q&A site with those at the end of messages in a micro blog. Finally, we show our results are useful to analyze users' messages in online communities. In this study, we used the data of Yahoo! chiebukuro, a widely-used Japanese Q&A site, and Twitter for observation and examination.**

***Keywords*—unsounded code string, micro blog, Twitter, Q&A site, Yahoo! chiebukuro.**

I. INTRODUCTION

We often find consecutive unsounded marks and characters are used at the end of online messages, such as mails, chattings, and questions and answers in Q&A sites. As a result, it is important to investigate how these expressions were used.

(exp 1) *sound recorder demo aru teido ha dekiru kedo, yappari Sound Engine ga osusume kana...* (You may be able to do a lot by using sound recorders, however, the one I would like to recommend is Sound Engine...)

(exp 1) is an answer submitted to a Japanese Q&A site, Yahoo! chiebukuro. In this case, periods are used consecutively at the end of it. It is probable that the answerer of (exp 1) used the three consecutive periods for expressing his/her opinion gently, in other words, for smooth communication. In this study, we define unsounded marks and characters as *unsounded codes*. Furthermore, we define three or more consecutive unsounded codes as *unsounded code strings*. For example, in Yahoo! chiebukuro, 25 % of answers have unsounded code strings, in other words, three or more consecutive unsounded codes at the end of them. Although unsounded code strings are popular, there are few studies on them. As a result, we investigated how we use unsounded code strings at the end of online messages [1]. In the report, we compared answers in a Q&A site with messages in a micro blog and discussed how we use unsounded code strings at the end of online messages. We used the data of Yahoo! chiebukuro [2], a widely-used Japanese Q&A site, and Twitter [3] for observation and examination. In this study, we review our previous report and show the new results of our study. Especially, we compare frequently used unsounded code strings at the end of answers in a Q&A site with those at the end of messages in a micro blog. The results of this study will give us a chance to understand not only the usage of unsounded code strings in online messages but the purposes and behaviors of users in online communities. Especially, the results can be useful to analyze the impacts of communication constraints on users' messages and communications. In this paper, we show our results are useful to analyze the impacts of the minimum length limit in Yahoo! chiebukuro on users' messages and communications.

The rest of this paper is organized as follows: In Section II, we surveys the related works. In Section III, we describe how unsounded code strings are used at the end of answers in a Q&A site. On the other hand, in Section IV, we describe how unsounded code strings are used at the end of messages in a micro blog. Furthermore, we compare unsounded code strings at the end of answers in a Q&A site with those at the end of messages in a micro blog. In Section V, we show our results are useful to analyze users' messages in online communities. Finally, in Section VI, we present our conclusions.

II. RELATED WORKS

Yamamoto pointed out that the number of users in communication media for exchanging short text messages has been increasing rapidly [4]. One good example is Twitter. Twitter has succeeded in winning the hearts and minds of many users. The reason is to limiting the message length to 140 characters. By limiting the message length to 140 characters, Twitter has succeeded in encouraging users to submit many messages quickly and enhancing their communications. As a result, in order to develop new communication media technology, we

TABLE I. THE NUMBERS OF QUESTIONERS AND ANSWERERS IN YAHOO! CHIEBUKURO (FROM APRIL/2004 TO OCTOBER/2005).

	number of questioners	number of answerers
the data of Yahoo! chiebukuro	165,064	183,242

TABLE II. THE NUMBERS OF QUESTIONS AND ANSWERS IN YAHOO! CHIEBUKURO (FROM APRIL/2004 TO OCTOBER/2005).

	number of questions	number of answers
the data of Yahoo! chiebukuro	3,116,009	13,477,785

should investigate short text messages. For example, online messages in Twitter consist of

- strings for reference (URL, username, hashtag, etc.),
- utterable words, and
- unsounded code strings.

Unsouded code strings are used frequently in short text messages. However, there are few studies on unsounded code strings, except emoticons. As a result, in order to develop new communication media technology, it is important to investigate unsounded code strings.

Emoticons, sometimes called face marks, are a kind of unsounded code strings. First emoticon, smiley face ";-)", was proposed by Scott Fahlman in September 1982 [5]. After his proposal, many emoticons have been used widely in online messages, such as email, chat, and newsgroup posts [6]. As a result, a large number of studies have been made on emoticons.

Many researchers in computational linguistics proposed methods of extracting and classifying emoticons in online messages. Inoue et al. analyzed 1,000 sentences in email messages and developed a system which extracted emotional expressions, especially emoticons, embedded in email messages [7]. Nakamura et al. proposed a method of learning emoticons for a natural language dialogue system from chat dialogue data in the Internet [8]. Tanaka et al. proposed methods for extracting emoticons in text and classifying them into some emotional categories [9]. Bedrick et al. proposed robust emoticon detection method based on weighted context-free grammars [10]. Hogenboom et al. showed that sentiment classification accuracy was improved by using manually created emoticon sentiment lexicon [11].

On the other hand, many researchers in social science analyzed how we use emoticons in online messages. Witmer and Katzman reported that women use more graphic accents (emoticons) than men do in their computer-mediated communication (CMC) [12]. Walther and D'Addario showed that emoticons' contributions were outweighed by verbal content [13]. Derks et al. reported emoticons are useful in strengthening the intensity of a verbal message [14]. Byron and Baldridge reported readers were likely to rate sender's emails more likeable if they used emoticons [15]. Harada discussed how Japanese speakers use emoticons for promoting communication smoothly from the viewpoint of politeness [16]. Kato et al. analyzed positive and negative emoticons and reported that negative emoticons are misinterpreted more frequently than positive ones [17]. Furthermore, Kato et al. reported that emoticons are used more frequently between close friends than ordinary acquaintances [18].

We think emoticons are a kind of unsounded code strings, however, there are few studies on other kinds of unsounded code strings. As a result, we should investigate not only emoticons but other kinds of unsounded code strings. The results of this study are useful to understand the usage of unsounded code strings in online messages. Furthermore, the results will give us a chance to understand the purposes and behaviors of users in online communities.

III. UNSOUNDED CODE STRINGS AT THE END OF ANSWERS IN A Q&A SITE

In this section, we discuss unsounded code strings at the end of answers submitted to a Q&A site.

Before we define a unsounded code string, we explain the data of Yahoo! chiebukuro, which we used for investigating unsounded code strings in a Q&A site. Yahoo! chiebukuro is a Japanese version of Yahoo! answers and one of the most popular Q&A sites in Japan. In Yahoo! chiebukuro, each user can submit his/her answer only one time to one question. Each questioner is requested to determine which answer to his/her question is best. The selected answer is called *best answer*. The data of Yahoo! chiebukuro was published by Yahoo! JAPAN via National Institute of Informatics in 2007 [19]. This data consists of about 3.11 million questions and 13.47 million answers which were posted on Yahoo! chiebukuro from April/2004 to October/2005. In the data, each question has at least one answer because questions with no answers were removed. In order to avoid identifying individuals, user accounts were replaced with unique ID numbers. By using these ID numbers, we can trace any user's questions and answers in the data. Table I shows the numbers of questioners and answerers in the data of Yahoo! chiebukuro. Table II shows the numbers of their questions and answers in the data of Yahoo! chiebukuro.

Next, we define an unsounded code and unsounded code strings. In this study, we define that an unsounded code string is three or more consecutive unsounded codes. In this study, unsounded codes are limited to the following marks and characters:

- punctuation marks,
- Greek characters,
- Cyrillic characters, and
- ruled lines.

These marks and characters are generally unsounded when they are used at the end of Japanese sentences. We observed unsounded code strings at the end of answers submitted to Yahoo! chiebukuro, and found they were used for

1) smooth communications

(exp 2) *koko ni kaki shirushita bunmen wo sonomama kanojyo ni misete ageru koto wo osusume shimasu. futari no aida ni shinrai kankei ga kizukete iru nara kitto daijyobu!!!* (You had better show what you described here to your girl friend with no change

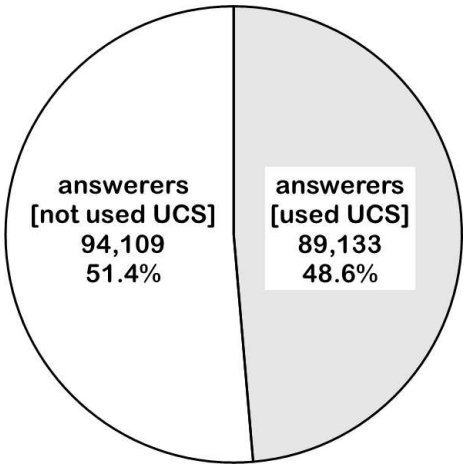


Fig. 1. The proportion and number of answerers who used unsounded code strings at the end of their answers (from April/2004 to October/2005). UCS means an unsounded code string.

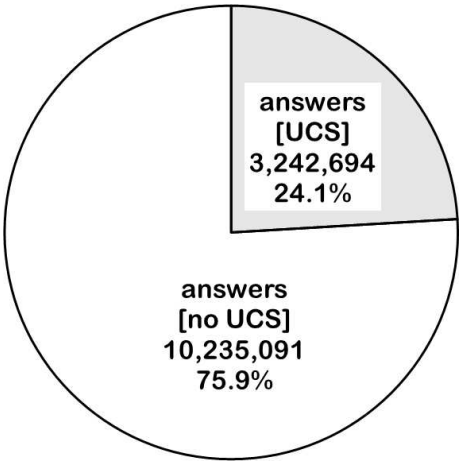


Fig. 2. The proportion and number of answers that have unsounded code strings at the end of them (from April/2004 to October/2005).

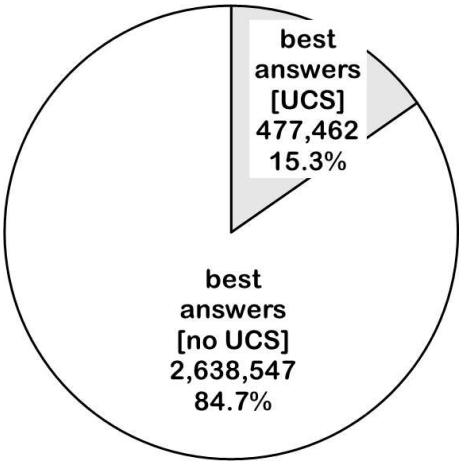


Fig. 3. The proportion and number of best answers that have unsounded code strings at the end of them (from April/2004 to October/2005).

at all. If you have a trust relationship with her, you don't worry!!!)

2) minimum length limit avoidance

(exp 3) *alumi foiru ni tsutsun de hi no naka ni pon!!!!!!!!!!!!!* (Wrap aluminum foil around and pop it into a fire!!!!!!!!!!!!!)

The minimum length limit was introduced into Yahoo! chiebukuro in May/2004. Due to this limit, users in Yahoo! chiebukuro are prohibited from submitting answers less than 25 multibyte characters long. Makoto Okamoto [20], a former producer of Yahoo! chiebukuro, said that this rule was introduced for avoiding less informative answer submissions. In this rule, one single byte character is counted as 0.5 multibyte character. In order to avoid this limit, the answerer of (exp 3) used 13 "!" at the end of his/her answer. We may note that, in case of Japanese texts, the length of words and sentences are generally counted by multibyte characters. In this study, single byte characters are counted as 0.5 multibyte characters. We count characters in the data of Yahoo! chiebukuro by using programming language Perl (version 5.14.2) [21] on Ubuntu linux (version 12.04) [22].

As shown in Table I, there are 183,242 users each of whom submitted at least one answer to Yahoo! chiebukuro. Figure 1 shows the proportion and number of users who used unsounded code strings at the end of their answers. As shown in Table II, 13,477,785 answers were submitted to Yahoo! chiebukuro, and 3,116,009 of them were selected as best answers. Figure 2 shows the proportion and number of answers that have unsounded code strings at the end of them. On the other hand, Figure 3 shows the proportion and number of best answers that have unsounded code strings at the end of them.

Figure 4 shows the cumulative relative frequency distribution of

- the length of all the answers,
- the length of answers that have unsounded code strings at the end of them, and
- the length of unsounded code strings.

As shown in Figure 4, the median of the length of unsounded code strings at the end of answers is 10 multibyte characters. This value is more than twice the length of unsounded code strings at the end of (exp 1) and (exp 2). We think that it is too long for smooth communication. As a result, we investigate the association between the length of

- unsounded code string at the end of answers and
- the other part of them.

The result is shown in Figure 5. In Figure 5, the heatmap shows the association between the length of unsounded code string at the end of answers and the other part of the answers. In the heatmap, darker color denotes more frequent data element. The heatmap shows long unsounded code strings at the end of answers are mainly used when the other part of the answers are less than 25 multibyte characters long. Furthermore, unsounded code strings at the end of the answers

TABLE III. THE NUMBER OF ANSWERERS, ANSWERS, AND BEST ANSWERS IN CASE OF ANSWERS THE LENGTH OF WHICH, EXCLUDING UNSOUNDED CODE STRINGS AT THE END OF THEM, WERE (1) LESS THAN 25 MULTIBYTE CHARACTERS AND (2) 25 MULTIBYTE CHARACTERS OR LONGER.

length of answers (excluding unsounded code strings at the end of them)	number of answerers	number of answers	number of best answers	best answer ratio
less than 25 multibyte characters	52,998	1,745,797	191,791	11.0
25 multibyte characters or longer	77,299	1,496,897	285,671	19.1

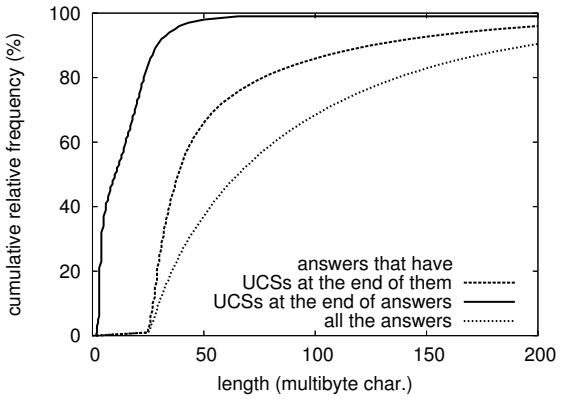


Fig. 4. The cumulative relative frequency distribution of the length of (1) all the answers, (2) answers that have unsounded code strings at the end of them, and (3) unsounded code strings at the end of them. UCS means an unsounded code string.

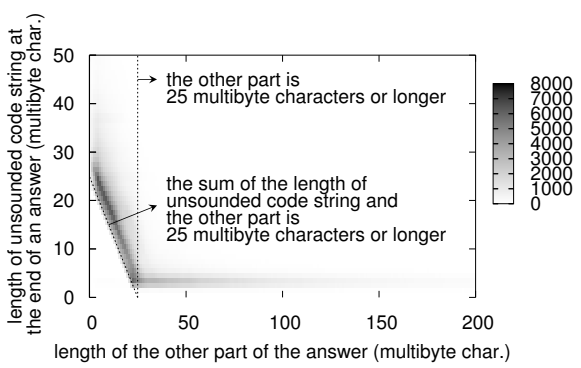


Fig. 5. The heatmap which shows the association between the length of unsounded code string at the end of answers and the other part of the answers.

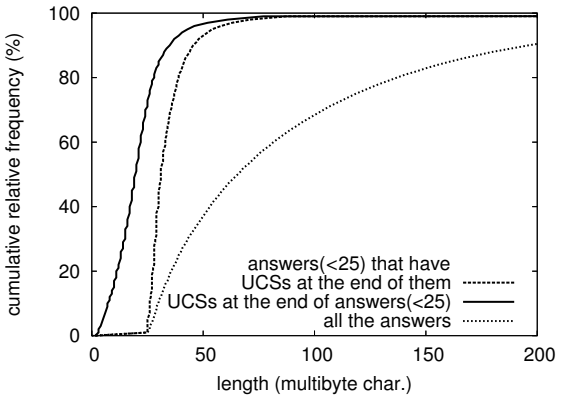


Fig. 6. The cumulative relative frequency distribution of the length of (1) all the answers, (2) answers that have unsounded code strings at the end of them and are less than 25 multibyte characters long (excluding unsounded code strings at the end of them), and (3) unsounded code strings at the end of them.

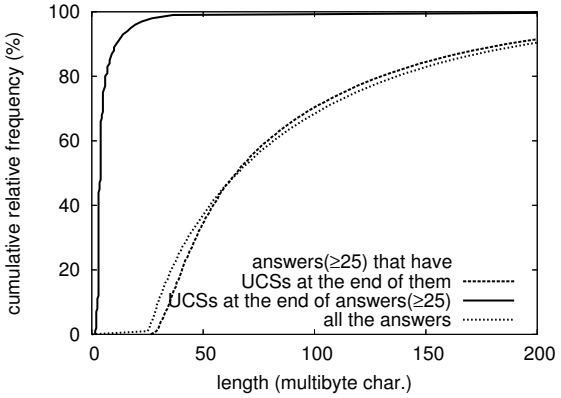


Fig. 7. The cumulative relative frequency distribution of the length of (1) all the answers, (2) answers that have unsounded code strings at the end of them and are 25 multibyte characters or longer (excluding unsounded code strings at the end of them), and (3) unsounded code strings at the end of them.

come in a variety of lengths, however, the sum of the length of unsounded code string at the end and the other part of them, in other words, the length of the answers are frequently 25–30 multibyte characters long. On the other hand, when the other part of answers are more than 25 multibyte characters long, unsounded code strings at the end of the answers are mainly 3–5 multibyte characters long, and the answers come in a variety of lengths. It may be said that the usage of unsounded code strings at the end of answers differs greatly depending on whether the other part of the answers are less than 25 multibyte characters long. As a result, we divided answers that have unsounded code strings at the end of them into

- answers the length of which are less than 25 multibyte characters (excluding unsounded code strings at the end of them)
- answers the length of which are 25 multibyte characters or longer (excluding unsounded code strings at the end of them)

and investigated them in the following points:

- the number of answerers, answers, and best answers

TABLE IV. THE TOP 40 MOST FREQUENTLY USED UNSOUNDED CODE STRINGS AT THE END OF ANSWERS THE LENGTH OF WHICH ARE LESS THAN 25 MULTIBYTE CHARACTERS (EXCLUDING UNSOUNDED CODE STRINGS AT THE END OF THEM)

unsounded code string	frequency
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	22,091
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	20,936
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	20,655
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	20,467
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	20,257
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	20,147
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	19,989
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	19,624
｡｡｡｡｡｡｡｡｡｡｡｡｡	18,722
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	18,552
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	17,718
｡｡｡｡｡｡｡｡｡｡｡	17,491
｡｡｡｡｡｡｡｡｡｡｡｡	17,475
｡｡｡｡｡｡｡｡｡｡	16,681
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	16,173
｡｡｡｡｡｡｡｡｡	15,299
｡｡｡｡｡｡｡｡	13,954
･･･････････････････	13,445
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	13,411
｡｡｡｡｡	13,163
･･･････････････････	13,134
････････････････････	13,122
｡｡｡｡｡｡｡	13,041
｡｡｡｡｡｡	13,003
･･･････････････････	12,901
･･･････････････････	12,867
･･･････････････････	12,148
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	12,113
･･･････････････････	12,099
･･･････････････････	12,051
｡｡｡	11,460
･･･････････････････	11,322
･･･････････････････	10,787
･･･････････････････	10,251
｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡｡	10,193
｡｡｡｡	10,133
･･･････････････	9,818
･･･････････････	9,423
･･･････････････････	9,284
･･･････････････	8,573

TABLE V. THE TOP 40 MOST FREQUENTLY USED UNSOUNDED CODE STRINGS AT THE END OF ANSWERS THE LENGTH OF WHICH ARE 25 MULTIBYTE CHARACTERS OR LONGER (EXCLUDING UNSOUNDED CODE STRINGS AT THE END OF THEM)

unsounded code string	frequency
･･･	205,483
･･･ ｡	137,534
｡｡｡	119,480
｡ ␣␣	65,237
････	42,212
!!!	33,206
･････	25,626
｡｡｡｡	25,306
???	21,107
･･ ｡	19,694
｡｡｡｡｡	18,012
････ ｡	17,387
･･･････	13,995
｡ ␣␣␣	12,453
｡｡｡｡｡｡	9,295
･････ ｡	9,094
!!!!	8,614
^^;	8,391
･･･････	8,268
･･･ ␣	7,905
･･･?	7,593
｡｡｡｡｡｡｡	6,943
``	6,764
････････	6,322
!!!!!	6,277
｡｡｡｡｡｡｡｡	6,082
? ␣␣	5,762
･･･ ｡ ␣	5,725
␣*･*･*･*･*	5649
｡｡｡｡｡｡｡｡｡	5453
･･･････････	5439
｡｡｡｡｡｡｡｡｡｡	4990
･･･)	4875
(^^)	4767
･･････････	4715
｡ ␣␣␣␣	4560
!! ␣	4537
????	4482
･･･････ ｡	4473
｡｡｡｡｡｡｡｡｡｡｡	4459

␣ means a single byte space.

(Table III),

- the length of answers and unsounded code strings at the end of them (Figure 6 and Figure 7), and
- frequently used unsounded code strings at the end of answers (Table IV and Table V).

First, we discuss answers the length of which are less than 25 multibyte characters (excluding unsounded code strings at the end of them). In case of these answers, unsounded code strings at the end of them were used for avoiding the minimum length limit. This limit is a special problem in Yahoo! chiebukuro, not introduced into Twitter. As a result, we do not compare unsounded code strings for avoiding the minimum length limit with those used at the end of online messages in Twitter.

Next, we discuss answers the length of which are 25 multibyte characters or longer (excluding unsounded code strings at the end of them). In case of these answers, unsounded code strings at the end of them were used for smooth communication, not for minimum length limit avoidance. As shown in

Figure 7, the length of these answers (excluding unsounded code strings at the end of them) have a distribution similar to those of all the answers submitted to Yahoo! chiebukuro. As a result, it may be said that, when the length of answers are 25 multibyte characters or longer (excluding unsounded code strings at the end of them), the length of these answers are less affected by whether unsounded code strings are used at the end of them. We compare these unsounded code strings with those used at the end of online messages in Twitter.

Finally, we discuss frequently used unsounded code strings at the end of answers in Yahoo! chiebukuro. Table IV shows the top 40 most frequently used unsounded code strings at the end of answers the length of which are less than 25 multibyte characters (excluding unsounded code strings at the end of them). As shown in Table IV, all of the top 40 most frequently used unsounded code strings are classified into two types of consecutive multibyte characters, consecutive Japanese periods, such as "。。。", or consecutive multibyte bullets, such as "・・・". These consecutive multibyte characters occupy 43.2 % of all the unsounded code strings at the end of answers the length of which are less than 25 multibyte characters. On the other hand, Table V shows the top 40 most frequently used unsounded code strings at the end of answers the length of which are 25 multibyte characters or longer (excluding unsounded code strings at the end of them). As shown in Table V, 17 of the top 40 most frequently used unsounded code strings are also these consecutive multibyte characters. These consecutive multibyte characters occupy 37.2 % of all the unsounded code strings at the end of answers the length of which are 25 multibyte characters or longer. As a result, in both cases, unsounded code strings composed only of these consecutive multibyte characters are used frequently. However, the distribution of the length of them differ greatly. The unsounded code strings at the end of answers the length of which are less than 25 multibyte characters are mainly long. Only 4.5 % of them are "。。。", "。。。。", "・・・", or "・・・・". On the other hand, the unsounded code strings at the end of answers the length of which are 25 multibyte characters or longer are mainly short. 70.7 % of them are "。。。", "。。。。", "・・・", or "・・・・". There is another thing to note. Emoticons are not popular at the end of answers in Yahoo! chiebukuro. As shown in Table V, only two types of emoticons, ^^; and (^^), are ranked in the top 40. We think that it is partially due to the prohibition of ASCII and Shift JIS artworks. ASCII and Shift JIS artworks are often found in blog comment forms, Q&A sites, and web-based bulletin boards. Figure 8 shows an typical example of Shift JIS artwork submitted to a Japanese web-based bulletin board, 2channel [23]. However, operators in Yahoo! chiebukuro thought this kind of submissions, especially, ASCII and Shift JIS artworks, such as Figure 8, were not good for communications in Yahoo! chiebukuro, and then, they prohibited users from submitting them. We think many users in Yahoo! chiebukuro reduced the use of emoticons because they think that emoticons are classified into ASCII and Shift JIS artworks. We will discuss another reason why many users in Yahoo! chiebukuro reduced the use of emoticons in next section.

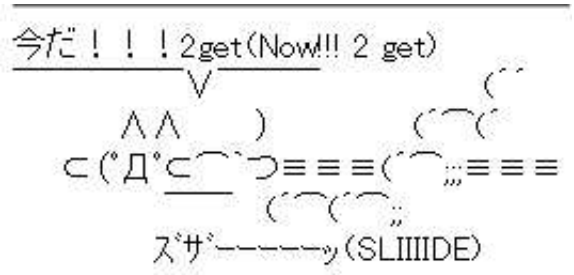


Fig. 8. An example of Shift JIS artwork to a Japanese web-based bulletin board, 2channel.

IV. UNSOUNDED CODE STRINGS AT THE END OF MESSAGES IN A MICRO BLOG

In order to compare with unsounded code strings at the end of answers in Yahoo! chiebukuro, we investigate unsounded code strings at the end of messages in Twitter. We obtained messages submitted to Twitter, in other words, tweets by using the streaming API. However, the streaming API allows us to obtain only 1% of all public streamed tweets because of API restriction. We used the streaming API and obtained 7,085,267 Japanese tweets in three weeks in November and December 2012. These tweets can be classified into three types:

- reply
  A reply to a particular user. It contains "@username" in the body of the tweet.
- retweet
  A retweet is a reply to a tweet that includes the original message.
- normal tweet
  A normal tweet is neither reply nor retweet.

Figure 9 shows the proportion and number of normal tweets, replies, and retweets in Twitter (from November/2012 to December/2012). From these tweets, we extracted 966,187 tweets that have unsounded code strings at the end of them. These 966,187 tweets are 13.6% of all the tweets. Figure 10 shows the proportion and number of normal tweets, replies, and retweets that have unsounded code strings at the end of them. As shown in Figure 9 and Figure 10, replies occupy 44.5% of tweets that have unsounded code strings at the end of them while replies occupy 35.7% of all the tweets. As a result, replies have unsounded code strings at the end of them more frequently than other kinds of tweets. It is because each reply is sent to a particular person. When we send a message to a particular person, we generally try to avoid unnecessary frictions with him/her. As a result, we use unsounded code strings at the end of our replies more frequently than other kinds of tweets.

Before we discuss unsounded code strings at the end of tweets, we remove retweets. It is because, messages in retweets are created not by submitters, but by other users. As a result, retweets are inadequate to investigate how we use unsounded code strings at the end of online messages. Figure 11 shows the cumulative relative frequency distribution of

- the length of all the tweets (excluding retweets),

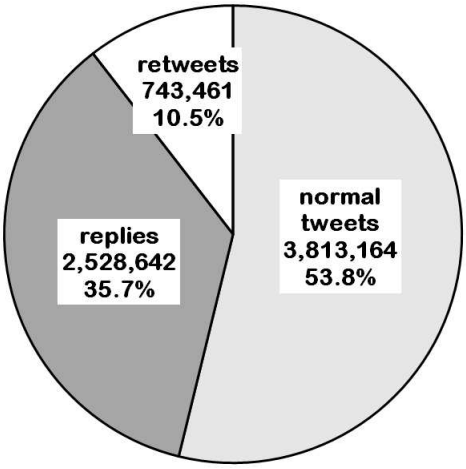


Fig. 9. The proportion and number of normal tweets, replies, and retweets in Twitter (from November/2012 to December/2012).

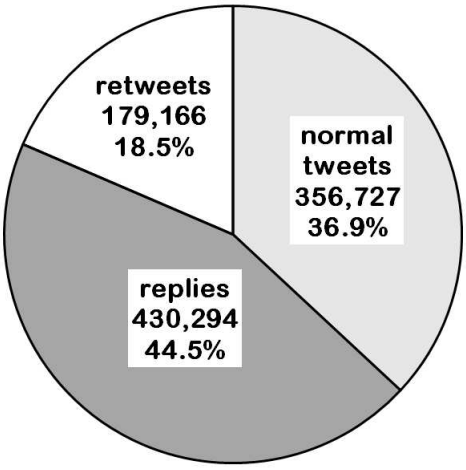


Fig. 10. The proportion and number of normal tweets, replies, and retweets in Twitter that have unsounded code strings at the end of them (from November/2012 to December/2012).

- the length of tweets (excluding retweets) that have unsounded code strings at the end of them, and
- the length of unsounded code strings at the end of tweets (excluding retweets).

In Figure 12, the heatmap shows the association between the length of unsounded code string at the end of tweets and the other part of the tweets. Figure 11 and Figure 12 show unsounded code strings at the end of the tweets are mainly 3–5 multibyte characters long, and the tweets come in a variety of lengths. The length of unsounded code strings at the end of tweets have a similar distribution pattern to those of answers in Yahoo! chiebukuro, which are 25 multibyte characters or longer (excluding unsounded code strings at the end of them). As a result, unsounded code strings at the end of online

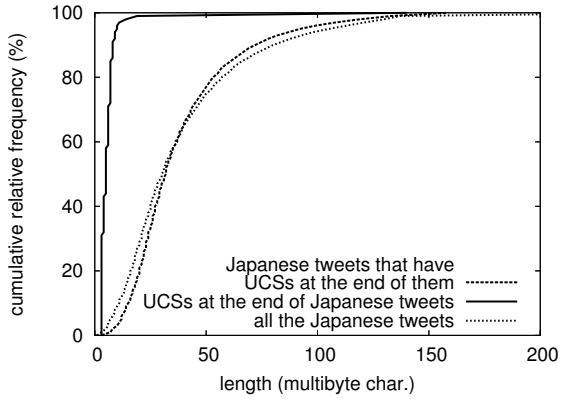


Fig. 11. The cumulative relative frequency distribution of the length of (1) all the Japanese tweets, (2) Japanese tweets that have unsounded code strings at the end of them, and (3) unsounded code strings at the end of them.

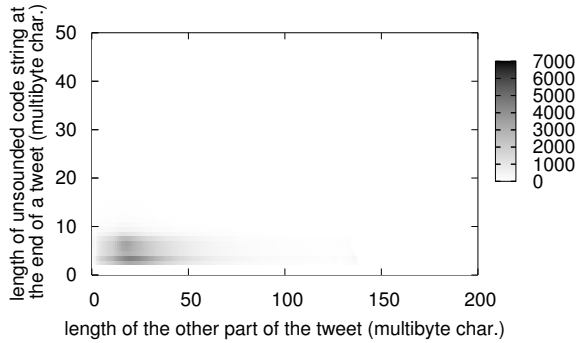


Fig. 12. The heatmap which shows the association between the length of unsounded code string at the end of Japanese tweets and the other part of the tweets.

messages are mainly 3–5 multibyte characters long when they are used for smooth communications with particular persons. Table VI shows the top 40 most frequently used unsounded code strings at the end of tweets (excluding retweets). As shown in Table VI, only one kind of consecutive Japanese periods, “。。。”, and two kinds of consecutive multibyte bullets, “・・・” and “・・・・”, are ranked in the top 40 most frequently used unsounded code strings at the end of tweets (excluding retweets). These consecutive multibyte characters occupy 5.6 % of all the unsounded code strings at the end of tweets (excluding retweets). As a result, these consecutive multibyte characters, such as “。。。” and “・・・”, are used less frequently at the end of tweets than at the end of answers in Yahoo! chiebukuro. On the other hand, emoticons are used more frequently at the end of tweets than at the end of answers in Yahoo! chiebukuro. 24 of the top 40 most frequently used unsounded code strings at the end of tweets (excluding retweets) are emoticons or parts of emoticons. One of the reasons why emoticons are used frequently at the end of tweets is that Twitter users often sent their tweets to

TABLE VI. THE TOP 40 MOST FREQUENTLY USED UNSOUNDED CODE STRINGS AT THE END OF TWEETS (EXCLUDING RETWEETS).

unsounded code string	frequency
^) ／	32518
•••	30710
!!!	26511
)))	13847
(*^^*)	11665
...	10603
!!!!	10307
ー !!	7396
(^^)	7096
°°°	7036
(´• ω •`)	6093
!!!!!	5962
///	5371
(￣▽￣)	5013
! ! !	4823
(^−^)	4281
^)/*	4128
(* ´ ω ` *)	4001
••• ｡	3644
(´; ω ;`)	3458
(´Д‘␣)	3429
(´• ω • ‘)	3361
(^−^)/	3333
!!!!!!	3218
‘*)	3200
(;_;)	3149
???	3123
••••	3028
^)/	3015
ーーー	2989
␣」 ∠)_	2943
^*)	2916
」 ∠)_	2905
(^_^;)	2886
ー !!!	2736
……｡	2695
……	2687
^_^	2680
``	2646
)))))))	2637

␣ means a single byte space.

familiar persons while answerers in Yahoo! chiebukuro almost always sent their answers to strangers. Kato et al. reported that emoticons are used more frequently between close friends than ordinary acquaintances [18]. As a result, we think that emoticons are used more frequently in replies than in normal tweets. It is because replies are sent to particular persons. On the other hand, normal tweets are sent to not only particular persons but general public.

Next, we discuss unsounded code strings at the end of normal tweets and replies, individually. Figure 13 shows the

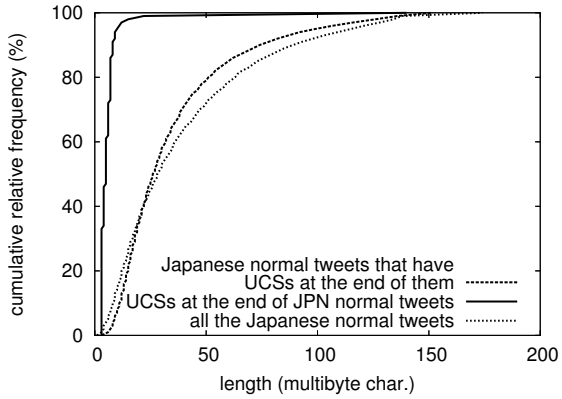


Fig. 13. The cumulative relative frequency distribution of the length of (1) all the Japanese normal tweets, (2) Japanese normal tweets that have unsounded code strings at the end of them (excluding unsounded code strings at the end of them), and (3) unsounded code strings at the end of them.

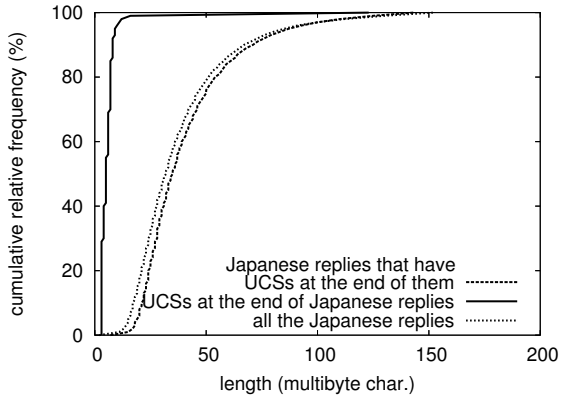


Fig. 14. The cumulative relative frequency distribution of the length of (1) all the Japanese replies, (2) Japanese replies that have unsounded code strings at the end of them (excluding unsounded code strings at the end of them), and (3) unsounded code strings at the end of them.

cumulative relative frequency distribution of

- the length of all the normal tweets,
- the length of normal tweets that have unsounded code strings at the end of them (excluding unsounded code strings at the end of them), and
- the length of unsounded code strings at the end of normal tweets.

Also, Figure 14 shows the cumulative relative frequency distribution of

- the length of all the replies,
- the length of replies that have unsounded code strings at the end of them (excluding unsounded code strings at the end of them), and
- the length of unsounded code strings at the end of replies.

As shown in Figure 14, there are few short replies, especially less than 5 multibyte long. It is because each reply includes "@username". Also, as shown in Figure 14, the length of

TABLE VII. THE TOP 40 MOST FREQUENTLY USED UNSOUNDED CODE STRINGS AT THE END OF NORMAL TWEETS.

unsounded code string	frequency
･･･	20699
!!!	13621
^) ／	11845
. . .	7188
!!!!	5691
。。。	4749
)))	4236
!!!!!	3463
￣ !!	3277
(´･ω･`)	3262
(*^^*)	2807
! ! !	2569
･･･。	2518
￣￣￣	2139
････	2055
␣」∠)_	1925
」∠)_	1922
(^^)	1904
!!!!!!	1890
……	1861
……。	1859
///	1817
(´･ω･‘)	1795
(´Д‘␣)	1736
(´;ω;`)	1661
‘*)	1647
``	1644
???	1557
(*´ω`*)	1490
(;_;)￣	1474
(　▽　)	1470
(^−^)	1411
￣ !!!	1370
! ! ! !	1303
(*_*)	1236
!!!!!!!	1192
^)/*	1165
(´Д`)	1155
･･･?	1102
)))))))	1065

␣ means a single byte space.

TABLE VIII. THE TOP 40 MOST FREQUENTLY USED UNSOUNDED CODE STRINGS AT THE END OF REPLIES.

unsounded code string	frequency
^) ／	20673
!!!	12890
･･･	10011
)))	9611
(*^^*)	8858
(^^)	5192
!!!!	4616
￣ !!	4119
///	3554
(￣▽￣)	3543
. . .	3415
^)/*	2963
(^−^)	2870
(´･ω･`)	2831
(^−^)/	2616
(*´ω`*)	2511
!!!!!	2499
^)/	2349
。。。	2287
! ! !	2254
^*)	2153
(^_^;)	1888
^ ^	1845
(￣;ω;`)	1797
(´Д‘␣)	1693
(;_;)	1675
(*^_^*)	1675
(*´▽`*)	1637
(␣^ω^␣)	1627
)))))))	1572
???	1566
(´･ω･‘)	1566
‘*)	1553
♪(␣´▽`)	1504
(*´∀`*)	1433
\(//▽//)\	1423
(^^)／	1378
￣ !!!_	1366
(￣￣￣)	1361
!!!!!!	1328

␣ means a single byte space.

replies that have unsounded code strings at the end of them have a similar distribution pattern to the length of all the replies. It may be said that the length of replies are less affected by whether unsounded code strings are used at the end of them. This result is similar to the result obtained when we investigated answers in Yahoo! chiebukuro. The length of answers in Yahoo! chiebukuro, which are 25 multibyte characters or longer (excluding unsounded code strings at the end of them), are less affected by whether unsounded code strings are used at the end of them. In both cases of Yahoo! chiebukuro and Twitter, unsounded code strings are used for smooth communication with particular persons. As a result, it may also be said that the length of online messages to particular persons are less affected by whether unsounded code strings for smooth communication are used at the end of them. On the other hand, as shown in Figure 13, the length of normal tweets that have unsounded code strings at the end of them have a slightly different distribution pattern to the length of all the normal tweets. It is because normal tweets were sent to not only particular persons but general public while replies

were sent to particular persons.

Table VII and Table VIII show the top 40 most frequently used unsounded code strings at the end of normal tweets and replies, respectively. As shown in Table VII, 20 kinds of emoticons or parts of emoticons are ranked in the top 40 most frequently used unsounded code strings at the end of normal tweets. These emoticons occupy 36.5 % of all the unsounded code strings ranked in the top 40 most frequently used unsounded code strings at the end of normal tweets. On the other hand, as shown in Table VIII, 29 kinds of emoticons or parts of emoticons are ranked in the top 40 most frequently used unsounded code strings at the end of replies. These emoticons occupy 67.3 % of all the unsounded code strings ranked in the top 40 most frequently used unsounded code strings at the end of replies. As a result, emoticons are used more frequently at the end of replies than at the end of normal tweets. In other words, emoticons are used more frequently at the end of tweets to particular persons than at the end of tweets to general public.

V. Discussions

In this section, we show our results are useful to analyze the impacts of communication constraints on users' messages and communications. We take the minimum length limit in Yahoo! chiebukuro for example.

After the minimum length limit was introduced, users in Yahoo! chiebukuro have been prohibited from submitting answers less than 25 multibyte characters long. However, our study showed 1,745,797 answers, that is, 13.0 % of all the answers in the data of Yahoo! chiebukuro, were less than 25 multibyte characters (excluding unsounded code strings at the end of them). These answers were submitted by 52,998 users, that is, 28.9 % of all the answerers in the data of Yahoo! chiebukuro. It shows that many users in Yahoo! chiebukuro wanted to submit short answers. Furthermore, our study showed unsounded code strings used for smooth communication are mainly 3–5 multibyte characters long. We therefore classify these 1,745,797 answers into two types:

- 1,642,866 answers the unsounded code strings at the end of which were more than 5 characters long, and
- 102,931 answers the unsounded code strings at the end of which were 3-5 characters long.

In the former case, most of the unsounded code strings were thought to be used for minimum length limit avoidance. These unsounded code strings were often unfit for the contents of answers and gave poor impressions to questioners. As a result, the best answer ratio of these 1,642,866 answers was 10.8 % while that of all the answers in the data of Yahoo! chiebukuro was 23.1 %. On the other hand, in the latter case, some of the unsounded code strings were thought to be used for smooth communication. However, the best answer ration of these 102,931 answers was 12.4 %. As a result, in both cases, the best answer ratios were lower than that of all the answers. This result shows short answers are often less informative than long answers, in other words, the minimum length limit is reasonable. However, we should not overlook the positive factor of short answers. Ohsawa et al. reported that short and less informative submissions sometimes promote constructive discussions in web-based bulletin boards [24]. As a result, it is probable that some short and less informative answers stimulate other answerers to submit their good answers and enhance communications in Yahoo! chiebukuro.

VI. Conclusion

In this study, we investigated unsounded code strings at the end of answers in Yahoo! chiebukuro and tweets in Twitter. Although unsounded code strings are popular, there were few studies on them.

In Twitter, unsounded code strings at the end of tweets are used for smooth communication. On the other hand, in Yahoo! chiebukuro, unsounded code strings at the end of answers are used for not only smooth communication but minimum length limit avoidance. The minimum length limit is a special problem in Yahoo! chiebukuro, not introduced into Twitter. We showed that the usage of unsounded code strings at the end of answers in Yahoo! chiebukuro differs greatly depending on whether answers are longer than the minimum length limit. When answers are longer than the minimum length limit, unsounded code strings at the end of them are used for smooth communication. In this case, the length of the unsounded code strings at the end of answers have a similar distribution pattern to the length of unsounded code strings at the end of tweets. Unsounded code strings at the end of the tweets in Twitter and answers in Yahoo! chiebukuro, which are longer than the minimum length limit, are mainly 3–5 multibyte characters long. In addition, we showed the length of replies in Twitter and answers in Yahoo! chiebukuro, which are larger than the minimum length limit, are less affected by whether unsounded code strings are used at the end of them. Furthermore, we showed frequently used unsounded code strings at the end of answers in Yahoo! chiebukuro and tweets in Twitter. We showed that emoticons were not used frequently at the end of answers in Yahoo! chiebukuro. On the other hand, they were used frequently at the end of tweets in Twitter, especially, replies. The difference is whether messages are submitted to familiar persons or not. In other words, emoticons are used more frequently at the end of messages which are sent to familiar persons than to strangers and general public. Finally, we took the minimum length limit in Yahoo! chiebukuro for example and showed our results could be useful to analyze the impacts of communication constraints on users' messages and communications.

In this study, we analyzed and compared unsounded code strings at the end of answers in Yahoo! chiebukuro and Japanese tweets in Twitter. However, it is not enough to obtain general knowledge about unsounded code strings. It is because we have found many unsounded code strings not only in Japanese tweets but also in other language tweets, for example, English, French, Spanish, Portuguese, and so on. We intend to study the usages of unsounded code strings in these languages and compare them with the usage of Japanese unsounded code strings. We think the results of our future work are useful to provide new multilingual communication services.

REFERENCES

[1] K. Nakajima, S. Nakayama, Y. Watanabe, K. Umemoto, R. Nishimura, and Y. Okada, "An analysis of unsounded code strings in online messages of a q&a site and a micro blog," in *Proc. INTERNET 2013*, 2013, pp. 32–37.

[2] *Yahoo! chiebukuro*, Yahoo! JAPAN, 2004. [Online]. Available: http://chiebukuro.yahoo.co.jp/ [retrieved: December, 2014]

[3] *Twitter*, Twitter, Inc., 2006. [Online]. Available: https://twitter.com/ [retrieved: December, 2014]

[4] Y. Yamamoto, "Next-generation communications media, considering the evolution," in *Technical Report of IEICE on Human Communication Science (HCS)*, vol. 109, no. 457, 2010, pp. 19–20.

[5] S. Fahlman, *SMILEY:31 YEARS OLD AND NEVER LOOKED HAPPIER!*, Carnegie Mellon School of Computer Science, 2012. [Online]. Available: http://www.cs.cmu.edu/smiley/ [retrieved: December, 2014]

[6] H. Nojima, "(smily face) as a mean for emotional communication in networks," in *Proc. IPSJ summer programming symposium*, 1989, pp. 41–48.

[7] M. Inoue, M. Fujimaki, and S. Ishizaki, "System for analyzing emotional expression in e-mail text: collection, classification, and analysis of emotional expressions," in *Technical Report of IEICE on Thought and Language (TL)*, vol. 96, no. 608, 1997, pp. 1–8.

[8] J. Nakamura, T. Ikeda, N. Inui, and Y. Kotani, "Learning face marks for natural language dialogue systems," in *Proc. 2003 International Conference on Natural Language Processing and Knowledge Engineering*, 2003, pp. 180–185.

[9] Y. Tanaka, H. Takamura, and M. Okumura, "Extraction and classification of facemarks," in *Proceedings of the 10th international conference on Intelligent user interfaces*, 2005, pp. 28–34.

[10] S. Bedrick, R. Beckley, B. Roark, and R. Sproat, "Robust kaomoji detection in Twitter," in *Proceedings of the Second Workshop on Language in Social Media*, 2012, pp. 56–64.

[11] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong, and U. Kaymak, "Exploiting emoticons in sentiment analysis," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, 2013, pp. 703–710.

[12] D. F. Witmer and S. L. Katzman, "On-line smiles: Does gender make a difference in the use of graphic accents?" *Journal of Computer-Mediated Communication*, vol. 2, no. 4, 1997. [Online]. Available: http://jcmc.indiana.edu/vol2/issue4/witmer1.html [retrieved: December, 2014]

[13] J. B. Walther and K. P. D'Addario, "The impacts of emoticons on message interpretation in computer-mediated communication," *Social Science Computer Review*, vol. 19, no. 3, pp. 324–347, 2001.

[14] D. Derks, A. E. R. Bos, and J. von Grumbkow, "Emoticons and online message interpretation," *Social Science Computer Review*, vol. 26, no. 3, pp. 379–388, 2008.

[15] K. Byron and D. C. Baldridge, "Email recipients' impressions of senders likeability," *Journal of Business Communication*, vol. 44, pp. 137–160, 2007.

[16] T. Harada, "The role of "face marks" in promoting smooth communication and expressing consideration and politeness in japanese," *the journal of the Institute for Language and Culture*, vol. 8, pp. 205–224, 2004.

[17] S. Kato, Y. Kato, M. Kobayashi, and M. Yanagisawa, "Analysis of the kinds of emotions interpreted from the emoticons used in e-mail," *the journal of Japan Society of Educational Information*, vol. 22, no. 4, pp. 31–39, 2007.

[18] S. Kato, Y. Kato, Y. Shimamine, and M. Yanagisawa, "Analysis of functions of emoticons in e-mail communication by mobile phone: Investigation of effects of degrees of intimacy with partners," *the journal of Japan Society of Educational Information*, vol. 24, no. 2, pp. 47–55, 2008.

[19] *The data of Yahoo! chiebukuro*, National Institute of Informatics, 2007. [Online]. Available: http://research.nii.ac.jp/tdc/chiebukuro.html [retrieved: December, 2014]

[20] *Makoto Okamoto*, ACADEMIC RESOURCE GUIDE(ARG), 2009. [Online]. Available: http://www.arg.ne.jp/node/7000 [retrieved: December, 2014]

[21] *The Perl Programming Language*, Perl.org, 2002. [Online]. Available: http://www.perl.org/ [retrieved: December, 2014]

[22] *Ubuntu Japanese Team*, Ubuntu Japanese Team, 2005. [Online]. Available: http://www.ubuntulinux.jp/japaneseteam/ [retrieved: December, 2014]

[23] *2channel*, PACKET MONSTER INC., 1999. [Online]. Available: http://www.2ch.net/ [retrieved: January, 2014]

[24] Y. Ohsawa, N. Matsumura, and Y. NAKAMURA, "Flaming promotes argumentation : A case study on of "2channel"," in *Technical Report of IEICE on Internet Architecture (IA)*, vol. 102, no. 143, 2002, pp. 55–60.

Focusing on Older Web Users: An Experience in Patagonia Argentina

Viviana Saldaño, Adriana Martin, Gabriela Gaetán, and Diego Vilte
Department of Exact Sciences, Caleta Olivia
University of Patagonia Austral (UNPA-UACO)
Santa Cruz, Argentina
e-mail: vivianas@uaco.unpa.edu.ar, adrianaelba.martin@gmail.com,
ggaetan@uaco.unpa.edu.ar, dvilte773@yahoo.com.ar

***Abstract*—Older Web users are now facing one of the most difficult challenges of their lives. The Web changes every day and they cannot keep up with it. As older age comes, individuals experience gradual and fluctuating decline in capabilities. These physical impairments make usage of the Web even more difficult. Web accessibility is an area devoted to solve accessibility problems of disabled people. However, as older people suffer disabilities, although less severe ones, they can profit from Web accessibility solutions. In this article, we review some of the most common impairments that affect older Web users, we analyze how these impairments are considered by Web Accessibility standards, and explore different approaches that improve Web user interface, in particular Email systems. Finally, we introduce our ideas to overcome unsolved Web accessibility barriers for older users describing an experience carried out at our University in Argentinean Patagonia.**

Keywords - Web Accessibility, Older Web users, User Interface (UI).

I. INTRODUCTION

This paper is an improved and expanded version of the ICSEA 2013 conference paper "Web Accessibility for Older Users: A Southern Argentinean View" [1].

The fact that the number and proportion of older people in the world population is progressively increasing arises as a critical factor to the present and future of civilizations, which are developing a strong dependence on Information and Communication Technologies (ICT). Particularly, the Web has become an essential tool and older Web users are struggling to keep up with technological changes and their demands of use in day-to-day life.

Most older adults experience age-related changes to their functional abilities (vision, hearing, cognition and mobility). These changes may complicate Web use [2], particularly for poorly designed sites. In Table I, we show some common functional impairments affecting older Web users, which we extracted from the literature review published by the W3C [3].

The study presented by Sayago and Blat [4] revealed that the accessibility barriers that had a more negative effect on the daily interactions of older people with the Web were remembering steps, understanding computer jargon and using the mouse.

Besides, from this study, we acknowledge that older Web users desire two conditions: independency and inclusiveness. Independency is the ability to use the Web on their own and inclusiveness is the need to interact with the Web using ordinary technology, as they do not intend to be different from the rest of users.

TABLE I. FUNCTIONAL IMPAIRMENTS AFFECTING OLDER WEB USERS

Ability	Impact	Difficulties
Vision	Screen Keyboard	1. Decreasing ability to focus on near tasks 2. Changing color perception and sensitivity 3. Pupil shrinkage and decreasing contrast sensitivity
Hearing	Audio Multimedia	4. Increasing inability to hear higher-pitched sounds
Motor skill	Mouse Keyboard	5. Slowness of movement, trembling
Cognitive	Overall Web use	6. Short term memory problems, concentration difficulties, distraction, change blindness

Another problem that older people have to face is social isolation [5]. Factors like diminished personal social networks, bereavement and health problems contribute to social isolation. Using the internet has significant value for elderly people, since it helps avoiding loneliness, boredom, helplessness, and decline of mental skills and it may increase the self-confidence, ability to learn, and memory retention.

Traditional communication technologies, such as the telephone, have played an important role in mitigating social isolation and supporting group gatherings. Also, the World Wide Web offers potential benefits for older adults, but its uptake is yet extremely limited.

One of the most used online communications is email. As of August 2011, Pew Research Center's report [6] observes that 34% of online seniors use social networking sites like Facebook and LinkedIn and 86% use email. We do not have detailed statistics in Argentina about older adults and online communications, the only fact is that 34.5% of adults who are ages 60 and older use internet.

There are many reasons why older adults do not use the Web [7]. Firstly, they tend to see the Internet as a tool to achieve functional goals such as bill payment, and not as a social or entertainment source [8]. Besides, they need an incentive to get and stay online [9]. It is often younger people who encourage technology use by older adults. Staying connected with geographically remote grandchildren is a major motivation for older adults in using technology

(such as email, chat, blog, video call and social network applications). An interesting finding was reported in [10], where it is suggested that given the right trigger many older people (even those previously uninterested) will make tentative steps towards some technology. In this case, the trigger was a disaster, the "ash cloud", which caused large scale disruption for air travel across Europe in 2010, and it motivated the need for computer usage.

Once older people are online they discover the advantages, such as being able to maintain existing social relationships and perhaps renew old ones that distance had precluded. Over two thirds of "silver surfers" say that using the Internet has improved their lives [11].

Other reasons for non-use of the Web include those involved with age-related impairments, such as the ones presented before in Table I.

In this paper, we analyze senior needs at the Web and explore different proposals to improve their Web interface experience [12]. Taking into account the state-of-the-art and the experience gained by our group while teaching computing to older people, we describe our ideas and show the improvements achieved during the delivery of the courses for elderly Web users. Since many fields are concerned on improving human-technology interaction, such as information retrieval and data mining, Human-Computer Interaction (HCI) and GUI, at this point, we have to clarify how we decided to face this work. We have been working for a while on accessible UI design to conform the W3C accessibility recommendations [13] [14]. Our knowledge gathered about UI design and Web Accessibility standards, permitted us to explore practical techniques to reinforce accessibility and usability and focus on the interaction between our seniors and the Web, using a real experience on Yahoo mail.

The rest of the paper is structured as follows: in Section II, we review Web accessibility standards and their relation with age related disabilities. Then, in Section III, we overview different useful approaches to improve older users' Web interface. After that, in Section IV, we examine three of the most popular email systems and analyze the accomplishment of their interfaces with desired characteristics of an older adult's oriented design. Then, in Section V we describe an experience performed at our University and explain our ideas for improvement. In Section VI, we introduce some discussion based on our experiences. Finally, in Section VII, we conclude and present some further work.

II. Web Accessibility Initiative Guidelines and Aging

The next few decades will see an unparalleled growth in the number of people becoming elderly compared with any other period in human history. The United Nations estimates that by 2050 one out of every five people will be over 60 years of age, and in some countries the proportion will be much higher than this [15].

There are some initiatives that provide advice addressing Web accessibility and usability for all people. As regards older users, many requirements are already considered by these initiatives.

The World Wide Web Consortium (W3C) Web Accessibility Initiative (WAI) [16] brings together people from industry, disability organizations, government, and research labs from around the world to develop guidelines and resources to help make the Web accessible to people with disabilities including auditory, cognitive, neurological, physical, speech, and visual disabilities.

Among these series of guidelines developed by WAI, widely regarded as the international standard for Web accessibility, are: Authoring Tool Accessibility Guidelines (ATAG), User Agent Accessibility Guidelines (UAAG) and Web Content Accessibility Guidelines (WCAG).

- The Authoring Tool Accessibility Guidelines (ATAG) documents define how authoring tools should help Web developers produce Web content that is accessible and complies with Web Content Accessibility Guidelines.
- The User Agent Accessibility Guidelines (UAAG) documents explain how to make user agents (Web browsers, media players, and assistive technologies) accessible to people with disabilities, particularly to increase accessibility to Web content.
- The WCAG documents explain how Web content can be made accessible for people with disabilities. The WCAG 2.0 [17] has twelve guidelines, grouped in four fundamental principles of accessibility: perceivable, operable, understandable, and robust. Each guideline is in turn decomposed in a set of success criteria, which are classified within three levels of conformance: A (lowest), AA, and AAA (highest).

As we have said previously, many requirements for older Web users are already covered by these initiatives. But recently, the WAI [16] has moved a step forward to provide valuable recommendations related to older Web users.

The WAI project, Web Accessibility Initiative: Ageing Education and Harmonization (WAI-AGE) project [18] analyzed the Web accessibility requirements of older Web users based on the research and investigation of many people.

WAI-AGE has identified that the existing WAI accessibility guidelines address the majority of requirements of older people for Web use [19]. It also identified that many Web designers and researchers are not considering the WAI guidelines when making recommendations about Website design for older people.

Although the guidelines developed by WAI were not written with older users' problems in mind, they provide solution to many of them. In Table II, we show the results of performing a matching analysis between most common older people accessibility barriers, presented before in Table I, and the corresponding guideline in WCAG 2.0.

TABLE II. OLDER WEB USERS DIFFICULTIES AND CORRESPONDING WCAG 2.0 GUIDELINES

Difficulty	WCAG 2.0 Guideline
1. Decreasing ability to focus on near tasks	1.4
2. Changing color perception and sensitivity	1.4
3. Pupil shrinkage and decreasing contrast sensitivity	1.4
4. Increasing inability to hear higher-pitched sounds	1.2 – 1.4
5. Slowness of movement, trembling	2.1 – 2.2
6. Short term memory problems, concentration difficulties, distraction, change blindness	2.2 – 2.4 – 3.2 – 3.3

We can see that the first three difficulties, which are visual impairments, are addressed by WCAG 2.0 in guideline 1.4. The fourth barrier, a hearing disability, is tackled by guidelines 1.2 and 1.4. The fifth difficulties, motor impairments, are addressed by guidelines 2.1 and 2.2. Finally, the sixth barriers, cognitive difficulties, are considered by guidelines 2.2, 2.4, 3.2, and 3.3.

The former considerations show that WCAG 2.0 guidelines meet most of older Web users' requirements. The problem is that few developers conceive Websites design with these guidelines in mind. In any case, there are already useful proposals to explore for providing better practices addressing senior needs, as we have highlighted when introducing the bases of our ideas for improvements (see key issue (2) in Section I)

However, having proposals to explore and apply for implementing/developing improvements, attends only part of the problem. Again, as we have highlighted (see key issue (1), finding best proposals to involve elders in the process of discover good improvements, is another challenge.

On this spot, (WAI-AGE) project [18] proposes a working draft, Web Accessibility for Older Users: A Literature Review [3], which is an output particularly focus on Europe but relatively speaking, this document could also apply internationally as well. In particular, Section 3.7 emphasizes older Web users' participation and provides some insights for involving the elderly in Web design and development. For example, some of these basic advises are: making the participants comfortable; keeping them on track; listen for their beliefs about computers and the Web; avoid computer jargon; give them time; thinking aloud; etc.

III. DIFFERENT WEB SOLUTIONS THAT IMPROVE SILVER SURFERS' EXPERIENCE

Many Web solutions have been developed to address cognitive, perceptual, and physical changes related to aging.

Aula et al. [20], designed a simplified search interface, called Etsin. This study showed that a simple design makes the search experience less problematic and more manageable for older adults. Dickinson et al. [7] developed a proof of concept search and navigation system focused on the usability of the interface, which demonstrated that appropriate software could provide a more positive initial experience of the web and could increase elderly persons' confidence in their ability to master the Internet.

In [21], an email system for older people with no experience of Internet use was developed. The system, named Cybrarian system, had reduced functionality and a simple interface. It was meant to attract older people, who were unconfident in the use of computers, to use the Internet and encourage them to progress to more sophisticated Internet use. Hawthorn [22] developed SeniorMail, an email system based on Microsoft's Outlook Express, which addressed the problems older novice have remembering how a system works by having a list of possible actions presented in a simplified menu system. Another solution developed to improve elders Web experience is ElderMail [23]. This system allows seniors to communicate easily with others via a Web based email system designed with a very simple interface that uses only three color coded buttons to send and receive email.

Nevertheless, older people's functional impairments are very different in type (vision, hearing, mobility, cognitive) and severity, and usually change over time. Thus, it is very difficult to specify a unique Web interface that meets the requirements for all of them [24]. So, the solution could be that each individual older user would be able to select the appropriate configuration by themselves.

There are some very interesting works related with this idea such as the IBM's Web Adaptation Technology [25], which develops a browser extension that allows manipulating Web content by combining and applying a number of page transforms and adaptations according to user preferences without requiring Web designers and developers to rewrite their Web content. A similar approach is proposed in [26], an ActiveX-based accessibility solution called Easy Web Browsing was developed to add senior-friendly features to existing Web sites. This solution does not require Web site owners to change their existing Web content. This design allows Web site owners to keep their Web content stylish for the younger people, while enabling senior citizens to access them with the assistive technologies provided by this ActiveX object.

Another tool is the Senior Citizen on the Web 2.0 (SCWeb2) Assistance tool [27], which is designed to assist older users as they use Web 2.0 content. For some users, dynamic content can be problematic due to the many updating components throughout the page, causing them hesitancy, stress, and frustration about unexpected situations. This tool provides help only when users require it, avoiding assistance and browsing the page in the usual manner when support is not needed.

There are many other solutions that provide Web accessibility not specifically oriented to older people. For example, Garrido et al. [28] propose improving Web accessibility in client browsers through interface refactorings. This approach is called Client-Side Web Refactoring (CSWR), it allows to automatically create different, personalized views of the same application. The refactorings proposed are compliant with W3C guidelines.

Besides, there are tools that allow users to change the way Web content is presented. GreaseMonkey [29] is a

Firefox extension that allows writing scripts to alter visited Web pages. It can be used to make a Website more readable or more usable, Web applications can be modified by adding content and/or controls to them. For instance, Mirri et al. [30] describe GAPforAPE (GreaseMonkey And Profiling for Accessible Pages Enhancement), an augment browsing system based on GreaseMonkey, which allows Web users to set up their preferences at client side and thus modifying content on the browser interface.

A. Summary

There are many methods that have been developed to facilitate access to the Internet to older people. Some of them are specialized solutions, such as simplified browsers or email systems. However, this approach does not consider inclusiveness, one of the two desired conditions in ageing. The second kind of solution is personalization of appropriate features in Web pages. They can be easily selected if the user needs to use them. There are repositories with many of these personalizations available, but it is necessary to find the most suitable solution.

In our work, given most elders prefer to use standard Web applications, although they have difficulties when they use them, we chose this second approach to improve the application interface. As we could not find a complete built solution in the repositories, we developed some personalizations that make older users' experience easier. These personalizations instantiate solutions to the barriers faced by older people.

IV. E-MAIL INTERFACES FOR OLDER PEOPLE

Although elders meet various barriers to computer use, there is considerable evidence that a significant number of older people do use email. Given the importance of positive experiences for allowing continuing computer use, appropriately designed interfaces would be very helpful.

In [21], a research on email for older novices was performed. As result, researchers concluded that such an email system should have a reduced set of functionality presented in a non-cluttered way without menu systems, page-specific help and instructions for the user, and a one-click paradigm where a mouse-click always led to a new screen in order to reduce user confusion. Besides, the default system appearance should have larger text and higher contrast between foreground and background.

With these constraints in mind we evaluated the interfaces of three email systems: Outlook, Gmail and Yahoo. Through this analysis we could observe the following: Outlook interface is the most simple one, good contrast, large text and big buttons, only the necessary options on each screen. One problem detected is the banner of publicity that may distract or confuse the user. On the other hand Gmail interface is the most cumbersome. In the main screen there is a lot of information, different tabs, and buttons without labels. When a message is selected to read, the layout of the screen is very confusing. To write a new mail, a new small window opens, leaving the previous screen partially behind. The button options do not have labels, unless the mouse is over them. Last, Yahoo interface is rather simple. Main screen layout is very similar to that of Outlook, although it shows more options and tabs. In screens to write a new mail or answering, there are unlabeled buttons. Besides, it also shows a publicity banner. A problem detected for all of the three email systems is about closing user session. There is not a button option for this task, it is implemented as a menu option on user menu.

Since we teach Yahoo email to older adults who take courses at University, we decided to make our own experience with our elders and evaluate the usability of this email system.

V. EVALUATION OF OLDER USERS' EXPERIENCE IN PATAGONIA

Since 2009, the National University of Patagonia Austral and the National Institute of Social Services for Pensioners (PAMI) have signed an agreement [31] for teaching computing, music, and theatre courses to older people.

These courses are taught twice a week and last three months. Computing courses are the most crowded, having about 20 pupils each.

Older people who assist to computing courses have expressed that they come to learn computing because they want to keep in touch with their families, with their grandchildren who live in other country regions.

Here, in Patagonia, distances between cities or towns are extremely long; besides, we are 1242 miles away from the capital city, Buenos Aires. Moreover, the weather is a critical factor, too. Winters are very long and cold, and strong winds blow. As a result, older people spend most of their time inside their houses, and they often feel lonely. Thus, getting online can have positive benefits for them. Tools like Email, FaceBook and Skype can empower older adults to stay connected with their friends and family.

In this study the purpose is to find out, which are the accessibility failures that the email's Web interface has got and evaluate if a more accessible interface would allow older people to utilize it more frequently and without suffering frustration for not remembering how to use the application.

A. Experiment 1

During the second half of 2012, teachers taught email classes. At the beginning of 2013, when computing classes started again, teachers noticed that most pupils did not use this communication tool. When asked for the reason of not using it, most pupils said that they did not remember how to use it, a few said that they were not interested in sending or receiving mails, and the rest, only some of them, said that they still used it. So, the purpose of this experiment is to investigate what accessibility difficulties has got the email's Web interface design.

1) Participants:

Eighteen older adults ranging in age from 64 to 73 years old (eleven women and seven men) were recruited for this activity. All of them took computing courses between April and June of 2013 and also during the second half of 2012.

2) Materials:

For this experiment, we used Yahoo mail application (Figure 1), which was also used during email classes.

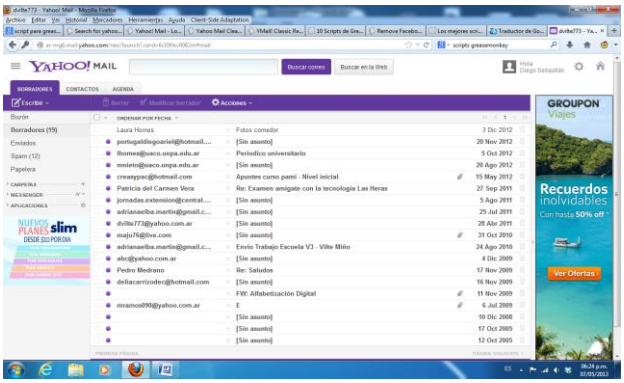


Figure 1. Yahoo mail inbox.

It is important to highlight that the courses are taught in a 25 desktops Lab equipped with 15 LCD monitors of 19-inch and 10 LCD monitors of 17-inch, whose resolutions are WXGA 1366 x 768 and XGA 1024x768, respectively. Although changing terminals settings (font sizes and colors) is posible, the Lab is used intensively every day to adopt this practice as usual.

3) Procedure:

Usability testing with the think-aloud method was conducted [32]. The evaluations were pair-based because older people feel more relaxed and confident about their work. Each evaluation was recorded, in order to analyze participants' behavior and comments.

4) Tasks:

Five tasks were proposed to explore the interface usability:

a) Read an email
b) Reply an email
c) Write a new email
d) Delete an email
e) Close user session

5) Results:

Of the 9 couples of participants, all could finish Tasks a) and c), 6 could not complete Task b), 2 could not conclude Task d) and 8 could not end Task e). These results are detailed in Table III.

From these results, we have found three problems throughout Tasks a)-e):

a) Problem 1: Advertisements

All participants complained about being distracted or even confused with the advertisements that appeared on the right side of the screen. They were afraid of clicking by error on these ads and causing an unexpected behavior of the email application, like closing, or losing the work being done.

b) Problem 2: Visual presentation difficulties

Besides, participants experienced other difficulties involving visual presentation of pages. Three couples of participants in Tasks a) and b) could not differentiate selected emails, because of light color contrast. Three couples of participants in Task a), three in Task b), and five in Task c) had difficulties in visualizing text because of font size, style, and inter-letter spacing. Also, 6 couples of participants in Task d) and 9 in Task e) made a great effort to distinguish available commands in menu bar.

TABLE III. RESULTS ACHIEVED BY OLDER USERS IN EMAIL USAGE *EXPERIMENT 1*

Task	Couples Error Ratio
a) Read an email	0/9
b) Reply an email	6/9
c) Write a new email	0/9
d) Delete an email	2/9
e) Close user session	8/9

c) Problem 3: Not understandable buttons

Participants also had trouble identifying buttons that represented email actions like "Reply" or "Forward". Eight couples of participants had difficulties identifying the button to conclude Task b), and 6 couples could not complete the task because of this problem. All participants had difficulties in Task e), remembering how to leave the application or "Sign Out", and only one couple could complete this task.

All the difficulties suffered by older users, are age-related issues like cognitive and visual impairment. Another factor involved is the lack of knowledge of technology and Web applications. Evaluating the WCAG 2.0 guidelines, we found that all these problems are considered within WCAG guidelines as we demonstrated before in Table II. Problems 1 and 3 correspond to difficulty number 6 detailed in Table II, which involves short term memory problems, concentration difficulties, distraction, and Problem 2 involves visual accessibility barriers shown as difficulties 1, 2, and 3 in Table II.

Hence, Yahoo email application is not compliant with this standard. However, this application provides solution to some of them, by setting appropriate configurations. But this is a very complex task to be performed by older users.

B. Experiment 2

The purpose of Experiment 2 is to evaluate an improvement to the email Website interface, which we developed to solve the problems found in Experiment 1.

In this improved interface, vertical banner ads have been deleted, and labels have been added for "Reply" and "Forward" buttons. Also, a button was added at the top of the form to allow users closing their sessions.

Figure 2 shows the modified interface of Yahoo mail inbox, including both adaptations: for Problem 1 vertical ads banner removement and for Problem 3 a button ("2" in Figure 2) labeled "Cerrar Sesión" to close user session, and the two labels "Responder" y "Reenviar" ("1" in Figure 2) for replying and forwarding, respectively.

1) Participants:

Fourteen older adults ranging in age from 66 to 74 years old (eight women and six men) were recruited for this activity. All of them took computing courses during the first half of 2012, and now they are taking theatre but not computing classes. However, they were willing to participate in this experiment.

Figure 2. Yahoo mail inbox after interface improvement.

2) Materials:

We modified Yahoo interface by applying two adaptations [33]. One of them is a script for deleting vertical ad banners that we downloaded from a scripts repository and the other one is a script developed for us in JavaScript to solve problems with buttons.

a) Problem 1: Advertisements

Although this vertical banner ad can be removed, this was not a permanent solution and became an annoyance to older pupils. In order to give solution to this problem, we chose GreaseMonkey. There are many add-ins that provide a number of features for visual and navigational enhancements to Web pages, which may fill usability gaps for older users.

Figure 2 shows the modified interface of Yahoo mail inbox where the vertical banner has been deleted. This modification was achieved by the installation of a GreaseMonkey script, CleanUp 1.1 that we downloaded from the scripts repository [34].

b) Problem 2: Visual presentation difficulties

Here, there are solutions provided by the browser and also by the operating system. The browser (Mozilla Firefox) allows modifying default settings for font size and style, and the operating system (Windows 7) provides an Accessibility Center that allows improving visual presentation, mouse setting and color contrast.

c) Problem 3: Not understandable buttons

At this point, we did not find any GreaseMonkey script, which solves difficulties with buttons' understanding or 'Sign Out' explicit inclusion in the application interface. So, we developed a script named "Oldie 1.0" that added labels to "Reply" and "Forward" buttons and a button to allow users closing their sessions.

3) Procedure and Tasks:

The same as for Experiment 1, detailed in Sections V.A.3) and V.A.4), respectively.

4) Results:

Of the 7 couples of participants, all could finish Tasks a), c) and e), 1 could not complete Task b), and 1 could not complete Task d). These results are detailed in Table IV. In this experiment, Problems 1, 2 and 3 detected previously have been eliminated. A couple of participants could not finish tasks b) and d) because they did not remember how to perform those tasks.

Therefore, we observe that this improved interface contributed to increasing task completion rates. But, for further analysis of the findings we performed statistical analysis of the number of participants that completed each task on each Yahoo interface.

TABLE IV. RESULTS ACHIEVED BY OLDER USERS IN EMAIL USAGE *EXPERIMENT 2*

Task Id	Task Description	Couples Error Ratio
a)	Read an email	0/7
b)	Reply an email	1/7
c)	Write a new email	0/7
d)	Delete an email	1/7
e)	Close user session	0/7

A significant Shapiro-Wilk test of normality indicated that the data were not normally distributed, in addition as the two samples are independent of each other, a Mann-Whitney U test was used. It demonstrated that there was not a significant difference in the median number of tasks completed on the original Yahoo mail application (Mdn = 7) and the modified Yahoo interface (Mdn = 7); the U-value is 11.5, the critical value of U at $p \leq 0.05$ is 4. Therefore, the result is not significant at $p \leq 0.05$.

Consequently, we decided to make a new experiment with a greater number of couples participating, in order to obtain more accurate results.

TABLE V. NUMBER OF PARTICIPANTS ACHIEVING TASKS

Task Id	Yahoo	Modified Yahoo
a)	9	7
b)	3	6
c)	9	7
d)	7	6
e)	1	7

C. Experiment 3

The purpose of Experiment 3 is to perform a new evaluation of partipants' accomplishments, as they successfully complete pre-defined tasks on both Yahoo mail application versions, the original one and the modified one. Besides, we are interested in comparing the usability provided by each interface.

1) Participants:

Fifty older adults ranging in age from 56 to 82 years old (thirty five women and fifteen men) were recruited for this activity. All of them participated in some UPAMI course previously, but they did not have the same computing level. Some of them had participated in several computing courses and others only once. Nevertheless, the idea of our experiment was to evaluate the usability in both versions of the application. If participants were not familiar with the original interface, then the new one would help them to accomplish tasks. In relation with the aging impairments, 42 participants suffered vision difficulties, 3 were affected with cognitive problems (short term memory) and 2 had a motor

disability (trembling). They were grouped in couples to perform the tasks.

2) Materials:

We used the two Yahoo interfaces that were already used in Experiment 1 and Experiment 2. Each Yahoo version run in a separate computer. Every couple of participants had two computers, one running the original Yahoo mail application and the other running the modified Yahoo mail application.

In addition, participants got a Likert questionnaire [35] asking what version they preferred completing each task on. This 5 point Likert question ranged from "Original" (original Yahoo) to "Modified" (Modified Yahoo) with a neutral point of "No Opinion". This resulted in five of these answers for each couple of participants (one for each task). The Likert questionnaire is detailed in Figure 3.

Besides, participants received two System Usability Scale (SUS) questionnaires [36], one per Yahoo interface version, which provided a 0-100 score relating to the usability of the corresponding Yahoo version (Figure 4).

3) Procedure and Tasks:

Each couple of participants was asked to attempt 5 tasks on both Yahoo versions:

a) Read an email
b) Reply an email
c) Write a new email
d) Delete an email
e) Close user session

These tasks were the same that the ones assigned in previous experiments. Participants attempted each task on one of the two Yahoo versions before moving on to the second Yahoo version.

After each task was performed, we noted whether a participant successfully completed the task, allowing completion rates to be investigated.

Then, after each task was completed on both systems, participants answered a Likert questionnaire on paper about what version they preferred completing the task on. This resulted in five of these answers for each couple of participants (one for each task). The Likert questionnaire is detailed in Figure 3.

After all tasks were attempted, participants completed two System Usability Scale (SUS) questionnaires [36], one by each Yahoo version, which provided a 0-100 score relating to the usability of the corresponding Yahoo version (Figure 4).

4) Results:

a) Task Completion Rates

In Table VI, we can observe the improvement of participants' performance when using the modified Yahoo mail application.

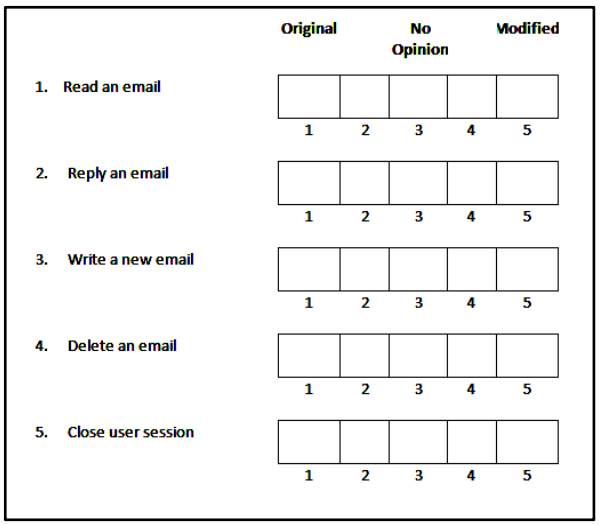


Figure 3. Likert questionnaire for expressing Yahoo version preference.

In Table VI, we can observe the improvement of participants' performance when using the modified Yahoo mail application.

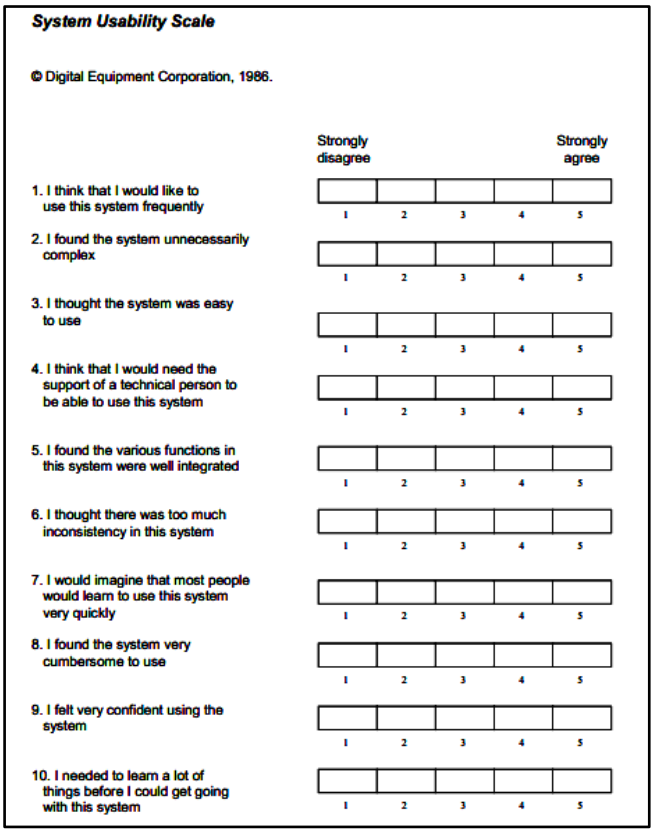


Figure 4. System Usability Scale (SUS) questionnaire.

A significant Shapiro-Wilk test of normality confirmed that the assumptions of a paired t-test were not met. Besides, as this experiment samples are paired, the Wilcoxon signed-rank test was used, which shows that a statistically significant difference exists between the median number of tasks completed on the original Yahoo application (Mdn=16) and the Modified Yahoo interface (Mdn=24). The W-value is 0. The critical value of W for N = 5 at $p \leq 0.05$ is 0. Therefore, the result is significant at $p \leq 0.05$.

TABLE VI. NUMBER OF PARTICIPANTS ACHIEVING TASKS IN EXPERIMENT 3. TOTAL NUMBER OF PARTICIPANTS: 25

Task Id	Yahoo	Modified Yahoo
a)	20	24
b)	12	18
c)	20	24
d)	16	22
e)	11	25

b) Preference Per Task

In Figure 5, we show the results of the Likert questionnaire, which illustrates participants' preference for Yahoo versions in each task.

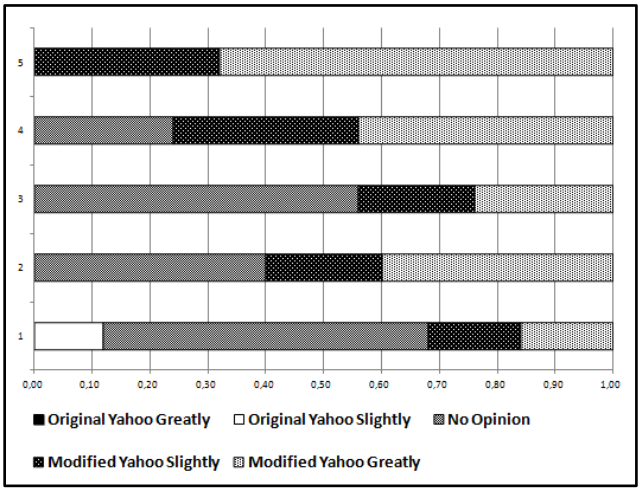


Figure 5. Bar chart showing Yahoo version preference responses for each task.

As we can see, no participant chose Yahoo in its original version for any task performed. The most preferred options were "Modified Yahoo" (Slightly or Greatly) and "No Opinion".

In Figure 6, a summary of Yahoo version preference responses is shown for all tasks performed.

The high frequency of "No Opinion" being selected may be partially explained by considering that some tasks were not directly affected by the modified Yahoo interface. This resulted in some tasks having the same steps to completion on both versions of the interface.

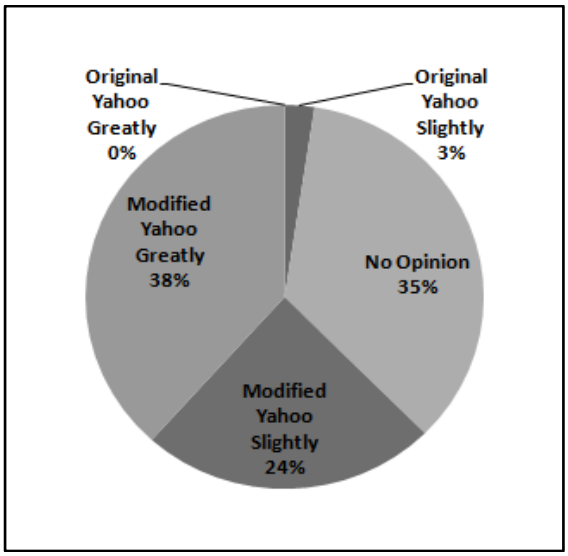


Figure 6. Pie chart summarizing Yahoo version preference responses.

For further analysis of the Likert-based preference questionnaire, we categorized answers into 3 categories: (1) Original Yahoo is preferred ("greatly"or "slightly"), (2) Modified Yahoo is preferred ("greatly"or "slightly") or (3) "No Opinion" is selected. We grouped the Likert points together to compare the three overall perspectives.

A Friedman test was used after the assumptions of a one-way repeated ANOVA were not met. The results of the Friedman test show that there was a statistically significant difference between the number of instances where participants preferred either the Original Yahoo interface (Mdn = 0), the Modified Yahoo interface (Mdn = 15) or "No Opinion" (Mdn = 10); $\chi^2(2. 5) = 6.3$, $p = 0.0429$. We can therefore conclude with considerable confidence that the observed differences among the mean rankings for the three categories of preferences reflect something more than mere random variability, they reflect a clear predilection for the Modified Yahoo interface.

c) System Usability Scale Scores

In this Experiment, we generated two SUS scores (one for each Yahoo version) per participant.

SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items (questions) are not meaningful on their own.

To calculate the SUS score, firstly the score contributions from each item must be added. Each item's score contribution will range from 0 to 4. For items 1,3,5,7, and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position.

Then, the sum of the scores must be multiplied by 2.5 to obtain the overall value of system usability. SUS scores have a range of 0 to 100.

In Table VII, we show the corresponding scores obtained from each of the 25 participants in each Yahoo version. It can be appreciated that the Modified interface has got better usability scores than the original one.

TABLE VII. CALCULATED SUS SCORES PER PARTICIPANT AND YAHOO VERSION

Participant	Original Yahoo SUS Scores	Modified Yahoo SUS Scores
1	45	75
2	62.5	75
3	60	100
4	50	70
5	62.5	90
6	67.5	82.5
7	67.5	82.5
8	47.5	87.5
9	60	85
10	67.5	82.5
11	67.5	82.5
12	45	75
13	62.5	75
14	62.5	75
15	60	85
16	75	75
17	55	87.5
18	55	87.5
19	42.5	90
20	50	75
21	50	75
22	55	87.5
23	67.5	82.5
24	63	90
25	68	82.5

Using a Wilcoxon signed-rank test, we can show that there was a significant difference in the SUS scores for the Original Yahoo interface (Mdn = 60) and the Modified Yahoo interface (Mdn = 82.5); Z = -4.2857, p = 0. The result is significant at $p \leq 0.05$.

Table VIII shows the median response to each SUS question per interface version. These detailed individual questions of the two SUS scales allow for more specific comparisons into the different components of the SUS's usability model over just the score alone. For example, participants strongly agreed that they would like to use the Modified Yahoo interface frequently, but had no opinion about the same question regarding the Original Yahoo interface.

VI. DISCUSSION

Many of the difficulties suffered by older Web users are already solved. However, as older people do not recognize their disabilities, they miss the opportunity to use the Web in a more comfortable way.

There are many accessibility tools provided by the operating systems and also by the Web browsers. But as they are classified as 'Accessibility Tools', most users believe that they are targeted to help people with severe disabilities that do not include the elderly.

Besides, there are some useful accessibility tools developed and available in Web repositories.

We have worked with some email accessibility requirements detected while teaching computing courses to older adults. Experiment 1 allowed for gaining a significant experience to develop our ideas, while Experiment 2 applied for testing these ideas on the field. Afterwards, in Experiment 3 we performed a more complete execution of both Experiments 1 and 2, including additional activities to allow usability comparison and evaluate interface preference between participants.

We found that some of the detected requirements could be solved by modifying the Web browser or the operating system configuration. Other requirements were accomplished by installing some scripts that provide the desired accessibility adaptations, like the scripts (CleanUp 1.1 and Oldie 1.0) we proposed and developed to solve Problems 1 and 3, respectively.

However, all these solutions require assistance from a computing specialist, or at least, from someone with the required skills, who must configure or install the appropriate add-ins.

Thus, we are working on a pragmatic research approach and applying an iterative incremental process to develop a tool that includes all the accessibility adaptations and allows older people select the appropriate configuration by themselves. Besides, this tool must be able to provide help to older users, who are not familiar with application concepts and hence avoiding hesitation and frustration. This will contribute to increasing quality of life of our Patagonian older Web users.

VII. CONCLUSION

Older adults represent the fastest growing portion of the world's population. Most older adults have got some declines that affect computer use, as difficulties with vision, hearing, mobility or cognition.

The World Wide Web Consortium (W3C) has got some initiatives like Web Accessibility Initiative (WAI) and Web Accessibility Initiative: Ageing Education and Harmonization (WAI-AGE), which provide solutions to many of the problems of older people. However, many Web designers do not consider WAI recommendations when designing Websites.

So, there are some approaches focused on improving Websites' accessibility. Some of them consist on Web adaptations that provide solution to a varying amount of accessibility issues.

TABLE VIII. TABLE SHOWING THE MEDIAN RESPONSES TO THE INDIVIDUAL LIKERT QUESTIONS IN THE SYSTEM USABILITY SCALE FOR EACH YAHOO INTERFACE VERSION.

SUS Question	Original Yahoo Median	Modified Yahoo Median
1. I think that I would like to use this system frequently	No Opinion 3	Strongly Agree 5
2. I found the system unnecessarily complex	No Opinion 3	Strongly Disagree 1
3. I thought the system was easy to use	Strongly Agree 5	Strongly Agree 5
4. I think that I would need the support of a technical person to be able to use the system	Disagree 2	Strongly Disagree 1
5. I found the various functions in this system were well integrated	No Opinion 3	Agree 4
6. I thought there was too much inconsistency in this system	Disagree 2	Disagree 2
7. I would imagine that most people would learn to use this system very quickly	No Opinion 3	Agree 4
8. I found the system very cumbersome to use	No Opinion 3	Disagree 2
9. I felt very confident using the system	Agree 4	Strongly Agree 5
10. I needed to learn a lot of things before I could get going with this system	No Opinion 3	Disagree 2

In this article, we showed different solutions provided to solve distinct older pupils' requirements. However, from our experience, we must highlight two issues about these solutions: (i) they do not cover all needs and, (ii) they are not usable enough for elderly citizens. Due to these reasons, new solutions should be developed and these solutions must prevent older people having to get help from someone else who can configure or install suitable accessibility settings to grant our seniors one of their main wishes: "independence".

As regards social requirements of our older students, our next goal is exploring difficulties experienced by them with social networks and finding appropriate solutions. This is a high priority requirement of our older citizens since our distant geographical situation and extreme weather conditions deprive them of enjoying many current activities that older people in other geographies can perform.

ACKNOWLEDGMENT

This work is supported by the UNPA project 29/B144 (Diseño y Evaluación de Portales Web) and project 29/B167 (Identificación, Desarrollo y Uso de Soluciones Web Centradas en el Usuario).

REFERENCES

[1] V. Saldaño, A. Martin, G. Gaetán, and D. Vilte, "Web Accessibility for Older Users: A Southern Argentinean View," The Eighth International Conference On Software Engineering Advances (ICSEA 2013), October 27-November 1, Venice, Italy. ISBN: 978-1-61208-304-9.

[2] V. Hanson, "Age and web access: the next generation," Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibililty (W4A '09), ACM, USA, April 2009, pp. 7-15.

[3] A. Arch, "Web accessibility for older users: a literature review," W3C Working Draft, World Wide Web Consortium (W3C), May 2008, accessed 1st December 2014. http://www.w3.org/TR/wai-age-literature/

[4] S. Sayago and J. Blat, "About the relevance of accessibility barriers in the everyday interactions of older people with the Web," Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibililty (W4A '09), ACM, USA, April 2009, pp. 104-113.

[5] S. Pedell, F. Vetere, L. Kulik, E. Ozanne, and A. Gruner, "Social isolation of older people: the role of domestic technologies," Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OZCHI '10), ACM, New York, NY, USA, November 2010, pp. 164-167.

[6] K. Zickuhr and M. Madden, "Older adults and internet use," Pew Research Center's Internet & American Life Project, 2012.

[7] A. Dickinson, M. Smith, J. Arnott, A. Newell, and R. Hill, "Approaches to web search and navigation for older computer novices," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07), ACM, New York, NY, USA, May 2007, pp. 281-290.

[8] N. Selwyn, "The information aged: A qualitative study of older adults' use of information and communications technology," Journal of Aging Studies, 18(4), November 2004, pp. 369-384.

[9] E. Hartnett, S. Minocha, J. Palmer, M. Petre, S. Evans, C. Middup, K. Dunn, B. Murphy, T. Heap, and D. Roberts, "Older people and online social interactions: an empirical investigation," The UKAIS International Conference on Information Systems (UKAIS), Worcester College, University of Oxford, March 2013.

[10] L. Gibson, P. Forbes, and V. Hanson, "What can the ash cloud tell us about older adults's technology adoption?," Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '10), ACM, New York, NY, USA, October 2010, pp. 301-302.

[11] L. Gibson, W. Moncur, P. Forbes, J. Arnott, C. Martin, and A. Bachu, "Designing social networking sites for older adults," Proceedings of the 24th BCS Interaction Specialist Group Conference (BCS '10), British Computer Society, Swinton, UK, September 2010, pp. 186-194.

[12] S. Sayago, L. Camacho, and J. Blat, "Evaluation of techniques defined in WCAG 2.0 with older people," Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibililty (W4A '09), ACM, New York, NY, USA, April 2009, pp. 79-82.

[13] A. Martin, V. Saldaño, G. Miranda, and G. Gaetan, "AO-WAD: A Generalized Approach for Accessible Design within the Development of Web-based Systems," Proceedings of The 7th International Conference on Software Engineering Advances, ICSEA 2012, IARIA, Portugal, November 2012, pp. 581-587.

[14] A. Martín, G. Rossi, A. Cechich, and S. Gordillo, "Engineering Accessible Web Applications. An Aspect-Oriented Approach," World Wide Web Journal, 13(4), 2010, pp. 419-440.

[15] United Nations Department of Economic and Social Affairs Population Division, "World population ageing, 1950-2050," UN, 2002, accessed 1st December 2014.
http://www.un.org/esa/population/publications/worldageing19502050/index.htm

[16] Web Accessibility Initiative, accessed 1st December 2014.
http://www.w3.org/WAI/

[17] Web Content Accessibility Guidelines (WCAG) 2.0, accessed 1st December 2014. http://www.w3.org/TR/WCAG/

[18] Web Accessibility Initiative: Ageing Education and Harmonisation (WAI-AGE), accessed 1st December 2014.
http://www.w3.org/WAI/WAI-AGE/

[19] A. Arch, "Web accessibility for older users: successes and opportunities," (keynote) Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibililty (W4A '09), ACM, USA, April 2009, pp. 1-6.

[20] A. Aula and M. Kaiki, "Less is more in Web search interfaces for older adults," First Monday [Online], Volume 10 Number 7, 2005.

[21] A. Dickinson, A. F. Newell, M. J. Smith, and R. L. Hill, "Introducing the Internet to the over-60s: Developing an email system for older novice computer users," Interacting with Computers, 17(6) (2005), 621-642.

[22] D. Hawthorn, "Possible implications of aging for interface designers," Interacting with Computer, 12, (2002), 507-528.

[23] Eldermail, accessed 1st December 2014.
http://www.eldermail.net/

[24] D. Sloan, M. Atkinson, C. Machin, and Y. Li, "The potential of adaptive interfaces as an accessibility aid for older Web users," Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A '10), ACM, New York, NY, USA, April 2010, Article 35.

[25] J. Richards and V. Hanson, "Web accessibility: a broader view," Proceedings of the 13th international conference on World Wide Web (WWW'04), ACM, USA, May 2004, pp. 72-79.

[26] H. Muta, T. Ohko, and H. Yoshinaga, "An activeX-based accessibility solution for senior citizens," Proceedings of the Center On Disabilities Technology And Persons With Disabilities Conference 2005, accessed 1st December 2014. http://www.csun.edu/cod/conf/2005/proceedings/2227.htm.

[27] D. Lunn and S. Harper, "Improving the accessibility of dynamic web content for older users," Proceedings of the 2011 International Cross-Disciplinary Conference on Web Accessibility (W4A '11), ACM, New York, NY, USA, March 2011, Article 16.

[28] A. Garrido, S. Firmenich, G. Rossi, J. Grigera, N. Medina-Medina, and I. Harari, "Personalized Web accessibility using client-side refactoring," Internet Computing, IEEE, 17(4), July-August 2013, pp. 58-66.

[29] Greasemonkey, accessed 1st December 2014.
http://www.greasespot.net/

[30] S. Mirri, P. Salomoni, and C. Prandi, "Augment browsing and standard profiling for enhancing Web accessibility," Proceedings of the 2011 International Cross-Disciplinary Conference on Web Accessibililty (W4A '11), ACM, New York, NY, USA, March 2011, Article 5.

[31] http://www.unpa.edu.ar/noticia/la-unpa-y-el-inssjp-firmaron-convenios-por-capacitacion-y-continuidad-del-programa-upami, accessed 1st December 2014.

[32] J. Dumas and J. C. Redish, "A Practical Guide to Usability Testing," (1st ed.), Intellect Books, Exeter, UK, 1999.

[33] P. Brusilovsky, A. Kobsa, and W. Nejdl, "The Adative Web: Methods and Strategies of Web Personalization," Springer, 2007.

[34] http://userscripts-mirror.org/, accessed 1st December 2014.

[35] W. Trochim, "Likert Scaling," Research Methods Knowledge Base, 2nd Edition, 2006, accessed 1st December 2014.
http://www.socialresearchmethods.net/kb/scallik.php

[36] J. Brooke, "SUS: A quick and dirty usability scale," In Usability evaluation in industry. Taylor and Francis, London, 1996.

Acceptance and Value of Mobile Payment Service Designs in Complex Ecosystems

Elisabeth Pergler
CAMPUS 02 University of Applied Sciences, Information Technologies & Business Informatics,
Graz, Austria
e-mail:
elisabeth.pergler@campus02.at

Daniela Glatz,
Christian Adelsberger
evolaris next level GmbH,
Graz, Austria
e-mail: daniela.glatz@evolaris.net;
christian.adelsberger@evolaris.net

Rainer Schamberger
PSA Payment Services Austria GmbH
Vienna, Austria
e-mail: rainer.schamberger@psa.at

***Abstract*—This paper presents the results of an acceptance analysis of existing mobile payment services (MPS) and MPS concepts. The analysis was conducted by means of technical documentation on features and functionalities, usage tests and interviews with experts from the MPS ecosystem. The results indicate high acceptance of wallet MPS that support additional functionalities such as loyalty card inclusion. In addition, card-based MPS obtain high values for ease of use, and thus, might serve as transitional solution until technical standards are implemented in the ecosystem. Subsequent to a short introduction and presentation of the state of the art, the development of the evaluation framework of this study will be presented, on which the analysis of the MPS at hand is based. The results of the analysis are further transferred into the context of value and a research model is elaborated. The paper concludes with the design of a field study that will evaluate the acceptance of the suggested MPS and the corresponding value-based research model in a real-life context.**

Keywords-mobile payment service; technology acceptance; external factors; complex ecosystem

I. INTRODUCTION

This paper is based on our contribution to ICDS 2014 - The Eighth International Conference on Digital Society [1]. Recent market research indicates a growing importance of mobile payment. At the beginning of 2013, Gartner [2] predicted that the value of the mobile payment transactions will increase by 44 percent in 2013 compared to 2012, to an estimated $235.4 billion worldwide by the end of the year.

In some regions, such as Japan and the US, mobile payment is already part of people's everyday lives. The development on the European markets, on the other hand, is still behind prior expectations. There exists high insecurity among many potential stakeholders within the complex ecosystem of mobile payment. The insecurity refers to technology standards as well as service designs and business models.

What makes the ecosystem of different MPS "complex" is the fact that different pre-conditions and circumstances are relevant for each solution. Some examples include relevant partners in the value creation/delivery and supply chain process, a variety of contract forms, agreements, legal aspects, and responsibilities. For the user of a single MPS, this complex ecosystem means that they might not be able to use the chosen MPS for the payment at a certain retailer, because the involved parties and companies are not in a contractual relationship that is necessary for a successful transaction at the point of sale.

Acceptance of mobile payment is an issue that has been addressed in various empirical studies. These resulted in interesting causal models of mobile payment acceptance with high explanatory power, e.g., [3], [4], and [5]. Acceptance by itself, defined as the decision to adopt or not adopt a MPS, is not sufficient to predict the market success of a particular payment service as their success is highly dependent on the ecosystem, in which they operate and their actual design. Thus, there is a need to connect theoretical foundations from acceptance research to practical design issues of actual mobile payment services and to context factors that arise from the complex ecosystem, in which mobile payment services operate.

The main objective of the present research project is therefore a systematic analysis of generic mobile payment services (MPS) within a novel acceptance evaluation framework that is derived from validated causal models of mobile payment acceptance. In a first step, it is necessary to develop the evaluation framework based on an extensive literature review. Mobile payment services are classified based on a market analysis and representative services are selected for each generic service. These are then analysed within the acceptance evaluation framework. Data for the analysis is obtained from service features and mobile payment service usage tests and expert interviews with service providers, banking and payment experts. The comparison of acceptance factors for each service results in a systematic assessment and enables conclusions regarding acceptance of the analysed mobile payment services.

The results are then enriched by the concept of Mobile Value Core (MVC) [6], which defines the realized value of mobile services and products from two different foci, the provider-oriented focus and the user-oriented focus. In a next step, the different concepts of value in the context of MPS are transferred into the scheme provided by MVC. In the following section a preliminary value-based research model is outlined in accordance with MVC.

The paper concludes with an outlook on a subsequent field study. The research design of the field study is based on the major findings of the presented project and is necessary

to evaluate the results of this research project in a real world context.

The paper is structured as follows: Section II presents the state of the art. In Section III, we describe how the evaluation framework was developed, which is the basis for the analysis of different payment services (Section IV). Section V elaborates the concept of value in the context of mobile payment services. The draft research model is outlined in Section VI. The paper concludes in Section VII, with an outlook to our future work and study design.

II. STATE OF THE ART IN MOBILE PAYMENT RESEARCH

Analysis of the state of the art will start with Section A, in which the technological implementations of mobile payment services will be presented, followed by Section B, which will provide an introduction to the acceptance factors of mobile payment services.

A. Technological implementations of mobile payment services

Mobile payment services can be classified according to technological designs and features that influence the payment process. The following classification is based on [7] and [8]:

- carrier medium,
- payment method,
- technology,
- type of payment system,
- payment process, and
- storage of sensitive customer data.

Mobile payment services are differentiated according to the carrier medium that is used. In this study, smart phones and Near Field Communication (NFC)-cards are considered as media types. The second criterion is payment method. Possible types are debit as well as credit cards, pre-paid mechanisms and direct debit processes. Debit card payment either initiates account debiting immediately after the transaction at the point of sale or a couple of days later. Credit card payment does not initiate immediate debiting of the account, but enables a loan without interest for the rest of the month. The amounts of several transactions are accumulated and account debiting takes place at the end of each month. Pre-paid payment requires money to be deposited on a card or smart phone in advance. The payment method is accepted at the point of sale as long as the account balance is positive.

An important issue is the technology that is used to communicate with the payment terminal at the point of sale. Common technologies are NFC, 2D-codes and bar codes. Payment systems operate either in form of so called open-loop systems or closed-loop systems. In an open-loop system a third party is involved that handles transactions on behalf of several banks. This party can be independent or can be the result of the collaboration of several banks. Closed-loop systems involve one single bank that processes the transactions whereas open-loop systems involve several banks in the transaction process. Payment processes are either offline or online. Online payment processes require input of a PIN by the user at the terminal. This is necessary for identification of the card holder. The card is checked online and the transaction will be completed only after successful verification. Offline payment, on the other hand, does not include verification of the available payment limit at the bank in charge of the account. There is no identification and card verification at the point of sale and communication takes place only between smart phone or card and the terminal. There exist five main types of sensitive customer data storage. The construction-wise inclusion of the secure element embedded in the smart phone is one technical option. A major disadvantage of this type is the connection of the secure element and its data to a particular phone that cannot be transferred to another device. Another option is usage of micro-SD cards that are equipped with a secure element. These can be put in the micro-SD slot of the smart phone and transferred in case of device changes. The secure element can also be stored on the SIM card. As these are bound to a certain mobile network operator this might hamper changes of the mobile network operator. The fourth option is storage of sensitive customer data on a card that is equipped with an NFC chip. The fifth option is a centralized, server-based solution that does not necessarily need to store sensitive customer data in the mobile payment system.

B. Acceptance factors of mobile payment services

Many acceptance research studies of mobile payment acceptance are based on technology acceptance model (TAM) [9], and thus, incorporate perceived usefulness (PU) and perceived ease of use (PEOU) as main factors influencing behavioral intention (BI) to use, e.g., [10], [11], [12], etc. A comparative study in different cultural settings [10] included technology readiness as a personality trait in the original TAM. Results of this study indicate a significant positive effect of technology readiness on PEOU and PU as well as BI in most cultural settings. Individual mobility as a personal requirement regarding technology characteristics and perceived security resulted in positive effects on BI or attitude towards mobile payment in [5]. Personal innovativeness is another personality factor that has been tested with significant positive effects within the TAM framework [12]. This study also included technology characteristics such as convenience and reachability that showed positive effects on either PEOU or PU.

Security is one of the most often tested technology characteristics. In most cases, it is operationalized as a perception of security [5]. It has also been empirically tested in the particular setting of mobile payment acceptance in tourism [13]. In some studies, security issues are regarded as aspects of perceived risk and operationalized within this construct [11] and [14].

Trust is a construct that obtained particular interest within mobile payment acceptance research. Trust has been tested as an antecedent of PEOU and PU [4] and it has been found that it is affected by characteristics of the mobile technology itself and characteristics of the service provider, such as

reputation. An examination of trust within the valence framework indicated highly dynamic effects of trust in internet payment and initial trust in mobile payment on negative valences (perceived cost and risk) and positive valence (relative advantage) that is affecting BI [3].

Other studies are based on unified theory of acceptance and use of technology (UTAUT) [15] and include social influence and other constructs in addition to PU and PEOU to explain BI. In [16], UTAUT was extended by the mobile payment specific factors trust and perceived security. Both factors resulted in significant effects on intention to use mobile wallets in the research model.

Contextual issues have been included in various studies in different forms. We apply the multidisciplinary context model from [17] in order to classify these constructs and variables in a systematic way.

- Social context refers to people around the subject, their relationships to the subject, and interactions with the subject. Social context includes, for example, subjective norm [5], reference group evaluation [14], friends' evaluation [14], etc.
- Task context considers the particular objective of the present usage situation. It is interpreted [18] as a breadth of mobile payment use situations [14] or circumstances in use situations.
- Physical context includes all objects that are surrounding the subject and their current status and direction. Examples for the inclusion are the construct individual mobility [5] and compatibility [19], [3].
- Temporal context is what gives the current usage situation a meaning like, e.g., past mobile payment use [14].

III. DEVELOPMENT OF EVALUATION FRAMEWORK

In this section, the evaluation framework will be described. First, an overview to the Evaluation Process (Section A) will be provided, followed by the selection of acceptance factors in Section B. These factors will be operationalized in Section C.

A. Evaluation process

The evaluation is illustrated in Figure 1 (Sequence Diagram Evaluation Process). It shows that the process was based on the identification of relevant acceptance factors through literature review. These factors were operationalized and applied to all selected MPS by the means of usage tests and expert interviews. MPS were selected based on a thorough desk research, in which all information and data available were collected. Further and deeper information was gathered through usage tests and expert interviews. As a result, for each MPS and each of the relevant acceptance factors, a classification was suggested, whether the potential of acceptance of the MPS at hand is to be considered high, medium, or low. The evaluation process was carried out from February to August 2013.

This classification was based on a discussion process within the project team and double-checked by external MPS experts. Usability tests were not part of the analysis, as it can be assumed that this aspect will be covered in time before market launch of the MPS.

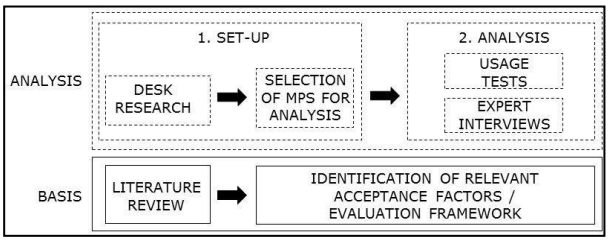


Figure 1. Sequence Diagram Evaluation Process

B. Selecting acceptance factors

Acceptance factors for the evaluation framework are derived from the literature review. PEOU and PU are the most widely used constructs to explain acceptance of mobile payment. Their concepts are provided in Table I.

TABLE I. PERCEIVED EASE OF USE / PERCEIVED USEFULNESS

Acceptance Factor	*Construct*	*Definition*
Perceived Ease of Use	PEOU [11], [13], [12], [10], [16]	The original definition from [9] "the extent to which using a new system is expected to be free of efforts"
	PEOU [5]	"Important aspects related to mobile payment services ease of use include, for example, clear symbols and function keys, few and simple payment process steps, graphic display, and help functions […]"
Perceived Usefulnes	PU [11], [13], [12], [10], [16]	The original definition from [9] "the degree to which a prospective adopter believes that by using a particular system would improve his or her job performance"
	Attitude [14]	"This construct can be taken to reflect an individual's attitude towards a MPS, ranging from a very positive to a very negative assessment of the system's utility."
	PU [5]	"[…] users are only willing to accept innovations if those innovations provide a unique advantage compared to existing solutions […]"
	Convenience [12]	"Convenience is nothing but a combination of time and place utilities, which are clearly principal characteristics of m-payment."

Trust, perceived risks and (perceived) security were also included in many studies. Table II lists the various tested concepts.

TABLE II. SECURITY-RELATED FACTORS

Construct	*Definition*
Perceived risk [3]	"[…] extent to which prospective users expect mobile payment services to be uncertain or risky."
Initial mobile payment trust [3]	"Trust is a subjective belief that a party will fulfill his or her obligations according to the expectations of the trusting party."
Perceived risk [11]	"[…] the expectation of losses related to purchase […]"
Perceived security [13]	"[…] a threat which creates circumstance, condition or event with the potential to cause economic hardship […]"
MPS risk [14]	"The MPS risk construct refers to the possible harmful consequences an individual expects from MPS use […]"
Consumer trust [4]	"[…] in the context of m-payments, the two dimensions of consumer trust are trust in mobile service provider and trust in technology facilitated by mobile service provider characteristics and mobile technology characteristics respectively."
Perceived environmental risk [4]	"[…] is the risk associated with the underlying technological infrastructure […]"
Perceived structural assurance [4]	"[…] the consumer's perception about the institutional environment […]
Perceived security [5]	"In the context of electronic services, security risk, conceptualized as the likelihood of privacy invasion, has been found to be a particularly critical concern […]"
Perceived security [16]	"[…] the degree to which a customer believes that using a particular mobile payment procedure will be secure."
Trust [16]	"[…] the belief that vendors will perform some activity in accordance with customers' expectations."

External factors, such as necessary hardware or software adaptations, are included in the analysis due to their influence on provider decisions whereas other factors, such as availability and provider characteristics, are excluded from this analysis as these are highly influenced by time and location of assessment, e.g., Google wallet is currently not available in Austria but might be in future. Personal character traits and social influence are also excluded for this analysis as they are strictly individual but will be included in a future field study. The same is true for the different concepts of value, which will be in the focus of the field study.

C. Operationalization of acceptance factors

In a next step, the four major constructs were operationalized in order to obtain measures for mobile payment service usage tests and issues for the expert interviews. These methods were necessary, as detailed desk research on the technical features and functionalities was only partly able to cover the complexity of the topic at hand and usage tests were only possible for existing MPS. Details and functionalities regarding conceptualized MPS were obtained from interviews.

The process of operationalization focused on mobile payment procedures and features of different services.

Ease of use is analyzed considering the steps a user needs to take before using the mobile payment service and the process of each transaction. Moreover, some additional processes are considered such as PIN changes and payment history or analysis features.

Usefulness is analyzed with regard to transaction speed, i.e., average time that is required per transaction, considering quicker transactions as more useful. Also, additional functionalities are examined, such as integration of loyalty cards or shop finder.

Security is analyzed considering storage of sensitive customer data and risks that occur in operation.

External factors that affect the ecosystem are considered in terms of required adaptations at the bank and point of sale in order to enable the MPS to operate.

IV. ANALYSIS OF MOBILE PAYMENT SYSTEMS

Ten different existing mobile payment services and feasible mobile payment concepts were included in the analysis. They cover different combinations of technical implementations and designs. As a limitation, it has to be stated that the selection of MPS is based on desk research, experimenting with existing solutions, and the project team's understanding of the most possible combinations of technology and designs, no study or literature exists in this regard to suggest a different mode of selection. The limitations are acceptable as this is only the first part of the analysis that will be followed by an extensive field study in the course of which actual users will test MPS and their acceptance will be analyzed:

1. NFC debit card in an open-loop system enabling online and offline payments (e.g., PSA Payment Services Austria GmbH with all Austrian banks)
2. NFC pre-paid card in a closed-loop system enabling offline payments (e.g., Quick by PayLife)
3. NFC credit card in an open-loop system enabling online and offline payments (e.g., Mastercard PayPass and Visa PayWave)
4. Debit/credit application for iPhones with additional NFC hardware in an open-loop system enabling online and offline payments (e.g., CardMobile) – exclusive iOS solution
5. Barcode debit application for smart phones in an open-loop system enabling online payments (e.g., Secure Payment Technologies GmbH - pilot test)
6. Account-based 2D-code application for smart phone in an open-loop system enabling online and offline payments (e.g., CellumPay)
7. NFC debit/credit wallet application for smart phone in an open-loop system enabling online payments (e.g., Google Wallet)
8. NFC credit wallet application for smart phone in an open-loop system enabling online payments (e.g., myWallet by Deutsche Telekom)
9. 2D-code debit/credit application for smart phone in a closed-loop system enabling online payments (e.g., Starbucks and Square)
10. NFC debit application for smart phone in an open-loop system enabling online and offline payments (concept only)

Table III provides an overview on relevant factors for assessing ease of use, taking into account aspects before usage, the process of transaction and additional aspects.

With regard to the required effort of users before usage and during each transaction, card-based MPS are most easy to use. Wallet applications are also easy to use and in most cases offer additional functionalities like in-application PIN changes that increase ease of use. Barcode-based MPS are least easy to use as they require additional activities in the course of each transaction process.

TABLE III. ANALYSIS OF EASE OF USE

MPS	*before usage*	*transactions*	*other aspects*
1	Existing card is replaced by NFC enabled card; no registration required	Amount appears on terminal display; card is put close to display; visual or audio signal; NFC chip information is read by terminal; card is removed; successful transaction indicated by visual or audio signal; random PIN requests	PIN is changed at the bank; no history or analysis available
2	Existing card is replaced by NFC enabled card; no registration required; top up money	Amount appears on terminal display; card is put close to display; visual or audio signal; NFC chip information is read by terminal; card is removed; successful transaction indicated by visual or audio signal; amount is debited immediately from prepaid account	No PIN; application for smart phone that reads NFC chip and provides transaction history and account balance
3	Existing card is replaced by NFC enabled card; no registration required; one contact payment required	Amount appears on terminal display; card is put close to display; visual or audio signal; NFC chip information is read by terminal; card is removed; successful transaction indicated by visual or audio signal; random PIN requests	No history or analysis available
4	Download iOS 5.0 or higher and application; additional hardware for iPhone; registration of card; top up money	Application is launched; smart phone is put close to display; visual or audio signal; amount is displayed; amounts from 20 Euro require individual passcode; transaction is confirmed	PIN can be changed via application; transaction history for 30 days and account balance, for iOS only
5	Online banking activation; application download; application and account activation via transaction number and activation number;	Application is launched; PIN authorization; payment code is provided; barcode on smart phone display is scanned at the terminal; transaction is verified online	PIN can be changed anytime
6	Application download; registration of application via text message; creation of mobile PIN; registr. credit card; activation of credit card	Phone number and payment ID are provided to cashier; cashier selects payment method; customer receives confirmation request; card is selected; PIN is entered; confirmation is sent as push notification; cashier receives confirmation	PIN can be changed via application; transaction history available;
7	Application download; account registration; activation of credit card; test transaction	Application is launched; smart phone is put close to display; payment information is transferred automatically; transaction is confirmed by customer	PIN can be changed via application; transaction history and payment analysis via Google account
8	Application download; replace existing SIM card by myWallet NFC SIM card; registration	Application is launched; login information is entered; customer selects card; smart phone is put close to display; transaction is initiated; amounts from 25 Euro require PIN	PIN can be changed anytime; transaction history available
9	Application download; card registration	Pay by square: Application is launched; card is selected and QR code appears; cashier scans QR code; invoice is sent via email Pay by face: application is launched; name and photo are assigned using GPS information; cashier confirms matching face and photo	PIN can be changed via application; transaction history and analysis available
10	no details available	no details available	no details available

Usefulness (see Table IV) ought to be highest for wallet solutions as they include additional functionalities. The same is true for code-based MPS, but there is no information available regarding transaction speed of these services. Card-based MPS are considered to be very fast considering transaction speed, and thus, increase user perceptions of usefulness but do not enable any additional functionalities.

TABLE IV. ANALYSIS OF USEFULNESS

MPS	*transaction speed*	*additional functionalities*
1	offline payment (up to 25 Euros) approximately 350 milliseconds; online payment takes longer as it requires a PIN	none
2	200 – 300 milliseconds at POS terminal; 500 milliseconds at ATM	none
3	approximately 1 second without PIN;	none
4	online payment approximately 1 second; offline payment less than 1 second	none
5	no details available	none
6	online approximately 4 to 7 seconds	loyalty card inclusion; sweepstakes; prepaid card handling; mobile ticketing; mobile commerce inclusion
7	depends on payment situation	personalization features; Google offers inclusion
8	no details available	individual daily transaction limits; loyalty card inclusion
9	no details available	shop finder; invoice via email
10	no details available	no details available

Security issues, which are analyzed in Table V, are rather balanced among MPS except for stored value technologies. These might cause actual loss of money for the customer. Storage of sensitive customer data can influence ease of use as mobile phone and mobile network operator respectively are not easy to change in case of embedded secure elements or SIM-based secure elements. Transaction limits increase security, but may also harm ease of use and, in some cases, even usefulness, e.g., when transactions are made impossible. A similar effect is caused by PIN requirements. They increase security of the MPS but decrease ease of use and transaction speed.

TABLE V. ANALYSIS OF SECURITY

MPS	*storage of sensitive data*	*countermeasures against risks in operation*
1	on NFC chip on the card	random PIN requests (after five transactions the latest) for low value transactions; PIN required for transactions from € 25s
2	on NFC chip on the card	stored value technology is a risk considering theft as money is stored on the card with no further authorization required
3	on NFC chip on the card	random PIN requests (after four transactions the latest) for low value transactions; PIN or signature required for transactions from 25 Euros
4	secure element on MicroSD	only service provider can access secure element; additional app login possible; stored value is limited to € 50
5	none	barcodes are valid only once and only for 4 minutes; limit of 10 transactions per day; limit of € 100 per day; limit of 4 payments per hour
6	data is split between smart phone and remote server	mobile PIN for each transaction; remote deactivation of application available
7	embedded secure element and Google Cloud	remote deactivation available; transaction limit of $1.000 per day for one device and $10.000 for more than one device
8	SIM-based secure element (or NFC sticker)	data encryption on NFC-SIM; card and smart phone can be locked; individual daily limits
9	not applicable	online deactivation of application available; pay by face: face authentication
10	SIM-based secure element	certificates to avoid fraud

Table VI provides an overview of the relevant external factors for MPS analysis. Considering the point of sale, most MPS require adaptations with regard to terminals and software. Some are based on cash desk software adaptations as well. The most intrusive MPS design (number 6) even requires a connection between the point of sale and the remote server of the MPS provider. Effects on participating banks are minor to those on participating retailers. Those that require adaptations of the bank-wise core system are less likely to succeed unless initiated by the bank.

TABLE VI. ANALYSIS OF EXTERNAL FACTORS

MPS	*bank*	*point of sale*
1	adaptations in backend system required	NFC terminals and software required; no changes with regard to business processes and interchange fee model
2	none	NFC terminals and software required; no changes with regard to business processes and interchange fee model
3	none	NFC terminals required; no changes with regard to software, business processes and interchange fee model
4	none	NFC terminals required; particular module for low value transactions required; no changes with regard to business processes and interchange fee model;
5	adaptation of core system required	particular barcode scanner required (smart phone display scan enabled); cash desk software required; no interchange fee
6	none	connection of point of sale system to backend system and remote server; QR code printer or display required; no interchange fee
7	none	NFC terminals required
8	none	NFC terminals required; no changes with regard to business processes and interchange fee model;
9	none	QR code reader required; display required; adaptation of network, terminal and software infrastructure; acceleration of business processes (order, payment); no interchange fee
10	mobile issuing infrastructure including mobile network operators and banks required	NFC terminals required; no changes with regard to business processes and interchange fee model

V. VALUE IN THE CONTEXT OF MOBILE PAYMENT SYSTEMS

In this section, the concepts of mobile value core and Expectation Confirmation Theory will be described in order to introduce the topic of value to the study at hand (Section A). In Section B, common concepts of "value" in m-payment and m-commerce research will be presented and an overview will be provided.

A. Mobile value core and "Expectation-Confirmation-Theory"

Value in terms of realized value is the central factor within the mobile value core concept [6]. The mobile value core (MVC) defines the realized value as the intersection between provider expectations/offer and user expectations/value. The Expectation Confirmation Theory (ECT) [20] builds the theoretical foundation, i.e., both, provider and user, have expectations regarding the value of a mobile product or service prior to its usage, which are either confirmed or disconfirmed.

ECT has already been successfully used to explain continued technology usage [21]. According to ECT confirmation results in satisfaction, and thus, continued usage. These relationships are being utilized to elaborate a provider-oriented focus as well as a user-oriented focus within the three layer model of the MVC concept. The three layers refer to different modes of user involvement that cause different expectations and also different experiences with the technological artifact, e.g., the third layer requires co-creation of the artifact in the actual later usage context to enable contextual expectations and will thus result in optimized expectation-experience fit. The provider-oriented focus is on market information (target groups, user segments, and technology penetration ratios), the artifact (functionalities and design), business model aspects (value proposition and pricing scheme) and additionalities (e.g., customer support and brand image). On the other hand, the user-oriented focus is on motivation (needs, current tasks, and alternative technologies), the artifact (expected functionalities and design perceptions), business model aspects (value for money perceptions) and additionalities (e.g., service quality, information quality, brand image).

MVC further emphasizes that the co-creation of mobile products and services will increase the realized value of these services and products. Furthermore, co-creation in real-life settings will moreover optimize the realized value. The following section puts emphasis on both, the analysis of the provider-oriented focus, and the expectations of users and their acceptance of MPS.

B. Value in m-payment & m-commerce research

Value is a somewhat neglected factor in empirical research on m-payment and m-commerce acceptance, although it is not only important in the context of user acceptance, but also in the bigger context of the eco-system. Sanayei and Ansari [22] included cost for the bank server and security as technology quality in an analytic hierarchy process. Both factors turned out to be of great importance within the context of technological mobile payment decisions [22].

The concept of value as means for user acceptance research will be analyzed based on a literature review as an essential perception in connection to the usage of an m-payment or m-commerce technology. This means that concepts of value that are not directly related to the use of a technology (like for example in [23] – value as the “value of the customer for the organization”) are not part of the review at hand, whereas concepts that consider the value of the use of a technology for a company’s business model would be included.

With regards to m-payment research, the concept of "value" so far has been included in a few studies, although its definition is varying on a rather broad spectrum, ranging from "value" as a clearly pre-defined construct within a research model, and "value" as an equivalent of specific factors, like for example in [5], where Schierz et al. developed a conceptual model for mobile payment, based on TAM. The model was tested in an online-survey with persons who use apps on their smartphones, and provided significant results regarding the influence of the independent constructs on attitude towards using - "the degree to which using a technology is positively or negatively valued by an individual." [5].

Finance-related risks, such as perceived costs, did not show significant effects on BI in an empirical study on acceptance of a card-based payment service [11]. A study on consumers’ willingness to pay for mobile payment services indicated that consumers are either not willing to pay any fee for using mobile payment, or the fee varies between different purchased goods [18].

Amoroso et al. [24] on the other hand use "perceived value" as a behavioral construct as part of their mobile-wallet research model (based on UTAUT) that was developed based on a Japanese m-payment solution. In their scientific work, perceived value "is defined as the trade-off between what customers receive, such as quality, benefits, and utilities, and what they sacrifice, such as price, opportunity cost, transaction cost, time, and efforts" [24].

Utilitarian definitions can also be found in m-commerce research, e.g., in [25]: Ko et al. also use the concept of “perceived value” in their research model that is based on TAM and was tested in an online-survey in Korea with users of mobile fashion shopping services. In this model, "perceived value" is the dependent variable, comparable to [24], "attitude towards using", both constructs are being defined similarly as "the trade-off between costs and benefits" [25].

Lu et al. [26] explain perceived value similarly as "the seller’s overall assessment of an auction website service utility based on perceptions of what is received (benefits) compared to what is given (costs)" [26]. In their study, a similar model, compared to [25] is used, based on TAM. The results of the survey of online auction sellers in Taiwan, which was conducted using a web-based questionnaire, show a significant influence of perceived network externalities on perceived costs and perceived benefits of the online auction service, both influencing the perceived value significantly, which furthermore is influencing the behavioral intention to use the auction service.

Kim et al. [27] conducted a similar study with users of mobile internet at a Singapore university, using a similar model (based on TAM), defining perceived value "as a consumer's overall perception of M-Internet based on the considerations of its benefits and sacrifices needed to acquire and/or use it" [27].

Other m-commerce studies also include the concept of value, but as independent variable, like for example in [28]. "Contextual perceived value" (CPV) is used as an independent construct within a TAM-based research model that is aiming to identify the effects of the contextual perceived value of marketing offers on the behavioral intention to use m-commerce. The survey was conducted with users and non-users of m-commerce, by carrying out personal on-site interviews in offices and a university in Seoul, and it showed a significant influence of CPV on the behavioral intention to use m-commerce. CPV is defined as "MC-specific additional benefits to understand consumer

acceptance of mobile commerce" [28]. The influence of CPV of marketing offers was higher for non-users of m-commerce.

Table VII below provides an overview of existing value concepts in mobile payment and mobile commerce acceptance literature from both foci of the MVC. The table includes information on the general perspective of value (provider or user focus) that is addressed in the study, but also on the actual value object (market information/ motivation, artifact, business model aspects, additionalities). Moreover, the different positions of value in the research model (dependent variable, independent variable, moderator variable, mediator variable) are considered.

TABLE VII. RESULTS ANALYSIS

reference	*general perspective*	*value object*	*position of value*
[5]	user	motivation	moderating variable "attitude towards use" on intention to use
[24]	user	motivation	moderating variable on "attitude towards using" and "behavioral intention to use"
[25]	user	motivation	moderating variable between independent variables and intention to use/behavioral intention to use/adoption intention
[26]	user	motivation, artifact, business model aspects	moderating variable between perceived costs/benefits and intention to pay
[28]	user	motivation, business model aspects, additionalities	independent variable, influencing behavioral intention to use m-commerce

Based on Table VII, two observations are rather obvious.

First of all, the dominant general perspective of the analyzed research is the perspective of the user ([5], [24], [26], [28]) and his/her perception of acceptance, based on a certain value that is associated with the technology at hand. Focusing on the user perspective means at the same time that adding "value" to a research model does not necessarily mean to include value for all actors of the ecosystem, but mainly for the user, which might lead to an important effect on the interpretation of study results from the providers' or decision makers' points of view.

The second issue, which is closely related to the first one (general perspective: user), is the value object "motivation" ([5], [24], [26], [28]) of the user to actually use the technology. This concept is mostly included into research models as a moderating factor between independent variables and the behavioral intention ([5], [24], [25], [26]). Empirical tests of "value" as a part of technology research models proof a very high explanatory power of this concept in order to understand technology acceptance of technological mobile payment and mobile commerce solutions ([5], [26], [28]).

A general conclusion based on the literature review is the observation of a bias with regard to value involvement. Most studies and research models treat value as a simple concept in a uniform way. Even though there is diversity among definitions of value and questionnaire items used for model tests, the concepts themselves are rather similar with regard to perspective and object of value.

VI. PRELIMINARY RESEARCH MODEL

The research model (see Figure 2) is based on ECT [20] and drafted in alignment with MVC [6]. The detailed value concepts are derived from the analysis in the previous sections.

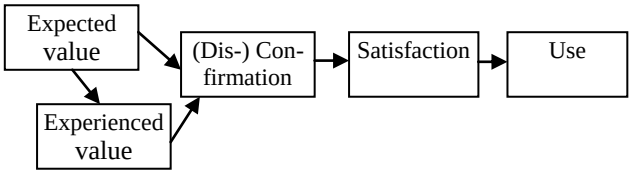


Figure 2. Preliminary research model based on ECT and MVC

According to ECT the main relationships included in the research model are:

- Expected value influences experienced value.
- The difference between expected value and experienced value defines confirmation or disconfirmation.
- Confirmation results in satisfaction, whereas disconfirmation harms satisfaction.
- The higher the satisfaction, the more likely is actual usage.

As the artifacts are (partly) co-created in a mobile Living Lab, there are some implications from MVC:

- Confirmed value should have positive effects on actual usage (in contrast to behavioral intention).
- Expectations and experiences of value should be close to equal regarding the parts that were co-created with users.

Further development of the model will be conducted based on the results of the presented research. Value will be deconstructed according to the findings from additional literature review and analysis. Preliminary result of the literature analysis indicates a preference of several different concepts of value, e.g., financial value, social value, and personal value instead of a single holistic value concept. The factors "expected value" and "experienced value" will be split in accordance with this differentiation. Thus, the final research model will consist of additional factors.

VII. CONCLUSION AND OUTLOOK ON FUTURE WORK (FIELD STUDY DESIGN)

Table VIII presents the results of the analysis that indicate high potential of acceptance for NFC-based wallet MPS (numbers 7 and 8) and NFC card-based MPS (numbers

1, 2 and 10). Face verification obtained optimistic results in the analysis, but requires very intrusive external adaptations, e.g., additional hardware at cash desk to enable photo-face comparison by staff and additional mobile phone recognition equipment within stores, and, moreover, does not support open-loop payment systems. Whereas high ease of use and high usefulness are positive indicators of overall acceptance, the effects of security on ease of use and usefulness can be either positive or negative.

TABLE VIII. RESULTS ANALYSIS

MPS		*ease of use*	*usefulness*	*security*	*security → EOU*	*security → U*
1		high	medium	medium	negative	negative
2		high	medium	low	positive	positive
3		high	medium	medium	negative	negative
4		medium	medium	low	none	none
5		medium	?	medium	negative	negative
6		low	medium	high	negative	none
7		high	high	high	none	none
8		medium	high	medium	negative	none
9	Pay by face	high	high	high	positive	none
	Pay by square	medium	high	medium	none	none
10		?	?	medium	negative	none

In the field study design, card-based solutions will be tested against wallet MPS according to the obtained analysis results, taking the complex eco-system of mobile payment solutions into consideration. Therefore, a central aim of the field study will be the identification of those factors that add specific value to mobile payment and how these factors could be implemented successfully. "Success" will not only be measured by the extent of technology acceptance, but also by the extent to which the solutions are suitable for different personalities, use situations, social constellations etc., hence, taking a variety of context factors into account. The research model will be tested by means of partial least squares analysis (PLS). In addition to testing the hypothesis of the value-based research model, the main research questions are:

- What kind of differences with regard to acceptance can be identified between card-based solutions and MPS?
- Are there acceptance differences between transaction types (debit vs. credit)?
- Differences could be stated regarding relative benefits, perception of value, perceived complexity, security, trustworthiness, and consequences of PIN requirements and the like.
- Which MPS is believed to be most successful (wisdom of crowds)?
- How is the concept of "wallet" perceived and rated and what are customers' associations and demands in this regard?
- Are there any influences/changes on daily routines expected? What kind of influences/changes are there? Are they the same for all MPS?

In order to tackle this huge variety of research questions and also taking the complex eco-system of MPS into account, the field study will consist of three parts, each applying different methods. In a field trial, 70 respondents will use two card-based solutions (debit, credit) and two mobile-phone-based MPS (debit, credit) over a period of two to three months complementing their common payment methods and provide feedback continually via standardized questionnaires, before, during, and after the survey period. In addition in situ feedback will be provided via mobile questionnaire after each purchase. In total, each participant will be using MPS between eight and ten times at least, using each solution at least once.

After the trial, a small number of participants will be invited to take part in a co-creation session in order to further optimize the identified most promising MPS and also in order to explore possible consequences for their daily lives.

Besides the users' point-of-view, the experiences and perspectives of the major stakeholders in the MPS eco-system providing the test-setting (financial institute, acquirers, issuers, and retailers) will be thoroughly analysed by means of expert interviews.

ACKNOWLEDGMENT

This research project is funded within COMET (Competence Centers for Excellent Technologies) program.

REFERENCES

[1] E. Pergler, A. Adelsberger, D. Glatz, R. Schamberger, "Acceptance of mobile payment service designs in complex ecosystems," Proceedings of The Eighth International Conference on Digital Society, March 23rd-27th 2014, Barcelona, Spain, pp. 135-142.

[2] S. Shen, "Forecast: Mobile Payment, Worldwide, 2013 Update," 2013.

[3] Y. Lu, S. Yang, P. Y. K. Chau, Y. Cao, "Dynamics between the trust transfer process and intention to use mobile payment services: a cross-environment perspective," Information & Management, vol. 48, 2011, pp. 393-403.

[4] S. Chandra, S. C. Sristava, and Y.-L. Theng, "Evaluating the role of trust in consumer adoption of mobile payment systems: an empirical analysis," Communications of the Association for Information Systems, vol. 27, 2010, Article 29, pp. 561-588.

[5] P. G. Schierz, O. Schilke, and B. W. Wirtz, "Understanding consumer acceptance of mobile payment services, an empirical analysis," Electronic Commerce Research and Applications, vol. 9, 2010, pp. 209-216.

[6] E. Pergler, A. Tarkus, "Reframing co-creation for mobile living labs: the mobile value core," Proceedings of the 26th Bled eConference, June 9-13 2013; Bled, Slovenia, pp. 314-32.

[7] C. Neger, "Types of cashles payment in Austria. Plastic money," (=Varianten des bargeldlosen Zahlungsverkehrs in Österreich. Plastikgeld), 2010, VDM, Saarbrücken.

[8] M. Verdier, "Retail payment systems: what can we learn from two-sided markets?," Communication & Strategies, vol. 61, 2006, pp. 37-56.

[9] F. D. Davis, "Perceived usefulness, perceived ease of use and user acceptance of information technology," MIS Quarterly, vol. 13, 1998, pp. 319-339.

[10] N. Guhr, T. Loi, R. Wiegard, and M. H. Breitner, "Technology readiness in customers' perception and acceptance of m(obile)-payment: an empirical study in Finnland, Germany, the USA and Japan," Proceedings of the 11th International Conference on Wirtschaftsinformatik, 27th February – 1st March 2013, Leipzig, Germany, pp. 119-133.

[11] G. W.-H. Tan, K.-B. Ooi, S.-C. Chong, and T.S. Hew, "NFC mobile credit card: the next frontier of mobile payment?," Telematics and Informatics, vol. 31, 2014, pp. 292-307.

[12] C. Kim, M. Mirusmonov, and I. Lee, "An empirical examination of factors influencing the intention to use mobile payment," Computers in Human Behavior, vol. 26, 2010, pp. 310-322.

[13] R. Peng, L. Xiong, and Z. Yang, "Exploring tourist adoption of tourism mobile payment: an empirical analysis," Journal of Theoretical and Applied Electronic Commerce Research, vol. 7, 2012, pp. 21-33.

[14] T. J. Gerpott and K. Kornmeier, "Determinants of customer acceptance of mobile payment systems," International Journal of Electronic Finance, vol. 3, 2009, pp. 1-30.

[15] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: toward a unified view," MIS Quarterly, vol. 27, 2003, pp. 425-478.

[16] D.-H. Shin, "Towards an understanding of the consumer acceptance of mobile wallet," Computers in Human Behavior, vol. 25, 2009, pp. 1343-1354.

[17] N. A. Bradley and M. D. Dunlop, "Toward a multidisciplinary model of context to support context-aware computing," Human-Computer Interaction, vol. 20, 2005, pp. 403-446.

[18] D. Viehland and R. S. Y. Leong, „Consumer willingness to use and pay for mobile payment services," International Journal of Prionciples and Applications of Information Science and Technology, vol. 3, 2010, pp. 35-46.

[19] C. Li and M. Zhang, "Research on the factors affecting consumers' willingness to the use of mobile payment," in Advances in D. Zheng (Ed.): Computer Science and Engineering, ASIC 141, pp. 575-580.

[20] R. L. Oliver, "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," JMR, Journal of Marketing Research, vol. 17, no. 3, 1980, pp. 460-469.

[21] A. Bhattacherjee, G. Prekumar, "Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test," MIS Quarterly, vol. 28, no. 2, 2004, pp. 229-254.

[22] A. Sanayei and A. Ansari, "Selection of the appropriate mobile payment technology in mobile banking, " International Journal of Information Science and Management, Special Issue 2, 2010, pp. 13-26.

[23] S. San Martín, B. López-Catalán, M.A. Ramón-Jerónimo, "Factors determining firms' perceived performance of mobile commerce," Industrial Management & Data Systems, vol. 112, no. 6, 2012, pp. 946-963.

[24] D. L. Amoroso and R. Magnier-Watanabe, "Building a research model for mobile wallet consumer adoption: the case of Mobile Suica in Japan," Journal of Theoretical and Applied Electronic Commerce Research, vol. 7, 2012, pp. 94-110.

[25] E. Ko, E.Y. Kim, E.K. Lee, "Modeling Consumer Adoption of Mobile Shopping for Fashion Products in Korea," Psychology & Marketing, vol. 26, no. 7, 2009, pp. 669-687.

[26] H.-P. Lu, K.-Y. Lin, "Factors influencing online auction sellers' intention to pay: An empirical study integrating network externalities with perceived value," Journal of electronic commerce research, vol. 13, no. 3, 2012, pp. 238-254

[27] H.-W. Kim, H.C. Chan, S. Gupta, "Value-based Adoption of Mobile Internet: An empirical investigation," Decision Support Systems, vol. 43, 2007, pp. 111-126.

[28] T. Lee, J. Jun, "The role of contextual marketing offer in Mobile Commerce acceptance: comparison between Mobile Commerce users and nonusers," Int. J. Mobile Communications, vol. 5, no. 3, 2007, pp. 339-356.

Dependency Parsing and Its Application using Hierarchical Structure in Japanese Language

Kazuki Ono

Division of Infrastructures Management
KLab Inc.
6-10-1 Roppingi, Minato, Tokyo 106–6122, Japan
Email: ono-ka@klab.com

Kenji Hatano

Faculty of Culture and Information Science
Doshisha University
1-3 Tatara-Miyakodani, Kyotanabe, Kyoto 610–0394, Japan
Email: khatano@mail.doshisha.ac.jp

***Abstract*—Conventional Japanese dependency parsing methods are primarily based on the bi-nominal model between phrases, which has a limitation related to the order of phrases. Accordingly, the length of a phrase, which depends on the language dependency, is limited. In this paper, we propose a novel dependency parsing method for Japanese based on an extended hierarchical language model simulated by the hierarchical Pitman-Yor process in order to overcome the abovementioned limitation. We also evaluate the accuracy of the proposed dependency parsing method as well as its practicality to detect the topics of each document. Experimental results show that the proposed method can parse dependencies in long, complex sentences and can allocate topics to each document relatively well compared with the conventional method. Consequently, it can be said that the proposed method is feasible in the research fields of both Japanese dependency parsing and topic modeling.**

Keywords-Dependency Parsing; Syntax Analysis; Syntax Tree Modeling; Topic Modeling.

I. INTRODUCTION

Asian languages are based on case grammar, and their sentences have the following word order: subject, object, and verb (SOV), leading to these languages being termed SOV languages. Japanese also has the same characteristics; however, its phrase order is relatively unfettered.

In Japanese, postpositions represent the syntactic function of each phrase, while in English, the syntactic function of each phrase is represented by the phrase order. Moreover, phrases that contain one or more postpositions following a noun, which play a similar role to the declension of nouns, have been devised and used for syntactically analyzing Japanese sentences. Hence, it may be said that a syntactic analysis is equivalent to parsing dependencies between phrases. This is because the semantic rules of Asian languages are explained as relationships of phrases. As a result, almost all dependency parsing methods in such languages extract these semantic rules and utilize them for analyzing relationships of phrases in Asian languages.

For instance, we propose a dependency parsing method for Japanese based on tree-substitution grammar (TSG) [1], which has originally attracted attention as a language model in the syntactic analysis used for English [2]. This is because TSG constructs a hierarchical structure defined as a set of rewrite rules of context-free grammar (CFG) and helps to establish the highest accuracy of English parsing by learning it [3], [4]. Alternatively, other dependency parsing methods utilize a bi-nominal dependency model to decide whether a dependency relation exists between pairs of phrases or not [5]–[7]. The reason why they did not employ a rule-based approach like us is that the computational cost of our approach is very expensive.

It is true that parsing dependencies based on TSG is computationally expensive; however, syntax trees generated by TSG describe the order of phrases, whereas those generated using the bi-nominal dependency model only represent relationships between phrases. In short, TSG-based syntax trees retain not only the relationships between phrases but also the orders of phrases; therefore, they facilitate a precise extraction of the dependencies between phrases. As a result, dependency parsing of Asian languages based on the bi-nominal dependency model is less successful. We believe that the accuracy of the dependency parsing of Asian languages can be improved by applying TSG to them.

In this paper, consequently, we propose a TSG-based dependency parsing method for Japanese. The proposed method is characterized by the handling of an exchangeable sequence of phrases in case grammar to utilize the extracted hierarchical structure using TSG. That is to say, we can generate a dependency parsing model by calculating the probabilities of integrating phrase dependencies from supervised training data with the abovementioned hierarchical structure. Thus, it can be said that the proposed mothod is a novel dependency parsing method, with hierarchical structure relations, for the Japanese because it can handle the relationships of each of the phrase dependencies.

Moreover, we discuss a possible beneficial effect of the proposed dependency parsing method: to estimate the topic of a document on the Web. Usually, the topic of a document is estimated using an n-gram model. In this approach, the topic model is constructed using the word distribution ignoring the order of words, and hence, it may not be accurate. Therefore,

our topic model is constructed using a topic distribution based on the modification relations in documents.

The remainder of this paper is organized as follows: In Section II, we explain the basic issues related to our study; in particular, we provide further details about the statistical language model and its application, latent Dirichlet allocation. In Section III, we discuss a conventional dependency parsing method for the Japanese language and its application to topic estimation. In Section IV, we describe the proposed dependency parsing method and its evaluation. In Section V, we describe the topic model built using the proposed dependency parsing method and its evaluation. Finally, in Section VI, we conclude our paper and present directions for future work.

II. Basic Issues

In this section, we will discuss the basic issues related to our study.

A. Statistical Language Model

The fundamental research purpose of linguistics is to find rules and characteristics through observations of an existing language [8]. For example, word segmentation of Asian languages is a crucial research topic, because these languages do not have any explicit boundary between words. However, most people can recognize the boundaries by using heuristics and can detect each word in a sentence. These heuristics are generalized into the rules and characteristics of the language, and are used as grammar rules.

On the other hand, there are concerns that such rules and characteristics should be statistically defined from an enormous document set. This is because the rules and characteristics found by linguists might not be objective outputs but subjective ones. Thus, some computational linguists have tried to statistically define certain rules and characteristics by using an enormous document set. In the case of the word segmentation described above, the probability of word segmentation can be calculated using the Dirichlet and the Pitman-Yor processes [9], [10]. Compared with the Dirichlet process, the Pitman-Yor process is more efficient in and effective for modeling the existing languages (i.e., n-gram language model); therefore, we propose a dependency parsing method based on the Pitman-Yor process.

The Pitman-Yor process is a non-parametric Bayesian model [11]. It is a stochastic process that generates an infinite discrete probability distribution G, which is denoted by $\mathrm{PY}(d, \theta, G_0)$. $\mathrm{PY}(d, \theta, G_0)$ has three parameters: d denotes a discount parameter with $0 \leq d \leq 1$, θ represents a strength parameter that meets the condition $\theta \geq -d$, and G_0 indicates a base distribution over a probability space. In the research field of natural language processing, the probability space is usually generated from the probabilities of phrase occurrences.

When d is zero, $\mathrm{PY}(d, \theta, G_0)$ becomes the Dirichlet process denoted by $\mathrm{DP}(\theta, G_0)$. In short, because $\mathrm{DP}(\theta, G_0)$ can generate an infinite-dimensional Dirichlet distribution, $\mathrm{PY}(d, \theta, G_0)$ can also support it. As outlined above, d denotes a concentration parameter for G_0; therefore, we can define the following equation:

$$G \sim \mathrm{PY}(d, \theta, G_0) \tag{1}$$

That is, $\mathrm{PY}(d, \theta, G_0)$ is an extended version of $\mathrm{DP}(\theta, G_0)$. Let W be a fixed and finite vocabulary of V phrases. The Pitman-Yor process can generate a vector of phrase probability $G(w)$ for each phrase $w \in W$. $\mathrm{DP}(\theta, G_0)$ is approximated by the Dirichlet distribution $D(\theta G_0(w_1), \ldots, \theta G_0(w_i), \ldots, \theta G_0(w_r))$ that expresses any division of the disintegration space, where the size of the phenomenon space is r for observing any phrase w_i, as follows:

$$\mathrm{DP}(\theta, G_0) \sim D(\theta G_0(w_1), \ldots, \theta G_0(w_i), \ldots, \theta G_0(w_r)) \tag{2}$$

Here, $G_0(w_i)$ denotes an integral of the base distribution G_0; therefore, it is equal to the sum of the probabilities of phrase occurrences in the disintegration space. In short, we can say that the Pitman-Yor process is a recursive stochastic process whose input is a set of phrases w_i and its base distribution is G_0.

In general, these procedures for generating an n-gram distribution drawn from G are often referred to as the Chinese restaurant process [12]. In this process, we fancifully imagine a restaurant with an infinite number of tables whose capacities are infinite. The existing distribution of customers (phrases) who come to the restaurant and the tables (discount strength) with a one-by-one dish (phrase vocabulary) is based on the n-gram model, denoted by G, and the hypothesis of the tables elicits the base distribution G_0. In the Pitman-Yor process, the customers are compared to phrases such that a customer continues to sit at the same table if the customer coming to the restaurant is the same. However, if it is another customer that had not previously entered the restaurant, the customer sits down at a new table. Given this scenario, we can express the Pitman-Yor process as follows:

$$\mathrm{PY}(d, \theta, G_0) = \frac{c_k - d}{\theta + c_\cdot} + \frac{\theta + dt}{\theta + c_\cdot} p(w_k) \tag{3}$$

where t denotes the number of customers under the base distribution G_0, c_k indicates the number of species, $c_\cdot$ represents the total number of customers, and $p(w_k)$ refers to the probability of a customer visiting the restaurant. Here, the parameters d and θ are non-parametric. Therefore, we have to generate the distribution approximated by the base distribution G_0 with these parameters using a Gibbs sampling algorithm, such as the Markov Chain Monte Carlo method [13].

Note that $p(w_k)$ in Equation (3) denotes the probability of word w_k in the document set in the case of computational linguistics. However, phrases are given various types of contexts so that $p(w_k)$ is not a uniform state in the document set. In short, we cannot directly apply the Pitman-Yor process to an n-gram model.

In fact, based on an n-gram language distribution over which the current phrase is given, various contexts can be formulated using an hierarchical Pitman-Yor process [14]. The hierarchical Pitman-Yor process is also a stochastic process that is based on a hierarchical extension of the Pitman-Yor process. Consequently, the n-gram distribution G_u can be generated by the Pitman-Yor process using the base distribution $G_{\pi(u)}$, which is generated in the given context $\mathbf{u}$ as follows:

$$G_{\mathbf{u}} \sim \mathrm{HPY}(d_{\pi|\mathbf{u}|}, \theta_{\pi|\mathbf{u}|}, G_{\pi(\mathbf{u})}) \tag{4}$$

where the strength and discount parameters are calculated on the basis of the length of context $\mathbf{u}$, $\pi(\mathbf{u})$ denotes the suffix of $\mathbf{u}$ consisting of all but the earliest phrase, and $G_{\pi(\mathbf{u})}$ represents a vector of probabilities of its context. When we recursively use a stochastic process such as $G_{\pi(\pi(\mathbf{u}))}$ over $G_{\pi(\mathbf{u})}$ using Equation (4), we can define the stochastic process that generates the n-gram distribution. This process is repeated until we get G_ϕ as the base distribution, the vector of probabilities over the current phrase given the empty context ϕ.

As described above, the structure of the stochastic process generated by the hierarchical Pitman-Yor process is expressed as a suffix tree with depth n. Each node in the suffix tree corresponds to a context consisting of up to $(n-1)$ phrases, and each child corresponds to the addition of a different word to the beginning of the context. The hierarchical Pitman-Yor process can generate an n-gram language model precisely as demonstrated in the language model based on Kneser-Ney smoothing [15]. Given the above explanation, we can express the hierarchical Pitman-Yor process as follows:

$$\begin{aligned}&\mathrm{HPY}(d_{\pi|\mathbf{u}|}, \theta_{\pi|\mathbf{u}|}, G_{\pi(\mathbf{u})}) \\ &\quad = \frac{c(w_k|\mathbf{u}) - d_{\pi|\mathbf{u}|}}{\theta_{\pi|\mathbf{u}|} + c_{\pi|\mathbf{u}|}} + \frac{\theta_{\pi|\mathbf{u}|} + d_{\pi|\mathbf{u}|} t_{\pi|\mathbf{u}|}}{\theta_{\pi|\mathbf{u}|} + c_{\pi|\mathbf{u}|}} p(w_k|\mathbf{u'}) \quad (5)\end{aligned}$$

where $t_{\pi|\mathbf{u}|}$ denotes the number of customers under the base distribution $G_{\pi(\mathbf{u})}$; $c(w_k|\mathbf{u})$, the number word w_k when the state has context $\mathbf{u}$; $c_{\pi|\mathbf{u}|}$, the total number of words given context $\mathbf{u}$; and $p(w_k|\mathbf{u'})$, the probability of generating a word when the state has context $\mathbf{u'}$.

B. Latent Dirichlet Allocation

Latent Dirichlet process (LDA) is one of the language models used in a one-document case, while the n-gram model is used in a one-sentence case. In LDA, we assume that the document consists of several topics; this is similar to a probabilistic latent semantic analysis (pLSA) [16]. Therefore, LDA is considered a pLSA model with a topic distribution by introducing variational Bayesian methods. Since its development, this model has attracted attention in the research field of topic models.

Both LDA and pLSA models create a word distribution by ignoring the order of words in a sentence and utilizing the idea of bag-of-words. In each topic model, the probabilistic models are built using a document set D containing a set of words W. However, their representation schemes are different.

In the earliest years, pLSA was called aspect model [17]. The aspect model is a latent variable model for co-occurrence data that associates an unobserved class variable $z \in Z = \{z_1, \ldots, z_K\}$ that accompanies a bag-of-words $w \in W = \{w_1, \ldots, w_M\}$ in a document $d \in D = \{d_1, \ldots, d_N\}$, where M denotes the number of bag-of-words, z represents a topic class, and K refers to the number of topics, because the topic model is based on the bag-of-words in pLSA. In other words, the generation process of the model is as follows:

- The document is selected on the basis of $P(d)$.
- The unobserved class variable z is also selected on the basis of $P(z|d)$.
- The word w is generated if we can calculate $P(w|z)$.

Consequently, we obtain two pairs of data (d, w), and the corresponding joint probability model can be expressed as follows:

$$\begin{aligned}P(d, w) &= P(d)P(w|d) \\ P(w|d) &= \sum_{z \in Z} P(w|z)P(z|d) \quad (6)\end{aligned}$$

This process is shown in Figure 1 using a graphical model, which is used for a graph-based representation. This is commonly utilized in fields such as probability theory and statistics.

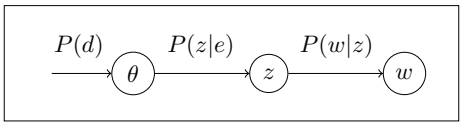


Figure 1. Graphical model representation of the pLSA model using asymmetric parameterization.

Here, we attempt to reverse the conditional probability $P(z|d)$ using Bayes' theorem and can define $P(z|d)$ as follows:

$$P(z|d) = \frac{P(z)P(d|z)}{P(d)} \quad (7)$$

Using Equations (6) and (7), we can obtain the model expressed by the following formula:

$$P(d, w) = \sum_{z \in Z} P(z)P(w|z)P(d|z) \quad (8)$$

Figure 2 shows a graphical model of Equation (8). This model shows that both d and w are independently conditioned and calculated using z. In Equation (8), $P(z)$, $P(w|z)$, and $P(z|d)$

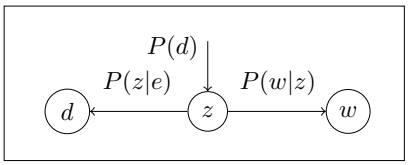


Figure 2. Graphical model representation of the pLSA model using symmetric parameterization.

can be calculated if the following log-likelihood function is maximized:

$$L = \sum_{d \in D} \sum_{w \in W} n(d, w) \log P(d, w) \quad (9)$$

where $n(d, w)$ denotes the number of bag-of-words w in the document d. In order to obtain an appropriate result of Equation (8), we have to maximize L in Equation (9). To do so, we use the expectation maximization (EM) algorithm [18]. Thus, we can build a topic model by using topic z corresponding to (d, w).

LDA also uses Z, which indicates a topic for each word in a document, as a latent variable. In LDA, z_{ij} denotes a latent topic of word w_{ij} in a document i. We can draw it as the following graphical model as Figure 3:

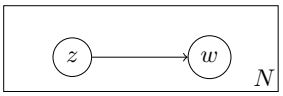


Figure 3. Initial LDA graphical model.

In addition, if the probability of generating word w_l in topic k is assumed to be ϕ_{kl}, the probability of generating w_{ij} in topic w_{ij} can be calculated; the corresponding graphical model is shown in Figure 4.

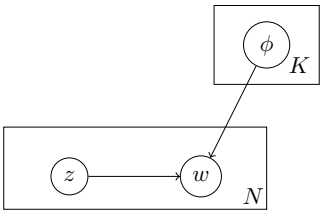


Figure 4. Adding parameter ϕ that denotes the generation probability of a word for a given topic.

Here, Z can be calculated using θ, because z_{ij} is approximated by θ_i, which is a topic distribution of document i. Therefore, we add this process to the graphical model described immediately before (see Figure 5).

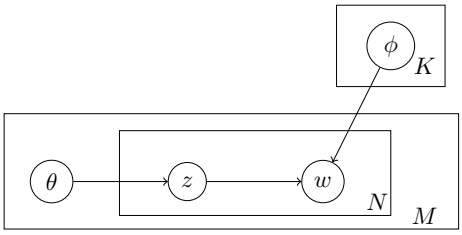


Figure 5. Adding parameter θ that denotes the topic distribution per documents.

Further, θ can also be approximated by using a Dirichlet distribution $Dir(\alpha)$ for a given parameter α. From the beginning, θ denotes the topic distribution of each document, expressing the word distribution composed of the bag-of-words; in short, θ satisfies $\{\theta_1, \theta_2, \ldots, \theta_K\} \in \theta$, where K denotes the number of topics. This means that α is composed of a set of α_k corresponding to θ_k, which is an element of θ, because θ_k denotes the topic k assigned to the documents. As is the case with θ, ϕ can be approximated by using the Dirichlet distribution $Dir(\beta)$, which is assigned a parameter β. That is, ϕ satisfies $\{\phi_1, \phi_2, \ldots, \phi_L\} \in \phi$, where L denotes the number of topics, and β consists of a set of β_l corresponding to ϕl, which is an element of ϕ because of the reason described above.

In light of the above explanation, LDA can be modeled using θ, which is a topic generation probability of each document, ϕ, which is a word generation probability of each topic, and Z, which is the topic of each word. The graphical model of LDA is shown in Figure 6.

In addition, α and β denote the parameters of the Dirichlet distributions θ and ϕ, and these can be estimated using the Markov Chain Monte Carlo method.

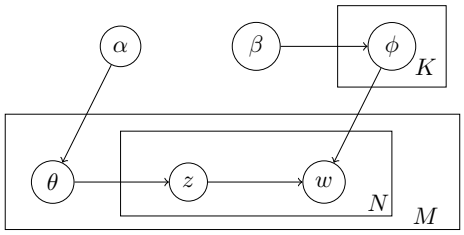


Figure 6. Graphical model of LDA.

III. Related Work

In this section, we first discuss a conventional dependency parsing method for Japanese and its limitations. The problem of the conventional methods can be solved by introducing the concept of TSG, which obtains highly precise English language parsing with a hierarchical structure of phrases. Therefore, we will now discuss a TSG-based dependency parsing method for the Japanese language.

We will also discuss an application of the proposed dependency parsing method and explain the importance of dependency parsing for natural language processing in Japanese.

A. Dependency Parsing for Japanese

As described in Section II-A, a Japanese sentence does not have an explicit boundary between words except punctuation marks. Therefore, in the first stage of computational linguistics for Japanese, single characters in a sentence and the relationships between two characters were focused upon. Currently, word segmentation in Japanese sentences is performed using an unsupervised approach [10].

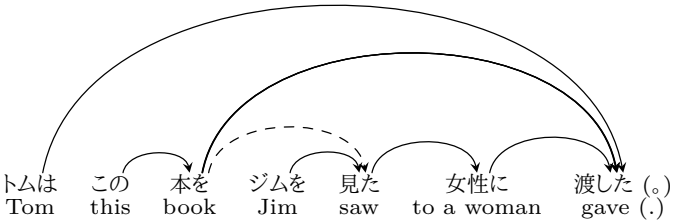
Dependency parsing in Japanese sentences also focuses upon relationships between two phrases in the beginning. This is called a bi-nominal approach, which generates a syntax tree for a sentence. This approach is a conventional method for parsing the dependencies between phrases in Japanese, and some methods using this approach have been developed thus far [5], [7]. These methods generate a syntax tree model to determine whether there is a dependency in Japanese between phrases in a sentence or not on the basis of features such as parts of speech and inflectional forms. They also conduct a syntactic analysis by using the generated syntax tree based on the following algorithm:

1) Check all the phrases in a sentence to ascertain whether they have a dependency with others located on the right side of the syntax tree.
2) Designate any phrase that has a dependency in Step 1 as the analysis result. At this time, the referrer of the dependency is excluded from the target of Step 1.
3) Repeat Steps 1 and 2 and keep the analysis results until all but the final phrase has a dependency.

The conventional methods described above consider only the relationship between two phrases in a sentence; however, do

not do otherwise. Therefore, they might misjudge some of the dependencies shown in Figure III-A.

The Japanese sentence "トムはこの本をジムを見た女性に渡した。" is shown in Figure 1. This sentence has a complex structure, and its English translation is "Tom gave this book to a woman who saw Jim." Originally, the phrase "本を (the book)" takes the dependency depicted by the solid line to another phrase "渡した (gave)," because "the book" is the object of "gave." However, sometimes the dependency from "本を (the book)" to "見た (saw)" as depicted by the dashed line in Figure III-A is extracted by the bi-nominal approaches. This is because the bi-nominal model cannot consider the hierarchical structure of the sentence. Thus, it is impossible to perform accurate parsing for a statement that has a complex structure.



(Tom gave this book to a woman who saw Jim.)

Figure 7. Example of parsing in Japanese.

Therefore, we propose a new method for parsing dependencies in Japanese on the basis of the hierarchical structure defined by the phrase relationships.

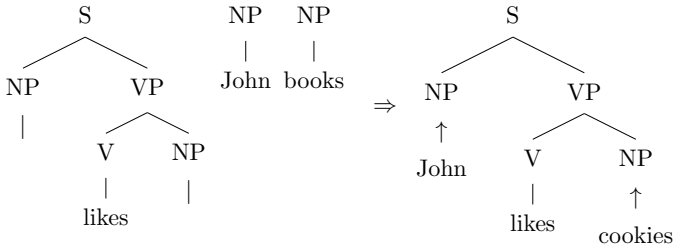
B. Tree-Substitution Grammar

Tree-substitution grammar (TSG) is an extension of context-free grammar (CFG) [3]. Both CFG and TSG have rewrite rules, called productions, which are used for constructing syntax trees by replacing nonterminal symbols with elementary trees, which are a part of the syntax tree.

The TSG production replaces a nonterminal symbol with an elementary tree whose depth is greater than one, while the CFG production does the same with an elementary tree whose depth is exactly one. As a result, the TSG production has a hierarchical structure based on a context. TSG is 4 tuple and is defined as $G = \{T, N, S, R\}$, where T denotes a set of terminal symbols; N, a set of nonterminal symbols; $S(\in N)$, a set of the distinguished root nonterminals; and R, a set of productions. In general, the syntax tree of an English sentence is described as a tree structure because of its phrase-structure rule.

Here, the root node is labeled with a nonterminal symbol, and its leaf nodes are labeled with either terminal symbols or nonterminal symbols. In short, the syntax tree of the input sentence is constructed by recursively replacing the nonterminal symbols with the elementary trees.

Figure 8 shows an example of parsing from the syntax tree, which is labeled on the left. For example, the S $\rightarrow$ (NP (VP (V like) NP)) production rewrites the nonterminal symbol S



Upper arrow "↑" indicate substitution sites in TSG.

Figure 8. A syntax tree consisting of three elementary trees.

with the fragment (S NP (VP (V likes) NP)). This syntax tree has two NPs as its nonterminal symbols, and the production rewrites these nonterminal symbols with the fragments (NP John) and (NP cookies).

In short, a derivation creates a tree by starting with the root symbol and rewriting (substituting) it with an elementary tree, then continuing to rewrite frontier nonterminals with elementary trees until there are no remaining nonterminals.

In TSG, the replacement of nonterminal symbols with elementary trees is performed in an arbitrary manner; however, we can calculate the probabilities of the rewrite rules in the syntax tree. Probabilistic tree-substitution grammar (PTSG) is an extension of TSG whose productions have their probabilities $P(e|c)$, where the elementary tree e replaces a nonterminal symbol c. $P(e|c)$ is statistically calculated using training data. The distribution of the PTSG production G_c can be generated by the hierarchical Pitman-Yor process described before using nonterminal symbol c insted of context $\mathbf{u}$ in Equation (4), so that G_c is defined as follows:

$$G_c \sim \mathrm{HPY}(d_c, \theta_c, G_{\pi(c)}) \tag{10}$$

where d_c and θ_c denote the hyperparameters of the hierarchical Pitman-Yor process when we use the nonterminal symbol c, and $G_{\pi(c)}$ represents a distribution over the infinite-dimensional distribution of the elementary tree with c.

To generate the elementary tree e, we now draw e_1 from G_ϕ, thereby obtaining an elementary tree with nonterminal symbol $c_1, \ldots, c_m$, and then, draw $e_2, \ldots, e_m$ in turn from the base measure $G_{\pi(c_1)}, \ldots, G_{\pi(c_m)}$. The above process is repeated until a full tree is generated.

However, a problem associated with the segmentation of elementary trees arises when PTSG is derived. Gibbs sampling, in which random variables are repeatedly sampled on the basis of the current values of all other random variables in the model, is usually used for solving this problem. Figure 9 shows an example of the segmentation of a syntax tree. In the procedures described above, we can perform dependency parsing based on the hierarchical structure. This method, state-of-the-art English dependency parsing [4], is known as symbol-refined TSG (SR-TSG). SR-TSG applies *symbol refinement* techniques [19], [20] to parse English sentences. These techniques can also be used in parsing methods that do not consider the hierarchical structure of a sentence, such as CFG. Thus, although not directly related to TSG, SR-TSG can be used for constructing symbol-refined syntax trees and is considered a state-of-the-art method.

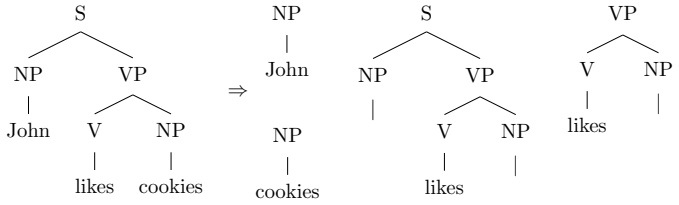


Figure 9. Elementary trees based on PTSG.

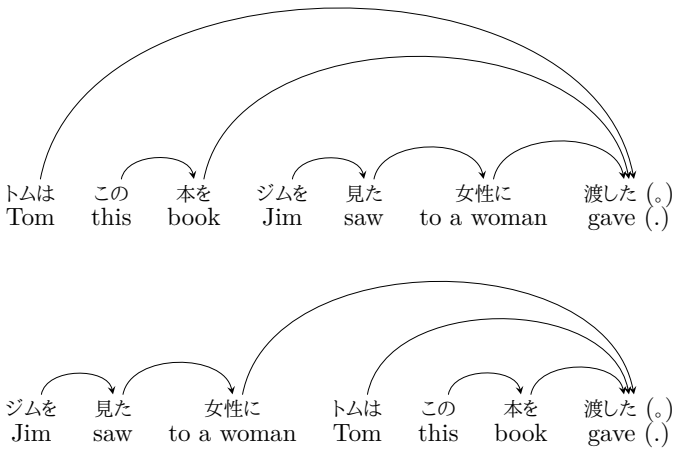


Figure 10. Two Japanese sentences that have the same structure.

However, TSG cannot support dependency parsing of Japanese sentences, because its syntax tree does not have a tree structure, which represents the syntactic function of each phrase, as required for the postpositions in the Japanese language. For example, two sentences are shown in Figure 10. The phrase order is different in these two sentences, while each phrase has the same dependencies. Therefore, it is impossible to conduct dependency parsing for Japanese sentences by using TSG, because TSG is only applicable to the language whose structure can be expressed as a syntax tree. As a result, we propose a novel dependency parsing method for the Japanese language that considers both the hierarchical structure and the order of phrases using an extension of TSG.

C. Topic Model

In general, a topic model called LDA or its derivations consider a bag-of-words to be an n-gram distribution in each document, and assign latent topics to each document if the generated topic distribution is the same as the observed n-gram distribution. However, the topics estimated by the generated topic model depend only on the word distribution. This is because such topics are characterized on the basis of the word distribution, which is constituted by the observed words, in each document. This is primarily caused by not considering grammatical factors such as the order of words and the dependency structures in a document. Therefore, a general topic model faces the problem of assigning incorrect topics to the documents.

For all of these reasons, modern topic models applicable to natural English language processing utilize the rules of grammar as their features, and some studies manage hierarchization of the syntactic structures and the syntax trees based on them [21], [22]. Similarly, the topic model applicable to natural Japanese language processing uses syntax rules as its features. For example, Kitajima et al. defined the features except an n-gram distribution of words as events that are mainly composed of phrase dependencies [23]. This topic model treats an event in units of not words but phrases to extract the events from each document. As a result, they estimate the topics of the documents on the basis of the distribution of the events.

Here, we would like to consider the *event* in [23]. We have already explained that this event is composed of a phrase dependency; in other words, it is an example of a phenomenon expressed by characters (e.g., What is happening? Who felt how?). From the standpoint of the research field of natural language processing, the event consists of pairs of two phrases with a dependency in each document. Consequently, they are expressed such as $(subject, predicate)$ or $(predicate_1, predicate_2)$. Because $(subject, predicate)$ denotes the movement or atmosphere of *subject* using *predicate*; moreover, $(predicate_1, predicate_2)$ describes the characteristic and atmosphere of *predicate*$_2$ using *predicate*$_1$. Therefore, we can detect events by using a dependency parsing technique if we take note of the pairs described above.

The part of speech of the subject is usually a noun or an unknown word, and that of the predicate is usually a verb, an adjective, or an adjectival noun, according to the idea presented in [23]. Therefore, a topic model based on the pairs of two phrases is constructed using the result of dependency parsing. Consequently, we can estimate the latent topics considering a grammatical factor by performing the topic estimation of every document.

IV. Hierarchical Dependency Parsing

As described in Section III, the dependency parsing of Japanese language should consider both the hierarchical structure and the order of phrases in a sentence. Existing works just consider the relationship between two phrases in a sentence, so that they might have the potential for extracting false dependencies. The hierarchical structure which is focused in this paper is called *weak context* in the research field of natural language processing. In this section, we explain our method for considering a weak context in a sentence by using an extension of TSG.

A. Syntax Tree Construction

In order to consider the dependencies between phrases, we generate a syntax tree based on the n-gram model made from the occurrence of dependencies. We also calculate the probabilities of the referrer phrases subject to the occurrence of the destruction phrase. In Figure 11, for example, the dependency between "トムは (Tom)" and "渡した (gave)" means that referrer phrase "トムは (Tom)" depends on the referenced phrase "渡した (gave)." If we can extract this dependency, we can calculate the occurrence probability of the dependency between phrase "トムは (Tom)" and "渡した (gave)" as the conditional probability P_D(トムは (Tom)| 渡した (gave)), which can be calculated using the prior probability P_D(渡した (gave)) in the bi-gram model.

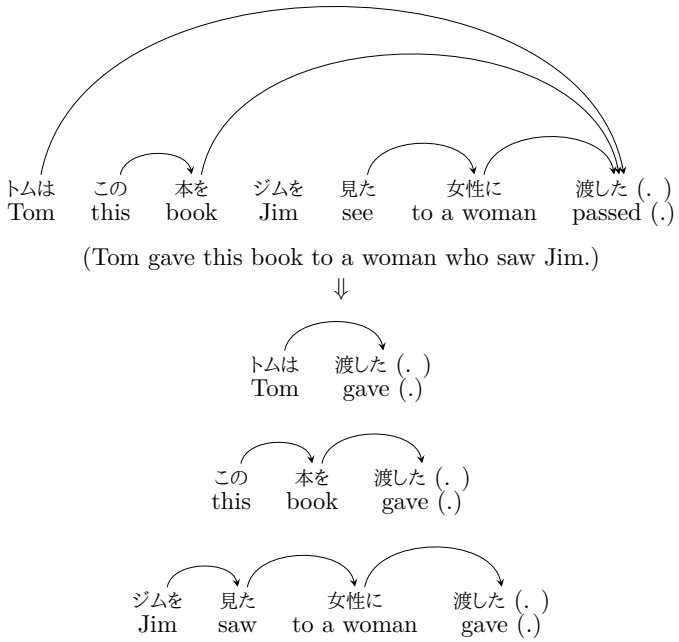


Figure 11. Hierarchical dependencies in the Japanese language.

At present, we cannot construct an n-gram model that reflects the occurrence of phrases if n is small. Conversely, the size of the n-gram model is large if n is large. In order to solve this problem, we use the variable-order Pitman-Yor language model [24], which is an extension of the hierarchical Pitman-Yor process [14]. This technique can help to treat n as a variable in the n-gram model. Using the variable-order Pitman-Yor language model, we can generate a syntax tree model considering the number of phrases. When we calculate the probability of the referrer phrase being the root node of the elementary tree by Equations (4) and (5) just like the hierarchical Pitman-Yor process.

For example, Figure 11 shows three elementary trees comprising the end of a sentence. In short, we obtain three elementary trees called the weak context and can calculate the probability $P_D(\cdot|$ 渡した (gave)) of the end of the sentence being the root node of the observed syntax tree.

Thus, we can precisely generate a syntax tree with a weak context dependency.

B. Hierarchical Dependency Parsing Algorithm

It is known that dependencies in the Japanese language have the following limitations:

- Japanese is a head-final language. Except for the rightmost one, thus, each segment modifies exactly one segment among the phrases appearing to its right.
- The dependencies do not cross one another.

Since the syntax trees are constructed using the method illustrated in Figure 12 (a), each tree whose root node is the phrase at the end of the sentence, is hierarchical; therefore, we parse the sentence in a bottom-up manner by using an algorithm called CYK [25], which is based on a depth-first search. The phrase at the end of the sentence gets the dependency from the phrase immediately before it; therefore, we regard the phrase at the end of the sentence and the phrase immediately before it as the weak context dependency and find another dependency in the sentence.

In Figure 12 (b), for example, the phrase "渡した (gave)" is the phrase at the end of the sentence; therefore, we obtain the dependency from "女性に (to a woman)." As a result, we can calculate the probability of the dependency P_D(女性に (to a woman)| 渡した (gave)). In the same manner, we can calculate different probabilities of dependencies P_D(見た (saw)| 女性に渡した (gave to a woman)) and P_D(見た (saw)| 渡した (gave)), if we assume that there is a dependency between "女性に (to a woman)" and "渡した (gave)". Presently, we compare P_D(見た (saw)| 女性に渡した (gave to a woman)) with P_D(見た (saw)| 渡した (gave)), and then, extract the dependency between the phrases. If P_D(見た (saw)| 女性に渡した (gave to a woman)) is larger than P_D(見た (saw)| 渡した (gave)), we decide that there is a dependency between "見た (saw)" and "女性に渡した (gave to a woman)." This process is repeated until we get to the phrase at the beginning of the sentence. In the end, all the phrases for which the relationships are adjudged to be dependencies form the syntax tree of the sentence.

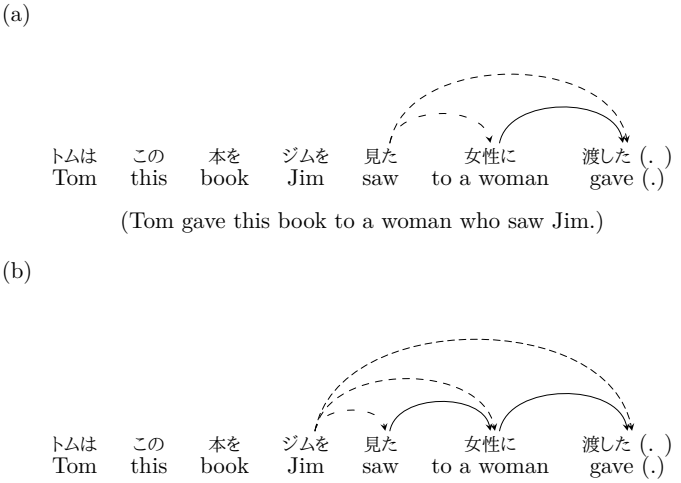


Figure 12. An example of dependency parsing processes using the hierarchical dependency model for Japanese.

C. Experimental Evaluation of Proposed Dependency Parsing

To evaluate the proposed method , which considers a weak context in the Japanese language, we compared it with a conventional method using the Kyoto University Text Corpus [26]. The Kyoto University Text Corpus is a Japanese text corpus for handling the dependencies of morphemes and segments in the Japanese language, and is commonly used for evaluating the effectiveness of the morphological analysis and dependency parsing of the Japanese language. It consists of newspaper articles published over a period of five days; however, we utilized only one day's articles (1,100 sentences) for the dependency parsing because it takes more time to generate syntax trees and we need to decrease the processing time.

We chose CaboCha [5] as the conventional method, because CaboCha is considered the most effective method for parsing dependencies in the Japanese language. The Kyoto

University Text Corpus was also used for the evaluation of CaboCha.

The following procedures are our experimental evaluation for obtaining dependencies in the Japanese language:

1) In order to create the syntax trees of sentences, we first extract the dependencies between phrases in each sentence.
2) We construct syntax trees focusing on the phrases at the end of the sentences assigned to the root node of each syntax tree.
3) We apply the hierarchical Pitman-Yor process and estimate parameter θ by using the Gibbs iteration.
4) We divide the sentences into segments after performing a morphological analysis of the Japanese language and check the segment sequences against our syntax trees.
5) We parse dependencies while continuing to extract the dependencies between phrases from the phrase at the end of an input sentence.

We also performed the same process for an experimental evaluation using CaboCha on the basis of a bi-nominal model [5]. To compare the proposed method with CaboCha, we utilized the following measurement denoted by R that is frequently used for evaluating the accuracy of the dependencies:

$$R = \frac{Y}{X} \tag{11}$$

where X denotes the number of dependencies in the test data, and Y represents the number of dependencies extracted by each parser (the proposed method and CaboCha). Of course, dependencies extracted by each parser are contained in the test data; therefore, R is usually termed recall in the field of information retrieval [27]. In order to calculate R by using each parser, we randomly selected 200 sentences from newspaper articles published on the other four days. In other words, we focused on the efficacy of the dependencies extracted by each method. Table I indicates that the method is, unfortunately, inferior to CaboCha from the viewpoint of effectiveness.

TABLE I. EXPERIMENTAL RESULTS 1

	CaboCha	Proposed method
R	0.871	0.769

This is because the proposed method utilizes syntax trees not only using a small quantity features for parsing dependencies in Japanese language. For which, CaboCha can parse dependencies in Japanese language with various features (word, lexical form of word, POS tag, and inflectional form), so that the accuracies of parsing dependencies in Japanese language are effective using CaboCha. Therefore, the precision of the proposed method can be improved further.

TABLE II. EXPERIMENTAL RESULTS 2

	Proposed method	CaboCha
Long, complex sentences	0.700	0.550

However, the proposed method has a competitive advantage for accurately extracting dependencies from long Japanese sentences whose syntax tree has a complex structure. In other words, we have to evaluate Japanese sentences from which CaboCha cannot extract dependencies. Therefore, we randomly selected 20 long sentences (10 percent of the test data used in the previous experiment) with relatively complex structures.

Table II shows that the proposed method could accurately extract dependencies from 14 of the 20 sentences, while CaboCha could do this only for 11 of the sentences. In short, the proposed method could parse dependencies of long, complex Japanese sentences better than CaboCha. This is attributed to the fact that it is possible for the parser to interpret the meaning of the postposition in some cases, even if the syntax tree model does not have a hierarchical structure. However, the syntax trees constructed using the proposed method contain not only the relationships between phrases but also the hierarchical structure of each phrase. Consequently, the proposed method can restrain a failed analysis of a complex Japanese sentence carried out using a conventional method.

V. APPLICATION OF PROPOSED PARSER

We evaluate not only the performance of proposed dependency parsing method but also its practicality to estimate topic modeling. Before we explain the contribution of our topic model generated using the proposed dependency parsing method, we describe the problem that the topic model generated using the conventional method has.

A. Problems of Event-based Topic Model

As described in Section III-C, the recent topic model is not built using word distribution. One such topic model proposed by Kitajima et al. is called an event-based topic model and is built using dependencies between phrases in each document [23]. In [23], the dependencies between phrases are extracted using a Japanese dependency parser called CaboCha [5], which is the most famous Japanese dependency parser. After extracting the dependencies between phrases, auxiliary verbs and postpositions are removed from the dependencies to extract the relationships between words. As a result, the event-based topic model is considered to be built using the dependencies between words.

However, in the case of the Japanese language, the auxiliary verbs and postpositions play a role in denoting case and tense to constitute grammar, whereas in other languages, they are not essential words. Therefore, we believe that both auxiliary verbs and postpositions should not be removed from the dependencies between phrases, and dependencies extracted by a Japanese dependency parser should be directly used for building an event-based topic model.

In the event-based topic model, we can extract the event related to the topic of the sentence, because we focus on its subject. However, we cannot accurately extract topics when we want to extract events from a sentence with a complex structure, because only $(subject, predicate)$ and $(predicate_1, predicate_2)$ are focused upon.

For example, if there is a Japanese noun phrase "海がきれいな景色の美しいホテル (the hotel with a magnificent view of the beautiful sea)," it is performed using a Japanese dependency parser, as shown in Figure 13. Therefore, from the noun phrase, we can extract four events, namely (海が, きれいな), (きれいな, ホテル), (景色の, 美しい), and (美しい, ホテル). These events are translated into English as follows:

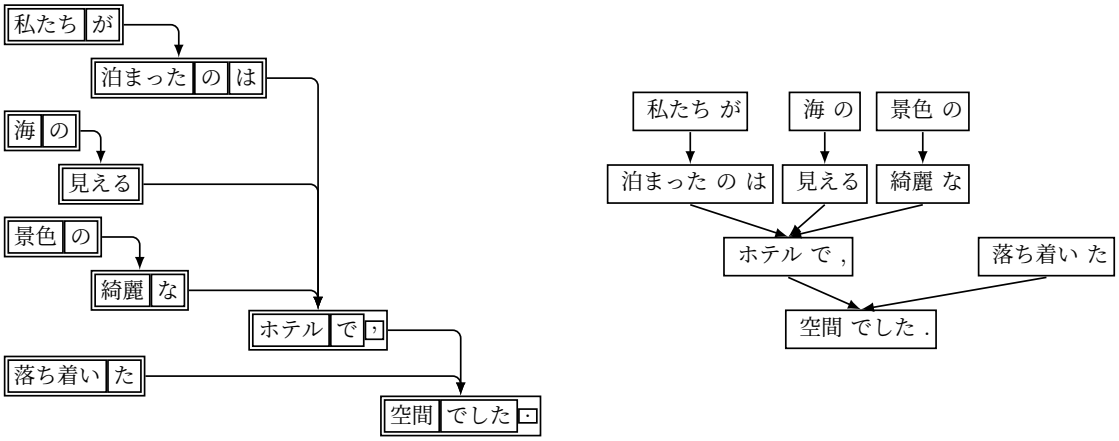


Figure 14. Example of the syntax tree of the parsing sentences.

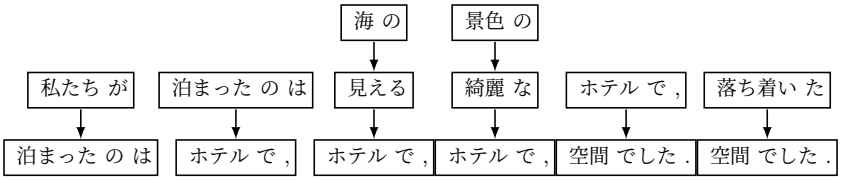


Figure 15. Extraction of phrases based on the hierarchical dependency syntax model.

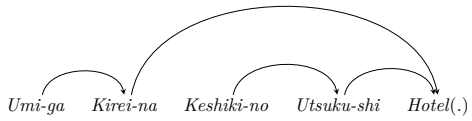


Figure 13. Example of a syntax tree of a phrase.

- (海が, きれいな): beautiful sea
- (きれいな, ホテル): beautiful hotel
- (景色の, 美しい): magnificent view
- (美しい, ホテル): magnificent hotel

From these events, we can find two meanings “きれいな (beautiful)” and “美しい (magnificent)” related to “ホテル (hotel).” However, only “美しい (magnificent)” qualifies “ホテル (hotel),” while “きれいな (beautiful)” qualifies “海が (sea).” As a result, there may be a vague relation between words if we generate an event-based topic model by using their dependencies. For these reasons, we have to detect the hierarchical structure of the events to extract them accurately.

This problem is caused by the application of a bi-nominal dependency parsing in a morphological analysis of Japanese sentences; therefore, we use the proposed dependency parsing, which can extract dependencies between phrases with complex structures, and utilize the dependencies as features to build our topic model. By using the proposed approach, we obtain variable orders of the dependencies; therefore, there is only a slight chance of finding a vague relation between phrases. Thus, the proposed approach has the potential to build a more appropriate topic model than the conventional method. This also helps us to evaluate the practical uses of the proposed dependency parsing.

B. Event Structure-based Topic Model

In order to construct a topic model using the proposed dependency parsing method for Japanese sentences, the following steps are executed:

1) Using the proposed dependency parsing method described in Section IV, we parse each sentence in a given document set and extract the dependencies between phrases in a variable order, and construct a syntax tree by using the extracted dependencies.
2) We divide the syntax tree into appropriate size of dependencies between phrases.
3) We can generate a topic model using the appropriate size of dependencies.

Using a Japanese sentence “私たちが泊まったのは，海の見える景色のきれいなホテルで，落ち着いた空間でした (The hotel with a magnificent view of beautiful sea which we stayed had a relief space),” we can construct a syntax tree shown on the left of Figure 14. The purpose of constructing the syntax tree is to obtain the appropriate size of dependencies between phrases, as shown in Figure 15, because the quality of the topic model depends on that of the extracted dependencies between phrases. Therefore, we extract them by reconstructing the tree structure whose root node denotes the last phrase “空間でした (had a relief space),” as shown on the right of Figure 14, by using the hierarchical dependency syntax model. Finally, we extract the set of phrases (element trees) shown in Figure 15.

Note that we do not use all the morphemes contained in the extracted phrases, as in the case of an event-based

topic model. That is, we remove only the phrases containing a postposition that expresses the case in each sentence, and use the remaining phrases (e.g., noun, verb, and adjective) as features of the proposed topic model. This procedure clarifies the topic boundaries.

As described above, the proposed topic model is built using the LDA shown in the graphical model that is replaced w with ph in Figure 6.

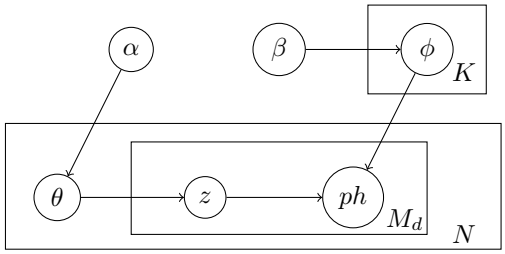


Figure 16. A graphical model of LDA with our extension.

In brief, a set of documents ph is expressed as the distribution of the appropriate size of phrases in each document. That is, θ is generated on the basis of the topic distribution corresponding to the phrases, and ϕ is generated on the basis of the phrase distribution corresponding to the topics. The parameters K, M_d, and N shown in Figure 16 denote the number of topics, the number of phrases with the most appropriate size of dependency in a document d, and the number of documents, respectively. As a result, this graphical model corresponds to a general LDA graphical model, replacing a word with a phrase having the most appropriate dependency size.

C. Experimental Evaluation of Proposed Topic Model

To evaluate the event structure-based topic model based on the proposed dependency parser, we compare our topic model with an event-based topic model. It is difficult to define the evaluation measure of this experiment; however, we evaluate each topic model in terms of whether it can explain the details of each detected topic or not.

First, we use the review data of *Rakuten Travel* [1], which is a travel website that provides information regarding hotel accommodation, to build the topic models. The use of these data allows us to score each review dataset on a five-point scale with respect to the quality of the hotel room and the hotel's geographical convenience. In short, there is a correlative relationship between the text of the review data and its score; therefore, we can allocate topics to each document on the basis of this correlative relationship. As a result, we can consider this correlation as the indication of each topic model because the generated topic model expresses this correlation through the elements in the documents.

We also consider that a document containing a number of topics is suitable for evaluating a topic model, because the text data are of varying lengths. As a result, we use review data whose length is approximately 100 words.

A large variety of parameters in each topic model are decided by reference to [23]. That is, the number of documents M is 2,000 which is randomly extracted from the entire review dataset, and the number of topics in the selected documents is 50. Hyperparameters α and β are also decided by a collapsed Gibbs sampler known as the Markov Chain Monte Carlo method. Table IV shows the results of this experiment obtained using conventional (left-hand side) and proposed topic models (right-hand side).

Here, we focus on two phrases: one is "チャペル–ある (be – a chapel)" at topic 3 on the left-hand side of Table III, and the other is "チャペル–ある–華やか (be – brilliant – a chapel)" at topic 4 on the right-hand side. Compared with these phrases, our topic model can extract a modifier "華やか (brilliant)" of a noun "チャペル (a chapel)." This should be valued because we can verify whether the topic has a positive aspect or not depending on whether the topic model can detect the modifier. From Table III, it seems that our topic model can detect a number of modifiers and can explain the contents of each topic easily by using them; therefore, our topic model can correctly classify documents compared with the event-based topic model, because our topic model can allocate appropriate phrases to each topic.

VI. Conclusion and Future Work

In this paper, we proposed a method for parsing dependencies in the Japanese language by using syntax trees based on a hierarchical structure, which are constructed by analyzing the modification relations in the considered documents. By considering these relations, the proposed method could solve the problems described in Section I and improve the accuracy of dependency parsing for long, complex Japanese sentences compared with the conventional method.

Moreover, topic estimation using the proposed dependency parsing method could be performed well compared with the existing method on the basis of the word distribution and by ignoring the order of words. That is, an appropriate dependency parsing method helps us not only to parse modification relations well but also to apply another research field of natural language processing; therefore, we could demonstrate the necessity for conducting basic research such as that on a dependency parsing method.

In the near future, we intend to utilize the order of words as well as the results of the syntactic and case structure analyses to improve the accuracy of the dependencies in the Japanese language. We also plan to conduct experimental evaluations using all the data in the Kyoto University Text Corpus.

Acknowledgment

We would like to express our sincere gratitude to all who commented on this work at the Eighth International Conference on Digital Society (ICDS 2014) for their assistance in identifying the weaknesses in our work and for giving suggestions on how to improve this study. We would also like to thank Rakuten, Inc. and the National Institute of Informatics for permission to use the resources of Rakuten Travel. This work was partly supported by a Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (A) (Grant Number 25240014) and Grant-in-Aid for Exploratory Research (Grant Number 25540150).

[1] http://travel.rakuten.co.jp/

TABLE III. EVENT BASED TOPIC MODEL

topic id	phrases with dependency relation
0	初めて–利用 泊まる–思う 西葛西–多い
1	部屋–狭い キャンペーン–着ける ツイン–宿泊 料金–宿泊
2	立地–良い ホテル–窓口 ツイン–部屋 とても–良い
3	駅–近い チャペル–ある 客室–忘れる 江坂–歩く
4	シングル–泊まる ホテル–泊る 以前–利用 カード–支払い
5	ホテル–部屋 ツイン–利用 シングル–快適 シングル–ある
6	ホテル–最寄り できる–ホテル ページ–見る 国道–ある
7	シングル–宿泊 ツイン–部屋 赤坂–向かう 京都–行く
8	以前–利用 皆さん–ご 繁華–困る 名古屋–出る
9	昨年–利用 交通–便利 まだ–おすすめ ９月–情報
10	フロント–対応 シングル–予約 ミナミ–中心 提供–地図
11	今回–宿泊 ９月–利用 部屋–広い 北大–いえる
12	部屋–狭い 私–利用 残念–ある 地元–住む
13	ホテル–窓口 シングル–予約 ツイン–宿泊 宿泊–教える
14	シングル–思う ・–点 部屋–きれい 部屋–狭い
15	静か–落ち着く ここ–数 除く–良い １月–連
16	先日–宿泊 ツイン–部屋 初めて–泊まる W–部屋
17	何–利用 シングル–泊まる 別館–ない 知人–ため
18	止まる–離れる 確か–広い 宿泊–考える 平成–前
19	シングル–予約 低層–なる ベッド–足元 窓口–宿泊
20	部屋–広い 部屋–狭い シングル–泊まる 日航–近く
21	今回–利用 私–泊まる 泊まる–する 昨年–利用
22	一般–ビジネス ．–広い 車–行う １０月–利用
23	駅–近い ツイン–部屋 ホテル–窓口 私–住む
24	ツイン–泊まる 駅–近い 泊まる–日 先日–利用
25	宿泊–ある ホテル–窓口 一泊–快適 シングル–泊まる
26	私–予約 ホテル–窓口 ここ–情報 ツイン–泊まる
27	ページ–予約 部屋–バス 部屋–広い 国際–便利
28	部屋–広い 駅–歩く 駅–近い 窓口–宿泊
29	ホテル–窓口 今回–宿泊 値段–割 会員–皆さん
30	駅–近い フロント–対応 ホテル–ある 宿泊–ホテル
31	シングル–泊まる ポイント–利用 先日–宿泊 今回–利用
32	結構–きれい 部屋–綺麗 ２月–利用 とにかく–安い
33	部屋–臭い 仕事–遅い 泊る–ツイン ３月–する
34	ずいぶん–訪れる 広い–情報 非常–する 念願–北海道
35	ホテル–窓口 宿泊–ある 窓口–利用 駅–便利
36	駅–便利 部屋–広い 利用–持てる 場所–わかる
37	ツイン–泊まる ツイン–部屋 ホテル–窓口 昨日–泊まる
38	ホテル–窓口 ここ–泊まる ダブル–利用 方–泊まる
39	ホテル–ある 大阪–慣れる 泊まる–感想 ホテル–近く
40	何–利用 部屋–宿泊 窓口–なる 髭–剃る
41	シングル–利用 年末–利用 皆さん–情報 初めて–泊まる
42	シングル–宿泊 ツイン–泊まる ９月–利用 フロント–対応
43	部屋–広い 今年–出来る 駅–近く 駅–近い
44	部屋–狭い 値段–割 シングル–宿泊 ツイン–シングル
45	近く–ある シングル–宿泊 ホテル–部屋 ネット–利用
46	初めて–利用 今回–利用 遊べる–ホテル 窓口–予約
47	部屋–きれい シングル–泊まる 値段–割 ここ–投稿
48	シングル–一泊 料金–安い 良い–良い 部屋–きれい
49	大変–感謝 ダブル–感じ 立派–ホテル 仕事–便利

TABLE IV. HIERARCHICAL EVENT BASED TOPIC MODEL

topic id	phrases with dependency relations
0	泊まる–思う–使える 西葛西–多い–抱く ベッド–快適–窮屈
1	部屋–狭い–きれい 安い–部屋 キャンペーン–着ける ロビー–作り–豪華
2	突然–止まる とても–良い–ホテル ホテル–窓口 ３月–よい–便利
3	以前–利用–とき 江坂–歩く 完成–きれい–助かる 中間–泊まる–部屋
4	カード–支払い–拒否 チャペル–ある–華やか ９月–中旬–利用
5	ホテル–泊る–トイレ 立地–良い–ある ツイン–ホテル–ある
6	ホテル–最寄り–駅 ＪＲ–線路–横 ページ–見る–予約
7	京都–行く 下–記述–もの 蒲田–駅–良い 観光–利用–持てる
8	価格–魅力–難 シングル–充分–思う ホテル–窓口–利用 ６月–下旬
9	まだ–おすすめ–思う シングル–出来る–思える おそらく–行う–点
10	ミナミ–中心 客室–改修–行う 立地–不便–便利 名古屋–ホテル–埋まる
11	ホテル–みる–行ける なかなか–きれい–ホテル すすきの–繁華–近く
12	シングル–思う–落ち着く 残念–ある–アクセス 地元–住む–する ５月–宿泊
13	ホテル–窓口–利用 以前–宿泊–時 部屋–入る–差し込む 町外れ–小高い–丘
14	ホテル–窓口–予約 ・–点–親切 別–予約–予約 部屋–狭い–いい
15	ここ–数–泊まる 私–泊まる–建物 駅–近い–利用 ちょっと–歩く–ある
16	W–部屋–つかう 泊まる–思う–泊まる 部屋–狭い–ありがたい
17	別館–ない–不便 一–なる–思う 部屋–バス–広め 新しい–ある–ホテル
18	ホテル–窓口–予約 止まる–離れる–眠れる 平成–前–言う 窓口–ある
19	ホテル–窓口–予約 相対–狭い–狭い 低層–なる–なる 昨年–９月–利用
20	ホテル–窓口–利用 みなさん–情報 宿泊–入る–心細い 議論–届く–届く
21	シングル–タイプ–宿泊 何–利用–利用 思う–ほど–奇麗
22	駅–便利–きれい ．–広い–朝食 メッセージ–ある–思う やはり–予定–観る
23	宿泊–ところ–思う コンビニ–ない–難点 クリスマス–土日–一泊
24	泊まる–日–雨 ロビー–大理石–ホテル 何–宿泊–最高 部屋–狭い–いい
25	午前–着く–ため 少し–前 私–予約–もらう 中–上–指摘
26	ホテル–窓口–予約 尾道–繁華–離れる 福岡ドーム–観戦–ため
27	部屋–バス–狭い ホテル–窓口–予約 私–泊まる–狭い 妻–東京–行く
28	名古屋–ある–立地 ホテル–ある–いい 昨年–夏–滞在 見かけ–古い–裏腹
29	ポイント–利用–いい 本–宿泊–感じる 去年–宿泊 立地–離れる–こたえる
30	ページ–予約 結構–きれい–コストパフォーマンス
31	ホテル–窓口–予約 温泉–予約–遠い 結構–きれい–ある 建物–古い–良い
32	宿泊–ホテル 妻–二–泊まる シングル–泊まる–狭い とにかく–安い
33	ずいぶん–訪れる–持てる 泊る–ツイン–気分 ３月–する–なる
34	場所–わかる–回る 広い–情報–みる 非常–する–ホテル 部屋–広い
35	ほんとに–ラブホテル–静か 朝食–バイキング 部屋–入る–差し込む
36	小松–右–清潔 １０月–ツイン–困る 部屋–綺麗–充実
37	大きな–執務 建物–言える–良い １月–週末–利用 地下–ある–いい
38	ツイン–部屋–宿泊 近く–ある–ない ダブル–利用–満足 連–欠点–厳しい
39	大阪–慣れる 宿泊–もてる–静か 泊まる–感想–する 非常–ホテル–滞在
40	部屋–用意–囚人 依然–宿泊–事 髭–剃る–入れる メッセージ–ある–思う
41	最寄り駅–行ける–ない 中洲–駅 先週–出張–際 観光–応対–ほう
42	年末–利用–思う 他–皆さん–情報 昨年–夏–滞在
43	今年–出来る 値段–割–高い 私–行く–ホテル シングル–する–ある
44	利用–なる ツイン–広い–気分 部屋–狭い 地下鉄–駅–約
45	ホテル–窓口–利用 ネット–利用–なる 広い–良い–ホテル 室内–する–部屋
46	新宿–駅–到着 近く–ある 窓口–予約–利用 いい–もと
47	ノート–もつ–出張 ここ–投稿–高い 利用–満足–感じ 中間–泊まる–部屋
48	立地–いく–ある 部屋–きれい–快適 良い–良い–広い 料金–安い–最高
49	大変–感謝–利用 ダブル–感じ–きれい 立派–ホテル–ゴージャスん

REFERENCES

[1] K. Ono and K. Hatano, “A dependency parsing method based on the hierarchical structure in Japanese language,” in The Eighth International Conference on Digital Society, 2014, pp. 203–208.

[2] M. Post and D. Gildea, “Weight pushing and binarization for fixed-grammar parsing,” in Proceedings of the 11th International Conference on Parsing Technologies, 2009, pp. 89–98.

[3] P. Blunsom and T. Cohn, “Unsupervised induction of tree substitution grammars for dependency parsing,” in Proceedings of Conference on Empirical Methods in Natural Language Processing, 2010, pp. 1204–1213.

[4] H. Shindo, Y. Miyao, A. Fujino, and M. Nagata, “Statistical Parsing with Probabilistic Symbol-Refined Tree Substitution Grammars,” in Proceedings of International Joint Conference on Artificial Intelligence, 2013, pp. 3082–2086.

[5] T. Kudo and Y. Matsumoto, “Japanese dependency analysis using cascaded chunking,” in Proceedings of Conference on Natural Language Learning, 2002, pp. 63–69.

[6] Y. Cheng, M. Asahara, and Y. Matsumoto, “Machine learning-based dependency analyzer for chinese,” Journal of Chinese Language and Computing, vol. 15, no. 1, 2005, pp. 13–24.

[7] M. Sassano, “Linear-time dependency analysis for japanese,” in Association for Computational Linguistics, 2004.

[8] N. Chomsky, The Logical Structure of Linguistic Theory. Springer, 1975.

[9] S. Goldwater, T. L. Griffiths, and M. Johnson, “Contextual Dependencies in Unsupervised Word Segmentation,” in ACL-44 Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics. ACL, 2006, pp. 673–680.

[10] D. Mochihashi, T. Yamada, and N. Ueda, “Bayesian unsupervised word segmentation with nested pitman-yor language modeling,” in ACĹ09 Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing. ACL, 2009, pp. 100–108.

[11] J. Pitman and M. Yor, “The Two-Parameter Poisson-Dirichlet Distribution Derived from a Stable Subordinator,” The Annals of Probability, vol. 25, no. 2, 1997, pp. 855–900.

[12] D. J. Aldous, Ecole d’Ete St Flour 1983, ser. Lecture Notes in Mathematics 1117. Springer, 1985, ch. Exchangeability and Related Topics, pp. 1–198.

[13] H. Ishwaran and L. F. James, “Generalized weighted Chinese restaurant processes for species sampling mixture models,” Statistica Sinica, vol. 13, 2003, pp. 1211–1235.

[14] Y. W. Teh, “A hierarchical Bayesian language model based on Pitman-Yor processes,” in Proceedings of Association for Computational Lin-

guistics, 2006, pp. 985–992.

[15] R. Kneser and H. Ney, "Improved backing-off for M-gram language modeling," in Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 1995, pp. 181–184.

[16] T. Hofmann, "Probabilistic latent semantic analysis," in Proceedings of Uncertainty in Artificial Intelligence, 1999, pp. 289–296.

[17] ——, "Probabilistic latent semantic indexing," in Proceedings of the Annual International ACM SIGIR Conference, 1999, pp. 50–57.

[18] G. J. McLachlan and T. Krishnan, The EM Algorithm and Its Extensions., 2nd ed. Wiley-Interscience, 2007.

[19] M. Collins, "Head-Driven Statistical Models for Natural Language Parsing," Association for Computational Linguistics, vol. 29, no. 4, 2003.

[20] T. Matsuzaki, Y. Miyao, and J. Tsujii, "Probabilistic CFG with Latent Annotations," in Proceedings of Association for Computational Linguistics, 2005, pp. 75–82.

[21] J. Boyd-graber and D. Blei, "Syntactic Topic Models," in Neural Information Processing Systems, 2008, pp. 185–192.

[22] Y. Hu and J. Boyd-graber, "Efficient Tree-Based Topic Modeling," in Association for Computational Linguistics, 2012, pp. 275–279.

[23] R. Kitajima and I. Kobayashi, "Latent topic estimation based on events in a document," Advanced Computational Intelligence and Intelligent Informatics, 2012, pp. 603–610.

[24] D. Mochihashi and E. Sumita, "The infinite markov model," in Proceedings of Annual Conference on Neural Information Processing Systems, 2007, pp. 1017–1024.

[25] N. Chomsky, "Three models for the description of language," IRE Transactions on Information Theory IT-2, vol. 2, no. 3, 1956, pp. 113–124.

[26] S. Kurohashi and M. Nagao, "Building a Japanese Parsed Corpus while Improving the Parsing System," in Proceedings of Language Resources and Evaluation Conference, 1998, pp. 719–724.

[27] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval: The Concepts and Technology behind Search (ACM Press Books), 2nd ed. Addison-Wesley Professional, Feb. 2011.

Development, Testing, and End-User Evaluation of Pervasive Community Signatures and Micro-Agreements Infrastructure

Architecture, Android Implementation, Performance Tests, Usage Examples, and User Evaluation

Mitja Vardjan, Helena Halas, Simon Jureša, and Jan Porekar
Research Department
SETCCE
Ljubljana, Slovenia
{mitja.vardjan, helena.halas, simon.juresa, jan.porekar}@setcce.si

Abstract—**Digital signatures are widely used for non-repudiation and other purposes. In various cases, there is a group of two or more parties that have to agree on a common set of data and digitally sign it in order to provide the other party or parties a proof of non-repudiation. A simple and scalable infrastructure for community signatures or groups of individual party signatures is described. It allows third party applications to simultaneously digitally sign arbitrary XML documents by any number of entities, for any purpose, using high level interfaces, not having to deal with digital signatures themselves. A dedicated backend server dynamically merges received documents and signatures from all parties. When a sufficient number of entities have signed the document, a signal is triggered to announce the document finalization. Despite the simple overall design, handling security issues and user control at appropriate spots are crucial for any business application. In the paper we present the performance and robustness tests of the current prototypal community signatures infrastructure. We also present the results of end user trials and measure the quality of experience perceived by end-users that are using a pervasive application that is interacting with the community signatures infrastructure.**

Keywords-community; agreement; digital signature; mobile environment, pervasive, e-business, infrastructure

I. INTRODUCTION

One of the most used aspects of digital signatures is non-repudiation. When electronic documents are digitally signed by one or more parties, the signatures can be used to verify the document integrity and, more importantly for this work, to prove that the parties have agreed on the document and stand behind it.

In many cases, only one valid digital signature is provided with the document at any time. The goal in such cases is usually to ensure document integrity, or to provide non-repudiation of a single entity. In case of signing contracts, agreements, and similar documents, two or more entities are to provide non-repudiation to each other. Some of these entities can be owners of internet connected pervasive services or internet connected objects. The signing process and distribution of digital signatures can easily get overly complex or even infeasible for the entities, especially if their number is large or arbitrary. This can be remedied in a business process where the document format and the order, in which it is signed by the entities, are determined by the application or protocol.

The infrastructural service described here allows for groups and communities to reach legally binding agreements in an ad-hoc manner. Third party services can offload any documents that need to be agreed over group of participants or even whole communities. These documents range from service level agreements, meeting minutes to non-disclosure agreements or even business contracts that may have rich content embedded. The work in this paper is a continuation and complement of [1] and [2].

The functionality reuses the concepts of digital identities, certificates and digital signatures. Documents are structured with Extensible Markup Language (XML) and agreements are signed using XMLDSig [3]. Both architecture and implementation target mobile and pervasive environments by providing an asynchronous and scalable solution that limits bandwidth usage, avoids unnecessary communication, and enables all user devices to be used from arbitrary local networks that are connected to the Internet intermittently and through firewalls.

Existing group signature and concurrent signature [4] solutions, especially the improved and multi-party versions [5][6][7] fit various purposes, but may not be most suitable for use by third party application developers who prefer well known solutions and expect fast and easy integration. Some existing designs for group signature use their own custom signatures and require additional solution-specific steps to sign the data and to verify a signature [8][9], or allow only community members to sign [10], which is not suitable for communities that are formed in an ad-hoc fashion. Such requirements can put additional burden to both implementation of third party applications that use the signature infrastructure, and to community administration. In terms of efficiency and optimization, additional network interactions are required, e.g., when the keystone is released in case of concurrent signatures. Moreover, both group signatures and concurrent signatures diverge even further from the traditional way of signing paper documents, still widely used. While the concept of fair exchange of signatures and decreased verification time are highly beneficial in some cases, the additional differences may

present an obstacle for adoption of the solution. For example, if the identity of the first signers is not known to all, subsequent signers may be less likely to be willing to sign the document. This may be because in case of known identities, they trust the party or parties who already signed the document, or simply because they have a proof that the party with known identity has already signed the document, e.g., when negotiating a service-level agreement (see [2] for examples). On the other hand, for communities where all members are equal and do not know or trust each other, the concurrent signatures are better in terms of fairness and non-exposure, but they are not used in the presented work.

The next section describes the initial document creation and its distribution to other users. The section is followed by descriptions of document signing and finalization procedure. Afterwards, various privacy and security aspects of the whole process are explained. Next, the implementation of community signatures infrastructure is presented and afterwards thorough tests of both performance and robustness to give an overview of its current capabilities and features. This section is followed with usage examples to illustrate a few implemented and suggested services that are using the presented community signature infrastructure. Special focus is dedicated to the "Pervasive Meeting Minutes" application and service that allows for unobtrusive capturing of meeting minutes. The application uses and interacts with the community signature infrastructure. At the end of the paper we present the results of the end-user trials focusing on perceived added value of end-users and on how the pervasive application, that is using the community signatures infrastructure, adds to quality of experience.

II. Document Creation and Distribution

Initially, an XML document with arbitrary schema and contents is created either by one party or in a collaborative manner by multiple members of a community. The document may hold a service level agreement, meeting minutes, non-disclosure agreement, or even business contracts that may have rich content embedded such as images, video or voice recording.

Regardless of what the document represents, the community members are expected to review it once it is finalized and confirm they agree with it. Their consent is formally expressed with their digital signature, appended to the document as a detached XMLDSig [3]. Depending on the application, a member may choose to sign the whole document, only some of its parts, or nothing and leave the document intact.

The initial document is distributed to the intended signers or members by uploading it to a dedicated Representational State Transfer (REST) server in a single HTTP PUT request. The REST server stores the document under the name, supplied by the client as resource name within the URL. The name is generated as a random string of a fixed length. The concept of resource name is similar to universally unique identifier (UUID) [11] but the name is shorter because it is checked for uniqueness at the server level when the resource is initially uploaded. Unless a resource with same name already exists on the server and the HTTP PUT request has to be repeated with a new name, the upload is a single step operation. The request includes the owner's serialized X.509 certificate [12] as part of the URL. This certificate is stored by the server for later authorization to access the document by others. It is never used to sign the document, unless the user chooses to do so. Therefore, it could be anonymous or generated ad-hoc by the initial document uploader. Its corresponding private key is used to sign the resource name. This signature is not supplied with the initial upload, but with another URL, generated by the community signature infrastructure.

Whenever a document is downloaded or a new version of existing document is uploaded, digital signature of resource name is passed as a URL parameter. The same URL is used for downloading and updating documents. The URL of the uploaded document is distributed to the members as an invitation for them to agree with and digitally sign the document.

Figure 1. Document creation and distribution.

The members list is usually application specific and the URL distribution is handled in the background by an app that is using the community signature infrastructure. If this is not the case, the URL and the document can still be accessed manually within the signature infrastructure itself (Figure 1). This lightweight and easy to implement process is suitable for the uploader device and signer devices, which are usually smart phones or tablet PCs. When a user chooses to reject or ignore the invitation to sign the document before he even reads it, bandwidth usage is negligible.

III. Micro-Agreements and Document Finalization

In the process of agreeing, the canonical form [13] of agreement document is digitally signed with a private key that is stored in participant's smart phone's secure storage (see Figure 2 for more) The meeting participants do not need

to sign the document immediately but can postpone the signing of the agreement.

After the agreement is signed by a participant it is uploaded back to the community sign service using the same URL that has been used to download it. The reasoning is that for community signatures, anyone who is authorized to download the document should be able to upload the signed version as well. If this is not the case, the concept of authorization signature in the URL can be easily expanded to include option to allow download only or both upload and download. An example solution is to sign document resource name, suffixed with an appropriate parameter, known to the service. The community sign service at the REST server verifies whether the digital signature is valid and whether the content of the agreement has not been modified in any way.

The community signature functionality allows third party services that are using it to specify the minimal number of community members that need to agree in either relative terms such as percentage of community or fixed threshold numbers. Every time the document with a new signature is uploaded to the community signature service backend node, this micro-agreement is merged into the main document stored on the server. Due to the nature of detached XMLDSig, the merges originating from various signers can be performed in any given order and the signers will experience a convenient and seemingly parallel signing procedure.

Figure 2. A community member receives invitation to sign a document.

The resulting document at any moment contains signatures from all parties that have signed the document and sent it back to the server so far. When number of parties that signed the document exceeds the given threshold, the community signature service backend server signals completion and participants can now download the final agreement, which now contains at least the required number of signatures (Figure 3) and represents a common and a legally valid agreement. Depending on the implementation, the document finalization can be signaled to the original document creator, e.g., meeting organizer, who can first inspect the document and the signers and then choose to signal document finalization to the other selected parties. At any point, the parties can see the current status of any document they have signed, or were invited to sign. Figure 4 shows the status of a document in the process of being signed (left) and the status of that same document at a later time, when one more party has signed it and the number of signers reached the required threshold (right). If concurrent signatures were used, full status with signers' identities could be displayed only after the keystone is released.

Unlike a group signature [8] where multiple individual signatures are replaced with a single group signature, individual signatures are preserved and any party can verify individual signatures using a standard verification procedure. Due to the nature of XMLDSig, any party can also get the list of all signers solely from the document.

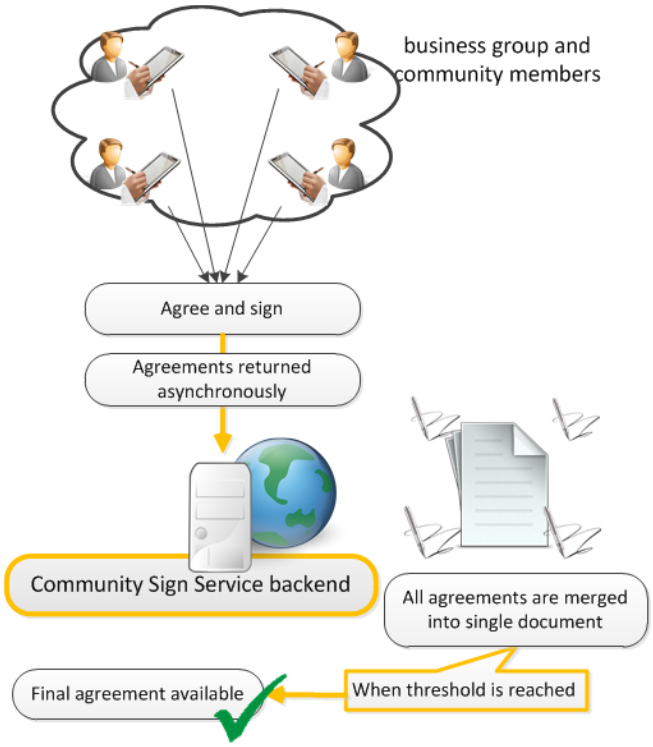


Figure 3. Community signature and document finalization.

The downside of not using the concept of group signature [8] is that processing power and time to verify all signatures increase with number of signatures in the final document. As the increase is only linear, this is usually not problematic in terms of scalability. If all parties can be forced to use a specific key-pair type, then verification of multiple signatures could be sped up [14][15], although care must be taken because some such solutions have issues [16].

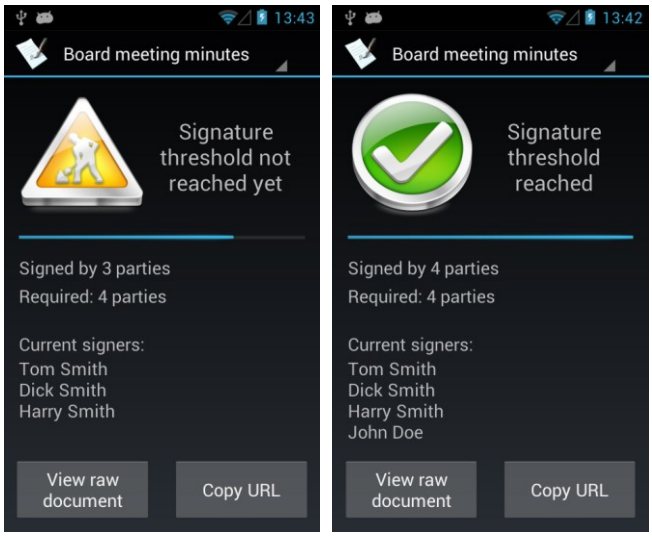


Figure 4. Viewing current status of the document signing process.

IV. Privacy and Security Aspects

The two main groups of information that could be treated as sensitive are the document contents and the list of entities who have signed the document. The document itself has to be made fully available to all entities that are given the option to sign it. Same applies to the list of signers because they all receive the final document in the end, leaving no alternative to ultimately trusting the entities not to disclose any sensitive information they receive.

Various notifications about document finalization do not carry any personal or document data and usually do not need to be secured. A few other points where it makes sense to take security into account are described below.

A. Document Distribution

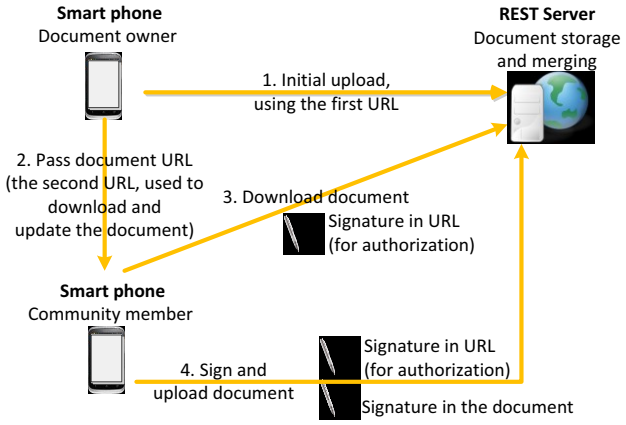


Figure 5. The two roles of signatures.

There are established protocols to encrypt the network traffic from eavesdropping. However, a custom solution described in Section II is used as a secure and convenient method to authorize the clients to download and upload the document. With the proposed solution, the clients (entities) are given only one URL that already contains all necessary tokens (Figure 5). As the digital signature of requested resource is part of the URL, the certificate owner can easily disable access by removing the public part of his certificate at the service backend (Figure 1 and Figure 5).

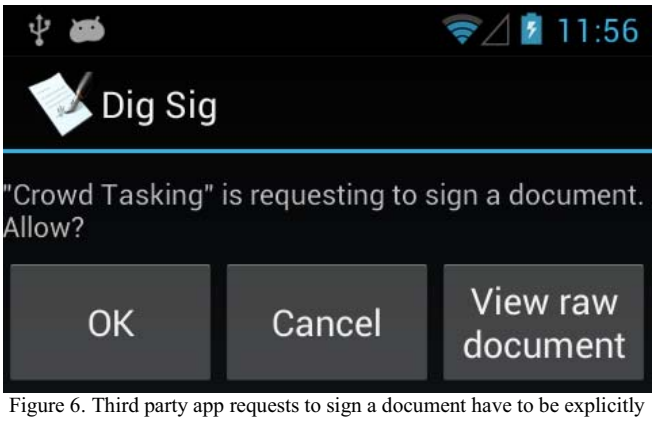


Figure 6. Third party app requests to sign a document have to be explicitly confirmed by the user.

Alternatively, when the certificate is revoked, access is automatically disabled, provided that the service backend implementation does check certificate revocation lists.

In any case, the number of network operations from mobile devices is limited and the authorization is integrated into the simple and widely used HTTP methods, so third party developers are not required to implement any authorization procedures.

B. Storage of Certificates on Android

With any digital signature based system, it is vital to protect the private keys from unauthorized use. The prototype has been implemented for Android where a secure storage is provided by the operating system. This storage is used for storing user's certificates and private keys. It is accessed in two significantly different ways, depending on Android version. For Android versions up to 4.2.2, the API is not public and the operating system grants requests to the storage based on the requestor process ID. The concept is described in [2]. For Android versions 4.3 and newer, the access to the secure storage is possible only through the new and official API for storing and accessing certificates and keys. To support all versions, the app implements both strategies and chooses the appropriate one dynamically.

C. Using the Securely Stored Private Keys on Android

To sign an arbitrary XML document, our prototype app can be used directly. However, in most cases it is to be used by other apps that parse the document and show the user a human readable and application specific document representation before the user authorizes signing. The problem is to access the user's private keys, which are not available to third party apps and not even to the operating system. As a solution, the third party app can simply invoke in the background our prototype app with access to private keys to sign the given document.

It is vital for the prototype app to show the user which app is trying to sign the document in the background, to prompt the user to authorize signing (Figure 6) and choose the identity to use (if multiple certificates are stored). The key itself is never exposed to third party apps, so only the data explicitly approved by the user are signed.

V. PERFORMANCE TESTS

Commonly, the primary reported problems after field release of software are not crashes or incorrect responses, but poor performance or inability to achieve required system throughput [17]. It is not uncommon that although the software has gone through extensive functionality testing, it has never been really tested to assess its expected performance. Such neglect of planning for performance issues often leads to performance problems once the software is released to the field, which in turn often significantly impacts the project's ultimate success or failure. The presented implementation has been continuously tested for functionality issues as its parts were developed. Once it reached sufficient level of completion, but still at early prototype phase, it has been tested also for performance issues that could possibly arise from the following two reasons:

- Multiple users are signing the same document and due to large number of users, thus, large number of signatures being appended and verified, the overall signing process takes more time to complete.
- Multiple users are signing the same document simultaneously. In addition to the reason above, the server load is increased because multiple HTTP requests and multiple database read/write requests are served in same time interval.

The tests were divided into two groups accordingly. Aspects of performance include latency, throughput, scalability, and reliability. In terms of reliability, the prototype performed successfully every time in each test. No errors were observed, except the out-of-memory error described in subsection C "Verifying signatures on the Android client". In some of real world tests and demonstrations, unreliable network and outbound port filtering used on some public networks caused additional delays and failures. While network connectivity related delays are unavoidable and the solution is designed and implemented to be used behind firewalls that filter out all incoming connections, outbound port filtering can usually be amended by configuring the REST server to listen on a more common port such as 80 or 443, which are usually not outbound-filtered.

With a powerful cloud computing infrastructure, the performance can be greatly improved, of course. However, to assess possibility of using a cheap server solution, the tests were carried out with a few years old everyday Core 2 Duo desktop PC running all necessary server software, including the database. For Android devices, Android software emulator and Google Nexus 10 were used.

The aim of performance tests is to realistically model expected common use cases of the implemented infrastructure under test. On the other hand, the tests were performed in a controlled environment so they could be repeated with same parameters. In the real world, the performance could be greatly affected also by various unavoidable difficulties like intermittent or extremely slow network, overloaded system resources on Android device, caused by other apps or malware, etc. The effect of such parameters is out of scope of these tests.

The following subsections focus on latency, throughput, and scalability aspects of performance.

A. Subsequently Signing a Document

For the first group of tests, an initial document version was uploaded to the REST server. Then, an Android client downloaded it from the REST server, locally signed it, and sent it back to the server. The server merged the signature and notified the client about successful finish of operation. For both the client and server, the real Android implementation was used. The Android client was run in test mode, which meant the user's clicks on appropriate buttons to approve the signing process and select the identity were performed automatically. After the operation finished, the whole operation (except the initial upload of the first document version) was repeated until the document contained 20 signatures. With every additional signature, the document got larger and the XML manipulation took longer. Besides, the signatures are verified at each step and as the number of signatures increases, the overall verification time at each step is increased. This effect is expected to be most profound in case of documents with large content to sign. In the tests, the whole document was being signed, except the existing signatures from previous steps.

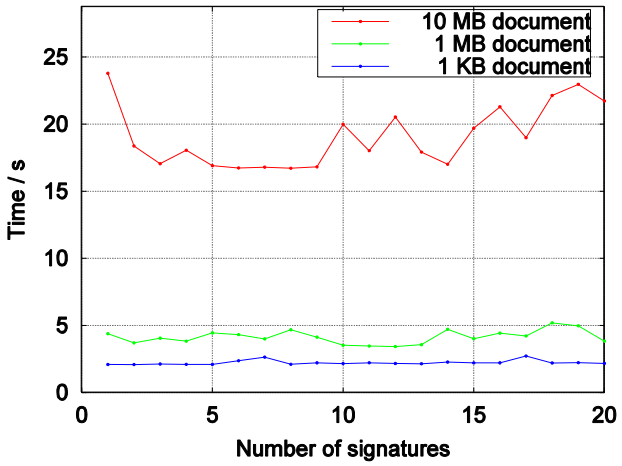


Figure 7. Time of subsequent signing.

Three different document sizes were used (Figure 7). The smallest document (1 KB size) contained minutes of a very short meeting. The larger 1 MB and 10 MB documents had additional text inserted. In both cases, the text was generated by a pseudo random generator, so the document could not be efficiently compressed. This modelled small and XML-embedded binary data that is already efficiently encoded or compressed, e.g., voice recording or other multimedia contents.

As expected and explained above, the time to verify the signatures increases with their number, and this effect is most noticeable with the largest document size.

Another increase in time is for the first signing process. This increase too, is most evident for the largest document size, but also noticeable for the medium size. It is expected that the later signing processes are faster because of data caching on the server and various connections and libraries already initialized during the first signing. Thus, the increased time of first signing procedure (or the decreased time of subsequent signing procedures) is probably unavoidable.

Median times of subsequently signing 1 KB, 1 MB, and 10 MB documents were 2.19, 4.09, and 18.21 seconds, respectively. Despite using a low-end hardware, shorter times were expected. Based on inspection of server logs, some interactions with the database consistently took at least half second and up to one second even for the smallest 1 KB files. Inserting a new XML node with signature and storing the new document also consistently took about one second for small XML documents. This indicates bottlenecks in our prototype that should be optimized first.

B. *Simultaneously Signing a Document*

The other test mimicked simultaneous signing of a document by multiple users. For the most realistic test, multiple Android devices should sign the document within a short time interval. However, a more feasible approach was used where the multiple Android devices or users were modelled by multiple threads running on a single Android emulator. This test is even more focused on the REST server than the previous tests, so the local signing on Android was skipped to avoid unnecessary load of already burdened client. The threads on Android only downloaded a signed document from the server and then uploaded it back. The server would still perform the usual request authorization, signature merging, and document update routines for each request.

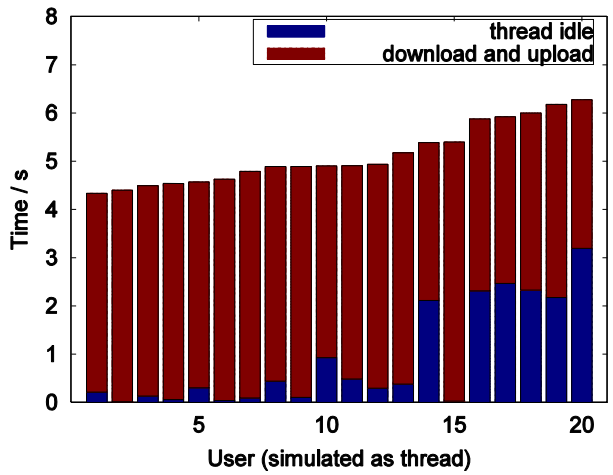


Figure 8. Time of simultaneous signing.

Figure 8 shows the time (from test start) until a thread actually started (blue bars) and the time a thread took to download the document, upload it to the server, wait for the server to perform the usual verification and signature merge, and receive notification about successful completion from the REST server (red bars). To assess primarily the concurrency issue, the 1 KB document was used here to eliminate influence of network bottlenecks and other consequences of large files. Time for simultaneously signing a 1 KB document is represented by the red bars in Figure 8. The median value was 4.37 seconds. The blue bars in Figure 8 mostly increase with thread index and some later threads reach 2 or even 3 seconds. This indicates client overload which possibly lead to over-estimated delays. However, based on server logs, most delays are caused at the server side, so the performance of a real setup would be only slightly better. It should be noted that the variable idle time of threads (blue bars in Figure 8) before they start the download and upload routines more realistically models an actual signing process than a strictly simultaneous start would, because it is unlikely that all users would sign a document at exactly the same time. So some randomness in thread start is necessary for realistic results, provided that at some point, all threads run simultaneously (between 3.2 s and 4.3 s in Figure 8), because it is the simultaneous signing that is being tested.

It should be stressed that the bar chart in Figure 8 is monotonically increasing only because the threads in Figure 8 are ordered by the time they received notification about successful completion from the server, not by the time they started.

Obviously, the results could be greatly improved not only by software optimization, but also by using more capable server-side hardware.

C. *Verifying signatures on the Android client*

After the document signing process is finalized and a user downloads the final document version, his Android client verifies all signatures in the document. The benefit is that the user does not have to ultimately trust the server, but this additional verification takes additional processing time.

Time of document signatures verification has been measured on Samsung Nexus 10 device with two different documents: a document where 1 KB of contents has been signed and a document where 10 MB of contents has been signed. One signature has been added to both documents and then same tests repeated with 10 signatures per document, which resulted in four distinct tests. Each of those four tests has been performed 20 times to account for variance in results. The measured time includes the time spent by invoking the digital signature app by sending it an intent and getting the results back to the original app, as it is this complete time that is relevant for a third party app that uses the infrastructure.

It has been observed that on old devices like Google Nexus S that have less memory than most modern Android devices, a fatal out-of-memory error occurs during verification of signatures for the 10 MB test document. The error occurs also on modern devices when the document is significantly larger, e.g., a 100 MB document on Nexus 10 tablet. The cause of this error is parsing XML with the

simple and easy to use Document Object Model (DOM) [17] parser interface which was required by previous versions of Apache XML Security for Java [19]. Since Apache XML Security for Java version 2.0.0, released in May 2014, StAX parser [20] can be used instead. Unlike the DOM parser, StAX is a streaming parser and does not load the whole document into memory. In case of large XML documents, it is usually more efficient, faster, and consumes less memory. It is expected that switching to StAX parser would not only solve the out-of-memory errors, but also optimize the signing and verification procedures on the Android client in terms of speed and resource usage, especially in case of large XML documents where only a small portion of the document is signed or being signed. According to severity classification of problems [17], the out-of-memory error is potentially a critical problem and should be among the first issues resolved.

TABLE I. VERIFICATION TIME ON ANDROID IN SECONDS

Document size	**1 KB**		**10 MB**	
Signatures	***1***	***10***	***1***	***10***
Time min	0.294	0.365	2.012	7.882
Time Q1	0.318	0.402	2.086	7.960
Time median	0.328	0.425	2.113	8.053
Time Q3	0.352	0.436	2.188	8.114
Time max	0.397	0.466	2.225	8.177

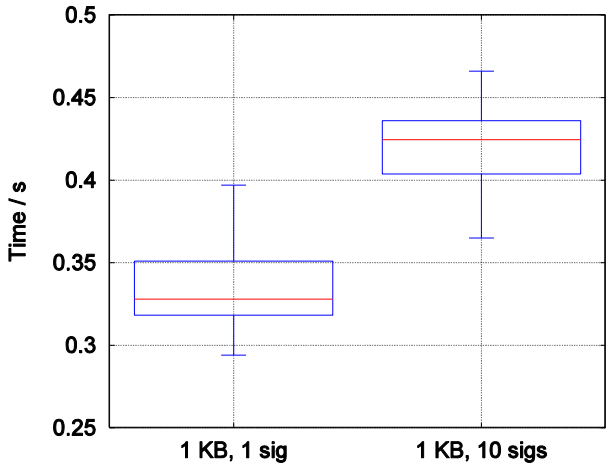


Figure 9. Verification time on Android client for 1 KB documents.

The statistical parameters of elapsed time for all four tests are listed in Table I and presented by box plots in Figure 9 and Figure 10. Minimum, the first quartile, median, the third quartile, and maximum are shown for verification time of each document. As expected, the time to verify signatures significantly increases with the size of signed data in the document (which approximately matches the document size). It also increases with number of signatures, but this increase is, as expected, more evident with large document sizes (Figure 9 and Figure 10). Unlike the server processing times, the Android client processing times are much more consistent and less variant. There is also no noticeable difference between the time of first verification after Android system restart and subsequent verifications. These results confirm that the user would experience small enough times of local verification in any case where the number of signatures and the document size are not too large. For cases where this assumption is not satisfied, an alternative solution with group signatures may be more appropriate.

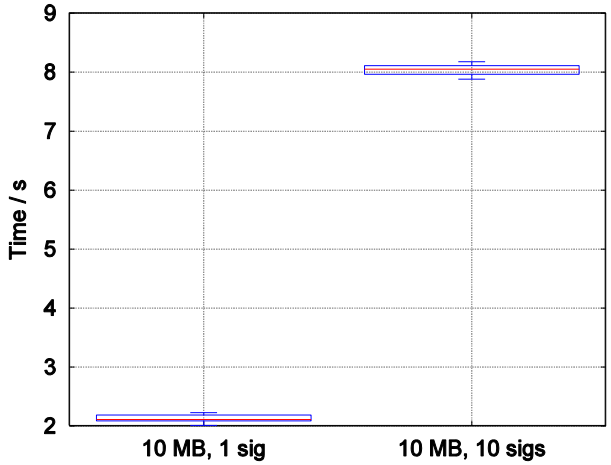


Figure 10. Verification time on Android client for 10 MB documents.

VI. USAGE EXAMPLES

Examples of usage are described below. The community micro-agreements are suited to also be used by applications and services that enable governance tools to communities.

A. *Pervasive Meeting Minutes*

"Pervasive Meeting Minutes" application allows for unobtrusive capturing of meeting minutes. The application uses and interacts with the community signature infrastructure. Community micro-agreements infrastructure allows business communities to capture meeting minutes and other meeting agreements in a legally valid and binding manner. The meeting organizer can choose whether the consensus is reached among only participants that are physically present during the meeting or the whole community.

Existing community signature prototype implementation has been used by an example app to capture meeting minutes. After users register to the meeting through this app, they can actively participate in the meeting. Their input is recorded by their Android devices and sent to a central Android device, which has the role of the document owner. When the meeting is finalized on that central device, the minutes are uploaded to the document storage server (Figure 12) and its URL is distributed to meeting participants. The REST servers which handle distribution of document URLs and receive notifications about document finalization (Figure 12) are application specific, i.e., implemented as part of the meeting minutes software, not the general community

signature software. Google Cloud Messaging (GCM) is used to relay the messages to Android phones of users who are to sign the minutes. At an earlier point, the meeting software automatically registers Android devices of community members with GCM to receive these messages. GCM is used by the meeting software as a convenient way to push small messages to Android devices, connected to the Internet through firewalls, with variable network addresses, etc. The community signature infrastructure does not require using neither GCM, nor the additional REST server to distribute document URL, but only to distribute the URL to community members. Therefore, any alternative distribution of the URL is valid. For example, the app on the central device embeds the URL into a Quick Response Code (QR code) and the physically present members can scan it (Figure 11), or – as the last resort – copy the URL that is displayed below the QR code (Figure 11). Again, this is only an alternative way of URL distribution and the primary way is application specific automatic distribution in the background, in this case through GCM.

Figure 11. Meeting attendants without the app can scan the QR code to download the signed document.

Arrows in Figure 12 indicate information flow for the implementation with GCM, starting with document upload by the document owner to the first REST server shown at the top center.

On the client side, directly seen by the users, the central Android device is running the main meeting app, seen by all users and shown in Figure 13. On the right, meeting details and minutes are shown. On the left, list of currently registered and manually added users is shown. Any number of users can use this central tablet as an alternative for their own Android devices to participate in the meeting. To provide input, a user only has to click his name in the list and speak or type his input. For the registered users, it is of course more convenient to use their own smart phones and tablets to provide their input, which is then automatically and instantly distributed to the other users and the central Android device in Figure 13 through dedicated REST server and GCM, in similar manner as the document URL is distributed in Figure 12.

Regardless of the implementation, the signatures are always in standard XMLDSig form, as in Figure 14. In the figure, XML nodes with signature and certificate values are collapsed but the highlighted text shows the signatures refer to the whole document, i.e., the whole meeting minutes. In case a participant agreed only with part of the document, his signature would refer to the relevant part only, provided that the application specific implementation allowed signing only a part of the document.

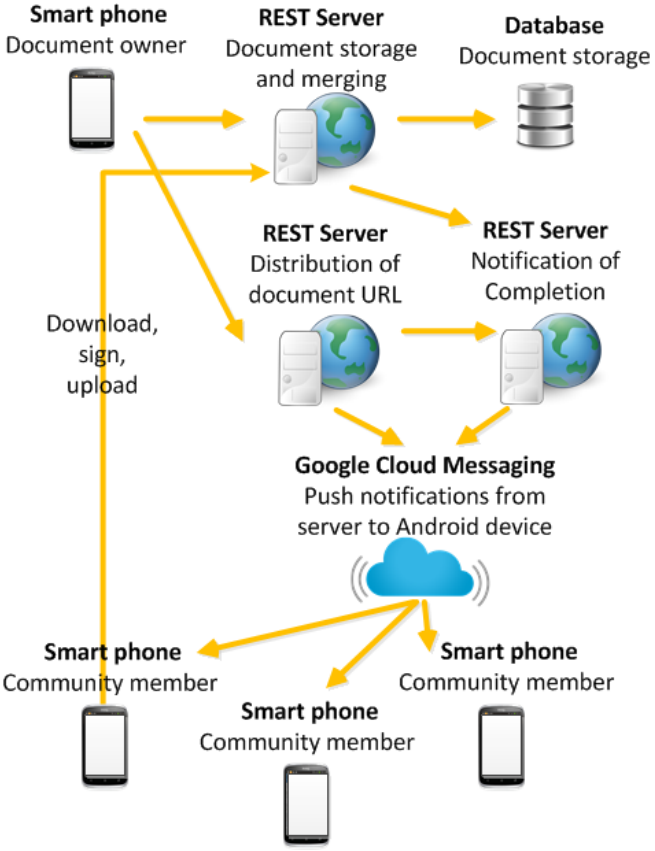


Figure 12. Process and information flow between devices in a chosen implementation for capturing meeting minutes.

In this example, the omnipresent issue of identity mapping is evident. Mapping between various identity types is essential for any legally binging document. Typical identity types relevant for community signatures are:

- Possible identities in the signed document. Figure 14 shows a case where identities are explicitly listed in the signed document. This is not always the case. The document could include only impersonal statements.

- Identities in encoded X.509 certificates, contained in the collapsed "ds:KeyInfo" nodes in Figure 14.
- Identities of the community members who signed the document.

Clearly, any implementation should check:

- mapping between the certificate filed values, e.g., common name, and the document identities, if any,
- certificate validity and whether it is issued by a trusted authority,
- mapping between certificate and real entity, e.g., by checking the entity listed in the certificate is actually a member of the community that is supposed to sign the document.

Figure 13. Pervasive Meeting minutes app with a meeting in progress.

For large communities, this can be far from trivial, as the certificate identities can be ambiguous and also because a single entity can be listed under different names in the certificate and community members list.

B. *Crowd Tasking*

A service called Crowd Tasking has been developed to enable community members to create tasks (an example is shown in Figure 15), propose solutions, post comments and solve tasks. These tasks usually involve some physical presence of people or physical work, which makes it inconvenient or impossible to post either the solution, or proof of the task solution to the service or to the Internet.

The service will integrate with the community signature infrastructure to enable task members to sign the agreements about the work to be done by each of them and to enable task creators to confirm the task completion by additional signature. As with any other usage of community signature, the interactions of third party service with community signature infrastructure and the document signing happen in the background, except prompting the user to confirm signing.

```
<meeting Id="Board001">
  <conversation>
    <user>
      <id>92001</id>
      <username>John Doe</username>
    </user>
    <minute>Hello and welcome to the meeting!</minute>
  </conversation>
  <conversation>
    <user>
      <id>97001</id>
      <username>David Smith</username>
    </user>
    <minute>Hi John. First things first: we need more drinks.</minute>
  </conversation>
  <conversation>
    <user>
      <id>92001</id>
      <username>John Doe</username>
    </user>
    <minute>Agreed.</minute>
  </conversation>
</meeting>

<ds:Signature xmlns:ds="http://www.w3.org/2000/09/xmldsig#" Id="Signature-a8
  <ds:SignedInfo>
    <ds:CanonicalizationMethod Algorithm="http://www.w3.org/TR/2001/REC-xr
    <ds:SignatureMethod Algorithm="http://www.w3.org/2000/09/xmldsig#rsa-s
    <ds:Reference URI="#Board001">
  </ds:SignedInfo>
  <ds:SignatureValue>
  <ds:KeyInfo>
</ds:Signature>

<ds:Signature xmlns:ds="http://www.w3.org/2000/09/xmldsig#" Id="Signature-5e
  <ds:SignedInfo>
    <ds:CanonicalizationMethod Algorithm="http://www.w3.org/TR/2001/REC-xr
    <ds:SignatureMethod Algorithm="http://www.w3.org/2000/09/xmldsig#rsa-s
    <ds:Reference URI="#Board001">
  </ds:SignedInfo>
  <ds:SignatureValue>
  <ds:KeyInfo>
</ds:Signature>
```
Figure 14. An example of meeting minutes document structure which is signed by two parties.

Third party apps can also take a common approach to integrate with the community signature solution and provide a single complete service. E.g., within Crowd Tasking, the user can initiate new meetings within a task or join existing meetings to discuss task management and work distribution with other users. The pervasive meeting minutes software described in previous section is then automatically initialized and shown on the central Android display (Figure 13). When the meeting is finished, the users receive notification to sign them within the Crowd Tasking app, which then uses community signature solution to sign the document.

C. *Service Sharing Within a Community*

The policy negotiation described in [2] could be extended by integrating with community signatures and micro agreements presented here. A service provider would negotiate a service level agreement (SLA) with a community instead of only a single service consumer. The community members would decide if a particular SLA is compatible with community's internal rules and sign the SLA so the service could be shared within the community.

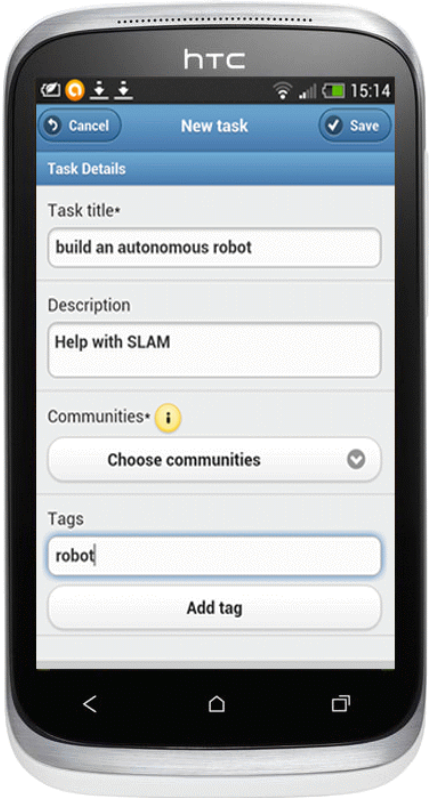


Figure 15. Crowd Tasking Service.

VII. User Evaluation

Separate user trials were organized in order to gain feedback about the implemented functionality and features, and how they could be applied to business communities. The concept of focus groups has been used. A focus group is a moderated discussion among a small group of people who discuss a topic under the direction of a moderator, whose role is to promote interaction and keep the discussion on the topic of interest [21][22]. A focus group technique stems from social research, but is now widely used to investigate new ideas in many research fields. In design science focus group technique can be useful as an exploratory method to achieve incremental improvements in artefact design or as confirmatory method to establish the utility of the design in a field use [22]. In our case purpose of focus groups was to collect ideas to improve user experience, which is very important in pervasive environment.

Two focus groups with external users took place in January and February 2014 at SETCCE research laboratory in Tehnološki Park 21, Ljubljana, Slovenia.

Prior to conducting the focus groups, the software was integrated with facilities in SETCCE research laboratory. Then an internal evaluation with SETCCE employees took place. The participants used software prototypes and gave their opinion and propositions, which have been taken into consideration when further improvements were made. The improved version was used for both first focus group with two participants and second focus group with three participants.

A. The Trial Participants

The external participants selected for evaluation have been chosen to represent the business communities. They are SETCCE's key customers coming mainly form Telecommunication and Insurance industries. The participants were all men, aged 31-60, and have from up to 5 years to over 10 years of experience in negotiation and participation in business meetings and assemblies. They are involved in decision making. They very frequently use mobile apps (most use Android), but had not participated in the design of the proposed solutions and were not familiar with the proposed infrastructure and concepts. Because of this, the evaluation was conducted as a focus group rather than a "trial" where participants are given tasks and interact with the software themselves. There are many possibilities how Crowd Tasking can be used so we had to focus usage to get relevant feedback.

B. Trial Objectives and Deployment

The meetings started with a demonstration video and presentations to explain Crowd Tasking service and the proposed infrastructure for community signatures. These were followed first by prototype demonstration and then by a moderated discussion and filling a questionnaire. During the trials, the participants were given the smart phones and tablets on which they were able to use the Crowd Tasking service, the pervasive meeting minutes service, and community signatures under moderated conditions. Four Android devices were used: a tablet to capture meeting minutes and serve as the initial document owner; three Android smartphones / tablets were used as normal user devices.

Demonstration consisted of the following steps:

- Creation of a new task and creation of a new meeting within the task.
- A user starts the meeting and the other users check in to the meeting.
- Holding a meeting.
- Meeting finalization.

Community signature solution is used in the last step. When the participants agreed on details about further task actions and solving, the meeting was finalized. Each participant received a notification to sign the meeting minutes. While one user signed the minutes, the others watched their phones to see how the document status changed when the first user signed it and how it changed when the process finished when the required number of users signed it.

The main objective is to get answers to the following questions:

- Are the implementation concepts of collaborative signature and micro-agreements suitable for use by other applications?
- Are the functionality and main concepts of collaborative signature easy to understand and be

accepted by typical business users? What are the obstacles for adoption by business communities?

C. Trial Staff and Rollout

The moderator was responsible to provide an introduction to the focus group, explain participants the code of behaviour and confidentiality (the names of the participants will not be made available publicly, no voice recording takes place, except during voice to text conversion during service demonstration, etc.), ask the questions as specified by the questioning route (see below), complement/clarify these questions if necessary, maintain the flow of conversation, make sure everyone has a chance to share their meanings, etc. He should create a comfortable, open atmosphere and avoid head nodding and verbal comments that signal approval of a meaning, avoid giving personal opinions.

The two presenters are responsible for providing an objective and balanced presentation of the concepts, demonstration of the services, aiding the participants in experimenting with the services themselves, etc. A presenter should not try to address criticism that comes up during the discussion. However, this was hard to avoid since both presenters had the best knowledge and understanding of the services and technologies used.

The assistant moderator is responsible to collect the list of participants, write down relevant ideas, comments and other parts of discussion, take notes throughout the session, and record any non-verbal activity that might help to correctly interpret the users' comments, or that might signal approval or disapproval.

D. Trial Questionnaire

Questionnaire consists of both open-ended questions (respondent formulates his own answer) and closed-ended questions (respondent picks an answer from a given number of options). The latter type includes polytomous (respondent has more than two options) and continuous (respondent presented with a continuous scale) questions. When preparing the questionnaire we strove that questions follow each other logically, from the least sensitive to the most sensitive, from the factual and behavioral to the attitudinal, and from the more general to the more specific.

E. User Feedback

The users provided separate feedback for the Crowd Tasking service and the underlying Micro-Agreements functionality. The complete questionnaire and aggregated answers are available in [23]. The items below summarize their answers and draw some conclusions.

1) Feedback Related to Crowd Tasking and Pervasive meeting minutes

Options to easily create new communities and invite members to meetings have been unanimously marked as useful.

Most of the questions about the usefulness and appropriateness of the "Crowd Tasking" was answered unanimously. Participants agree that the app has value in the business world/business processes, that by using the app some aspects of the business could be optimized and their work could be simplified. The app is easy to use. The app works as you would like, user experience is a positive one and the app is not too complex. Different functionalities in the app are well integrated, the app seems useful to them and they would use it. They would recommend the app to a colleague, customer, and partner. The app seems to be appropriate to confirm the agreements reached in the negotiations and to confirm decisions made at the Assembly meetings.

The participants did not agree if the app supports all the expected functionality and that the app is suitable for all levels of users.

Participants would like to see support for different workflows and better document visualization.

As positive aspects of the app they see mobility and virtual business environment, interoperability of different devices, simplicity and ease of use. A negative aspect is that the app works only on Android platform. In particular community there will be also someone with iPhone, Win device. Participants said that their experience of this app was wonderful, easy, satisfying and stimulating. Although most of this feedback is not directly related to the solution described in this paper, the feedback shows it is possible to create a simple to use and satisfying app based on the micro-agreements solution described here.

2) Feedback Related to Micro-Agreements platform

The existing and presented functionality of micro-agreements has been estimated as useful and important by the participants. Equally important was the ability to be easily integrated and used by third party services. However, suggested integration with other document repositories in the cloud (e.g., DropBox and Google Docs) has been marked as less important.

When signing contracts or agreements, trust in the other party or parties is particularity important. The participants feel that an assessment of trust level for other parties would be very beneficial. The participants think that they would benefit from clear visual presentation of trust levels at the time of signing. Nevertheless, only some would be willing to sign an agreement with users they are not familiar with, even if they had their trust level displayed. This is yet another lead to conclusion that the business users are still somewhat conservative and cautious when adopting new solutions and technologies that do not strictly mimic the established paper based business practices.

In general, the participants did not bring up any big privacy concerns or issues. They were more concerned about the trust in other signers, as already described above.

An issue that does not seem to bother business users too much, or not at all, is fair exchange [5][6][7] of signatures. The signing process of micro-agreements uses the common digital signatures with identity information embedded in XML document (XMLDSig). When a user signs an agreement, his identity is added to the signature and subsequent signers of that document can see the list of all previous signers. This is in contrast to the newer concept of "fair exchange" of digital signatures where the identities are not disclosed until a so called keystone is released. At that

moment, all identities become known to all users. While the fair exchange concept is indeed fairer to members of communities where all members are equal, the participants valued the micro-agreements feature of showing the list of identities that have signed the document so far. The business users, especially the more traditional ones, may be reluctant to sign a document where they cannot see the existing signatures from previous signers. Premature exposure of early signers' identities seems to be necessary, even though it does not result in a fair exchange of signatures.

Surprisingly, not all users considered asynchronous signing (without a pre-determined order) very useful, which again shows some users prefer traditional ways of signing documents where the signers sign a document in a sequential manner.

VIII. Conclusion and Further Work

An infrastructure and prototype implementation of community signatures and micro-agreements has been presented, followed by usage examples. The design uses digital signatures to sign XML documents, which can serve as legally binding agreements. It is based on REST servers, a database or other storage system, and Android devices. The simple, scalable and generic main concepts allow for fast integration of various third party services with it. Network communication is optimized for mobile devices with limited and intermittent bandwidth, but at least occasionally working network connection is still required for all devices. Compared to concurrent signatures, the presented approach requires slightly less network interactions, is more similar to traditional signing process of paper documents, and as such does not exchange signatures between parties in a fair manner, which has both advantages and disadvantages. Ideally, the solution could offer both signature options to cover additional possible scenarios. Other services are planned to use the implemented community signature infrastructure in an application specific manner.

The tests of implemented prototype showed it is usable for small documents even if only a single low cost server is used. With the exception of 10 MB documents, the signing times are low enough for most realistic use cases, but should still be significantly improved for a production version.

The focus group and observation meetings with the help of business community users from telecom and insurance industries were an enriching experience. We learnt about their work process and extensively explored their needs. The participants were open for new technologies and further cooperation, and provided several suggestions and ideas how their own business processes could be improved using such technologies.

The trial participants expressed clearly that they would prefer to use a system tailored to their business process and business needs. They informed us that solutions built on SOCIETIES results would need to reuse and reintegrate the elements and approaches from SOCIETIES project [24] and adapt them to their business context, rather than reuse the whole integrated SOCIETIES platform. By conducting the user evaluation we were able to gain information on how to upgrade and improve the pervasive meeting minutes application as well as the community micro-agreements infrastructure in order the end-users' experience.

Next steps will be to separate the community micro-agreements infrastructure from the integrated SOCIEITES open source platform. The focus will be to make a stand alone prototype that will only feature the functionalities that support the e-business community requirements and scenarios. During this process we will improve the implemented functionalities according to the trial results. We will integrate it with different mobile e-business applications that require digital signing capabilities to gain valuable feedback from 3rd party application developers.

We plan to extend micro-agreements infrastructure to support Stork2 eID based cross-border authentication [25] and due to its open-source nature to promote is as one of main candidates for smart phone authentication building blocks for eSENS project [26] .

Additionally we plan commercially reuse parts of community micro-agreements infrastructure. To achieve this we will need i.) to port the technology to others mobile platforms such as Windows Phone and iOS, ii.) to improve the verification of correct binding and mapping between community member identities and X509 certificate based identities used for digital signing, iii.) to extend the infrastructure with e-signing and e-archiving capabilities and finally to iv.) integrate results into SETCCE's commercial services and products, such as ePero®Start [27], eNvoices® [28] and ProXSign® [29].

Acknowledgment

Authors would like to thank the SOCIETIES project [24] consortium, the STORK2 [25] project consortium, the eSENS [26] project consortium and the EC for sponsoring the projects.

References

[1] M. Vardjan and J. Porekar, "An infrastructure for community signatures and micro-agreements," ICDS 2014: The Eighth International Conference on Digital Society, ISBN: 978-1-61208-324-7, March 2014, pp. 13-19.

[2] M. Vardjan, M. Pavleski, and J. Porekar, "Securing policy negotiation for socio-pervasive business microinteractions," SECURWARE 2012: The Sixth International Conference on Emerging Security Information, Systems and Technologies, ISBN: 978-1-61208-209-7, Aug. 2012, pp. 142-147.

[3] XML-DSig, XML Signature Syntax and Processing, 2nd Edition http://www.w3.org/TR/xmldsig-core/, [retrieved November, 2014].

[4] L. Chen, C. Kudla, and K. G. Paterson, "Concurrent signatures," Advances in cryptology - EUROCRYPT 2004, Vol. 3027, May 2004, pp. 287-305.

[5] X. Tan, Q. Huang, and D. S. Wong, "Concurrent signature without random oracles," IACR Cryptology ePrint Archive, 2012.

[6] C. Shieh, H. Lin, and S. Yen, "Fair multi-party concurrent signatures," Proc. of 18th Cryptology and Information Security Conference, 2008, pp. 108-118.

[7] J. Xushuai, Z. Zhou, W. Qin, Q. Jiang, and N. Zhou, "Multi-party concurrent signature scheme based on designated verifiers," Journal of Computers, Vol. 8, No. 11, Nov. 2013, pp. 2823-2830.

[8] L. Harn, "Group-oriented (t, n) threshold digital signature scheme and digital multisignature," IEE Proceedings - Computers and Digital Techniques, Volume 141, Issue 5, Sep. 1994, p p. 307-313, DOI: 10.1049/ip-cdt:19941293.

[9] L. Harn, C. H. Lin, and C. W. Hu, "Contract signatures in e-commerce applications," International Conference on Broadband, Wireless Computing, Communication and Applications, Nov. 2010, pp. 384-388, DOI: 10.1109/BWCCA.2010.101.

[10] C. M. Hsu, S. H. Twu, and H. M. Chao, "A group digital signature technique for authentication," IEEE 37th Annual 2003 International Carnahan Conference on Security Technology, ISBN: 0-7803-7882-2, Oct. 2003, pp. 253 – 256.

[11] ITU Recommendation X.667 (09/04), http://www.itu.int/rec/T-REC-X.667-200409-S/en [retrieved November, 2014].

[12] X.509 standard recommendation, http://www.itu.int/rec/T-REC-X.509/en [retrieved November, 2014].

[13] Canonical XML 1.1, W3C recommendation, http://www.w3.org/TR/xml-c14n11/, [retrieved November, 2014].

[14] C. Li, M. Hwang, and S. Chen, "A batch verifying and detecting the illegal signatures," International Journal of Innovative Computing, Information and Control, Dec. 2010, pp. 5311-5320.

[15] A. Atanasiu, "A new batch verifying scheme for identifying illegal signatures," Journal of Computer Science and Technology, Vol. 28, Issue 1, Jan. 2013, pp. 144-151.

[16] M. S. Hwang and C. C. Lee, "Research issues and challenges for multiple digital signatures," International Journal of Network Security, Vol.1, No.1, Jul. 2005, pp.1-7.

[17] E. J. Weyuker and Filippos I. Vokolos, "Experience with performance testing of software systems: issues, an approach, and case study," IEEE Transactions on Software Engineering, Vol. 26, No. 12, pp. 1147-1156, Dec. 2000, doi:10.1109/32.888628.

[18] Document Object Model (DOM), http://www.w3.org/TR/2000/REC-DOM-Level-2-Core-20001113, [retrieved November, 2014].

[19] Apache XML Security for Java, http://santuario.apache.org/javaindex.html, [retrieved November, 2014].

[20] StAX Java XML parser, https://sjsxp.java.net/, [retrieved November, 2014].

[21] D. W. Stewart, P. N. Shamdasani, and D. W. Rook, "Focus groups: theory and practice," 2nd ed., vol. 20, Sage Publications, Newbury Park, CA, 2007.

[22] M. C. Tremblay, A. R. Hevner, and D. J. Berndt, "The use of focus groups in design science research," Design Research in Information Systems, vol. 22, Integrated Series in Information Systems, DOI 10.1007/978-1-4419-5653-8_10, Springer US, 2010.

[23] SOCIETIES FP7 EU project deliverable D8.9, "Final evaluation report," 2014.

[24] Self Orchestrating CommunIty ambiEnT IntelligEnce Spaces (SOCIETIES), EU FP7 project, Information and Communication Technologies, Grant Agreement Number 257493. http://www.ict-societies.eu/ [retrieved November, 2014].

[25] Secure idenTity acrOss boRders linKed 2.0 (STORK2), EU cofunded project INSFO-ICT-PSP-297263, https://www.eid-stork2.eu/ [retirieved November, 2014].

[26] Electronic Simple European Networked Services (e-SENS) is a large-scale project that embodies the idea of European Digital Market development through innovative ICT solutions, http://www.esens.eu [retrieved November, 2014].

[27] SETCCE ePeroStart @ http://www.epero.si/ [retrieved November, 2014].

[28] SETCCE e-invoicing platform (eNvoices) @ http://www.setcce.si/item.php?catId=22&itemId=70&section=4 [retrieved November, 2014].

[29] SETCCE ProXSign @ http://www.proxsign.com/ [retrieved November, 2014].

A Survey on Modelling Historical Administrative Information on the Semantic Web

Jesper Zedlitz and Norbert Luttenberger
Christian-Albrechts-Universität
Kiel, Germany
Email: {j.zedlitz, n.luttenberger}@email.uni-kiel.de

***Abstract*—Identifying and referencing places is important for many fields of research. Very different approaches of how to represent administrative structures on the Semantic Web can be found. This survey attempts to provide a broad overview of systems that work on (historic) administrative information. We present a classification for such systems, with special attention to the difference that arise from the processing of historic data. We also describe a sample of systems which approach the problem in very different ways. We conclude by evaluating which of the presented characteristics make a system universal and future-proof.**

***Keywords*—conceptual modelling; administrative affiliation; semantic web; linked open data; historical information**

I. Introduction

When working with location information, it is often not enough to provide only the name of a place. One would like to uniquely identify the place. This identification opens up new possibilities. Let us give two examples:

(1) An obvious use case is the indication of a place's position on a map. This would not be possible having only the places' name.

(2) During a search, one can aggregate places. Imagine a data collection with census information. For each person a number of information including the occupation and the name of the place of residence is recorded. With a unique identification and administrative information linked to this identified place it becomes possible to search through all places within a province ("Show me all clock makers in Bavaria.") even if the places' names do not contain the province name.

However, this identification is not as simple as it might seem at a first glance. Usually, at the lowest level you have settlements such as villages, groups of houses, hamlets etc. These settlements are embedded in an administrative structure. At least in a large part of the world there are parallel political, ecclesiastical, and judicial administrative structures—"administrative objects" for the sake of brevity. It is easy to see that the affiliation of settlements with these administrative objects is needed to provide the desired functionality—such as the already mentioned aggregated search.

Changing administrative structures makes it difficult for the content editor to specify correct references. Also for the end-user changing structures are difficult understand. And, as a consequence, it is difficult for a user to formulate queries that include the expected results. Only a small example is given in Figure 1 showing only the affiliations for the late 19th and 20th century of the German village Suchsdorf as depicted by the GOV (see Section V. I).

Systems providing comprehensive historic administrative information help both, the content editors and the users, to navigate through these complex changing structures. For this reason, there are a number of projects that provide such kind of information on the Semantic Web in the form of Linked Open Data (LOD).

This survey attempts to provide an overview of the different modelling approaches that are used to publish (historical) administrative information on the Semantic Web. It is the extended version of the paper published at WEB 2014 [1]. It is organized as follows. The difficulties involved with place identification (not only) on the Semantic Web are described in Section II. In Section III we discuss related work. In Section IV, a classification for systems providing (historical) administrative information is given. A sample of ten of such systems is presented in Section V and characterized according to this classification. Section VI concludes by summarizing which of the presented characteristics make a system universal and future-proof.

II. Problems of place identification

For common place names, the name alone is obviously not sufficient for identification—just think of "Neustadt" in Germany. If only the names of places were available, it would be impossible to distinguish between entries from different places with identical names.

A common place name such as "Berlin" quickly leads to a presumption ("Berlin=capital of Germany") that may turn out to be wrong for the specific source. Not only is there a settlement called "Berlin" in the municipality of Seedorf in Schleswig-Holstein, Germany, but also numerous other places called "Berlin" exist in the United States.

Especially in Central Europe with its eventful history, it has often been the case that the name of a place changed over time. The problem is exacerbated in cases where different historic sources mention different names for the same place. Considering only the place name, one might wrongly assume that events have taken place at different locations, when in reality only the name of the place had changed over time. For example the German town "Wuppertal" carried the name "Barmen-Elberfeld" until 1930 (c.f. Section C.5). For a search one would like to see all results for one place, regardless of any name changes.

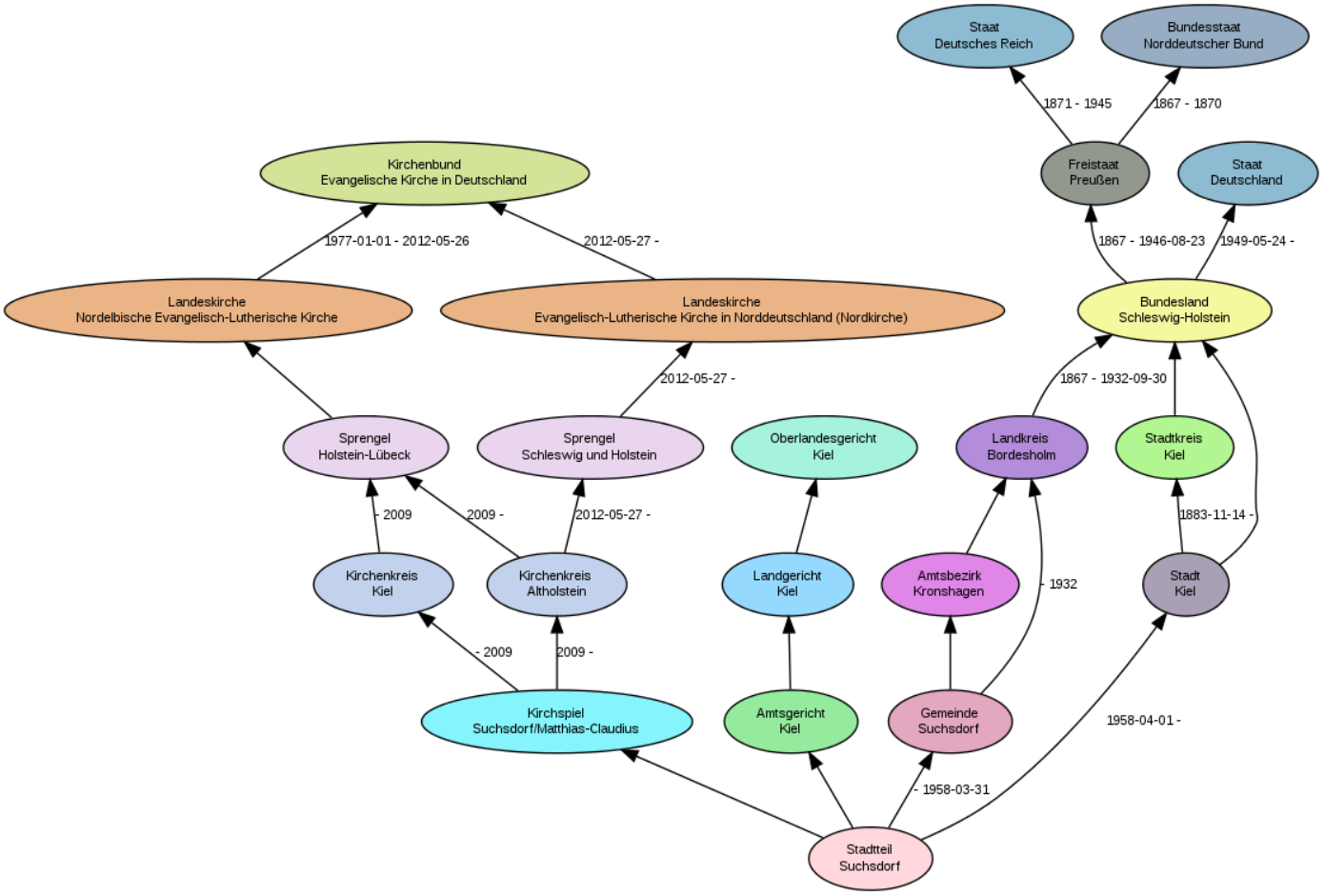


Fig. 1. Affiliations for the late 19th and 20th century of the German village Suchsdorf as depicted by the GOV.

You have to analyse a source closely to understand what is meant exactly by a place's name. In the simplest case it is the name of a settlement such as a village, a group of houses, a hamlet etc. However, it could also be the name of municipality or a parish. In that case the place's name might refer to a settlement—maybe with a different name—within that municipality/parish. Today, the settlement might belong to another parish or another community. The correct solution in that case is to use a reference to the administrative object. Otherwise one would at least pretend accuracy that is not given by the source—or even create a reference to the wrong place.

We illustrate this problem with the help of the map in Figure 2. The three polygons show—from smallest to largest—the village Schönberg, the municipality Schönberg, and the parish Schönberg. If a source states that an event took place in the municipality of Schönberg, it would be wrong to use the identification for the village Schönberg: It is possible that the event took place in the village "Neu Schönberg". However, if a source states that a person came from the parish Schönberg, the parish has to be identified not the municipality or the village Schönberg. At least, the parish contains nine municipalities and even more different villages, which come into consideration.

An often proposed solution for the identification of places is the usage of geographic coordinates. However, that does not solve the problem.

- Which coordinate do you use for a large city—the coordinate of the town hall or the coordinate of the church? There might be no church in the village or there might be several to choose from.
- Given two slightly different coordinates it is not possible to tell that they point to the same place.
- What does the coordinate point to? There are probably several objects at that position: the church, the village, the municipality, the parish, the county, etc.

The map in Figure 2 illustrates this problem. Approximately in the middle of the map you can find the geographic coordinate 54.396°N, 10.370°E. However, what is identified by this coordinate? The Lutheran church of Schönberg is located there. However, also the village of Schönberg, the municipality of Schönberg and the parish Schönberg contain this coordinate.

Therefore, geographic coordinates do not solve the problem. A unique identifier for settlements and administrative objects is needed. Such a unique identifier is associated with each resource on the Semantic Web—its Uniform Resource Identifier (URI). For this reason, there are a number of projects that provide such kind of information on the Semantic Web with approaches that may be very different in other technical details.

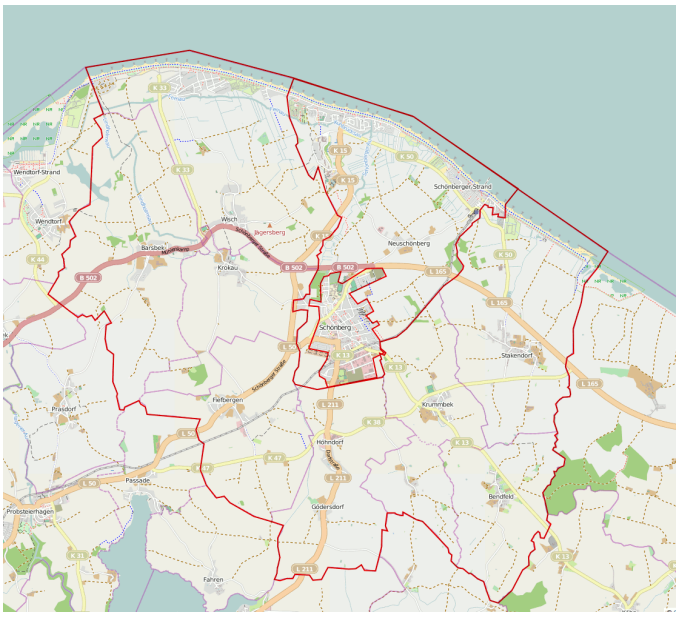


Fig. 2. Example for different administrative objects in the same region.

III. Related Work

The two technologies/languages that are typically used for publishing information on the Semantic Web are the Resource Description Framework (RDF) [2] and the Web Ontology Language (OWL) [3].

The representation of both spatial and temporal information with RDF or OWL is not trivial [4][5][6][7]. The problem of representing historical administrative information may be considered as a subproblem of modelling time in RDF and the Semantic Web. [8] contains a survey on approaches proposed to model temporal information, highlighting pros and cons of the different approaches.

Approaches such as named graphs [9] or contextual reasoning [10] can be used to enhance information with a time dimension. However, as the survey will show, these techniques have not yet found their way into existing systems that provide administrative information.

IV. Classification of Model Approaches

We have discovered several characteristics by which systems providing information about (historical) administrative structures can be classified. These characteristics fall into three groups: (A) fundamental decisions which are made regardless of the technology of the Semantic Web and which affect all systems, (B) characteristics which affect only systems in which the time is a concern, i.e., which also contain historical information. (C) characteristics which are based on the use of the techniques of RDF or OWL.

To some extent, the characteristics influence each other. Here is an example: If on the one hand you cannot or do not want to use the technique of reification (see Section C.3), but on the other hand you want to give population numbers for different years, you are almost forced to work with different individuals for different points in time. Otherwise, population numbers for different year could not be distinguished.

A. Basic characteristics relevant for all directories

1) Source citations: A simple characterization is the fact of whether source citations are provided for the published information or not. Especially in cases of contradicting information (e.g., different population numbers for the same point in time), source citations allow to judge the quality of the data.

2) Number of hierarchy levels: Some models have a limitation in the number of possible hierarchy levels. A typical example for three levels in the civil administration is a county belonging to a state which in turn belongs to a country. The administrative affiliation can be represented more generally if there is no restriction on the levels of the hierarchy. This is usually achieved by the definition of a general `isPartOf` or a `belongsTo` relationship.

3) Only current values or complete history: Some projects make the decision to provide only the latest data and no historical information. Therefore, no time-dependent values are needed. When a value or a name changes, the previous value is overwritten. One problem is that overwriting the values makes it difficult to reproduce reasoning. The value valid at the time of reasoning might be no longer available.

B. Basic characteristics for historical information

The following characteristics are only relevant for systems that also contain historical information. For systems that only provide current information, these characteristic do not apply.

1) Topology vs. specification of time: Instead of making specific indications of time when an administrative structure was established or dissolved, one may state topological relationships (i.e., predecessors, successors, etc.) instead. Figure 3 shows an example. When working with such a topology, one does not have to deal with all the problems that accompany the treatment of time on the Semantic Web.

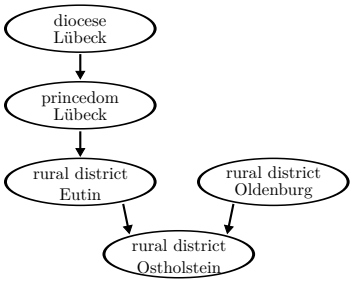


Fig. 3. Example for modelling changing administrative objects as a topology.

2) Activities or results: Two approaches exist to model changes in the administrative structure. The activities, i.e., the processes of changing (incorporation of one local authority by another, renaming, etc.), can be modelled, or the results of these changes. Figure 4 shows an example for both approaches. The left half shows the results of an incorporation. In the right half, the process of the incorporation is represented by an additional individual in the ontology.

3) Time-slices vs. individual times: To avoid problems that accompany the treatment of time on the Semantic Web, several ontologies can be used. Each of these ontologies contains information about just one single point in time. Within each

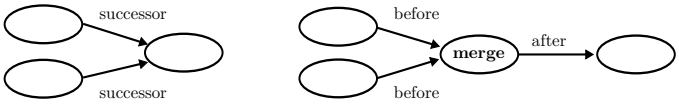


Fig. 4. Modelling a change as a result (left) or an an activity (right).

of the ontologies, all relations and numerical values (e.g., population numbers) are specified, as they were valid at the selected point in time. One ontology represents one time-slice. An administrative object that exists across multiple time-slices will appear in each of the ontologies as an independent individual. It is possible to connect the different individuals belonging to the same administrative object by using additional object properties. Figure 5 shows a sketch of this approach.

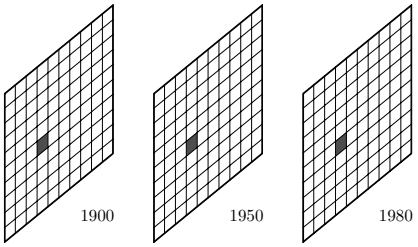


Fig. 5. Visualization of three "time-sliced" ontologies.

One disadvantage of the time-slice approach is that a completely new ontology for each considered point in time is needed. This solution is therefore only practical if one want to process a relatively small number of points in time. Throughout Germany changes at the municipal level occur about once a month. This would lead to a vast and hardly manageable number of ontologies.

4) One or multiple individuals: There are different approaches on how to model an administrative object during its entire lifespan. It can be represented by a single individual for its entire existence. In contrast, a new individual is created for every point in time—similar to time-slicing. As an intermediate form, a new individual is created only in cases where the administrative object is subject to changes.

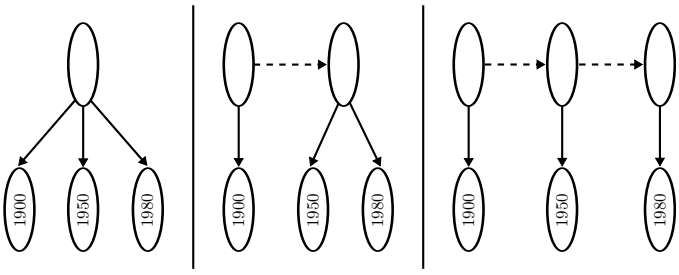


Fig. 6. Different approaches of using one or multiple individuals per administrative unit.

Figure 6 sketches the different approaches. The variant when a single individual is used for the entire lifetime of an administrative object is shown on the left. All three time-dependent values are associated with that individual. At the center, the variant where a new individual is created only in case of the change of a value is illustrated. In this example, the value has changed between 1900 and 1950. Between 1950 and 1980, it has remained the same. Therefore, two different individuals are required. The dashed line shows a possible object property connection between the individuals that represent the same administrative object. The variant where a new individual is created for every point in time is shown on the right. Each individual is connected with exactly one value—regardless of whether the value has changed or not.

C. RDF-/OWL-specific characteristics

1) Text vs. objects: One way to provide information about administrative structures is to model only settlements as individuals. Information about affiliations is added to these individuals as literals (data properties) using the names of the administrative objects. An example of such modelling is shown in the upper half of Figure 7.

With this type of modelling, it is very easy to search for settlements which are subordinate to a higher administrative level (e.g., a state)—only simple comparisons of data properties are necessary. A navigation within the administrative hierarchy is not possible with this type of specification. Also, a clear identification of the administrative objects is not given: If two superordinate administrative objects have the same name they cannot be distinguished.

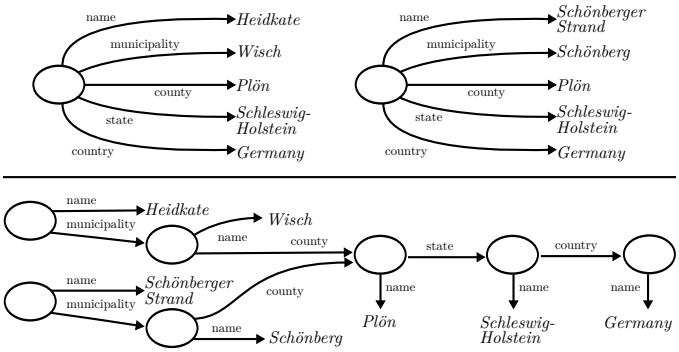


Fig. 7. Modeling with data properties only (top) and modelling with data and object properties (bottom).

This problem can be avoided by creating its own individual for each administrative object, which can be referenced by a URI—as usual on the Semantic Web. A link between the settlement and superordinate administrative object is created by using object properties. An example of this kind of modelling with individuals and object properties can be found in the lower half of Figure 7.

Since the objects of higher administrative levels (counties, states, etc.) occur only once and are named only once, one has to maintain less properties in total. The more objects are subordinate to another administrative object, the clearer this advantage becomes.

2) Types as classes, individuals, or literals: There are three ways to represent the type of an administrative object. Figure 8 illustrates these different modelling approaches:

1) an OWL class—the individual representing the administrative object is instance of that class

2) a reference to an individual—the individual representing the administrative object has an object property that specifies the type
3) a literal—the individual representing the administrative object has a data property that contains the type as literal

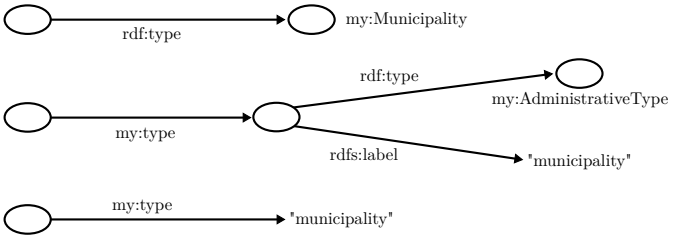


Fig. 8. Three different approaches to model the type of an administrative object.

The use of OWL classes (the first approach) has significant disadvantage when dealing with historic information: A municipality would be modelled as an instance of a class "Municipality". If the municipality gets town privileges later, it also becomes instance of a class "City". Here, a problem of this approach becomes evident: RDF or OWL have no time-dependent memberships in a class. Both class memberships are valid indefinitely. The change from municipality to city is no longer visible. Therefore, this approach is really only useful if you use multiple individuals for the representation of an administrative object over its lifetime.

Using an individual (the second approach) has the obvious advantage that labels can be specified for different languages using RDF's language tags. In addition to the RDF/OWL-2 class-hierarchy an individual offers the possibility to create a hierarchy of types using a specialized object property. Since the relation between the administrative object and it's type object is an object property it is possible to use it for inference rules. Data properties (as used in the third approach) do not offer such a flexibility when defining inference rules.

3) Reification: In logic in general, and the Semantic Web in particular, the term "reification" has several meanings [11]:

a) an encoding of n-ary relations/properties as individuals
b) the possibility RDF offers to assign URIs to statements and treat them as resources
c) the use of classes as individuals
d) the usage of RDF as metalanguage for other logics

For the context of this article only the meanings (a) and (b) are relevant.

Chaning administrative affiliations can be modelled as n-ary relations. Not only the two individuals representing the involved administrative objects are relation members, but also temporal information, source citations, etc. are members of the relation. Instead of defining an object property between two individuals directly, the connection itself is represented by an individual. Figure 9 illustrates how an administrative affiliation can be enriched with a time period by using reification.

This encoding of n-ary relations into individuals works both for RDF and OWL-2 DL. However, RDF has a concept called

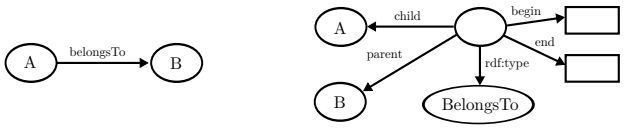


Fig. 9. Reification of an object property.

"higher-order statements" that can be used for making statements about other RDF statements. RDF offers the possibility to assign a URI to a statement and treat it as a resource. This resource can again be the subject of other statements. Here is an example. The first line and the last four lines have an equivalent meaning.

```
1 :A :belongsTo :B .

3 :s1 a rdf:Statement ;
4    rdf:subject :A ;
5    rdf:predicate :belongsTo ;
6    rdf:object :B .
```

The main difference between these two approaches is the fact that in the latter case the reified statement (the object property connection) is still part of the model. However, these statements about other statements do not exist in OWL-2 DL.

4) Specification of names: To specify names there is the choice to use the existing `rdfs:label` property or to define a separate specialized property. The range of the `rdfs:label` property includes the data type `rdf:PlainLiteral`. In RDF, plain literals have an optional language tag as defined by [12]. Therefore, it is possible to specify names in different language variants, e.g., `"München"@de`, `"Munich"@en`. Using these language tags, even the specification of very specific languages or dialects is possible. For example, the language tag `sl-rozaj-biske` indicates the "San Giorgio dialect of Resian dialect of Slovenian" [12, p. 80].

Another observed possibility to specify names for different languages is the usage of a specialized data property for each language, e.g., `englishName` or `germanName`. This approach limits the number of supported languages, which might be an advantage for the implementation. However, the universality of the system is reduced.

In order to manage different name variants for one language, the relatively often used Simple Knowledge Organization System (SKOS) [13] provides several properties for names. To supply further information, such as a period of time or references for a name, it is necessary to define a specialized property. Of course, this property can be defined as sub-property of SKOS properties.

5) Specification of time as standalone property: Indications of time can be specified either as a standalone property (see lines 1–4 in the example) or within literals—e.g., changing names of a German city from "Barmen-Elberfeld (-1930)" to "Wuppertal (1930-)" (see lines 6–9 in the example).

```
ex:Place1
  a ex:Municipality ;
  ex:name [ ex:value "Barmen-Elberfeld"; ex:end "1930"^^xsd
       :gYear ];
  ex:name [ ex:value "Wuppertal"; ex:begin "1930"^^xsd:
       gYear ] .

ex:Place2
  a ex:Municipality ;
  ex:name "Barmen-Elberfeld (-1930)" ;
  ex:name "Wuppertal (1930-)" .
```

A human user is likely to read and understand indications of time within literals. However, for machine processing (reasoning or SPARQL queries), they are not suitable. For example it would not be possible to select the name of the place for a given year.

V. PROJECTS IN DETAIL

For this article we selected projects that offer information about administrative structures on the Semantic Web. There are other approaches and projects for the publication of administrative structures. However, since they do not target the Semantic Web and its technologies, they are not covered in this article, e.g., the OpenGeoDB project (http://opengeodb.org). We also did not include ontologies or other data collections on the Semantic Web that only offer information for a single type or very few type of administrative objects, such as collections of statistical values for the countries of the world. The models for these kind of data are very simple and do not have to struggle with the problems discussed in this article.

We classify each modelling approach on the basis of the characteristics presented above. For most cases we show a visualisation of the project's model (or the relevant parts of the model) to provide the reader with a quick overview. Additionally, we give an example from the project's data written in Turtle[14]—a much more readable syntax than other RDF representations such as RDF/XML or NTriples.

Table I shows an overview of all ten projects that are discussed in this article. Each column of the table represents one of the characteristics listed above—in the same order as in this article. The table shows two rows for the SAPO (see Section V. H). The lower row represents an assumption about the ontologies used internally, based on the publications on SAPO.

A. schema.org

Schema.org [15] is an initiative of several search engine operators. It provides vocabulary and the TBox of an ontology for semantic annotation of HTML pages. Thus, in contrasts to the other systems presented in this paper, the schema.org ontology does not contain any individuals. The ontology contains a part that deals with the description of administrative structures. The relevant classes and their relations are shown in Figure 10.

The model is very simple. There are only three specialized types of administrative objects—cities, counties, and states. Additionally, there is a generic `AdministrativeUnit` class. However, it is not possible to specify the type of such a generic administrative object. In addition to the properties inherited from the class `Thing` to specify a name and a description, the geographic position can be specified by using two classes defined within the schema.org ontology.

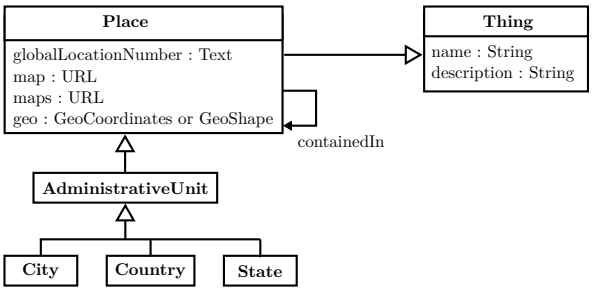


Fig. 10. Excerpt from the schema.org model that deals with administrative information.

The model provides no indication of time. Dependencies are represented using the object property `containedIn`. Due to the existence of the general `containedIn` relation, the representation of an arbitrary number of hierarchy levels is possible. Source citations are not possible.

B. DBPedia

The DBpedia project [16][17] extracts information from the various language variants of Wikipedia and publishes it as part of the LOD cloud as a RDF knowledge base. DBpedia's ontology contains a part that deals with the description of administrative structures. The currently most recent version of DBpedia's ontology can be found at [18]. The relevant sub-classes of `PopulatedPlace` and their connections via object properties are shown in Figure 11.

By taking a look a the ABox of the DBPedia ontology, it turned out that many of the classes in the `PopulatedPlace` class-hierarchy are currently not in use. The unused classes are depicted with gray rectangles in Figure 11.

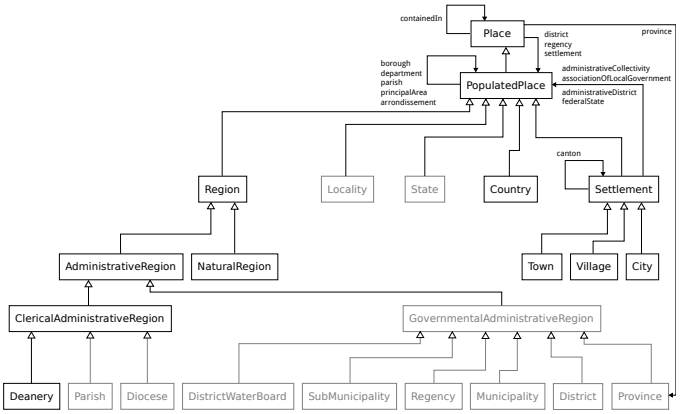


Fig. 11. Excerpt from the DBpedia ontology showing sub-classes of PopulatedPlace

The ontology also contains several other classes, which are not part of the `PopulatedPlace` class-hierarchy. However, the class names (e.g., "AustrianSettlement", "FrenchSettlement", "GermanSettlement") suggest that these classes should

TABLE I. OVERVIEW OVER THE CHARACTERISTICS OF THE PRESENTED SYSTEMS.

		all systems			systems with time				RDF/OWL technology				
	individuals	source citations	hierarchy levels	temporal information	time/topology	activity/result	time-slices/ individual times	one/multiple individuals per adm. object	administrative object type	property for names	time as property or within texts	adm. hierarchies as text or object	reification
schema.org	-	-	∞	-	*These systems do not contain temporal information.*				class	own	-	obj.	-
DBPedia	+	-	∞	-						rdfs:label	-	obj.	-
GeoNames	+	-	pract. 5	-					indiv.	own	-	obj.	-
LinkedGeoData	+	-	-	-						rdfs:label	-	text	-
GB Ordnance Service	+	-	4	-					class	rdfs:label + SKOS	-	obj.	-
SHV	+	-	pract. 5	+	topol.	r	?	ind.	class	rdfs:label	-	obj.	-
GND	+	-	∞	+	topol.	r	-	-	indiv.	own	-	obj.	-
SAPO	+	-	3	+	time	r	ind.	mult.	class	rdfs:label	text	obj.	-
SAPO (interal)	+	-		+	time	a	t.s.	mult.	?	?	?	obj.	-
GOV	+	+	∞	+	time	r	ind.	one	indiv.	own	prop.	obj.	+
TGN	+	+	∞	+	time	r	ind.	one	indiv.	own	prop.	obj.	+

actually be subclasses of `Settlement`, too. The ontology also contains properties for these classes. However, currently neither the classes nor the properties are used.

In general, DBPedia contains no historical information. Only for a single population number per individual an indication of time can be given by using the data property `populationAsOf`. Source citations are rudimentary possible. The object property `http://www.w3.org/ns/prov#wasDerivedFrom` is used to stated from which Wikipedia article the data was taken. For population numbers, the source citation can be specified as a note using the data property `http://dbpedia.org/property/populationNote`.

The type of an administrative object is specified by using OWL classes (C.2), see Figure 11. In addition, however, there is the data property `settlementType`. It contains the type of the administrative object as literal (C.2).

Basically, the ontology provides the distinction between settlements (`Settlement` and sub-classes) and administrative objects (`AdministrativeRegion` and sub-classes). The classes `Country`, `State`, and others show that this distinction has not been carried out systematically. It can be observed that in some regions of the world, there is a fusion of settlements and administrative objects. This is problematic because they are actually different individuals—in the semantics of RDF. This is the case particularly with the information taken from German Wikipedia.

Administrative affiliations are represented by the object property `isPartOf`. Additionally, there are a couple of object properties (e.g., `country`, `federalState`, `municipality`), which allows a direct connection to higher level administrative objects. Due to the existence of the general `isPartOf` relation, the representation of an arbitrary number of hierarchy levels is therotically possible (A.2).

In practice, administrative affiliations are represented very differently. Figures 12–14 show similar administrative structures for three different countries, a settlement with four (Poland and Germany) resp. three (Denmark) administrative levels. As you can see, different object properties are used in each case.

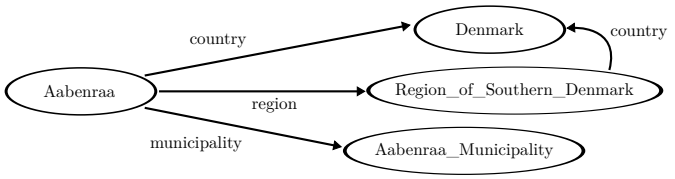


Fig. 12. Example from the DBPedia ontology showing a Danish settlement and its administrative objects.

For the Danish settlement (Figure 12) the object properties `municipality`, `region`, and `country` are used to connect it directly to all higher-level objects. However, the municipality is not connected to any superordinate objects.

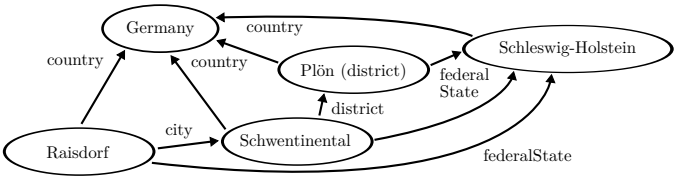


Fig. 13. Example from the DBPedia ontology showing a German settlement and its administrative objects.

The German settlement (Fig. 13) is also directly connected its higher-level objects, although not all of them. Different object properties are used: `city` and `federalState`. The municipality (Schwentinental) is directly connected to the rural district, the federal state and the country.

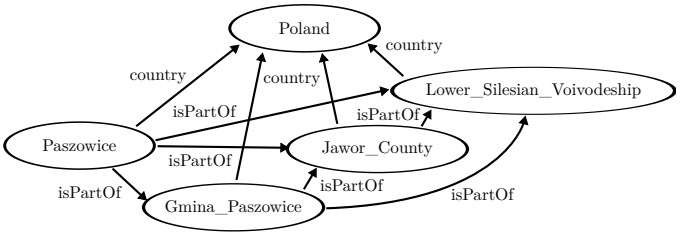


Fig. 14. Example from the DBPedia ontology showing a Polish settlement and its administrative objects.

For the Polish village (Figure 14) a more generic approach

using the isPartOf object property is used. Only for the connection to the country a specialized object property country is used. However, country is not a sub-property of isPartOf. All individuals are directly connected with its higher-level objects.

These different modelling approaches within the same ontology makes it quite difficult to work with the DBPedia ontology.

```
@prefix schema: <http://schema.org/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix virtrdf: <http://www.openlinksw.com/schemas/virtrdf
    #> .
@prefix dbpo: <http://dbpedia.org/ontology/> .
@prefix dbpedia: <http://dbpedia.org/resource/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix georss: <http://www.georss.org/georss/> .

dbpedia:Raisdorf
 a dbpo:Settlement, dbpo:Place,
  dbpo:PopulatedPlace, schema:Place,
  <http://www.opengis.net/gml/_Feature> ;
 foaf:homepage <http://www.raisdorf.de/> ;
 foaf:isPrimaryTopicOf dbpedia:Raisdorf ;
 foaf:name "Raisdorf"@en ;
 georss:point "54.266666666666666 10.216666666666667"@en ;
 geo:geometry "POINT(10.216666 54.266666)"^^virtrdf:
     Geometry ;
 geo:lat "54.2667"^^xsd:float ;
 geo:long "10.2167"^^xsd:float ;
 dbpo:areaCode "04307, 04342"@en ;
 dbpo:areaTotal 1.129000e+7 ;
 dbpo:city dbpedia:Schwentinental ;
 dbpo:country dbpedia:Germany ;
 dbpo:elevation 36 ;
 dbpo:federalState dbpedia:Schleswig-Holstein ;
 dbpo:populationAsOf "2006-09-29+02:00"^^xsd:date ;
 dbpo:populationTotal 7675 ;
 dbpo:postalCode "24223"@en ;
 prov:wasDerivedFrom
  <http://en.wikipedia.org/wiki/Raisdorf?oldid=540674771>
      ;
 rdfs:label "Raisdorf"@nl, "Raisdorf"@en, "Raisdorf"@de .
```

C. GeoNames

GeoNames [19] is a worldwide database containing information for more than 8 million settlements and administrative objects. It is probably the most commonly used gazetteer within the LOD cloud.

As it can be seen in Figure 15 the main class of GeoNames' data model is Feature. Using the properties featureClass and featureCode the type of the settlement or administrative object can be described in a very detailed way. Apart from a number of names, the geographical location can be specified. For this purpose, use is made of data properties from the WGS84 vocabulary [20].

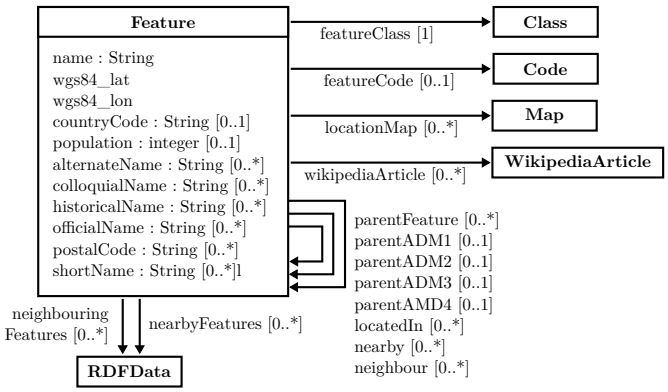


Fig. 15. GeoNames' main class Feature.

The model provides no indication of times (A.3). Dependencies are represented by using the object property parentFeature (C.1). Thus, the representation of an arbitrary number of hierarchy levels is possible (A.2). Additionally, there are five specialized hierarchy levels which are represented by the object property parentCountry, parentAdm1 ...parentAdm4. Therefore, the specification is practically limited to these five hierarchical levels. Source citations are not possible (A.1).

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix gn: <http://www.geonames.org/ontology#> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .

<http://sws.geonames.org/2825253/>
 a gn:Feature ;
 rdfs:isDefinedBy <http://sws.geonames.org/2825253/about.
     rdf> ;
 gn:name "Suchsdorf" ;
 gn:featureClass gn:P ;
 gn:featureCode gn:P.PPLX ;
 gn:countryCode "DE" ;
 geo:lat "54.35937" ;
 geo:long "10.07947" ;
 gn:parentFeature <http://sws.geonames.org/2862623/> ;
 gn:parentCountry <http://sws.geonames.org/2921044/> ;
 gn:parentADM1 <http://sws.geonames.org/2838632/> ;
 gn:nearbyFeatures <http://sws.geonames.org/2825253/nearby
     .rdf> ;
 gn:locationMap <http://www.geonames.org/2825253/suchsdorf
     .html> .
```

D. LinkedGeoData

The aim of LinkedGeoData [21] is to make the information collected in the OpenStreetMap [22] project available as a RDF knowledge base within the LOD cloud. The TBox of the ontology is very large, due to the types taken over from OpenStreetMap. However, the relevant part for the description of administrative structures is limited to the class Place and 16 sub-classes (City, Continent, Country, County, Hamlet, Island, Islet, IsloatedDwelling, Locality, Municipality, Region, State, Subdivision, Suburb, Town, and Village).

In contrast to all other systems presented in this paper, the specification of higher-level administrative objects is provided as literals (C.1) using the data property `isIn`, e.g., `"Kiel,Schleswig-Holstein,Bundesrepublik Deutschland,Europe"`. Therefore, navigation within the administrative hierarchy is not possible.

The model provides no indication of time (A.3). The type of an administrative object is represented via OWL classes (C.2). The specification of names is done by using `rdfs:label` (C.4). Source citations are not possible (A.1).

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix lgd: <http://linkedgeodata.org/ontology/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix geom: <http://geovocab.org/geometry#> .
@prefix data: <http://linkedgeodata.org/triplify/> .

data:node1454999119
  a <http://geovocab.org/spatial#Feature>, lgd:Place,
    <http://linkedgeodata.org/meta/Node>, lgd:Suburb ;
  lgd:changeset "9470245"^^xsd:int ;
  lgd:version "1"^^xsd:int ;
  rdfs:label "Suchsdorf" ;
  geo:lat 5.435505e+1 ;
  geo:long 1.008036e+1 ;
  lgd:isIn "Kiel,Schleswig-Holstein,Bundesrepublik
      Deutschland,Europe" ;
  dc:contributor data:user472256 ;
  dc:modified "2011-10-04T17:16:49"^^xsd:dateTime ;
  geom:geometry <http://linkedgeodata.org/geometry/
      node1454999119> .
```

E. GB Ordnance Survey

Great Britain's national mapping agency publishes information about settlements and administrative objects in Great Britain as linked data. [23][24] In addition to geographical information, the published data also contain information on the administration.

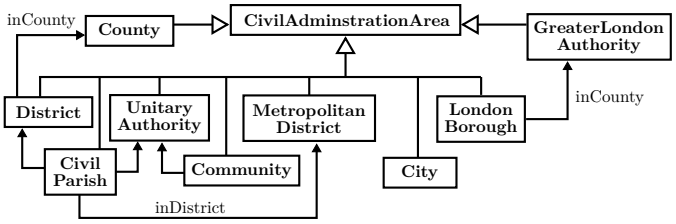


Fig. 16. Excerpt from the GB Ordnance Survey ontology.

Administrative objects are modelled as instances of the nine disjoint sub-classes of `CivilAdministrationArea` shown in Figure 16. The ontology specifies the relations between these classes very rigorously.

Via the `inDistrict`, the `inCounty` and the `inEuropeanRegion` object properties up to four hierarchy levels can be specified (A.2). For the specification of names `rdfs:label` and SKOS data properties are used (C.4). The model provides no indication of time (A.3). Source citations are not possible (A.1).

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix spatial: <http://data.ordnancesurvey.co.uk/ontology
    /spatialrelations/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix admingeo: <http://data.ordnancesurvey.co.uk/
    ontology/admingeo/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix georss: <http://www.georss.org/georss/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix geom: <http://data.ordnancesurvey.co.uk/ontology/
    geometry/> .
@prefix gbos: <http://data.ordnancesurvey.co.uk/id/> .
@prefix gbstat: <http://statistics.data.gov.uk/id/
    statistical-geography/> .

gbos:7000000000015052
  a admingeo:County ;
  spatial:contains gbos:7000000000014797, [...] gbos
      :7000000000014892 ;
  spatial:touches gbos:7000000000023423, [...] gbos
      :7000000000043822 ;
  spatial:northing 325253 ;
  skos:prefLabel "Staffordshire" ;
  admingeo:district gbos:7000000000014727, [...] gbos
      :7000000000014869 ;
  geo:lat 52.8248 ;
  rdfs:label "Staffordshire" ;
  admingeo:hasUnitID "15052" ;
  admingeo:hasAreaCode "CTY" ;
  admingeo:inEuropeanRegion gbos:7000000000041426 ;
  admingeo:gssCode "E10000028" ;
  georss:point "52.824807 -2.006617" ;
  owl:sameAs gbstat:E10000028 ;
  spatial:easting 399651 ;
  geom:extent geom:15052-16 ;
  spatial:within gbos:7000000000041426 ;
  geo:long -2.00662 .
```

F. Spatial Hierarchy Vocabulary

The Spatial Hierarchy Vocabulary [25][26] was created as part of the Leipzig professor catalogue. The structure of the model is very similar to the model of schema.org. Its classes and relations are shown in Figure 17.

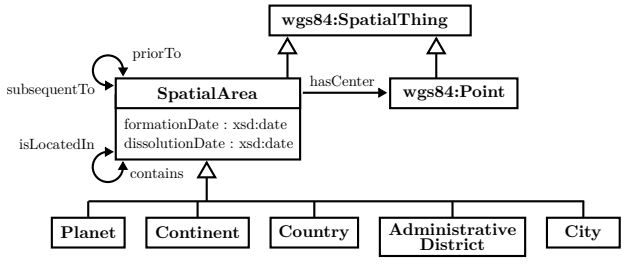


Fig. 17. The classes and relations of the Spatial Hierarchy Vocabulary

Dependencies are represented using the object property `isLocatedIn` and the inverse property `contains` (C.1). Thus, the representation of an arbitrary number of hierarchy levels is theoretically possible. The type of an administrative object is represented via OWL classes. However, since the model contains only five of these classes, the number of hierarchy levels is practically limited to five levels (A.2).

The model provides basic indication of time: It is possible to specify the establishment and the termination date of an object using data properties. Topological relations (B.1) between administrative objects are given by using the inverse object properties `priorTo` and `formationDate`. The specification of names is done by using `rdfs:label` (C.4). Source citations are not possible (A.1).

G. Gemeinsame Normdatei (GND)

Within the "Gemeinsame Normdatei" (GND) the German National Library also publishes information about geographical objects. It lists both civil and ecclesiastical administrative structures. However, only in exceptional cases settlements are associated with them.

Figure 18 shows an excerpt of classes and properties from the GND ontology that are relevant for the modelling of administrative structures. Administrative objects are represented by individuals that are instances of the class `PlaceOrGeographicName` and its sub-classes.

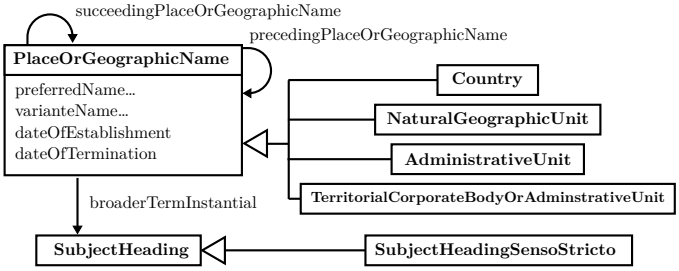


Fig. 18. Excerpt from the GND ontology that deals with administrative information.

The GND provides topological relations (B.1) between administrative objects by using the inverse object properties `succeedingPlaceOrGeographicName` and `preceedingPlaceOfGeographicName`. Additionally, establishment and the termination date of an object can be specified with data properties.

The type of an administrative object is represented by individuals that are members of the class `SubjectHeading` (C.2). They are connected to the individuals representing administrative objects via the object property `broaderTermInstantial`. For the specification of names two data properties have been defined (C.4). In some cases the website contains source citations (A.1). However, they are currently not available as Linked Data.

```
@prefix gndo: <http://d-nb.info/standards/elementset/gnd#>
    .
@prefix gnd: <http://d-nb.info/gnd/> .
@prefix area: <http://d-nb.info/standards/vocab/gnd/
     geographic-area-code#> .

<http://d-nb.info/gnd/4091969-9>
  a gndo:TerritorialCorporateBodyOrAdministrativeUnit,
    gndo:AdministrativeUnit ;
  gndo:gndIdentifier "4091969-9" ;
  gndo:oldAuthorityNumber "(DE-588c)4091969-9" ;
  gndo:broaderTermInstantial gnd:4073976-4 ;
  gndo:geographicAreaCode area:XA-DE-SH ;
  gndo:variantNameForThePlaceOrGeographicName
     "Eutin <Kreis>" ;
  gndo:preferredNameForThePlaceOrGeographicName
     "Landkreis Eutin" ;
  gndo:precedingPlaceOrGeographicName gnd:4115317-0 ;
  gndo:succeedingPlaceOrGeographicName gnd:4044079-5 .
```

H. Suomen Ajallinen Paikka Ontologia (SAPO)

In [27] and [28] Kauppinen, Hyvönen et al. describe how data with time reference is stored in multiple ontologies as time-slices. In case of changes in the administrative structures between the times represented in the ontology, the process of change is modelled with the help of "change bridge classes". Among the systems presented in this article, it is the only application of time-slices (B.3) and the modelling of activities (B.2). Apparently, this modelling approach was used only internally in preparation of the published ontology. In today's publicly accessible version of the "Suomen Ajallinen Paikka Ontologia" (SAPO) [29] these two ideas cannot be found.

For the lifetime of a administrative object, multiple individuals exist (B.4). After every change in the size (area) of the administrative object, a new individual is created. These individuals are combined into so-called "spaceworms". Figure 19 shows an example of such a union.

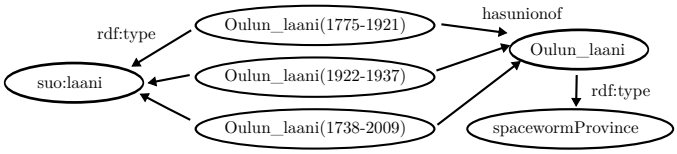


Fig. 19. Example for a "spaceworm" in the SAPO.

Indications of time are specified within `rdfs:label` values only (C.5). Therefore, they are not machine-interpretable and cannot be used for reasoning. There is no specialized data property for the specification of names—also `rdfs:label` is used (C.4). The type of an administrative object is specified by using OWL classes (C.2). The model contains three of these classes (`valtio`, `laani`, `kunta`) and hence three hierarchy levels (A.2). One can suspect that the internal ontologies of SAPO contain only these three classes, too.

```
@prefix dc: <http://purl.org/dc/terms/> .
@prefix sapo: <http://www.yso.fi/onto/sapo/> .
@prefix suo: <http://www.yso.fi/onto/suo/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://www.yso.fi/onto/sapo/Ypaja(1910-)>
   a suo:kunta ;
   dc:isPartOf
     <http://www.yso.fi/onto/sapo/Etela-Suomen_laani
          (1997-2009)>,
     <http://www.yso.fi/onto/sapo/Hameen_laani(1870-1959)
          >,
     <http://www.yso.fi/onto/sapo/Hameen_laani(1969-1972)
          >,
     <http://www.yso.fi/onto/sapo/Hameen_laani(1993-1996)>
          ;
   rdfs:label "Ypäjä(1910-)" .

sapo:Ypaja
   a sapo:spaceworm ;
   rdfs:label "Ypäjä" .
```

I. Genealogisches Orts-Verzeichnis (GOV)

One of the most extensive data models in this survey is provided by "Genealogisches Orts-Verzeichnis" (GOV) [30][31], a project of the German genealogical association "Verein für Computergenealogie". The focus of the dataset is on Central Europe, but also data from the U.S. and Australia is included. In addition to structures of political administration, ecclesiastical and legal administrative structures can be found.

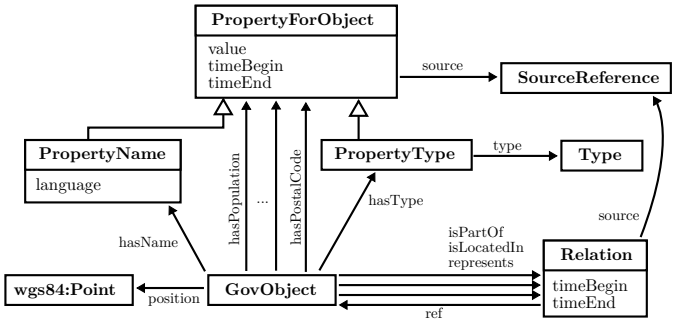


Fig. 20. Excerpt from the classes and properties of the Genealogisches Orts-Verzeichnis. Elements to handle references are left out.

In the GOV, the results of changes (B.2) are modelled. A single individual is used for the entire lifetime of an administrative object (B.4). Therefore, each administrative object is associated with exactly one URI. Both historical affiliations as well as time-dependent values are given (A.3).

Figure 20 shows an excerpt from the GOV ontology. In contrast to most systems presented in this article, reification—encoding of n-ary properties as individuals— (C.3) is used for relations and values to specify indications of time. The reification also allows to give source citations (A.1). These source citations are listed as object properties, not in text form only. That enables queries and reasoning over these source citations.

Dependencies are modelled with object properties `isPartOf`, `isLocatedIn`, and `represents` and the class `Relation` (C.1). Using the general `isPartOf` relation and the class `Relation`, the representation of parallel hierarchies with an arbitrary number of levels is possible (A.2).

Names are specified using the object property `hasName` and the class `PropertyName` (C.4), which has a data property indicating the language as ISO-639-2 code. In combination with the aforementioned reification, it is possible to specify different language variants of the name as well as different names in the same language.

The type of an administrative object is represented by individuals that are connected via the object property `hasType` and the class `PropertyType` (C.2). Again, the reification makes it possible to model type changes with an indication of time and source citations.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix gov: <http://gov.genealogy.net/ontology.owl#> .
@prefix type: <http://gov.genealogy.net/types.owl#> .

<http://gov.genealogy.net/SUCORFJO54AI>
 a gov:GovObject ;
 rdfs:isDefinedBy <http://gov.genealogy.net/SUCORFJO54AI/
      about.rdf> ;
 gov:hasPopulation [ a gov:PropertyForObject ;
    gov:value "8441" ;
    gov:timeBegin "2004-12-31" ;
    gov:timeEnd "2004-12-31" ] ;
 gov:property [ a gov:Property ; gov:value "8.04" ] ;
 gov:hasName [ a gov:PropertyName ;
    gov:value "Suchsdorf" ;
    gov:language "deu" ] ;
 gov:hasType [ a gov:PropertyType ;
    gov:type type:55 ;
    gov:timeEnd "1958" ;
       gov:source [ a gov:SourceReference ;
            gov:sourceRef <http://gov.genealogy.net/
                 source_387809> ;
            gov:note "S. 152" ]
      ],
    [ a gov:PropertyType ; gov:type type:54 ; gov:
        timeBegin "1958" ] ;
 gov:isPartOf
     [ a gov:Relation ;
       gov:ref <http://gov.genealogy.net/object_386988> ;
       gov:timeEnd "1958-03-31" ],
     [ a gov:Relation ; gov:ref <http://gov.genealogy.net/
         KIEIELJO54BI> ;
      gov:timeBegin "1958-04-01" ] ,
     [ a gov:Relation ;
       gov:ref <http://gov.genealogy.net/object_1042608>
            ],
     [ a gov:Relation ;
       gov:ref <http://gov.genealogy.net/object_285109> ;
       gov:source [a gov:SourceReference ;
            gov:sourceRef <http://gov.genealogy.net/
                 source_387809> ;
            gov:note "S. 152" ]
     ] .
```

J. Thesaurus of Geographic Names (TGN)

The second extensive data model in this survey is provided by the "Thesaurus of Geographic Names" [32]. It is a project of the Getty Research Institute. TGN is part of the Getty Vocabulary Program (GVP) [33] ontology. Therefore, most of its properties have a broad domain and range to be compatible with other Getty vocabularies such as the Art & Architecture Thesaurus (AAT). The ontology extensively reuses components from many other ontologies such as SKOS [13], W3C's PROV [34], Dublin Core (DC), schema.org and others.

Due to the aforementioned reuse of ontology components and the generic definitions of domains and ranges of object properties, it is hard to comprehend the key model design criteria for the TGN. We tried our best to compose the various elements from the different ontologies and observed the domain and range restrictions found in the data to make the structure visible. Figure 21 shows this structure.

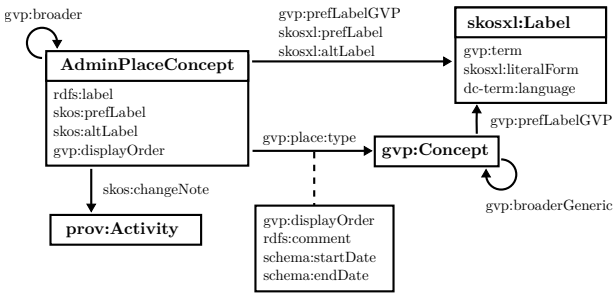


Fig. 21. Structure of an individual in the TGN.

In the TGN, the results of changes (B.2) are modelled. A single individual is used for the entire lifetime of an administrative object (B.4). Since the TGN uses a dual practice to separate between Concept and Thing [33, p. 48] the ontology also contains a second individual for each administrative object. A user of the TGN has to distinguish carefully, which individual (respective URI) to use. However, a discussion on the "Concept vs Place Duality" is beyond the scope of this article.

The TGN uses two different kinds of reification (C.3). Place names are modelled as encoded n-ary relations resulting in an individual of the class `skosxl:Label` (lines 42–48 in the example). TGN also uses RDF's higher-order statements for making statements about other RDF statements. Types of administrative objects and dependencies are mostly modelled that way in the TGN (lines 50–54 in the example). Source citations, indications of time and display orders can be added to these named statements (line 54 in the example).

Dependencies are modelled with several object properties, which are sub-properties of each other (C.1). The most generic ones are `gvp:broader` and `skos:broader`. Using the general `gvp:broader` relation, the representation of parallel hierarchies with an arbitrary number of levels is theoretically possible (A.2). However, in practice—at least for central Europe—only a few fragmentary administrative affiliations are listed. In addition to the object properties one hierarchy of administrative objects is given as data property `gvp:parentString` (line 24 in the example). If a change at any position of the hierarchy structure occurs, the value of this data property has to be changed for all subordinate objects. There is a risk that this will be forgotten and the ontology will contain contradicting information.

The type of an administrative object is represented by individuals (C.2) that are connected via the object property `gvp:placeType` and sub-properties (lines 31 and 50–54 in the example). These properties links to an individual from the AAT. These individuals are put into a separate complex type hierarchy via object properties).

Names are specified using the object properties `gvp:prefLabelGVP` and `skosxl:prefLabel` and the class `skosxl:Label` (lines 42–48 in the example). The language of the name is indicated in two ways, a RDF language tag at the Label's data property `gvp:term` and an object property from DC `dcterms:language`. In combination with the aforementioned reification, it is possible to specify different language variants of the name as well as different names in the same language.

One irritating fact in the context of names is the custom data property `displayOrder` (line 26 in the example). According to [33, p. 31] one can use this property to sort the places alphabetically. However, if a new place is added to the ontology the property values of all individuals following in alphabetical order have be to changed.

Source citations (A.1) as well as revision history play an important role in TGN and are modelled extensively (lines 37–40 in the example). Source citations are listed as individuals, not in text form only. That enables queries and reasoning over these source citations.

The TGN renders outstanding services to the modelling of provenance. However, the introduction of a semantic beyond OWL-2 is problematic.

```
1  @prefix dc: <http://purl.org/dc/elements/1.1/> .
2  @prefix dc-term: <http://purl.org/dc/terms/> .
3  @prefix foaf: <http://xmlns.com/foaf/0.1/> .
4  @prefix gvp: <http://vocab.getty.edu/ontology#> .
5  @prefix prov: <http://www.w3.org/ns/prov#> .
6  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
       .
7  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
8  @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
9  @prefix skosxl: <http://www.w3.org/2008/05/skos-xl#> .
10 @prefix tgn: <http://vocab.getty.edu/tgn/> .
11 @prefix tgn_rel: <http://vocab.getty.edu/tgn/rel/> .
12 @prefix tgn_rev: <http://vocab.getty.edu/tgn/rev/> .
13 @prefix tgn_term: <http://vocab.getty.edu/tgn/term/> .

15 tgn:7074121 a skos:Concept, gvp:Subject,
16             gvp:AdminPlaceConcept ;
17       rdfs:label "Suchsdorf" ;
18       rdfs:seeAlso <http://www.getty.edu/vow/TGNFullDisplay
             ?find=&place=&nation=&subjectid=7074121> ;
19       dc-term:contributor tgn_contrib:10000000 ;
20       skos:inScheme <http://vocab.getty.edu/tgn/> ;
21       skos:changeNote tgn_rev:5011369370 , tgn_rev
             :5011369369 , tgn_rev:5019010426 , tgn_rev
             :5019010427 ;
22       dc-term:created "2010-10-16T13:23:22"^^xsd:dateTime ;
```

```
23        dc-term:modified "2010-10-16T13:23:22"^^xsd:dateTime
              , "2014-06-04T11:15:03"^^xsd:dateTime ;
24        gvp:parentString "Schleswig-Holstein, Deutschland,
              Europe, World" ;
25        gvp:parentStringAbbrev "Schleswig-Holstein,
              Deutschland, ... World" ;
26        gvp:displayOrder "7314"^^xsd:positiveInteger ;
27        gvp:broaderPartitive tgn:7003688 ;
28        gvp:broaderPreferred tgn:7003688 ;
29        gvp:prefLabelGVP tgn_term:1001287042 ;
30        dc-term:source tgn_source:2009007144-subject-7074121
              ;
31        gvp:placeTypePreferred aat:300000745 ;
32        dc:identifier "7074121" ;
33        dc-term:issued "2014-06-04T11:15:03"^^xsd:dateTime ;
34        prov:wasGeneratedBy tgn_rev:5011369370 ;
35        foaf:focus tgn:7074121-place .

37  tgn_rev:5011369370 a prov:Activity , prov:Create ;
38        dc:type "created" ;
39        prov:startedAtTime "2010-10-16"^^xsd:date;
40        dc:description "new:Suchsdorf" .

42  tgn_term:1001287042 a skosxl:Label ;
43        skosxl:literalForm "Suchsdorf" ;
44        gvp:term "Suchsdorf" ;
45        gvp:displayOrder "1"^^xsd:positiveInteger ;
46        gvp:termFlag <http://vocab.getty.edu/term/flag/
              Vernacular> ;
47        gvp:termPOS <http://vocab.getty.edu/term/POS/Noun> ;
48        dc:identifier "1001287042" .

50  tgn_rel:7074121-placeType-300000745 a rdf:Statement ;
51        rdf:subject tgn:7074121 ;
52        rdf:predicate gvp:placeTypePreferred ;
53        rdf:object aat:300000745 ;
54        gvp:displayOrder "1"^^xsd:positiveInteger .
```

VI. Conclusion

There is a variety of different approaches on how to model historical administrative structures on the Semantic Web. In this article, ten systems were presented with their different approaches. In order to assess the differences better, a classification of systems has been developed which is divided into three main groups, each with different characteristics.

What characteristics make a system universal and future-proof? For some characteristics, this can be clearly stated. Especially for places from historical sources, time-dependent values are essential. An unlimited number of parallel affiliations enables the representation of the complex historical reality of administrative structures. The representation of the administrative structure should be done by using object properties—otherwise the key factor of unique identification will be lost. Indication of time should be modelled as separate properties to make them machine-interpretable.

Both, source citations and revision history allow quality control—usually difficult on the Semantic Web.

To provide information on time, sources, and the language used, the use of reification seems inevitable in the current state of RDF and OWL. Reification in the sense of making statements about other RDF statements is problematic because OWL-2 DL does not support these "higher-order statements" and a formal modelling is difficult using `rdf:Statement` individuals.

The use of time-slices is poorly scalable. However, during a preparation phase in the processing of sources (e.g., topographies which relate to a specific date), they might be useful. It is not future-proof to create a new individual for every point in time: There will be an unmanageable number of individuals (e.g., an annual addition of population numbers). The correct referencing of a single administrative object becomes difficult—but not impossible if additional union-individuals are defined.

Currently, the modelling of results instead of activities is dominant. However, modelling activities could provide a better way to represent complex administrative processes. Particularly for processes that affect a multitude of objects, the correlation of the changes will be better understood.

References

[1] J. Zedlitz and N. Luttenberger, "Modelling (historical) administrative information on the semantic web," in *WEB 2014, The Second International Conference on Building and Exploring Web Based Environments*, 2014, pp. 33–39.

[2] (2014) RDF 1.1 Concepts and Abstract Syntax. (last access: 2014-11-27). [Online]. Available: http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/

[3] (2012) OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. (last access: 2014-11-27). [Online]. Available: http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/

[4] J. R. Hobbs and F. Pan, "An ontology of time for the semantic web," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 3, no. 1, pp. 66–85, 2004.

[5] C. Gutierrez, C. A. Hurtado, and A. Vaisman, "Introducing time into rdf," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 19, no. 2, pp. 207–218, 2007.

[6] F. Frasincar, V. Milea, and U. Kaymak, "towl: Integrating time in owl," in *Semantic Web Information Management*. Springer, 2010, pp. 225–246.

[7] B. Motik, "Representing and querying validity time in rdf and owl: A logic-based approach," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 12, pp. 3–21, 2012.

[8] A. Rula, M. Palmonari, A. Harth, S. Stadtmüller, and A. Maurino, "On the diversity and availability of temporal information in linked open data," in *International Semantic Web Conference*, 2012, pp. 492–507.

[9] J. J. Carroll, C. Bizer, P. Hayes, and P. Stickler, "Named graphs, provenance and trust," in *Proceedings of the 14th international conference on World Wide Web*. ACM, 2005, pp. 613–622. [Online]. Available: http://www2005.org/docs/p613.pdf

[10] M. Homola and L. Serafini, "Contextualized knowledge repositories for the semantic web," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 12, no. 0, 2012. [Online]. Available: http://www.websemanticsjournal.org/index.php/ps/article/view/231

[11] P. Severi, J. Fiadeiro, and D. Ekserdjian, "Guiding reification in owl through aggregation," in *23rd International Workshop on Description Logics DL2010*, 2010, p. 408.

[12] M. Davis, "Tags for the Identification of Language (BCP47)," 2009.

[13] (2009) SKOS Simple Knowledge Organization System Reference. (last access: 2014-11-27). [Online]. Available: http://www.w3.org/TR/2009/REC-skos-reference-20090818/

[14] (2014) RDF 1.1 Turtle: Terse RDF Triple Language. (last access: 2014-11-27). [Online]. Available: http://www.w3.org/TR/2014/REC-turtle-20140225/

[15] (2014) The schema.org website. (last access: 2014-11-27). [Online]. Available: http://schema.org

[16] (2014) The dbpedia.org website. (last access: 2014-11-27). [Online]. Available: http://www.dbpedia.org

[17] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "Dbpedia-a crystallization point for the web of data," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, no. 3, pp. 154–165, 2009.

[18] (2014) DBpedia Ontology version 2014. (last access: 2014-11-27). [Online]. Available: http://downloads.dbpedia.org/2014/dbpedia_2014.owl

[19] (2014) The GeoNames Website. (last access: 2014-11-27). [Online]. Available: http://www.geonames.org/

[20] (2009) WGS84 Geo Positioning: an RDF vocabulary. (last access: 2014-11-27). [Online]. Available: http://www.w3.org/2003/01/geo/wgs84_pos

[21] (2014) The LinkedGeoData Website. (last access: 2014-11-27). [Online]. Available: http://linkedgeodata.org

[22] (2014) The OpenStreetMap Website. (last access: 2014-11-27). [Online]. Available: http://www.openstreetmap.org

[23] (2014) Ordnance Survey Linked Data. (last access: 2014-11-27). [Online]. Available: http://data.ordnancesurvey.co.uk/datasets/os-linked-data

[24] J. Goodwin, C. Dolbear, and G. Hart, "Geographical linked data: The administrative geography of great britain on the semantic web," *Transactions in GIS*, vol. 12, no. s1, pp. 19–30, 2008.

[25] M. Martin and T. Riechert, "Ortsbezogene navigation basierend auf einem vokabular zur erzeugung geographischer hierarchien," in *Catalogus Professorum Lipsiensis*. Universität Leipzig, 2010, pp. 107–116.

[26] T. Riechert, U. Morgenstern, S. Auer, S. Tramp, and M. Martin, "Knowledge engineering for historians on the example of the catalogus professorum lipsiensis," in *The Semantic Web–ISWC 2010*. Springer, 2010, pp. 225–240.

[27] T. Kauppinen, R. Henriksson, J. Väätäinen, C. Deichstetter, and E. Hyvönen, "Ontology-based modeling and visualization of cultural spatio-temporal knowledge," in *Developments in Artificial Intelligence and the Semantic Web-Proceedings of the 12th Finnish AI Conference STeP 2006*, 2006, pp. 26–27.

[28] T. Kauppinen and E. Hyvönen, "Modeling and reasoning about changes in ontology time series," in *Ontologies*. Springer, 2007, pp. 319–338.

[29] (2014) The Finnish Spatio-temporal Ontology. (last access: 2014-11-27). [Online]. Available: http://onki.fi/en/browser/search/sapo?&os=sapo

[30] (2014) The GOV Website. (last access: 2014-11-27). [Online]. Available: http://gov.genealogy.net

[31] J. Zedlitz and T. Kluttig, "Das Genealogische Ortsverzeichsnis (GOV)," *Der Archivar*, vol. 3/2014, pp. 289–292, 2014.

[32] (2014) Getty Thesaurus of Geographic Names. (last access: 2014-11-27). [Online]. Available: http://www.getty.edu/research/tools/vocabularies/tgn/

[33] (2014) Getty Vocabularies: Linked Open Data – Semantic Representation. (last access: 2014-11-27). [Online]. Available: http://vocab.getty.edu/doc/gvp-lod.pdf

[34] (2013) PROV-Overview: An Overview of the PROV Family of Documents. (last access: 2014-11-27). [Online]. Available: http://www.w3.org/TR/2013/NOTE-prov-overview-20130430/

Efficient and Adaptive Web-native Live Video Streaming

Luigi Lo Iacono* and Silvia Santano Guillén†
*Cologne University of Applied Sciences, Cologne, Germany
Email: luigi.lo_iacono@fh-koeln.de
†G&L Geißendörfer & Leschinsky GmbH, Cologne, Germany
Email: silvia.santano@gl-systemhaus.de

***Abstract*—The usage of the Web has experienced a vertiginous growth in the last few years. Watching video online has been one major driving force for this growth lately. Until the appearance of the HTML5 agglomerate of (still draft) specifications, the access and consumption of multimedia content in the Web has not been standardized. Hence, the use of proprietary Web browser plugins flourished as intermediate solution. With the introduction of the HTML5 VideoElement, Web browser plugins are replaced with a standardized alternative. Still, HTML5 Video is currently limited in many respects, including the access to only file-based media. This paper investigates on approaches to develop video live streaming solutions based on available Web standards. Besides a pull-based design based on HTTP, a push-based architecture is introduced, making use of the WebSocket protocol being part of the HTML5 standards family as well. The evaluation results of both conceptual principles emphasize, that push-based approaches have a higher potential of providing resource and cost efficient solutions as their pull-based counterparts. In addition, initial approaches to instrument the proposed push-based architecture with adaptiveness to network conditions have been developed.**

Keywords-HTML5, Video, Live Streaming, DASH, WebSockets, Adaptive Streaming

I. INTRODUCTION

The access of video content in the Web is evolving rapidly, as the internet traffic increases, with live video streaming becoming web-native as well [1]. According to the Cisco Visual Networking Index Global Forecast and Service Adoption for 2013 to 2018 [2], consumer Internet traffic has increased enormously on the last years and the expectations are that this trend continues growing due to more users and devices, faster broadband speeds and more video viewing. Already today, monthly web traffic is at 62 exabytes a month, meaning a 62% of the whole traffic. The forecast includes concrete numbers on Internet traffic predictions, indicating that the annual global IP traffic will surpass the zettabyte (1000 exabytes) threshold in 2016. In other words, it will reach 91.3 exabytes (one billion gigabytes) per month in 2016, going up to 131.6 exabytes per month in 2018. The latter figure means that it would take an individual over 5 million years to watch the amount of video that will be crossing global IP Networks in one month in 2018.

The increase in Internet usage is mainly driven by online video consumption, for which the expected proportion of all consumer Internet traffic will account 79% in 2018, up from 66% in 2013. Moreover, adding related forms of video distribution to this calculation such as video on demand and video exchanged through peer-to-peer file sharing would mean between 80 to 90% of global consumer traffic. Regarding to the number of users online video will then be the fastest growing Internet service with a Compound Annual Growth Rate (CAGR) of 10 percent from 2013 to 2018, growing from 1.2 billion users to 1.9 billion users by 2018, as shown in Figure 1.

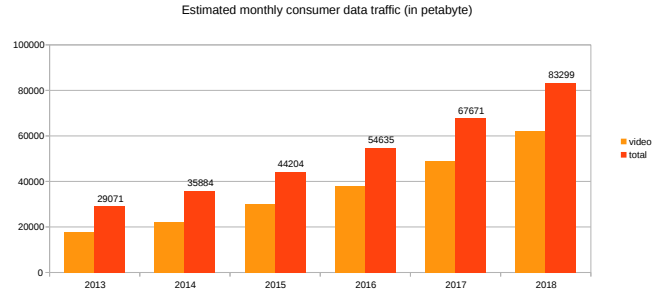


Figure 1: Estimated monthly consumer video traffic and total data traffic on the Internet [2]

In the early days of the Internet, video content has been delivered by specific streaming protocols such as Real-Time Protocol (RTP) [3] or Real-Time Streaming Protocol (RTSP) [4] in conjunction with specialized server-side software to handle the stream. These protocols break up the streams – it can be more than one, such as a video and multiple audio channels – into very small chunks and send them from the server to the client. This method is also denoted as *push-based delivery*.

Such streaming protocols suffered, however, from unfavourable firewall configurations restricting in many cases the access to media data. HTTP progressive download [5] has been developed partially to overcome this issue and to get multimedia streams past firewalls. The basic concept behind HTTP progressive download is to play back the media content while the resource is being downloaded from the Web server. This approach is also known as *pull-based delivery*, since the file containing the media data needs to be pulled from the server by a client's request.

While capable of reliably finding the path from a requesting client to a responding Web server, HTTP progressive download still not offers true streaming capabilities. This lack motivated

the introduction of methods for adaptive streaming over HTTP. To provide a streaming behaviour, adaptive streaming over HTTP segments the media stream into small, easy-to-download chunks. The adaptiveness is realized by encoding the media content at multiple distinct bitrates and resolutions, creating chunks of the same media fragment in different qualities and sizes. The available encodings enable the client to choose between the provided quality levels and then adapt to lower or higher definition automatically as network conditions keep changing. In order to inform the client about the offered video quality levels and the corresponding names of the resources, a meta file containing this information is provided by the server. The client then chooses a suitable quality level and starts requesting the corresponding chunks in order and with the file named specified in the meta file. This pull of media data needs to be performed by the client in a continuous manner in order to construct an enduring stream out of the obtained chunks. In an equivalent fashion an updated version of the meta file needs to be requested periodically as well, so that the client retrieves information on upcoming chunks to request.

The arena of technologies for adaptive streaming over HTTP has been dominated by proprietary vendor-proposed solutions, as will be discussed in the subsequent Section II. To harmonize the scattered picture a standardized approach known as MPEG Dynamic Adaptive Streaming over HTTP (DASH) has been ratified in December 2011 and published by the International Organization for Standards (ISO) in April 2012 [6]. Although, adaptive streaming over HTTP has been standardized lately and largely build upon Web standards, the play back still requires propietary extensions to be included into the Web browsers. Thus, from a perspective of a live video streaming that is native to the Web, the following set of requirements can be defined as necessary foundation:

- *Live content support*
  Delivering live content by the concept of chunk-based distribution.
- *Web-native*
  Building solely upon Web standards, so that no additional components are needed to develop and use the streaming services (e.g., by being HTML5-compliant on the client-side).
- *Minimal meta data exchange*
  Avoiding of extra message exchanges required for media stream control (e.g., by the adoption of communication patterns following the push model instead of the pull model).
- *Low protocol and processing overhead*
  Reducing overheads introduced by communication and processing means.

In the following section, available technologies will be reviewed in the light of these requirements. After that, in Section III, the proposed approach of a Web-native live Video Streaming will be introduced in terms of an architecture. Section IV then introduces an implementation of the introduced architecture, which serves as foundation for building various evaluation test beds as described in Section V. Finally, a detailed discussion of the evaluation results obtained from performed test runs will conclude the contribution of the present paper.

II. Related work

Microsoft Smooth Streaming (MSS) [7] has been one of the first adaptive media streaming over HTTP announced in October 2008 as part of the Silverlight [8] architecture. MSS is an extension for the Microsoft HTTP server IIS (Internet Information Server) [9] that enables HTTP media streaming of H.264 [10] video and AAC [11] audio to Silverlight and other clients. Smooth Streaming has all typical characteristics of adaptive streaming. The video content is segmented into small chunks that are delivered over HTTP. As transport format of the chunks, MSS uses fragmented ISO MPEG-4 [10] files. To address the unique chunks Smooth Streaming uses time codes in the requests and thus the client does not have to repeatedly download a meta file containing the file names of the chunks. This minimizes the number of meta file downloads that in turn allows to have small chunk durations of five seconds and less. This approach introduces, however, additional processing costs on the server-side for translating URL requests into byte-range offsets within the MPEG-4 file.

Apple's HTTP Live Streaming (HLS) [12] came next as a proposed standard to the Internet Engineering Task Force (IETF). As MSS it enables adaptive media streaming of H.264 video and AAC audio. At the beginning of a session, the HLS client downloads a play list containing the meta data for the available media streams, which use MPEG-2 TS (Transport Stream) [13] as wire format. This meta data document will be repeatedly downloaded, every time a chunk is played back. The media content is embedded into a Web page using the HTML5 VideoElement [14], whose source is the m3u8 manifest file [15], so that both the parsing of the manifest and the download of the chunks are handled by the browser. Due to the periodic retrieval of the manifest file, there exists a lower bound for the minimal duration of the chunks, which is commonly about ten seconds. A drawback of HLS is the current lack of client platforms support, with an availability at the moment mostly restricted to iOS devices and desktop computers with Mac OS X [16]. At the moment, the support for Android Operative System is only available on a few Android devices running Android 4.x and above, although still presenting several inconsistencies and difficulties. Moreover, for desktop no other web browser than Safari has native support for HLS and specific player plugins are therefore necessary in other browsers to play the streams back [17].

With the announcement of HTTP Dynamic Streaming (HDS) [18] Adobe entered the adaptive streaming arena in late 2009. Like MSS and HLS, HDS breaks up video content into small chunks and delivers them over HTTP. The client downloads a manifest file in binary format, the Flash Media Manifest (F4M) [19], at the beginning of the session and periodically during its life time. As in MSS, segments are encoded as fragmented MP4 files that contain both audio and video information in one file. It, however, differs from MSS with respect to the use a single metadata file from which the MPEG file container fragments are determined and then

delivered. In this respect, HDS follows the principle used in HLS instead, which requests and transmits individual chunks via a unique name.

These three major adaptive streaming protocols have much in common. Most importantly, all three streaming platforms use HTTP streaming for their underlying delivery method, relying on standard HTTP Web servers instead of specialized streaming servers. They all use a combination of encoded media files and manifest files that identify the main and alternative streams and their respective URLs for the player. And their respective players all monitor either buffer status or CPU utilization and switch streams as necessary, locating the alternative streams from the URLs specified in the manifest. The overarching problem with MSS, HLS and HDS is that these three different streaming protocols, while quite similar to each other in many ways, are different enough not to be technically compatible. Indeed, each of the three proprietary commercial platforms is a closed system with its own type of manifest format, content formats, encryption methods and streaming protocols, making it impossible for them to work together.

As introduced in Section I, it is a well-known fact that the consumption of video on the Web is growing every day and, moreover, consumers are moving from desktop computers to smartphones, tablets, and other mobile devices to watch video. All these devices present huge differences in compatibility. Despite that, the same experience is expected on all of them, maintaining the high quality and availability. To enable the delivery of video to any platform, a number of streaming protocols and different applications have to be supported. The situation would change greatly if it was possible to have a single distribution method and a single cross-platform client application. On the other hand, removing the requirement of installing plugins on the client side removes a significant obstacle for many users. Furthermore, for cross-platform compatibility, security and stability, many browser vendors have already decided they are not supporting plugins in the future. All those reasons have as a consequence the intention of avoiding solutions that involve plugins and opt for a Web-browser-native approach.

Recognizing this need for a universal standard for the delivery of adaptive streaming media over HTTP, the MPEG standardisation group decided to step into. MPEG DASH (Dynamic Adaptive Streaming over HTTP) [6] is an international standard for HTTP streaming of multimedia content that allows standard-based clients to retrieve content from any standard-based server. It offers the advantage that it can be deployed using standard Web servers. Its principle is to provide formats that enable efficient and high-quality delivery of streaming services over the Internet to provide very high user-experience (low start-up, no rebuffering, trick modes). To accomplish this, it proposes the reuse of existing technologies (containers, codecs, DRM, etc.) and the deployment on top of Content Distribution Networks (CDN). It specifies the use of either MPEG-4 or MPEG-2 TS chunks and an XML manifest file, the so-called Media Presentation Description (MPD), that is repeatedly downloaded to the client making it aware of which chunks are available.

	Support Live Streaming	Use HTML5 video element	Push delivery	Low overhead
HDS	✓	✗	✗	✗
HLS	✓	✓	✗	✗
MSS	✓	✗	✗	✗
DASH	✓	✓	✗	✗

Figure 2: Characteristics of HTTP-based adaptive live streaming platforms

Although the DASH standard may become the format of choice in the future, there is a lack of native Web browser integration. Initial steps towards a browser-native integration of MPEG-DASH using the HTML5 VideoElement have been undertaken. In these implementations, the browser is in charge of parsing, decoding and rendering the media data while, traditionally, applications like Adobe Flash or Microsoft Silverlight have been used for these features, in form of plugins. With JavaScript most of them can be achieved but still remains as a pitfall how to feed media data to the HTML5 VideoElement, a core issue to enable live streaming and adaptive streaming, where the source to play is a sequence of chunks and therefore, the 'src' parameter needs to be updated accordingly.

On this research several different solutions have been investigated in order to overcome the problem of the integration of DASH with HTML5 VideoElement, described in detail in Section IV.

The DASH-JS [20] project from the University of Klagenfurt introduces an approach to overcome this lack of native Web browser support. It proposes a seamless integration of the DASH standard into Web browsers using the HTML5 VideoElement and the MediaSource extensions [21]. The MediaSource extensions enable a seamless playback of a chunk-based stream, by defining a JavaScript API, which allows media streams to be constructed dynamically. They are still a W3C working draft and are currently supported by some of the major browsers: Chrome, Firefox Internet Explorer, Safari, Windows Phone 8.1 and Chrome for Android. This API solves the problem simplifying the process, taking care of the playing as the segments are downloaded and creating a sequence that is played back by feeding it chunk-wise into the HTML5 VideoElement. The media segments are downloaded and appended to a MediaSource buffer and this sequence will then be used as source for the HTML5 VideoElement. The first of the chunks consists of initialization data, which has to be appended to the buffer on the first place. After this initialization data has been loaded, the media segments will be retrieved and played back in the required sequence. Listing 1 shows the major parts of the MediaSource API and their usage in order to construct a steady media stream constructed out of the media fragments downloaded from the Web server via continuous

requests issued by the XMLHttpRequest (XHR) [22] API. The second part is repeated as long as the session is open.

Figure 2 summarizes the characteristics of the discussed adaptive live streaming platforms over HTTP in the light of the requirements defined for a Web-native live video streaming. As can be observed, none of the currently available platforms covers all of these characteristics, leaving space for further research and development.

Listing 1: Usage of the MediaSource API for fragmented media access.

```
//URL of next chunk
var url = (...);

var mediaSource = new MediaSource();
var video = document.querySelector('video');
video.src =
   window.URL.createObjectURL(mediaSource);

(...)

var sourceBuffer =
   mediaSource.addSourceBuffer(
   'video/mp4; codecs="avc1.42c00d"');
var xhr = new XMLHttpRequest();
xhr.open('GET', url, true);
xhr.responseType = 'arraybuffer';
xhr.onload = function(e)
{
   data = new Uint8Array(this.response);
   if (data == "false")
      mediaSource.endOfStream();
   else
      videoSource.appendBuffer(data);
};

xhr.send();
```

III. ARCHITECTURE

The basic idea of the proposed architecture is to ground the live streaming approach on a distinct communication protocol other than HTTP, which is still native to the Web but allows for a different communications design.

The WebSocket protocol was standardized by the IETF as RFC 6455 in 2011 [23]. As a flanking W3C standard, the WebSocket JavaScript API [24] provides an entirely event-driven interface for browser applications to use the WebSocket protocol. WebSockets are supported by all major browsers such as Chrome, Internet Explorer, Firefox, Safari and Opera in their desktop as well as mobile occurrence, as shown in Figure 3.

The protocol operates on top of a standard TCP socket and offers a bidirectional communication channel between a Web browser and a WebSocket server. The WebSocket is established by a HTTP-based opening handshake commonly operated on port 80, which preserves firewall-friendliness.

The code running on the browser side acts as client while there must be a server program running waiting for connections, usually installed on a Web server.

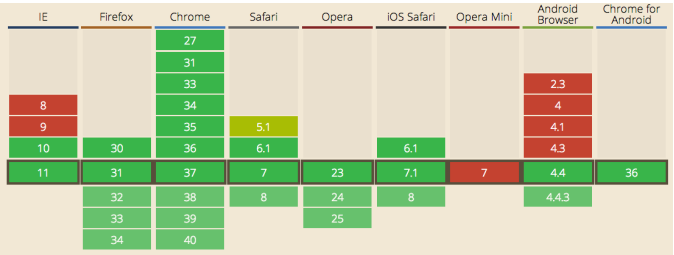


Figure 3: WebSocket browser support [25]

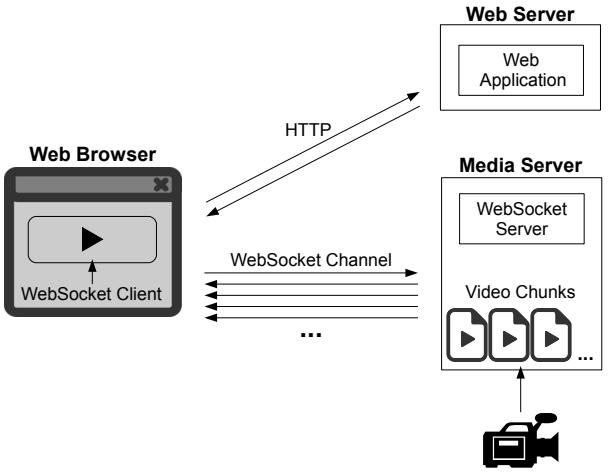


Figure 4: Proposed push-based Web-native live Video Streaming architecture

Figure 4 illustrates the architecture of the developed system, where the two different communication protocols used are represented, as well as a sample of the message exchange.

The communication between the Web browser and the Web server will be the first to be executed, as for every Website, via HTTP. After the Web browser has downloaded the Website, the JavaScript code on the Web Application will attempt to start the communication via WebSocket with the media server.

The communication between client and media server starts with a two-way handshake, as can be seen in Figure 4, before the actual data transmission. The way the data transmission between the two parts takes place, facilitates its use for live content and real-time applications. This is achieved by enabling the server to send content without the need of the client asking first for it, creating a real bidirectional connection that remains open for both parts to send data at any time.

The fact of being able to follow a push model is the core principle of this architecture, where a lot of real-time data needs to be sent, and will be sent from the server periodically, as soon as it is available instead of using a request-response procedure.

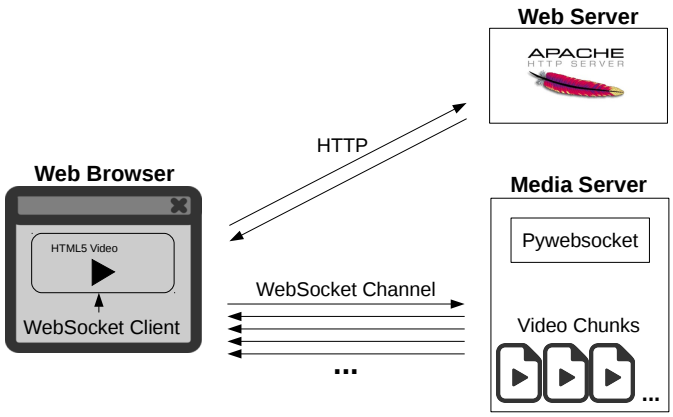


Figure 5: Prototype implementation of the proposed push-based Web-native live Video Streaming architecture

IV. Implementation

A prototype implementation of the proposed pushed-based architecture has been developed as a foundation for analysing the properties of the introduced approach. The technologies and components used for developing the prototype are depicted in Figure 5.

The initial Web page is delivered by an HTTP server and contains a JavaScript program, which gets downloaded by the browser. The browser will also be in charge of rendering the video while it is being downloaded.

As introduced at the beginning of this paper, for this research several different solutions have been investigated in order to overcome the problem of the integration of DASH with the HTML5 VideoElement. The trivial approach of updating the source of the VideoElement when the previous chunk has come to an end gives as a result quite a noticeable gap between sources for the user watching. Another approach, from which a better seamless switch could be expected, utilizes two VideoElements at the same position. This can be understood as one on top of the other in a certain way, letting at every moment just one of them to be visible. In this case, the next chunk to play would always be loaded on the hidden element in advance, carrying out the switch at the exact end of the playing chunk. The behaviour this technique produces is in fact very similar to the previous approach, showing gaps between the chunks to play for a short time during the switch.

Furthermore, the use of some publicly available APIs to reproduce a list of files using the VideoElement, such as SeamlessLoop 2.0 for JavaScript [26] and an own version of it replacing the audio by a VideoElement has also been considered, without success solving the problem.

As a consequence, from all options investigated the only remaining possibility to overcome this issue at the moment of implementing this system is to use the already mentioned MediaSource API for the implementation.

Afterwards, the JavaScript code, which will be executed on the client after downloading, creates an HTML5 VideoElement object and a MediaSource object and connects them using the API. This API allows the construction of media stream objects for the HTML5 VideoElement through which the media segments can be passed to the HTML5 VideoElement for play back. Thus, the decoding and rendering parts will be natively handled by the browser.

In what follows, the client needs to create the WebSocket connection and to assign the according event listeners to specific functions waiting for the next content chunks to arrive so that they can be added to the corresponding MediaSource buffer. This will be performed until the end of the session, which is reached either when the server has no more content to deliver or when the user decides to stop watching.

The WebSocket server application is implemented in Python language, using Pywebsocket [27], an extension for the Apache HTTP Server. This API makes possible to develop a server for the test, which resulted consuming very low RAM memory even for a large amount of clients connected, which is actually translated to a large amount of threads for the operative system. Just like most server applications, it does not start connections by itself but waits for connection requests. After the establishment, the client applications emit a starting signal, with which the video session begins and remains open as long as there is more content available.

V. Evaluation

To evaluate the proposed approach two distinct test beds have been implemented. Test bed A (browser-based, in JavaScript) is targeting the amount of metadata, i.e., data not part of the video, required to be exchanged between client and server. Test bed A (not browser-based, in Python) is concerned with the processing overhead on the server-side and the number of simultaneous clients servable from one server instance. These test beds have been realized for both a DASH-like HTTP transfer and the proposed WebSocket-based approach.

To perform the first evaluation (Test bed A), two browser-based clients have been developed: a version over HTTP, which avails itself of Apache HTTP server and another one over WebSocket, which, after establishing the connection, connects to our WebSocket server application.

To perform the second evaluation (Test bed B), all components have been implemented in Python language. For the clients, the modules used are websocket-client [28] and httplib [29], respectively. The server-side of the HTTP approach is programmed on top of the HTTP protocol implementations provided by the Python modules BaseHTTPServer [30] and SocketServer [31]. Based on these components, the implementation of a multi-threaded HTTP and WebSocket server has been undertaken. The server-side of the WebSocket approach is the same described in previous section.

The video used to perform the evaluation is the open source movie *Big Buck Bunny* [32], which was produced by the Blender Foundation and has been released under Creative

Commons License Attribution 3.0 [33]. The AVC codec is used in an MP4 container. The test video's bitrate is 100 kbps, the duration is 9' 56" and the total file size is 6.7 MB (6,656,763 bytes).

To simulate a live stream, the movie has been chunked into separate segment files according to the MP4 standard. These segments contain each a short portion of two seconds of duration and are stored in the media server. Since the chunk length is approximately two seconds, the number of chunks is 300.

A. *Communication overhead*

To gather the overhead introduced by each one of the two investigated communication alternatives, the network traffic has been captured, analysed and contrasted with theoretical thoughts. The network packets exchanged in both scenarios have been captured using Wireshark [34].

Each layer of the TCP/IP model introduces its own metadata in form of a header and in some cases even a trailer, but since Ethernet, IP and TCP are common to both compared approaches, only the protocol elements of the application-level are taken into account, which are the HTTP messages and the WebSocket frames, respectively.

```
⊞ Transmission Control Protocol, Src Port | Len: 440
⊟ Hypertext Transfer Protocol             |
  ⊞ GET /videos/bunny/bunny4.m4s HTTP/1.1 |
    Host: ec2-54-228-4-210.eu-west-1.comp |
    Connection: keep-alive\r\n            |
    User-Agent: Mozilla/5.0 (X11; Linux x | 8.0.1500.
    Accept: */*\r\n                       |
    DNT: 1\r\n                            |
    Referer: http://ec2-54-228-4-210.eu-w | \r\n
    Accept-Encoding: gzip,deflate,sdch\r\ |
    Accept-Language: en-US,en;q=0.8,es;q= |
    \r\n                                  |
```

Figure 6: Captured HTTP request asking for video chunk #4

Figure 6 shows the typical size of an HTTP GET request for retrieving the next video chunk, which has in this particular case a size of 440 bytes.

```
⊞ Transmission Control Protocol, Src Port: http (80)
⊞ [16 Reassembled TCP Segments (22912 bytes): #144(14
⊟ Hypertext Transfer Protocol
  ⊞ HTTP/1.1 200 OK\r\n
    Date: Sun, 18 Aug 2013 16:36:46 GMT\r\n
    Server: Apache/2.2.20 (Ubuntu)\r\n
    Last-Modified: Sun, 02 Jun 2013 15:17:07 GMT\r\n
    ETag: "2586c-5859-4de2d576475c3"\r\n
    Accept-Ranges: bytes\r\n
  ⊞ Content-Length: 22617\r\n
    Keep-Alive: timeout=5, max=100\r\n
    Connection: Keep-Alive\r\n
```

Figure 7: Captured HTTP response containing video chunk #4

Figure 7 presents the size of an corresponding HTTP response packet. The upper-most mark in the figure shows that a total of 22,912 bytes have been transmitted in the HTTP response. From the HTTP content-length header the amount of video bytes contained in this chunk can be retrieved, which is 22,617 bytes. With these two values, the size of the HTTP response header can be calculated (300 bytes). This makes a final amount of metadata of 740 bytes per chunk (440 bytes for the whole request and 300 bytes for the response header). This again sums up to an overall overhead of 222,000 bytes when considering all of the 300 chunks. For transmitting the test video of the size of 6,656,763 bytes, this method introduces an overhead of 3.3% in relation to the media content.

```
⊞ Transmission Control Protocol, Src Port: http-alt
⊞ [16 Reassembled TCP Segments (22627 bytes): #883(14
⊟ WebSocket
    1... .... = Fin: True
    .000 .... = Reserved: 0x00
    .... 0010 = Opcode: Binary (2)
    0... .... = Mask: False
    .111 1110 = Payload length: 126 Extended Payload
    Extended Payload length (16 bits): 22623
  ⊞ Payload
```

Figure 8: Captured WebSocket frame containing video chunk #4

The WebSocket protocol specification defines the header as a variable size structure ranging from a size of at least two bytes to a maximum of 8 bytes. This mainly depends on the size of the payload carried by the WebSocket packet, since this is encoded in a length field in the header, which grows depending on the actual content size. In case of a minimal two bytes header, the payload of the WebSocket frame can contain a maximum of 125 bytes. Since all of the 300 two seconds video segments are in any case larger than this mark, the resulting WebSocket packets do all have a header of four bytes, as can be observed from the captured WebSocket frame shown in Figure 8. This is due to a required extended payload length header field, which introduces additional two bytes. With this two byte extended payload length header field a maximum of 65,662 bytes of payload can be specified, which is large enough for all of the 300 video chunks.

Since there are no requests required to retrieve a next video chunk, this communication overhead from the DASH-like approach is not inherent to the proposed WebSocket-based transmission. Thus, the total amount of metadata introduced per chunk is four bytes (zero bytes for the request since it does not exist and four bytes for the header in the WebSocket frame). For all of the 300 chunks this sums up to a total of 1,200 bytes for transferring the video from the server to the Web client. This represents an overhead of around 0.02% in relation to the plain multimedia content of 6,656,763 bytes.

When observing carefully the numbers given in the Figures 7 and 8 it appears that the sizes of the payloads found in the HTTP response and the WebSocket frame differ by six bytes. This constant six byte offset can be found in any WebSocket frame in comparison to the corresponding HTTP response. This is due to additional meta data added by the WebSocket implementation used in this test bed (binaryjs [35]). Thus,

the concrete WebSocket framework and libraries used for development need to be examined whether they add additional metadata to the payload, since this has an influence on the overall efficiency. In this particular case, the exchanged meta-data sums up to a total of 3,000 bytes, which represents an overhead of around 0.05% in relation to the plain multimedia content of 6,656,763 bytes.

B. Processing overhead

To further examine the potential benefits of the proposed approach of using WebSockets as communication means for video live streaming in the Web, an additional test bed has been developed and operated; aiming at finding out the total quantity of clients that one server is able to handle simultaneously. Again, two equivalent instantiations of the test bed have been deployed for the DASH-like and for the WebSocket-based live video streaming.

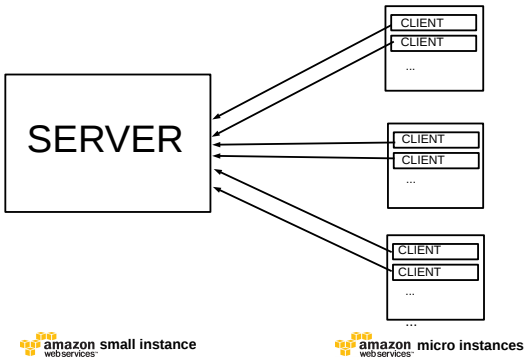


Figure 9: Architecture of the processing overhead test bed.

The machine used for this evaluation is an Amazon EC2 small instance server composed of one 64 bit ECU (EC2 Compute Unit), which provides the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor and 1.7 GB of RAM [36]. To simulate a large number of clients a set of 15 distinct and distributed EC2 micro instances have been deployed. An EC2 micro instance is equipped with up to 2 ECUs for short periodic bursts and 613 MB of RAM. The architecture of this test bed can be observed on Figure 9. The developed components described in Section IV have been installed on these systems in order to setup and operate the test beds. When building such a large scale test bed, the OS settings for the maximum number of open files per user, the maximum number of threads and the maximum number of TCP connections need to be modified accordingly.

The clients are all set up at the same time. At the moment the last of them connects to the server, all of them start being simultaneously served with the test video. After each client instance has received all content, it measures its own duration time; measured from the moment it started receiving content, to calculate the bitrate as follows:

Bitrate [bps] = Video size [bits] / Transfer time [s].

As mentioned previously, the video encoding bitrate is around 100 kbps. Hence, as long as the receiving bit rate is higher than the video bitrate, the user will be able to watch the video without encountering any disturbance. The moment in time when the number of clients is so big that the majority of them cannot be served anymore at the required minimum bitrate will be considered as the inflexion point. The expected theoretical results of these tests are shown in Figure 10, with a red dot symbolizing the defined inflexion point.

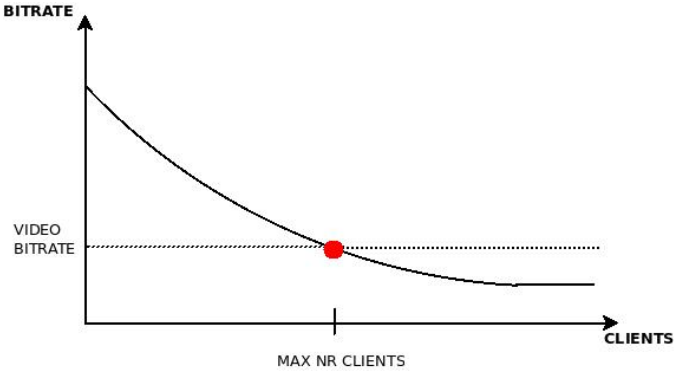


Figure 10: Expected curve of transmission bitrate

The number of clients has been increased stepwise starting from 100 clients. On each run, all clients have been equally distributed on 15 separate machine instances. Each run has been repeated 10 times to obtain a mean value. In each additional run, the server is restarted and the number of concurrent clients is increased by 100, until reaching 2,000 clients in the final run.

Figure 11 shows the results obtained from the DASH-like live streaming test bed. It can be observed that the graph for HTTP transmission bitrate shows a corresponding shape as theoretically expected and depicted in Figure 10.

The bitrate decreases from an average of 1,228 kbps, when there are 100 simultaneous clients to an average of 49 kbps, when the number of connected clients increases to 2,000. The red point indicates the inflexion point, which lies between 1,000 and 1,100 active clients. This denotes the largest quantity of simultaneous clients for this server, so that the minimum required video bitrate can still be served to the connected clients.

Figure 12 summarizes the results obtained from the WebSocket-based live streaming test bed. The bitrate decreases from an average of 4,067 kbps, when there are 100 simultaneous clients to an average of 170 kbps, when the number of active clients increases to 2,000. Thus, the WebSocket-based video server can still handle as much as 2,000 simultaneous clients and provide each with a video stream that comes with a bitrate still above the required encoding bitrate of 100 kbps.

The tests runs have been performed in both cases until 2,000 concurrent clients have been reached. Further measurements in the WebSocket-based test bed have not been performed. When extrapolating the obtained results, then the inflexion point will be located at around 2,300 clients (see Figure 12).

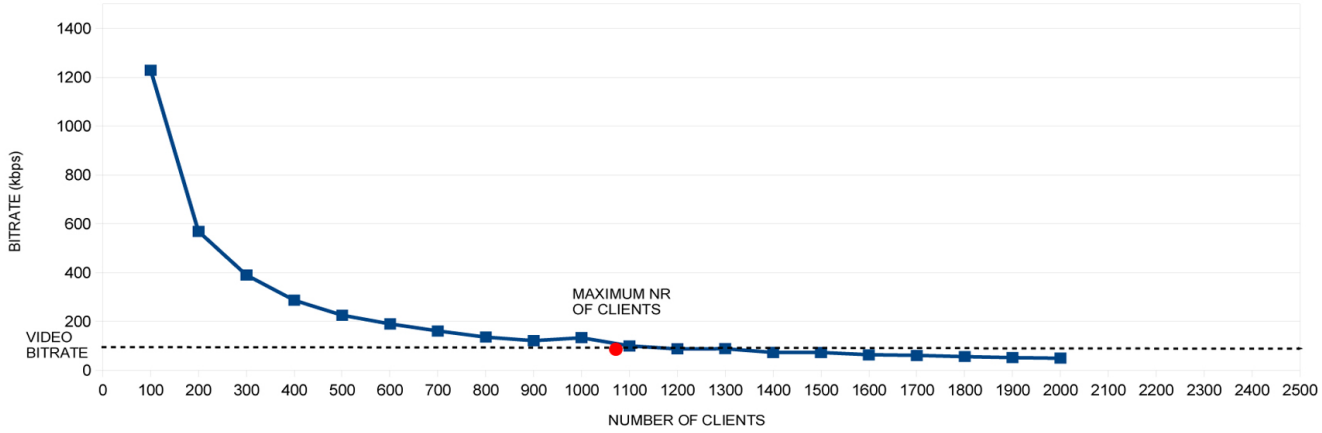


Figure 11: Average transmission bitrate for DASH-like streaming

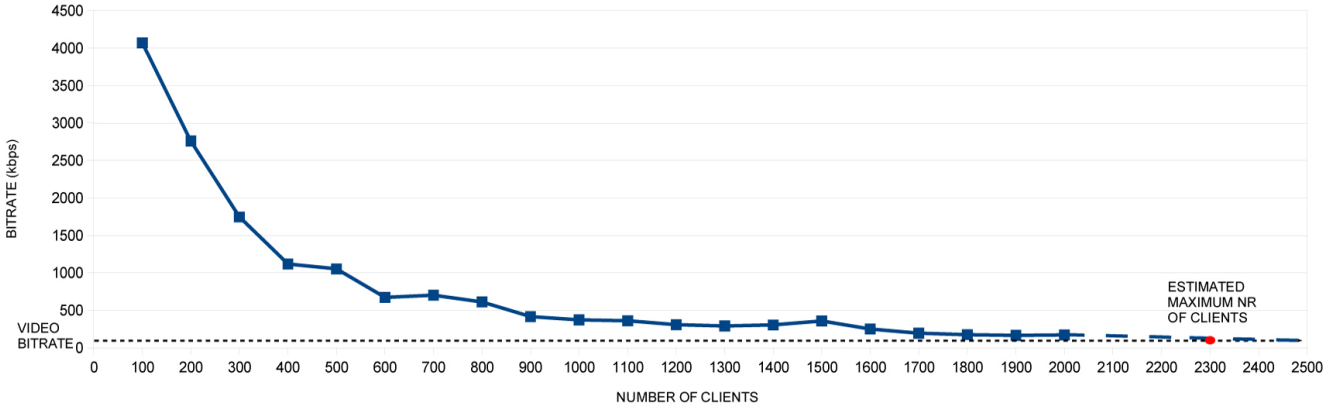


Figure 12: Average transmission bitrate for WebSocket-based streaming

From these experiments it can be deduced, that besides the communication overhead advantages, the proposed WebSocket-based live streaming approach has additional benefits in terms of processing costs. These efficiency advantages result in a larger user base being servable with the same amount of infrastructure resources.

To facilitate the task of evaluating the performance of this implementation, a video encoded at very low quality has been chosen. However, the obtained results are also to be applied for different video qualities.

VI. Further Considerations

Future research activities in respect to the proposed push-based web-native live streaming approach need to focus on further aspects. One is concerned with the adaptiveness to the underlying network conditions, in order to provide a better user experience in the presence of changing network properties. Another aspect to focus on in further research and development activities is the relation of Content Delivery Networks (CDN) and connection-oriented protocols, such as the WebSocket protocol.

A. *Adaptiveness to the network conditions*

Although adaptive bitrate streaming solutions are significantly more complex than constant bitrate streaming technologies, still, a method that determines the data throughput capabilities of each user in real time and controls the quality of the media stream accordingly, provides consumers with the best experience available, depending on their specific network and playback conditions. This can be achieved for the proposed architecture by including an encoder, which generates multiple distinct video encodings out of a a single source. The switch between the different video qualities occurs when necessary, attending to the network conditions. This can be achieved in two different ways. The first one is to monitor the video playback on the client-side and have the client notify the server

when it should reduce or increase the quality. The second one monitors on the server-side and takes the decision whether to change to another video quality level directly there. Although the first approach is currently used by most of the adaptive video streaming systems, it introduces an additional overhead due to the exchange of control messages between the client and the server. Therefore, it would not be reasonable to adopt such an approach for this architecture, where the focus has been to reduce overhead coming from periodical control requests in the first place. The second option, instead, is more reasonable in the light of the proposed push-based approach. The status of the WebSocket connection can be leveraged to facilitate this. More concretely, the TCP channel underlying the WebSocket can be used to monitor the current network conditions in terms of the available data throughput rate and to determine the most appropriate video delivery rate.

A first implementation of this concept has been developed, with three video qualities:

- Low definition video (320x240, 100 kbps)
- Standard definition video (1280x720, 1,200 kbps)
- High definition video (1920x1080, 5,000 kbps)

It provides the adaptively by checking whether the fill level of the output buffer increases, remains static or is empty during a certain observation time. In the first case the server stops sending data to the client until the buffer is drained and then continues sending with a lower video quality, if available. The second case does not require any specific action, since the current streaming settings fit to the current network conditions. In the last case, however, it seems that the client might be able to consume a higher quality level, which will be delivered, if available.

To verify this concept, a test environment has been implemented. The adaptive streaming server has been based on the Vert.x [37] application platform and concretely its modules for WebSockets and Flow Control. To simulate different network states a delay is introduced on every packet sent using the tool Netem [38]. It is controlled by the command 'tc', part of the iproute2 package of tools. The command to add a fixed amount of delay of n ms to the outgoing packages is:

```
tc qdisc add dev eth0 root netem delay n ms
```

It has been observed that the server adapts to the current state and selects the source video accordingly. For a case where the client is connected through a high-bandwidth connection with no delay, the client receives the High Definition video on the browser. When there is a delay higher than n ms, the client receives the standard definition quality and for any delay higher than m ms the low definition video.

With this prototype implementation the server-side data throughput monitoring and control has been proofed as a low-overhead and seamless extension for the push-based live video streaming system proposed in this paper. However, future work should continue in this direction to provide a fully adaptive implementation, which does not only reduce the quality for a low-bandwidth or high-latency connection but also applies more precise algorithms taking into account all parameters taking part on the data throughput and delivery to select the higher quality available for each user.

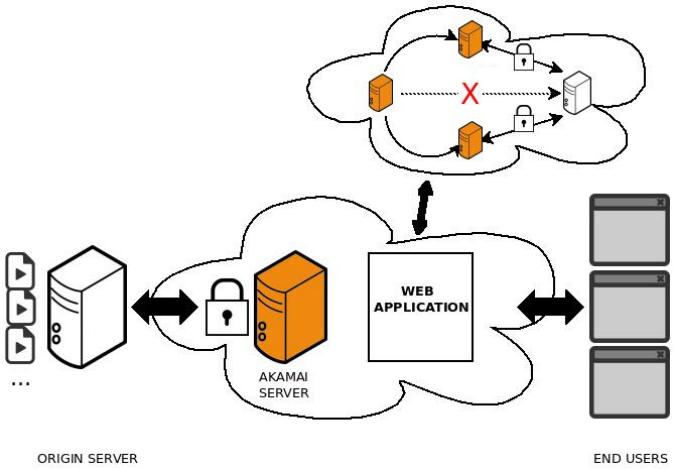


Figure 13: Akamai CDN architecture [39]

B. CDNs and push-based live video streaming

Currently, Content Distribution Networks (CDN) play a significant role when referring to web-based video streaming. The term refers to a large geographically distributed system of servers deployed in multiple data centres across the Internet, which has as a goal to serve content to end-users with high availability and high performance. This is accomplished by transparently mirroring content from customer servers, replicated all over the world. Thus, users receive the content from a server part of the CDN, which will be automatically picked depending on the kind of content and the user's location. The architecture of Akamai, one of the current top CDNs is depicted in Figure 13.

Content providers, and in particularly media companies, require these services for delivering their content to end-users. Referring back to the Cisco Visual Networking Index Global Forecast and Service Adoption for 2013 to 2018 mentioned in Section I, CDNs carried over 36% of the Internet traffic in 2013.

Therefore, the main issue they solve is the latency, the amount of time it takes the server to receive, process, and deliver a resource for a request by this mechanism, which leads to low download-times, enabling that a live event can be transmitted to every part of the world in real time as it is being consumed, as well as to decreasing the vulnerability to network congestion.

However, when proposing such a switch from HTTP to WebSocket protocol for live video streaming, the real-time nature of such content stream poses impediments concerning caching and load balancing systems, the main advantages of CDNs.

Web caching is used to store content for a certain amount of time. A situation where this is extremely useful is one

where a file will be repeatedly request by a big amount of users. CDNs are equipped with a cache of static resources like static images, CSS files, JavaScripts, as well as bigger files like video and audio media to reduce latency and network traffic. Certainly, most objects in the cache do not stay there permanently but expire so that new content can be served. How long the resources stay in the cache can vary very much depending on the content from some minutes to years. This mechanism is however not applicable to live content whichthat needs to be consumed in real-time and for which this does not provide an advantage. On the other hand, load balancing systems distribute all requests over multiple servers in order to avoid that a single server becomes overwhelmed and provide the maximum availability. This target is also difficult to accomplish with this mechanism when the media content to be delivered is being produced in real-time and the protocol used for delivering is WebSockets instead of the usual HTTP.

Further research should focus in finding options to take advantage of the infrastructure of CDNs and investigate if some changes would need to be made to use them for such a live streaming implementation over WebSockets as the one presented on this paper.

VII. Conclusion and Future Work

Video content distribution in the Web is evolving greatly. The adoption of HTTP for video streaming in the Web has its pros and cons.

For the on-demand retrieval of file-based videos the comprehensive and pervasive HTTP guarantees a broad accessibility of the content. This approach also fits well with the current deployment and usages of CDNs, ensuring the necessary global scaling of such an approach. However, these advantages do not apply to live streaming of video content. First, CDNs cannot exploit their strength, since the feeding of the content to the distributed cache servers does not adhere to the real-time character of live video streams. The idempotence of the HTTP GET method is henceforth less relevant for live casts and brings other drawbacks of HTTP back in focus. The client-initiated request-response communication pattern is one major source of issues when push-based communications need to be implemented as it is the case for the transmission of media content.

Currently, although file-based video content is still dominating, the consumption of live streams is on the raise. However, the available standards and technologies for enjoying live video content in a Web-native manner are still in their infancy. The HTTP-based DASH is a first step in this direction.

This paper examined the possibility of developing a live video streaming solution in a Web-native manner by means of standards belonging to the HTML5 standards family. Such an approach has been realized based on the HTML5 video element and WebSockets as real-time communication means. The performed evaluation of the developed video streaming solution demonstrates that this approach is much more efficient compared to methods relying on HTTP. Both, the communication as well as the processing overheads can be significantly reduced by the proposed WebSocket-based solution in comparison to HTTP-relying methods such as DASH.

First steps towards an adaptive live streaming architecture have been undertaken, proposing a mechanism situated on the server-side to introduce adaptiveness to the underlying network conditions to the proposed push-based web-native approach, which has to be further developed in future research activities, to enable a better user experience in the presence of changing network properties.

Another significant issue for future research relies on the relation of connection-oriented protocols, such as WebSocket and CDNs, in order to investigate options to capitalize on the infrastructure of CDNs for their use in a live streaming implementation over WebSockets as the proposed.

References

[1] L. Lo Iacono and S. Santano Guillén, "Web-native video live streaming," in WEB 2014 - The Second International Conference on Building and Exploring Web Based Environments, ISSN: 2308-4421, ISBN: 978-1-61208-333-9, 2014, pp. 14–19.

[2] Cisco, "Cisco visual networking index: Forecast and methodology, 20132018," White Paper, 2014, online available at www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf (last accessed: Nov. 2014).

[3] Audio-Video Transport Working Group, "Rtp: A transport protocol for real-time applications," IETF, RFC 1889, 1996, online available at www.ietf.org/rfc/rfc1889 (last accessed: Nov. 2014).

[4] H. Schulzrinne, A. Rao, and R. Lanphier, "Real time streaming protocol (rtsp)," IETF, RFC 2326, 1998, online available at www.tools.ietf.org/html/rfc2326 (last accessed: Nov. 2014).

[5] J. Follansbee, Get Streaming!: Quick Steps to Delivering Audio and Video Online. Elsevier, 2004.

[6] ISO/IEC Moving Picture Experts Group (MPEG), "Dynamic adaptive streaming over http," ISO/IEC, Tech. Rep., 2013, online available at www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57623 (last accessed: Nov. 2014).

[7] Microsoft, "Microsoft smooth streaming," www.iis.net/downloads/microsoft/smooth-streaming (last accessed: Nov. 2014).

[8] Microsoft Corporation, "Silverlight 5.1," 2013, www.microsoft.com/silverlight (last accessed: Nov. 2014).

[9] Microsoft, "Internet information services," www.iis.net (last accessed: Nov. 2014).

[10] ISO/IEC Moving Picture Experts Group (MPEG), "Iso mpeg-4," ISO/IEC, International Standard, May 2012, online available at www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=61490 (last accessed: Nov. 2014).

[11] ISO/IEC Moving Picture Experts Group (MPEG), "Advanced audio coding," ISO/IEC, International Standard, 2004, online available at www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=62074 (last accessed: Nov. 2014).

[12] Apple Inc., "Http live streaming," IETF, Internet-Draft, 2013, online available at www.tools.ietf.org/html/draft-pantos-http-live-streaming-12 (last accessed: Nov. 2014).

[13] ISO/IEC Moving Picture Experts Group (MPEG), "Iso mpeg-ts," ISO/IEC, International Standard, 2013, online available at www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=62074 (last accessed: Nov. 2014).

[14] R. Berjon, S. Faulkner, T. Leithead, E. D. Navara, E. O'Connor, S. Pfeiffer, and I. Hickson, Tech. Rep.

[15] Apple Inc., "Http live streaming - section 3: Playlist file," IETF, Internet-Draft, 2013, www.tools.ietf.org/html/draft-pantos-http-live-streaming-12 - Section 3 (last accessed: Nov. 2014).

[16] D. Minoli, Linear and Non-Linear Video and TV Applications: Using IPv6 and IPv6 Multicast (7.1.2). John Wiley & Sons, 2012.

[17] Encoding.com, "The definitive guide to hls," July 2014, "http://www.encoding.com/http-live-streaming-hls/" (last accessed: Nov. 2014).

[18] Adobe, "Adobe http dynamic streaming," www.adobe.com/products/hds-dynamic-streaming.html (last accessed: Nov. 2014).

[19] Adobe, "Flash media manifest file format specification 1.01," 2010, online available at osmf.org/dev/osmf/specpdfs/FlashMediaManifestFileFormatSpecification.pdf base(last accessed: Nov. 2014).

[20] B. Rainer, S. Lederer, C. Müller, and C. Timmerer, "A seamless web integration of adaptive http streaming," in Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European, 2012, pp. 1519–1523.

[21] A. Colwell, A. Bateman, and M. Watson, "Media source extensions," W3C, Last Call Working Draft, 2013, online available at dvcs.w3.org/hg/html-media/raw-file/tip/media-source/media-source.html (last accessed: Nov. 2014).

[22] A. van Kesteren, J. Aubourg, J. Song, and H. R. M. Steen, Working Draft, 2014, online available at http://www.w3.org/TR/XMLHttpRequest (last accessed: Nov. 2014).

[23] I. Fette and A. Melnikov, "The websocket protocol," IETF, RFC 6455, 2011, online available at www.tools.ietf.org/html/rfc6455 (last accessed: Nov. 2014).

[24] I. Hickson, "The web sockets api," W3C, Candidate Recommendation, 2012, online available at www.w3.org/TR/websockets (last accessed: Nov. 2014).

[25] A. Deveria, "Can i use web sockets," April 2005, "http://caniuse.com/websockets" (last accessed: Nov. 2014).

[26] D. T. Rico, "Seamlessloop," 2007, "github.com/Hivenfour/SeamlessLoop" (last accessed: Nov. 2014).

[27] T. Yoshino, "Pywebsocket," pypi.python.org/pypi/mod_pywebsocket (last accessed: Nov. 2014).

[28] H. Ohtani, "websocket-client," pypi.python.org/pypi/websocket-client/0.7.0 (last accessed: Nov. 2014).

[29] Python Software Foundation, "Http protocol client," docs.python.org/2/library/httplib.html (last accessed: Nov. 2014).

[30] Python Software Foundation, "Basehttpserver," docs.python.org/2/library/basehttpserver.html (last accessed: Nov. 2014).

[31] Python Software Foundation, "Socketserver," docs.python.org/2/library/socketserver.html (last accessed: Nov. 2014).

[32] Blender Foundation, "Big buck bunny," 2008, peach.blender.org (last accessed: Nov. 2014).

[33] Creative Commons, "Creative commons license attribution," 2007, "www.creativecommons.org/licenses/by/3.0/us/legalcode" (last accessed: Nov. 2014).

[34] Open Source, "Wireshark 1.10.2," 2013, www.wireshark.org (last accessed: Nov. 2014).

[35] E. Zhang, "Binaryjs," 2013, www.binaryjs.com (last accessed: Nov. 2014).

[36] Amazon Web Services, "Amazon elastic compute cloud (ec2)," www.aws.amazon.com/en/ec2 (last accessed: Nov. 2014).

[37] Eclipse Foundation, "Vert.x," 2013, "http://vertx.io/" (last accessed: Nov. 2014).

[38] Linux Foundation, "Netem," April 2005.

[39] E. Nygren, R. Sitaraman, and J. Sun, "The akamai network: A platform for high-performance internet applications," in ACM SIGOPS Operating Systems Review, Vol. 44, No.3, 2010, online available at http://www.akamai.com/dl/technical_publications/network_overview_osr.pdf (last accessed: Nov. 2014).

Improving Web Accessibility: Computing New Web Page Design with NSGA-II for People with Low Vision

Yoann Bonavero, Marianne Huchard, and Michel Meynard
LIRMM, CNRS and Université de Montpellier, Montpellier, France
Email: yoann.bonavero@lirmm.fr, marianne.huchard@lirmm.fr,
michel.meynard@lirmm.fr

***Abstract*—As society becomes increasingly aware of the need to take disabilities into account, new information technologies and intensive use of computers can be a chance or create new barriers. In the specific case of people with low vision, efforts to improve e-accessibility are mainly focused on the provision of third-party tools. Assistive technologies like screen magnifiers adapt graphical user interfaces to increase the quality of the perceived information. However, when these technologies deal with the Web, they are not able to meet all specific needs of people with low vision. In this paper, we propose an approach to make Web pages more accessible for users with specific needs. User preferences can concern font size, font family, text color, word and letter spacing, link color and decoration or even more complex features regarding brightness, relative size or contrast. We also take into account and encode the designer's graphical choices as designer preferences. Solving preferences of the user and of the designer to obtain a new Web page design is an optimization problem that we deal with Non-dominated Sorting Genetic Algorithm II (NSGA-II), a polynomial Multi-Objective Genetic Algorithm. We conducted detailed tests and evaluated the running time and quality of results of our tool on real Web pages. The results show that our approach for adapting Web page designs to specific user needs with NSGA II is worthwhile on real Web pages.**

Keywords*– *e-accessibility, Web page personalization, visually impaired, low vision, evolutionary algorithm, NSGA-II.

I. INTRODUCTION

In this paper, we deepen the research work presented in [1], where we developed an approach to improve Web page accessibility for people with low vision.

Many countries are adopting laws or treaties for enhancing digital accessibility. Some countries consider e-accessibility as a very important issue and even as a citizens' right. According to recent estimates, about 285 million people are considered visually impaired worldwide. 39 millions of them are blind and 246 million have low vision [2]. These figures are constantly growing, mainly because of the increased life expectancy.

ICT (Information and communication technologies) are increasingly used by everyone in everyday life. Unfortunately, this can be a double-edged issue for people with visual impairment, because these new technologies, which are able to compensate for user disabilities, can also be a new source of exclusion and discrimination. On the one hand, these technologies offer many solutions for everyday life activities. For example, they allow online purchasing, dealing with administrative documents, managing bank accounts, or locating places and finding routes. Beyond these services, ICT also bring a social dimension. They potentially offer access to information that was previously inaccessible for visually impaired people. On the other hand, many issues remain, due to the technologies used to design and develop websites.

Websites are composed of different kinds of data, including text-based documents, images, videos, and sounds. These data are displayed on pages formatted with respect to a visual style. This visual style, often given by CSS (Cascading Style Sheets) is written or used by the page author. The different choices made by the designer create the graphical context of the Web page. The graphical design of a Website reflects the brand or organization, and constitutes a landmark for people. It is also intended to influence the reader to recognize, assimilate, memorize a page and associate it with the related brand or organization. Moreover, it is intended to help users in their tasks by describing a navigation template, an information hierarchy and thus, it helps in understanding the page.

The publication language mainly used for the Web is HTML (HyperText Markup Language). This publication language is a very flexible and easy to understand language. Unfortunately, this flexibility gives us many ways to do the same basic things. For example, we can build the same (in terms of rendering) navigation menu with only list tags such as UL (Unordered List), LI (List Item), A (link to Another file), or with block tags such as DIV (DIVision) or SPAN (to span).

The W3C (World Wide Web Consortium) and other organizations publish sets of technical specifications in order to frame the development of websites. The W3C also provides a set of specifications to make accessible websites. The compliance of websites to these specifications assumes that they can be used by assistive technologies. Tools and guidelines are provided to developers and end-users, such as the WCAG 2.0 (Web Content Accessibility Guideline [3]), the UAAG (User Agent Accessibility Guideline [4]), the ATAG (Authoring tool Accessibility Guideline [5]) and the WAI - ARIA (Web Accessibility Initiative - Accessible Rich Internet Application [6]). Organizations like BrailleNet have been created to op-

erationalize the different standards of the W3C guidelines, including "AccessiWeb".

Unfortunately, e-accessibility is not a main concern of web-site designers and developers. It is often considered as a waste of time or an additional development cost, giving unsightly results that only target a small part of the population. Users can use third-party assistive tools to cope with visual difficulties when developers do not care about accessibility. Assistive technologies have existed for several years and are widely used by disabled people. Screen readers allow the user to get information in another communication way: vocal synthesis or braille display are used to vocalize information or display it in braille. Although this technology is designed for blind people, people with low vision also use it as a supplement to another assistive technology. Visually impaired people with low vision often use their partial sight as the principle means to access information. Screen magnifiers are applications that improve visual comfort and increase information acquisition. With these tools, it is possible to use zoom and color filters to compensate for visual issues. These tools are useful but frequently not sufficient to ensure e-accessibility because they have a general purpose, and they are not adapted to specific needs. This is mainly due to the high maintenance cost to maintain the compatibility with some applications like browsers, and to deal with after-sales technical issues. Nevertheless, e-accessibility is essential to ensure a quality of access to a large amount of Web services and contents. Recently, e-accessibility understanding is about to become a recognized professional skill for Web developers.

In this paper, we address the problems of adapting Web page design to the specific needs of a visually impaired person. Our approach proposes to replace the current pixel-level treatment process (in magnification filters) by an adaptation process based on knowledge of the HTML elements. Each HTML element has its own type and properties (color, size, position, etc.). The adaptation is performed from a set of user wishes, also called preferences. User's preferences can be font size, font family, text color, word and letter spacing, link color and decoration or even more complex wishes regarding brightness, relative size or contrast. We also take into account Web page designer's graphical choices as designer preferences. Solving these user and designer preferences to get a new Web page design is an optimization problem, that we manage with NSGA-II (Non-dominated Sorting Genetic Algorithm II), a polynomial Multi-Objective Genetic Algorithm.

In Section II, we explain how existing visual tools and assistive technologies work, and we highlight their main drawbacks regarding Web page context. We also present existing approaches that are intended to adapt user interfaces or Web pages in a personalized way. Section III presents how we represent Web page elements to be adapted as well as user (or designer) wishes. In Section IV, we describe how we are using and tuning NSGA-II. Section V reports the results that we obtained on several Web pages during our research. We conclude, in Section VI, by giving some perspectives of this work.

II. Existing work and problem statement

In this section, we explore hardware and software solutions developed to improve or provide website accessibility for visually impaired people with low vision. In all of these solutions we distinguish between two main kinds of tools and approaches. Some of them are used at development time as developer tools while others are used by end-users mainly on the client-side (filters and style-sheet redefinition). Besides, we describe more advanced proposals that try to partially automate some personalized adaptations.

A. Standards, guidelines, tools for developers

The W3C (World Wide Web Consortium) is at the origin of HTML and CSS standardization. It also works on accessibility via several initiatives, including WAI-ARIA (Web Accessibility Initiative - Accessible Rich Internet Application [6]). These standards have evolved through many versions to address the emergence of new technologies. They have two main objectives. The first is to ensure that resources can be parsed and used by external assistive technologies. The second is to provide minimal access to content for people who do not use, for different reasons, assistive technologies. One aim of these standards is to be independent of languages like HTML, JavaScript or CSS. This ensures the definition of robust standards regarding language diversity and evolution. The separation between the content and the page display style is the most important feature offered by HTML 4 and CSS 2. This separation should provide easier access to content.

In addition to the standards, guidelines like WCAG (Web Content Accessibility Guidelines [3]), UAAG (User Agent Accessibility Guidelines [4]) or ATAG (Authoring Tools Accessiblity Guidelines [5]), frameworks and tools are published to ease the use of the standards. We can mention "AccessiWeb", developed by the BrailleNet organization, which provides a simple operational interpretation of standards. WAI references a set of evaluation tools [7]. WCAG contains all standards about page content rendering, including the way the content is displayed to the user in terms of size, contrast, etc. UAAG gathers many required features of tools that browse Web pages. Finally, ATAG is concerned with tools that generate source code for Web content. The rules and standards are classified according to their importance to make websites accessible. Three increasing accessibility levels have been defined (A, AA and AAA). The first level (A) gives basic mandatory advice to ensure information accessibility. The second level (AA) provides important recommendations to be respected to avoid difficulties in accessing information. The third level (AAA) is about additional and optional ways to improve information access quality. When Web designers and developers include accessibility dimension in their websites, they mainly try to reach the intermediary AA level. Only a few very specialized websites require the highest level AAA.

Many tools exist that allow developers to make accessible websites or simply to get accessible existing websites. These tools analyze the HTML source code and either automatically rewrite it, or assist the developer to correct it, for example through suggestions in accordance with the standards [8].

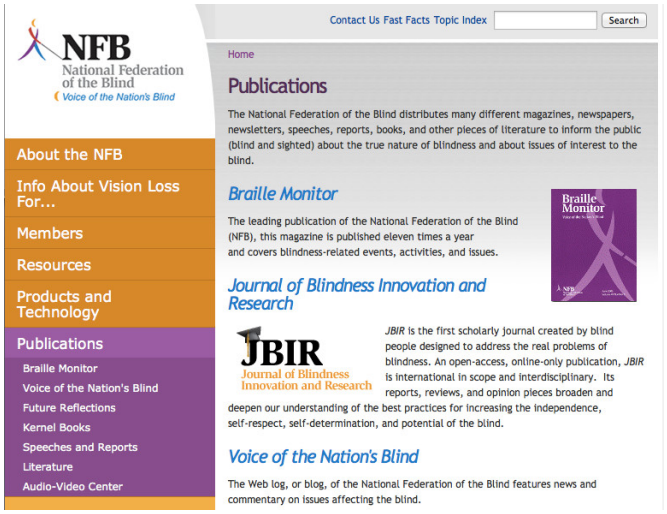


Figure 1. Original publication page of the National Federation of the Blind.

Figure 2. Applying zoom from a magnifier on the NFB publications Web page.

These tools can be separated into two categories: evaluation tools and transformation tools. The main drawback of these tools is that they do not enable adaptation for very different needs coming from various visually impaired people. Some user needs can contradict each other. Conflicts can arise due to dependencies between needs. For instance, high brightness contrast (for readability) and low light emission (for reducing the dazzle effect) can lead to conflictual needs (light emitting elements are linked to brightness contrast between them). Consequently, automated evaluation and transformation tools can only assist developers to meet a general accessibility requirement but are limited to implementation of the minimum recommended by standards.

B. Improving accessibility tools for the end user (magnification, browser options and extensions)

Some kinds of accessibility tools are available to get information from websites and report it to the user through another communication protocol. For example, for users with low vision, it is possible to retrieve information by transforming visual output with magnification applications or accessibility browser options and extensions.

Magnification tools allow the user to zoom on windows (e.g., Figure 2 zooms on Figure 1). Some of them propose font smoothing to avoid blurred characters, and mouse pointer modification to improve tracking movements. As another example, magnification tools can apply filters on the window. Filters include “gray scale”, “one color scale”, “black and white”, or “color inversion” (see Figure 4 for three of these filters). One widely used color filter is the color inversion filter (Figure 3). Its main purpose is to considerably reduce light emission when the user is on pages with a light background.

Another way for adapting Web pages consists of using the browser options and extensions that enable us to manipulate style sheets. It is possible to completely remove style sheets or to define a unique style sheet that will be applied to all Web

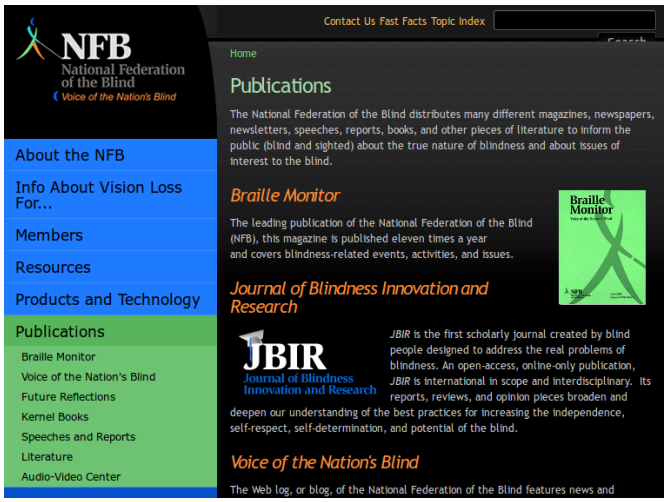


Figure 3. Applying a color inversion filter from a magnifier on the NFB publications Web page.

pages. To facilitate modifications, many browsers provide a graphical interface to help in the configuration of properties such as background color, text color, text size or link color.

With the two solutions defined previously (screen magnifier and style sheet redefinition), we can theoretically adapt almost all pages to be suitable for a large part of the impaired population. However, it is more complex in practice. These pages could be even better modified for the population, and there is a requirement for additional tools to meet more needs of visually impaired people. For various reasons, the existing solutions are not suitable for everyone. In the following, we highlight problems that occur when using filters (treatments on global images), then issues related to style sheet modification.

1) Filters: Figure 4 illustrates three filter applications. In the original *A* case, we represent a dark colored text background.

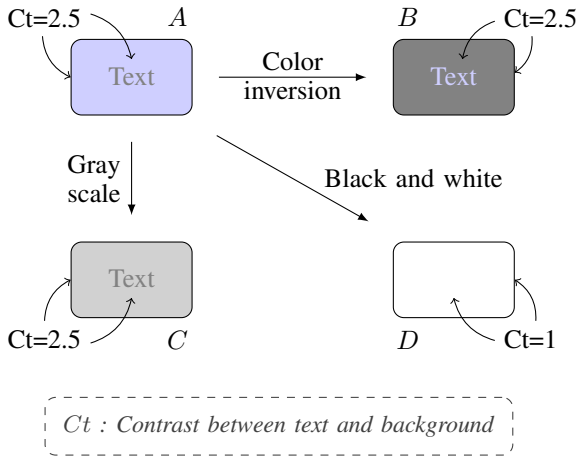


Figure 4. Filter application.

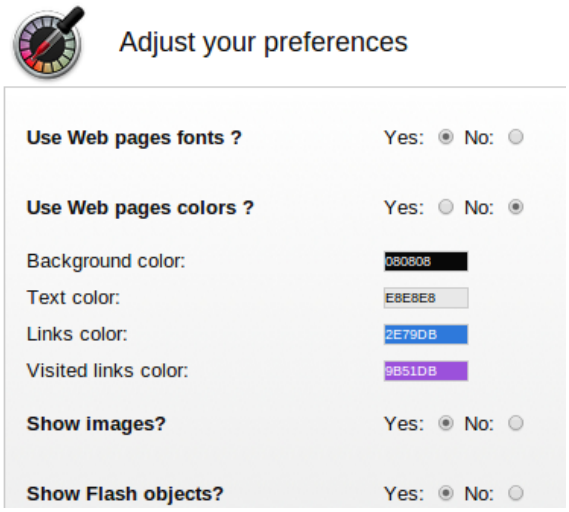


Figure 5. Chrome change colors extension.

From this first original case we apply some classical filters encountered in magnifier tools. In B, we apply to A the widely used filter inverting color of the display. We apply to A a gray scale color filter to give C, and we obtain D by applying a black and white color filter on A. For each case we associate a brightness contrast between the text color and the background color. Brightness contrast is the property that people with low vision mainly use to quantify the difference between text and direct background when they require an acceptable readability. It is measured by a ratio, which is a floating number between 1 and 21. It often ranges from 1:1 to 21:1. 1 (1:1) denotes the null contrast while 21 (21:1) denotes the highest contrast between two elements. This representation comes from the WCAG contrast computation. Regarding our filter, if the contrast is originally low (A case), filters like color inversion and gray scale filters cannot significantly improve brightness contrast and therefore the readability. Black and white filters, which are often efficient to increase the contrast, are no longer relevant in this case. These filters use a threshold to separate elements by their light emitting into two groups. Each element in the group of darkest elements is assigned a black color while each element in the second group is assigned a white color. If both text and background have two light or dark close colors (both below the threshold or both above the threshold), both elements are assigned the same color (in Figure 4, they both become white). Obviously, if both elements have the same color we lose all readability and contrast falls to 1:1. To avoid this, the filter must include a threshold, which would be customizable either automatically or manually by the end-user. However, in practice, this is not often the case, and when the threshold can be changed, this is a complex task for a non-expert user.

2) Style sheet definition: Most current browsers have extensions 5, modules or simply accessibility options to transform and adapt Web pages. These transformations are based on the manipulation of the Web page style sheet. In old browsers, options are often available to simply disable the original style sheet. This has the effect of only keeping the content without any graphical style. In more recent browsers, there are more

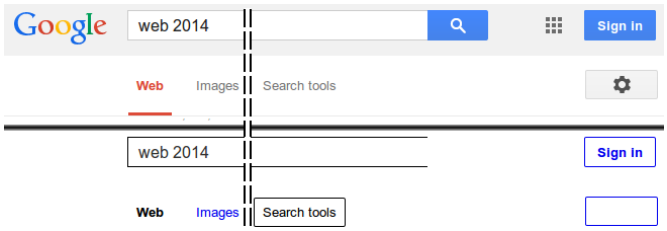


Figure 6. Google website with browser accessibility options enabled (top: original, bottom: with some accessibility options).

options to individually select to display images, keep original text font, etc. Sometimes this manipulation leads to huge information loss. Consequently, we can lose the global page context and also the brand or the organization design. This loss is unavoidable if Web pages are not properly developed in respect with the content and style separation rule given by the W3C.

An intermediary approach is also provided by browsers. Instead of disabling the entire style sheet, we can change a part of it. Rewriting a section of the page style sheet allows us to improve the readability of some elements. The end-user can define some properties like text color, text size, link color or background color. Some browsers allow more advanced users to define and provide their own style sheet. This alternative allows users to change all object properties. With style-sheet rewriting, the original context is more or less kept depending on the applied modifications. The global context (layout, colors, brand chart, etc.) is inevitably lost if there are major changes to compensate for complex disabilities.

Figure 5 shows how preferences can be adjusted in Chrome. Figure 6 presents the result on the Google result page when selecting some colors in accessibility options like background color. On the top, we have the original page, the middle of the page is removed to only show parts where alterations occurred. The result is displayed below the original. As a consequence,

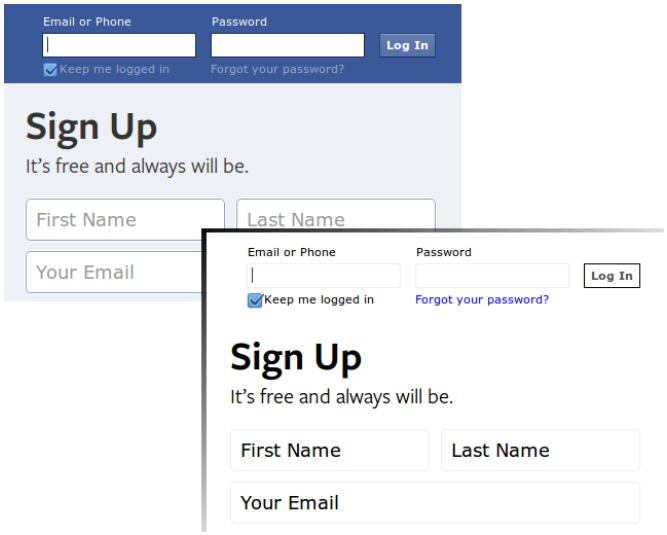


Figure 7. Facebook website with browser accessibility options enabled.

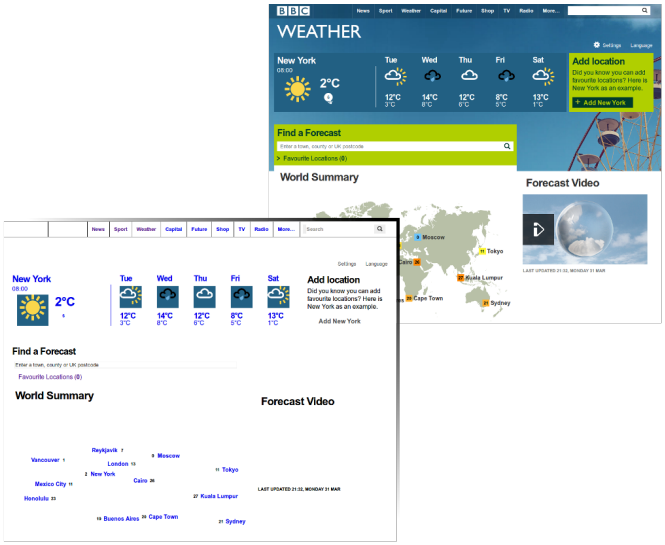


Figure 8. BBC website with browser accessibility options enabled.

buttons represented by an image included by the CSS file as a background image disappear and are replaced by the default selected color. Moreover, the main Google logo also disappears.

The Facebook sign in page (Figure 7) and the BBC (British Broadcasting Corporation) weather page (Figure 8) are deeply altered by accessibility options. Important information is lost. On the Facebook page, a recent version of HTML and styles sheets are used to make many visual effects. Main modifications concern fields of forms. Borders of login and sign in form fields become nearly invisible. Moreover, because of the removal of the top header background, login fields cannot be distinguished. Only the blinking caret is visible and allows the user to detect them. Concerning the sign in form, nearly invisible borders combined with lighter default text (text that disappears when clicking or filling the field) give to a user with low vision the impression that this text is the label of the field. Thus, the user has to click on the text to see that it is a field default text and not a label. The BBC weather page also has many alterations. This time this is not an interaction button that disappears but image justifying the position of many links. Indeed, the map at the page bottom contains text links to select a city. However, the world map seems to be inserted as a background image and not as an image. Thus, when accessibility options are set, the user is faced with a link disposition that is not logical. On this website, we also observe that the logo disappears and the search field is hard to find. Furthermore, here again, the location field appears with a nearly invisible border, which may generate ambiguity.

With style-sheet modification, the user often has to set many properties, including text color, link color, visited link color, hovered link color, title level 1 color, title level 2 color, etc. As a result, he has to manipulate a set of technical terms and a lot of options, and this task often is cumbersome and time consuming. Furthermore, preferences defined through this interface are the same for all pages. They are applied on all Web pages independently of their original style. Unfortunately, a single global configuration made by the user may not be relevant for all Web pages.

C. More advanced personalization approaches

Beyond the modules or extensions mentioned above, some more evolved proposals provide greater physical characteristic configuration support [9]. These end-user side applications allow the user to configure text properties such as size, letter spacing, or line spacing, colors of the text, background and links. They also allow the user to configure the image display (show or hide) and table display. Once the modifications are applied, almost all information about website colors may disappear. Then the original site ambiance may be lost.

User actions on websites can be used to adapt the Web page or the navigation: In [10], the authors propose to configure Web page display according to user actions and behavior. Depending on the clicked links and interactions, content can be hidden to highlight important content. This approach also allows to module menu content for adding potentially useful content. However, this approach must be taken into account at the development time. In [11], the authors propose to personalize Web display (shopping gallery) to a specific user or user group. The analysis of user usage on existing websites allows Web pages to be shown with different structure and navigation.

Several research studies have dealt with the generation of adapted user interfaces (UI). The SUPPLE++ systems [12] targets people with low vision or motor disability. For people with low vision, they propose users to control only the visual cue size. The tool is based on an optimization algorithm to combine adaptations to both low vision and motor disability. In [13], an abstract description of the UI and an ontology modeling the context (user capabilities, devices, etc.) are used

to automatically generate adapted accessible mobile user interfaces (for ATMs — Automatic Teller Machines, information kiosk, etc.) for people with disabilities. Although we do not focus on UI, here these works may be a source of inspiration, especially [13], which uses abstract models of the UI and of user preferences.

Some approaches focus on a particular visual problem, like the modification of colors for dichromate users while preserving the contrast [14].

Other approaches concentrate on accessing the structure of the page because it helps to understand and use the content. In [15], a configuration interface helps the user to specify elements of the Web page (title, content, navigation menu) that he wants to be shown depending the platform used, namely a Personal Computer (PC) or a Personal Digital Assistant (PDA). The KAI (Accessibility Kit for the Internet) system [16] considers both the developer and the user point of view. The user is able to choose what interests him the most in the accessed Web pages. As HTML cannot be used to know the real components due to its permissive syntax, BML (Blind Markup Language), a new markup language, is provided to the developer to annotate components of the page to be used by assistive tools. This marking operation can be done automatically for existing pages. Then a new HTML code is built using the markup and the user's preferences. It is used either with a normal browser or with an audio-touch platform that helps the user to access the page structure. During the transformation process, metrics are computed to rank Web pages according their accessibility. The transformation applied to improve the Web page focuses on making the structure accessible and not on specific visual problems.

D. Discussion

Standards, guidelines and evaluation tools are very useful, but it is hard to force developers to follow their advice. Statistics tell us that less than 10% of public websites are fully accessible [17].

We saw on detailed examples that user-side tools that apply global filters on the Web page, may improve some parts of the page, at the cost of degradation of other parts and that they can be totally inefficient in some situations. Using a style-sheet redefinition approach or configuration tools does not ensure that the original Web page design is kept. Besides, accessibility of this approach is not evident, because people with low vision may also have difficulty in editing or configuring style-sheet content.

From the existing work on personalized approaches, we learn that several issues are related to the Web page adaptation: user preference elicitation, modeling structure and web page design, modeling platform context (PC/PDA) and producing a new Web page whose design is a trade-off between the respect of these user preferences and the respect of the structure and initial design of the page.

Our objective is to adapt a page in accordance with its original appearance, with its structure and with the user preferences. We aim at proposing a general method that is able to take into account various visual needs. In our current work, we do not deal with the user preference elicitation, that we suppose is described in a simple formal language, as well as initial developer choices. User preference and design choice elicitation (explicitly or via a learning phase) will be studied in a future work. We also do not take into account various platforms at this stage, and we do not address content and navigation adaptation, that we want to keep as in the original page. We apply adaptation on the client side, using the original Web page. We separately process page elements according to their semantics, rather than applying a global strategy. We propose a method for computing a compromise between user preferences and developer choices on Web page elements. In our previous paper [1], we showed that classical representation and algorithms of Preference theory [18] did not scale in our context, even with small preference sets composed of basic preferences. In this paper, we focus on use of the meta-heuristic NSGA II that demonstrated better performance.

III. Our approach

We aim to develop a global approach as independent as possible from any specific Web page and also that would be able to consider any specific user's wish to find an adaptation. The previous section highlighted some problems related to the global treatment by magnifier software or to the simplistic preferences expressed via style sheet manipulation (browser accessibility options or dynamic tools). Our approach addresses these problems by using an Artificial Intelligence based approach.

In Section II, we illustrated the problem of the global treatment by magnification tools. We have shown that filters provided by a magnification tool can only be applied on the image rendered by the graphic card. Therefore, we can only apply filters on the entire screen or on an entire window in some operating systems. To achieve a better adaptation by taking all user needs into account, our approach is based on smaller elements than the entire screen, and it considers the following four components:

- Objects and properties (HTML elements and style) of the page written in the HTML and CSS files;
- Variation points (for example the color of a specific element or the size of the second level title);
- User's wishes put forward to compensate for his disabilities;
- Algorithms for finding an adaptation, from the initial Web page, according to the user's wishes, and such that the adaptation also integrates designer initial choices.

We develop these elements in the following subsections.

A. Preference representation

Designers connect different HTML elements to build Web pages. The element organization creates a tree structure. Each tree node represents an element with its physical properties like the position, appearance (size, color) or more abstract properties describing the element type (menu, content, image, link, etc.). In HTML5, which is the direct follow up of HTML4, there is a set of new tags. These tags are used to

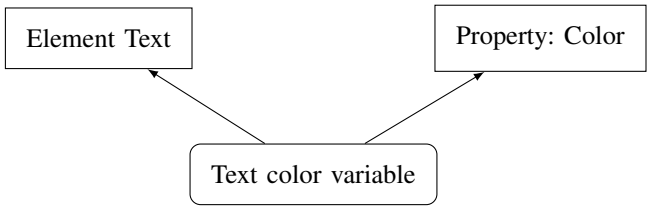


Figure 9. Association of one element and one property to make a variable.

describe element types with more semantics like navigation menu, article, section, complementary, etc. The old version of HTML (HTML4) only provides some tags without important semantics, like blocks. As a result, when a Web page is developed in HTML5, we have much information allowing us to express preferences and compute adaptations. However, if the page is in an older version of HTML, it is necessary to previously detect some important parts of the page, particularly the menu, the main content and sections, before computing an adaptation.

We define *a set of objects* (1) that represent all HTML elements that are important and useful in the page modification process. In other words, elements that will not be updated or that will not be used in computing are excluded from the object set. Excluding unnecessary elements allows us to reduce as much as possible the number of combinations and finally the search space size. HTML elements included in this set are, for example: first level titles (h1), paragraphs (p), anchors (a), images (img), articles (article), navigation bars (nav) etc.

$$Objects = \{O_1, O_2, .., O_n\} \tag{1}$$

The *variation points* are a set of variables (2) induced by properties. Properties can be either basic properties written in the HTML or CSS files, or computed from these basic properties. For example, height, width, position or color are basic properties found in the HTML or CSS files, while area is a computed property derived from both the height and width of an element. To summarize, a variable is a specific property of a specific object. An object gives rise to the creation of x variables, when x is the number of properties taken into account for this object. To be able to change the value of the red color component of an object, a variable representing this red component is created and added to the variation points. The domain of this variable is the value set of the red color component of the considered object. Thus, a variable is a pair composed of an object and a property (Figure 9).

$$VariationPoints = \{V_1, V_2, .., V_m\} \tag{2}$$

From the list of defined variables, a user can define different constraints (3). These constraints are also called *preferences* or *wishes*.

$$Constraints = \{C_1, C_2, .., C_k\} \tag{3}$$

For instance a user can say: "I prefer a dark color to a light color for titles". This preference means the user prefers to have a dark color for all titles in the page rather than a light color. This preference only concerns page titles and not the other texts in the page. With this page element segmentation, we can define different preferences for each object or each kind of object. All choices made on one or more variables constitute user preferences or user wishes. There is a main difference with existing work using user preferences. Tools using user preferences to adapt Web pages like in [9] use literal values for object properties like *red* or *blue* for the color or *14 pt* for a text size. We instead use constraints to compute such values.

We represent two different levels of preferences: basic or more complex preferences. The basic preferences, that come from Preference Theory [18], are represented as in (4).

$$V_i \ op \ x_i \ >_p \ V_j \ op \ y_j \tag{4}$$

where V_i and V_j are two variables (with possibly $V_i = V_j$), $>_p$ the preference symbol ($A >_p B$ means A preferred to B), op is a Boolean operator like $=$ and x_i (resp. x_j) is a value in the domain of V_i (resp. V_j). To represent the user's wish "I prefer black text to blue text", we use the variable c_T to represent the color of a text object T. The domain of c_T is $\{white, red, blue, black\}$. The user's wish is expressed as in (5).

$$c_T = black >_p c_T = blue \tag{5}$$

The basic representation also allows us to express conditional preferences. With conditional preferences we are able to represent preferences like "I prefer bold font to normal font if the font color is yellow". Here, we introduce a new variable w_T for representing the weight of the text concerned by the preference. We also introduce a new operator ':' to separate the condition from the remainder on the expression. A conditional wish is shown in (6). This representation was considered in our previous paper [1].

$$c_T = yellow : w_T = bold >_p w_T = normal \tag{6}$$

Here we explore more complex preferences in which we consider any complex function on variables and their domain values. This allows us to express wishes like "I would like to have a text size greater than or equal to 14 pt", "I would like to have bold text rather than regular text when text size is less than 14 pt" or "I would like to have a contrast between text and direct background greater than or equal to 50%". In this last example, the contrast is a binary function. It represents a distance between the colors of two objects: with the textual element T and the object B providing a background to T. To model this preference, we introduce two new variables, c_T that represents the text object color and c_B that represents the text background color. We define a contrast function $contrast(x, y)$ that returns the computed contrast between x and y. The result of this computation is compared to a user specified threshold. To determine the satisfaction of the preference, we evaluate (7) where l is the required threshold.

$$contrast(c_T, c_B) \geq l \tag{7}$$

B. Resolution algorithm

Different resolution algorithms exist. Choosing an optimization algorithm is justified by the search space width. Properties like contrast or brightness are not hard to compute on a given solution. Nevertheless, the search space width in real cases makes impossible the use of exact algorithms. For example, with 9 color variables (our smallest experiment case), even if we drastically reduce the domain to only 27 colors (3 values for red, green and blue components), we get a search space of about 7.6×10^{12} solutions. Let us remark that with such reduction, the search space may not contain any good solution.

Our problem consists of choosing an adaptation that satisfies several preferences, which can be modeled as a multi-objective optimization problem. In the general case, such problems have a set of solutions, known as Pareto-optimal solutions We are interested in finding a subset of these optimal solutions if any exists, or solutions that approach optimality. For solving such problems, several multi-objective evolutionary algorithms [19], [20], [21] have been proposed as an alternative to costly deterministic methods. Among them, we choose NSGA-II (Non-dominated Sorting Genetic Algorithm-II [22]), which is popular in search based software engineering [23] due to its performances in this domain. Evolutionary algorithms (EAs) mimic the biological evolution of a population with the use of evolution operators that select, cross or mutate individuals.

NSGA-II works as follows (efficient implementation is described in [22]). The algorithm begins with an initial population P_0 of N solutions (or individuals), which can be randomly built. Figure 10 presents the evolution of the population at $t+1$ iteration. P_t corresponds to the population of step t. An offspring population Q_t of size N is created from P_t individuals using selection, crossover and mutation operators. P_t and Q_t are combined to form the population R_t. The N best individuals of R_t in terms of non-dominance and diversity are kept to form P_{t+1}, using the following principles. Several groups of solutions, called fronts, are calculated. The first non-dominated front (F_1) groups non-dominated individuals, corresponding to the best known solutions, with regard to at least one objective. A solution s_1 *dominates* another solution s_2 if: (i) s_1 is no worse than s_2 in all objectives, and (ii) s_1 is strictly better than s_2 in at least one objective. The second non-dominated front (F_2) groups the non-dominated individuals of $R_t \setminus F_1$. The third non-dominated front (F_3) groups the non-dominated individuals of $R_t \setminus (F_1 \cup F_2)$, and so on. The k first fronts whose union has less than N elements are included in the P_{t+1} population. To complete the P_{t+1} population to get N elements, individuals are selected in $k+1$ front, based on a crowding distance [24] and a binary tournament selection operator. The crowding distance helps to select solutions that have the lowest densities of surrounding solutions. For a given solution s, this is measured as the average distance of the nearest solutions (neighbors of s) along each of the objectives. The resulting crowded-Comparison operator helps to select scattered solutions. These steps are repeated until some termination criteria are satisfied, for example when a maximum number of generations has been reached.

In the next section, we present how we tuned NSGA II to solve a set of complex preferences in the context of Web pages.

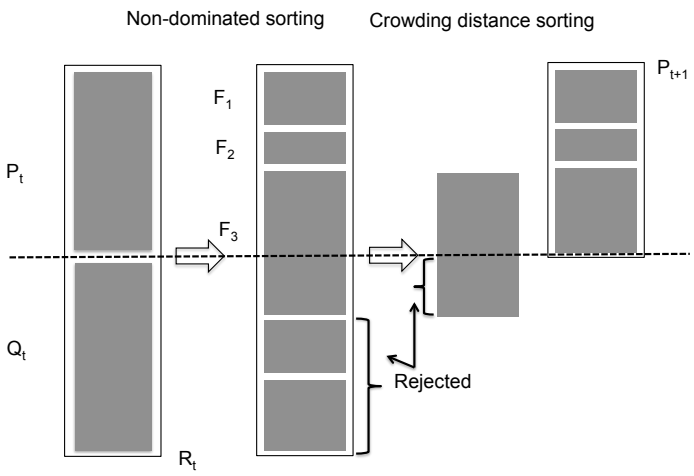


Figure 10. NSGA-II iteration, taken from [22].

IV. Experimental setup

In this section, we describe our experimental setup. As explained in the previous section, due to the nature of the optimization problem, we choose to use the Non-dominated Sorting Genetic Algorithm NSGA-II [22], which we implemented in C++. The feasibility of finding an adaptation, in accordance with user preferences, was already proved in a previous (limited) case study [1]. In the current paper, we extend our case study to five very different real Web pages. We also consider an enhanced set of preferences and we run the algorithms on a larger configuration set. In this experiment, we aim to determine cases where this approach adequately works, while giving trends about the running time, the number of generations and the number of satisfied preferences.

The NSGA-II algorithm relies on three operators. Crossover, mutation and selection operators are applied on a population to generate an offspring. The population size can be parameterized. The crossover and mutation operators are called according to a defined probability. The algorithm iterates through generations until some end criterion is satisfied.

In this setup, we configure the NSGA-II algorithm as follows (Table I). In the literature, the chosen population is commonly about 200. We decided to enlarge the population interval size to analyze the impact of this parameter on our specific problem. The population size can range from 100 to 300 individuals by steps of 50. This gives 5 population sizes for each tested website adaptation. Due to the $O(MN^2)$ complexity of this algorithm, where M is the number of objectives and N is the population size, we need to choose an upper bound for N that is not too high (300). Besides, in the algorithm, N represents the number of simultaneously treated individuals to generate the offspring. Thus, if we have an excessively low N, we considerably decrease interactions between individuals and reduce the population diversity. This is why we choose $N \geq 100$. We choose to set the probability to make a mutation rather

TABLE I. EXPERIMENT SUMMARY.

Parameter	Value(s)
Population Size	{100, 150, 200, 250, 300}
Crossover (resp. Mutation) prob.	0.94 (resp. 0.06)
Max exec time	10s
Max generations	unlimited
Preference sets	4
Maximal brightness difference	40%
Minimal color distance	40°
Minimal contrast	30%
Repeat	60 times

than making a crossover to an invariant value. This choice is based on the wish to have a more readable experimentation. We conduct some very simple tests on different input data, and we chose a probability of 94% for crossover (resp. 6% for mutation).

The non-deterministic base of NSGA-II, and of genetic algorithms in general, potentially induces local stagnation into the search space. Consequently, a part of the executions may stay blocked within a local optimum and will not return a solution in acceptable time. To avoid time explosion of an execution and disturbance in average running time, we limit the maximum execution time for an execution to 10 seconds. This time is justified by the practical end-user context. Adaptations are possibly computed each time the user changes to another page or another website. The non-terminated executions (time over or equal to 10 seconds) are included in time or generation statistics. Thus, when the number of non-terminated executions is high, the average running time tends towards 10 seconds. Besides, these non terminated executions are counted to get a terminated execution ratio and other statistics. The limited execution time is one of the two end criteria in this setup. The second end criterion is the global preference satisfaction level. For this experiment, we also decided to exit from the algorithm when we obtain a solution on which each preference is satisfied. Here we do not use the number of generations as an end criterion, we only get it for statistics. To summarize, we stop the algorithm if we find a good solution or if the execution exceeds 10 seconds.

NSGA-II allows us to have non-dominated solutions regarding several objectives. For this experiment, we define three general preferences, which correspond to practical problems of people with low vision:

- GP_1 : Uniform background color brightness.
- GP_2 : Minimal contrast between the text and its direct background.
- GP_3 : Keep original color for modified elements.

Here GP_1, GP_2 and GP_3 are general preferences. When we associate a general preference GP_i to a specific website, we obtain a set of preferences P_i. For example, for the Parempuyre website, which is part of our experiment, we have variables for the three backgrounds, including $inputBckClr$ (input background color), $leftHdBckClr$ (left header background color), and $bodyBckClr$ (body background color), and variables for text colors including $leftLkClr$ (left link color) and $hdClr$(text header color). The preference sets corresponding to the general preferences are as follows:

$$P_1 = \{proxBrightness(leftHdBckClr, bodyBckClr, inputBckClr)\}$$
$$P_2 = \{contrast(leftHdBckClr, hdClr) >= 30, contrast(bodyBckClr, leftLkClr) >= 30, ...\}$$
$$P_3 = \{distOrigClr(leftHdBckClr) < 40, distOrigClr(bodyBckClr) < 40, ...\}$$

The P_1 preference set contains a global preference given by the user to have a brightness that is comfortable for the eyes for all backgrounds on the Web page. This global preference is represented by an objective function, which refers to all backgrounds. The objective function may concern a large part of the variable set and often leads to huge dependencies between many variables. If we represent preferences and their dependencies with a graph, with only unary or binary objective functions, we may often have several connected components. When using a global objective function, we tend to group many connected components to a larger single connected component. To compute this objective function, we begin by computing the brightness of every background using indications of [3] (brightness values are in $[0, 1]$). Then we compare the difference $d = max - min$ between the minimal brightness min and maximal brightness max to a defined threshold $maxDiff$. $prefNb$ is the number of preferences used in the current preference set in a case study.

The objective function for P_1 is thus computed as follows:

$$If\ d >= maxDiff$$
$$return\ 0 + (1 - d)/prefNb$$
$$else\ return\ 1 + (1 - d)/prefNb$$

The two other preference sets are also based on colors. The P_2 preference set includes pairs composed of a text color and background color. These pairs of colors are linked to ensure minimal brightness contrast, suitable for readability. To compute the associated objective function, we compute the brightness of both colors and the contrast c using indications of [3]. The contrast is computed from the relative perceived brightness of the two colors. The WCAG computed values are in $[1, 21]$ and we normalize them to have values in $[0, 1]$. We compare the contrast to a minimum $minContrast$, which is set at 0.3. This value is slightly higher than the standard requirement in the Web Content Accessibility Guidelines (WCAG 2.0) for the AA level.

The objective function for P_2 is thus computed as follows:

$$If\ c >= minContrast$$
$$return\ 1 + (c - minContrast)/prefNb$$
$$else\ return\ 0 + (c - minContrast)/prefNb$$

The last preference set, P_3, allows the algorithm to keep as much as possible the original color context of the page after the adaptation. These preferences concern only one variable. P_3 corresponds to the designer preferences, because these preferences are concerned by the proximity between colors in adapted pages and initial colors. These preferences are binary

preferences, because they refer to two different color variables to be able to compute the contrast. For color distance preference (P_3), we compare the hue of the two colors $h_{original}$ and $h_{current}$. From the representation of the two colors in the chromatic circle, the color distance is the angle between the two hues. We set the maximal accepted angle $angleMax$ at 40°. To obtain the hue (in $[0, 360]$) of a given color from its RGB (Red Green Blue color model) components, we use the computation of the hue from the classical conversion function from RGB to HSV (Hue-Saturation-Value) color spaces [25].

The objective function for P_3 is thus computed as follows:

$$
\begin{aligned}
&hd = |h_{original} - h_{current}| \\
&if\ hd\ >= 180\ then\ hd = 360 - hd \\
&if\ hd\ <\ angleMax \\
&return\ 1 + (hd/180)/prefNb \\
&else\ return\ 0 + (hd/180)/prefNb
\end{aligned}
$$

We compose the three initial preference sets to obtain the four preference sets of our experiment.

$$S_1 = P_2 \tag{8}$$

$$S_2 = P_2 \cup P_3 \tag{9}$$

$$S_3 = P_1 \cup P_2 \tag{10}$$

$$S_4 = P_1 \cup P_2 \cup P_3 \tag{11}$$

The first preference set S_1 corresponds to only one general preference. It corresponds to a user that has issues with readability when contrast is low. The second preference set S_2 groups P_2 and P_3 preferences. In this configuration, the user possibly wants to increase contrast on the page (if the current contrast is not suitable) but he does not want to completely change the original page colors. The user may want to keep the color context close to the original one in order to recognize the browsed Web page or to avoid getting lost in the navigation (because he memorizes colors assigned to some parts of the web page). The third preference set S_3 is suitable for people who have a disease involving major light sensitivity. They need to have minimal contrast between the text and the direct background to improve readability, and also have similar background brightness to avoid dazzle. The last preference set S_4 is a complex case of adaptation, which combines the three preference sets. It also corresponds to existing real low-vision user needs.

We implement the selection, crossover and mutation operators to make the population evolve. The selection operator is based on the classical crowded-comparison operator. This comparison is used after applying the fast non-domination sort and with crowding distances assigned to each individual [22]. The selection operator keeps the diversity in the population.

We implement classical crossover and mutation operators. Mutation modifies part of an individual or an entire individual. Each component of the child individual has a ½ probability of getting a random value. In other ½ cases, the child keeps the parent value for this component. For an individual with x components we have a probability of $(½)^x$ either keeping all of them or changing all of them. For example, an individual with 4 components has a 0.0625 probability of completely changing or of being identical to the parent, with 8 components we have a 0.0039 probability. This probability is approximated, we do not care about the probability of getting the original value when we get a random value (in huge domains like color space, it is very low). Thereby in most cases a mutation keeps some parent components in the child.

The crossover operator also uses probabilities. However, instead of getting a random value for some components, we only get parent values. From two parents, the crossover operator generates one child. Each component of this child has a ½ probability of come-ing from the first parent. In other cases, the component value comes from the second parent. As in the mutation operator, the probability of entirely copying one parent is low when the component number is high.

Domains of variables may have several dimensions. For example, the text size variable domain generally has about 10 or 20 values. By contrast, the color variable domain can reach 2^{24} values in a true context. The number of variables depends on the type and complexity of the given user's preferences. All preferences defined in this setup are based on color variables. Each color variable is implemented with an RGB color space. This choice is based on the large size of the color domain. Each component of the implemented RGB color space is set with the sRGB (standard RGB) values of the Website. The domain of the web sRGB color space is about 16.7 million colors, but we use only 32.768 colors in the domain. Many picture editing software systems have a Web color mode. This mode contains substantially fewer colors in the domain. However, experimentation shows that limiting the color domain may have a negative impact on the resolution by the algorithm. If the number of colors is too low, some problems may have no solutions. Research also shows that the domain size has no significant effect. Increasing or reducing the search space does not change execution times because it is an optimization problem, thus it does not examine the entire solution set. We decide to have 32 values in each component (R, G and B) to not limit the number of solutions, while avoiding a huge search space. This size can be discussed and adapted to optimize color features.

To study the effect of the number of objectives, we tested two different objective function sets. We recall that the NSGA-II algorithm is a multi-objective algorithm that exploits non-domination sorting and has a complexity of MN^2, where M is the number of objectives functions. P_2 and P_3 preference sets are used in two different manners to define the objective functions. In the first case, for each preference, we make one objective function, which returns a floating point value equal to or higher than one when the preference is satisfied. In the second case, for each preference set (P_2 or P_3), we make two objective functions. Each of these two objective functions groups half of the preferences, which are randomly chosen.

Like the elementary objective functions, an aggregated objective function returns a floating value in $[0, 1 + 1/prefNb]$, which evaluate the quality of one solution accordingly to all embedded preferences. An aggregated objective function computes the average objective functions associated with the embedded preferences. The formulas of objective functions for P_1, P_2, P_3 have been designed to prevent compensation for

objective function values and prevent acceptance of a solution as a good one when one of its elementary objective functions is not satisfied.

To illustrate our chosen representation, the "Facebook" registration page will be represented by about 40 variables to implement the preference: "contrast between text and direct background higher or equal to x". For this preference we have to extract all text elements and their direct backgrounds, making color variables. For the Google search page (not the result page), we obtain about 17 variables, while for the "BBC News" home page we obtain about 200 variables. These values are rounded because large parts of websites are dynamic and regularly change (these computations were done on November, 14, 2013). The five websites were selected for the diversity of their architecture, their number of objects and the colors used.

For the first tested website, we combine the two possible variations on population size and objective function number. For the other websites, we set the population size at 250. In a $configuration$, we set a population size and a number of objective functions for a specific website. The execution of each configuration is repeated 60 times. This number of executions may seem high, but this algorithm is highly stochastic: the higher the execution number, the more precise the results are. Of course, we need to make a trade-off because a higher execution number leads to a smoother result, but it also corresponds to a very long experimentation time.

In the next section, we report and discuss the results obtained with all the previously described configurations on the selected websites.

V. Experiment results

In this section, we present the results obtained using the experimental framework presented above (Section IV), with our C++ implementation of NSGA-II. We selected websites for their diversity regarding the number of objects and colors on the page. These websites have no specific global theme or architecture. The chosen websites have been highlighted by users with visual impairment as websites with accessibility issues.

A. Parempuyre website

The "Parempuyre" website is our first example. It is a rather small website with a dozen color variables. The variables are text colors or background colors. They are selected from the previously defined preferences. We present the results for the non-aggregated case, and then for the aggregated case.

1) Parempuyre website without aggregation: For S_1 preferences that concern the minimal brightness contrast between a text color and its direct background color, we get all relevant textual elements from the Web page. One textual element associated with its foreground color property gives one variable. Other variables are also associated with background colors (three in this case). We run the algorithm with all preference sets and all configurations. The detailed results for the S_1 preference set is shown in Table II.

We recall that the presented figures are the result of the average of 60 executions. We have 3 backgrounds and 9 texts, giving 12 variables when associated with their color property. We have 9 contrast constraints. All executions terminate before 10 seconds. This means that the algorithm always finds a solution satisfying all defined preferences in the S_1 set. The population size has an impact on the number of generations needed to find these solutions: the higher the population size, the lower the number of generations. We find a satisfactory solution on average in less than 430 milliseconds. The maximal time for discovering a good adaptation is less than one second, and in the best cases only 50 milliseconds are needed. The generation number goes from 7 to 88, and is less than 37 on average. In this case study, the minimal requested contrast is 30%, which gives, for each constraint, 70% of satisfactory values in the search space.

TABLE II. "PAREMPUYRE" S_1 - EXPERIMENT FIGURES.

Variables	Prefs – obj. functions	search space
12	9 – 9	$\simeq 2 \times 10^{54}$

Pop. size	100	150	200	250	300
Exec. time (s)					
Average	0.18	0.24	0.29	0.42	0.43
Standard deviation	0.08	0.09	0.10	0.19	0.12
Min	0.05	0.06	0.12	0.18	0.22
Max	0.42	0.52	0.57	1.11	0.78
Generations					
Average	37	27	21	20	15
Standard deviation	17.1	10.5	7.0	9.1	4.4
Min	11	7	9	9	8
Max	88	58	40	54	28
Non cpltd. exec.					
Percentage	0	0	0	0	0
Satisfied obj. fct (avg.)	–	–	–	–	–
Standard deviation	–	–	–	–	–
Min	–	–	–	–	–
Max	–	–	–	–	–

On the contrary, the executions with the S_2 preference set give worse results (Table III). We still have the 12 variables representing objects associated with color properties. In addition to the minimal contrast preferences, we have a constraint in order to keep colors as close as possible to the original colors. There are 21 preferences on variables: 9 contrast preferences and 12 color closeness preferences. In this configuration, all preferences are implemented as objective functions, thus there are also 21 objective functions. With this constraint of staying close to the original colors, we considerably reduce the number of good solutions in the entire search space.

In this configuration, with these 21 objective functions, no execution returns a good solution. This is mainly due to the low number of good solutions regarding the search space. Adjusting the contrast means changing the color of one variable that also impacts its closeness to the original color. Even if the algorithm does not return a good adaptation, about 6 of the 20 objective functions are completely satisfied when we reach 10 seconds. The other objective functions (non-satisfied objective functions) may not have such bad values. This phenomenon is due to the fact that, NSGA-II is an elitist

TABLE III. "PAREMPUYRE" S_2 - EXPERIMENT FIGURES.

Variables	Prefs – obj. functions	search space
12	21 – 21	$\simeq 2 \times 10^{54}$

Pop. size	100	150	200	250	300
Exec. time (s)					
Average	10.00	10.00	10.00	10.00	10.00
Standard deviation	0.00	0.00	0.00	0.00	0.00
Min	10.00	10.00	10.00	10.00	10.00
Max	10.00	10.00	10.00	10.00	10.00
Generations					
Average	1,123	582	374	274	211
Standard deviation	14.8	13.5	12.6	12.8	15.4
Min	1,093	554	346	244	179
Max	1,156	614	409	313	256
Non cpltd. exec.					
Percentage	100	100	100	100	100
Satisfied obj. fct (avg.)	5.7	6.2	6.4	6.6	6.8
Standard deviation	1.7	1.7	1.9	1.6	1.7
Min	3	3	3	3	3
Max	11	10	11	11	11

TABLE IV. "PAREMPUYRE" S_3 - EXPERIMENT FIGURES.

Variables	Prefs – obj. functions	search space
12	10 – 10	$\simeq 2 \times 10^{54}$

Pop. size	100	150	200	250	300
Exec. time (s)					
Average	1.10	0.49	0.56	0.64	0.70
Standard deviation	2.15	0.47	0.46	0.41	0.36
Min	0.04	0.09	0.17	0.18	0.22
Max	10.00	2.29	2.58	3.49	2.72
Generations					
Average	196	53	38	30	24
Standard deviation	339.1	49.6	29.3	18.6	12.0
Min	8	10	12	9	8
Max	1,599	240	168	157	91
Non cpltd. exec.					
Percentage	5	0	0	0	0
Satisfied obj. fct (avg.)	4.7	–	–	–	–
Standard deviation	1.2	–	–	–	–
Min	3	–	–	–	–
Max	6	–	–	–	–

algorithm, it selects and keeps the best evaluated solutions for all objectives even if they are not satisfied.

In Table IV, we present figures of executions with the S_3 preference set. In this preference set, we introduce a "global" preference. By "global" preference, we mean a preference that covers all background color variables. Here this preference is an n-ary predicate involving the three background color variables. Introducing this preference concerning the uniformness of the three background color brightnesses results in adding dependencies with the contrast preferences. In this "Parempuyre" website configuration, the NSGA-II algorithm returns good solutions before 10 seconds for nearly all executions, even if the number of good solutions in the search space is reduced by the constraints.

TABLE V. "PAREMPUYRE" S_4 - EXPERIMENT FIGURES.

Variables	Prefs – obj. functions	search space
12	22 – 22	$\simeq 2 \times 10^{54}$

Pop. size	100	150	200	250	300
Exec. time (s)					
Average	10.00	10.00	10.00	10.00	10.00
Standard deviation	0.00	0.00	0.00	0.00	0.00
Min	10.00	10.00	10.00	10.00	10.00
Max	10.00	10.00	10.00	10.00	10.00
Generations					
Average	944	495	322	234	181
Standard deviation	149.3	76.4	47.8	31.1	26.8
Min	755	391	247	174	136
Max	1,100	600	389	280	234
Non cpltd. exec.					
Percentage	100	100	100	100	100
Satisfied obj. fct (avg.)	6.4	6.4	6.9	7.3	6.9
Standard deviation	1.5	1.5	1.7	1.8	1.6
Min	4	4	3	4	4
Max	10	11	12	12	12

Most of the executions return a good solution in less than 1 second on average. For a population of 100 individuals, we exceed 1 second, due to the 5% of non-terminated executions that add 10 s to the average time computation. The best executions return a good adaptation only in a few generations (less than 53). However, while with S_1 preference set less than 100 generations are needed in the worst execution to find a good solution, here the worst executions need up to 1,599 generations. That has an impact on the average execution time. It tends to be seven times more regarding the values obtained with S_1. Returning a good adaptation solution may take up to a second compared to less than 450 milliseconds in S_1 configuration (on average).

In the last case (with the S_4 preference set), we apply all preferences: minimal brightness color contrast, preservation of the original color context and uniformity of background brightness color (Table V). Independently of the chosen website, this configuration is the most complex configuration of our experiment. The association of many preferences may give many dependencies between variables. Here again, the global background uniform color brightness constraint tends to group many variables mainly if there are many background variables.

It is worth noting that even when we complicated the preference set compared to S_2, the number of satisfied objective functions is slightly higher. We apply three preferences instead of only two but, on average, when the maximal execution time is reached, we have satisfied one more objective function. Moreover, the worst executions discovered solutions with up to 12 satisfied objective functions among the 22 of this problem. In other words, while adding one more preference, we increase the number of satisfied objective functions. In some cases, the fact of having more preferences and finally more objective functions may help and guide the algorithm to find a good adaptation solution more rapidly. Some implicit rules are somehow given to the algorithm to converge to a good solution.

In all of the previous cases, each preference was represented

TABLE VI. "PAREMPUYRE" S_{1aggr} - EXPERIMENT FIGURES.

Variables	Prefs – obj. functions	search space
12	9 – 2	$\simeq 2 \times 10^{54}$

Pop. size	100	150	200	250	300
Exec. time (s)					
Average	0.08	0.26	0.15	0.20	0.27
Standard deviation	0.09	1.27	0.03	0.04	0.05
Min	0.03	0.07	0.09	0.07	0.17
Max	0.74	10.00	0.22	0.29	0.43
Generations					
Average	12	21	8	7	7
Standard deviation	14.1	101.7	1.6	1.4	1.2
Min	5	6	5	3	5
Max	114	803	12	11	12
Non cpltd. exec.					
Percentage	0	2	0	0	0
Satisfied obj. fct (avg.)	–	1.0	–	–	–
Standard deviation	–	0.0	–	–	–
Min	–	1	–	–	–
Max	–	1	–	–	–

TABLE VII. "PAREMPUYRE" S_{2aggr} - EXPERIMENT FIGURES.

Variables	Prefs – obj. functions	search space
12	21 – 4	$\simeq 2 \times 10^{54}$

Pop. size	100	150	200	250	300
Exec. time (s)					
Average	10.00	9.35	7.67	6.83	5.63
Standard deviation	0.00	2.43	4.08	4.49	4.67
Min	10.00	0.15	0.26	0.37	0.43
Max	10.00	10.00	10.00	10.00	10.00
Generations					
Average	1,819	900	471	292	166
Standard deviation	12.3	233.6	249.3	190.5	135.2
Min	1,797	16	18	18	15
Max	1,859	979	621	433	300
Non cpltd. exec.					
Percentage	100	94	75	67	54
Satisfied obj. fct (avg.)	1.4	1.6	1.6	1.9	1.9
Standard deviation	0.5	0.6	0.6	0.7	0.7
Min	1	1	1	1	1
Max	2	3	3	3	3

by one objective function. For the contrast constraint between the text and its direct background, each pair of text color variable and its associated background color variable corresponds to one objective function. We may obtain different results if we aggregate (see Section IV) objective functions. Instead of having one objective function for each contrast constraint, we group them into aggregated objective functions.

2) Parempuyre website with aggregation: The next tables present the same four preference sets, but now with aggregations of objective functions.

There are significant differences in the results obtained with the first preference set S_1, as shown in Table VI. The algorithm spends up to 270 milliseconds to return a good solution instead of up to 430 milliseconds. The time tends to be doubled when we do not use an aggregation configuration (with aggregated objective functions). Note that the number of generations is really stable, except for the 100 and 150 individual population. Note also that there is 2% of non-completed executions with a population of 150 individuals. The number of objective functions is reduced to only 2, which has an important role in the time needed to compute each generation. Excluding the population of 100 and 150 individuals, we only need 3 generations in the best cases and 12 generations in worst cases. In general, executions spend less time than in the non-aggregated S_1 configuration and need only a few generations.

In the S_2 preference set, the aggregation is much more useful: let us recall that in non-aggregated configurations, no execution terminates before 10 seconds.

Aggregation allows the algorithm to find a good adaptation in up to 46% of executions (with a population of 300 individuals). Whereas with the non-aggregated configuration (Table III), all executions reach 10 seconds without returning any solution that completely satisfies all preferences. With the aggregation, we highly improve the situation for the S_2 configuration. This example highlights the importance of the population size. This size plays an important role in the number of terminated executions. With a population of 300 individuals, we saw that 46% of executions returned one good solution, whereas with a population of 200 we have 25% of terminated executions, and we have 0% with a population of 100 individuals. Terminated executions return a good adaptation in about half a second.

TABLE VIII. "PAREMPUYRE" S_{3aggr} - EXPERIMENT FIGURES.

Variables	Prefs – obj. functions	search space
12	10 – 3	$\simeq 2 \times 10^{54}$

Pop. size	100	150	200	250	300
Exec. time (s)					
Average	0.05	0.08	0.11	0.16	0.19
Standard deviation	0.03	0.02	0.02	0.05	0.03
Min	0.03	0.05	0.04	0.09	0.10
Max	0.25	0.21	0.16	0.54	0.26
Generations					
Average	10	9	8	8	7
Standard deviation	6.8	2.6	1.6	2.7	1.2
Min	6	6	3	5	4
Max	52	24	12	27	10
Non cpltd. exec.					
Percentage	0	0	0	0	0
Satisfied obj. fct (avg.)	–	–	–	–	–
Standard deviation	–	–	–	–	–
Min	–	–	–	–	–
Max	–	–	–	–	–

Table VIII shows aggregation results for the S_3 preference set on the "Parempuyre" website. These executions are made with objective function aggregations.

While in the case of the aggregation of the S_1 preference set we reduced the execution time by twofold, here the impact is even more significant. On average, the aggregation allows us to reduce the execution times by threefold. The aggregation of objective functions for S_3 has a very positive impact (Table

TABLE IX. "PAREMPUYRE" S_{4aggr} - EXPERIMENT FIGURES.

Variables	Prefs – obj. functions	search space
12	22 – 5	$\simeq 2 \times 10^{54}$

Pop. size	100	150	200	250	300
Exec. time (s)					
Average	9.84	9.03	7.67	7.03	5.78
Standard deviation	1.26	2.91	4.07	4.37	4.54
Min	0.18	0.23	0.33	0.45	0.61
Max	10.00	10.00	10.00	10.00	10.00
Generations					
Average	1,676	812	440	281	161
Standard deviation	214.2	261.0	231.6	172.0	122.8
Min	34	24	22	21	21
Max	1,741	917	581	403	281
Non cpltd. exec.					
Percentage	99	90	75	69	54
Satisfied obj. fct (avg.)	2.3	2.7	2.8	2.9	3.0
Standard deviation	0.7	0.7	0.6	0.6	0.5
Min	1	2	2	2	2
Max	4	4	4	4	4

VIII).

In the case of S_4, the aggregation now allows us to have terminated executions leading to good adaptation, as is already the case for the S_2 aggregation case. Moreover, the population size seems to again have its importance in the number of terminated executions. With a population of 100 individuals, only 1% of executions are not interrupted before 10 seconds. A population of 200 individuals allows us to obtain 25% of terminated executions and we reach 46% when the population size is equal to 300 individuals. However, many executions remain not completed. In this last case, nearly 60% of objective functions are completely satisfied.

B. Godaddy website

The "Godaddy" website has more objects on the page and finally has more variables. From previously chosen preferences, 22 color variables are created to solve the problem. Table X gives results of the application of all the preference sets. For each preference set, we use a population size of 250 individuals. The population size has an impact on the computation efficiency, but this impact is not necessarily similar for all configurations. A population of 100 individuals sometimes leads to unusual results. The population of 250 seems to be a good trade-off between the execution time and the number of completed executions. We associate with each preference set its aggregation version in order to easily compare them.

For the S_1 preferences set, we can improve the computation efficiency and reach 100% of completely satisfactory solutions returned with the aggregation of objective functions. Without aggregation, not all executions give a completely satisfying solution. Some of them are interrupted by the end time criterion. When executions are not completed, on average, returned solutions satisfy a good part of all objective functions (10 out of 14). In best cases, up to 12 objective functions out of 14 are satisfied. The aggregation configuration allows us to reach 100% of executions, leading to good solutions. Moreover, the execution time is highly reduced. The algorithm returns a good adaptation solution in about 300 milliseconds instead of 5 seconds. The number of generations is closely linked to the running time. The generation number is also, like the execution time, highly reduced. We need less than 20 generations to get one good adaptation solution. Without aggregation, it takes 215 generations instead on average.

Similar to this case, with the S_3 preference set, we get a major improvement. From 18% of completed executions without any aggregation, we get 98% of completed executions with an aggregation configuration. Moreover, the running time for the executions is considerably reduced. We only need 500 milliseconds on average to get one good adaptation solution as compared to 9.67 seconds on average without an aggregation configuration. As in the previous example, the number of generations is also highly reduced on average. We need about 23 generations on average as compared to 358.

For the S_1 and S_3 preference sets, the aggregation achieves a substantial improvement in execution time and in the quality of the adaptation returned. It allows us to obtain adaptation solutions for a real-time use. Unfortunately, it does not improve all configurations. The complexity of the two other configurations is too high to get good adaptation solutions with or without aggregation.

In the S_2 preference set configuration, none of the executions gives a completely satisfactory solution even with aggregation. Without aggregation, we satisfy at least 5 of the 36 objective functions and up to 11 of the 36 objective functions in the best cases. With aggregation, we do not satisfy any objective function in the worst cases, whereas in the best cases we satisfy 3 of the 4 objective functions. Similarly, in the S_4 preference set configuration, we satisfy on average about 8 out of 37 and 1.6 out of 5 objective functions respectively, for non-aggregation and aggregation configurations. In the best cases, we satisfy 12 out of 37 objective functions without any aggregation as compared to 2 out of 5 with aggregation. Finally, in the worst cases, only 1 out of 5 objective functions are completely satisfied in the aggregation configuration and 4 out of 37 with aggregation.

In these two last cases (S_2, S_4), from the results obtained, the aggregation of objective functions seems to have no positive impact on the execution of the NSGA-II algorithm. However, these figures do not especially mean that the aggregation only improves simple configurations. Another aggregation (objective function repartition) with more aggregated objective functions for example, may give different results. In all cases, with less complex configurations, aggregation has a notable positive impact on the quantity of the solutions returned.

C. ReasonFrance and UBOLEA websites

The "ReasonFrance" website (Table XII) and the "UBOLEA" website (Table XI) are the two largest websites in our experiment. The two tables show the results for all preference sets and configurations with a population size of 250, with and without aggregation. With 27 variables, the "UBOLEA" website returns some good adaptations for the

TABLE X. "GODADDY" ALL PREFERENCES - EXPERIMENT FIGURES.

Variables		search space
22		2×10^{99}

Prefs. set	S_1	S_{1aggr}	S_2	S_{2aggr}	S_3	S_{3aggr}	S_4	S_{4aggr}
Preferences	14	14	36	36	15	15	37	37
Objective functions	14	2	36	4	15	3	37	5
Exec. time (s)								
Average	5.28	0.32	10.00	10.00	9.67	0.50	10.00	10.00
Standard deviation	1.61	0.03	0.00	0.00	0.98	1.24	0.00	0.00
Min	3.13	0.24	10.00	10.00	5.18	0.25	10.00	10.00
Max	10.00	0.39	10.00	10.00	10.00	10.00	10.00	10.00
Generations								
Average	215	15	197	400	358	23	191	372
Standard deviation	61.7	1.5	8.3	4.2	38.0	52.5	10.1	2.5
Min	133	12	184	373	206	12	176	365
Max	398	19	222	407	407	426	232	378
Non cpltd. exec.								
Percentage	5	0	100	100	82	2	100	100
Satisfied obj. fct (avg.)	10.0	–	7.8	1.0	9.9	2.0	7.8	1.6
Standard deviation	2.8	–	1.5	0.9	3.9	0.0	1.7	0.5
Min	6	–	5	0	3	2	4	1
Max	12	–	11	3	14	2	12	2

S_1 case. For the other preference sets, we do not get any good adaptations. In the S_1 case, the aggregation provides substantial improvement. It allows us to increase the number of terminated executions by up to 86%, whereas without any aggregation it is about 6%. Moreover, it allows us to considerably reduce the execution time from nearly 10 seconds to less than 2 seconds. The execution times that enable good solutions is around the minimal execution time (about 500 milliseconds). The minimal execution time in the best cases is about 7 seconds, but only about 350 milliseconds when aggregation is used. In the same way, the average number of generations is reduced from 323 to 77 generations. The minimal number of generations is also highly reduced when we use aggregation. The maximal number is still high because at least one execution is not completed and gives a high number of generations.

The other three preference sets (S_2, S_3, S_4) do not return any good adaptation in less than 10 seconds. Nevertheless, aggregation gives a slight improvement in the number of satisfied objective functions. For instance, in the S_2 preference set, on average, 8.9 of the objective functions are satisfied out of the 48 objective functions in the non-aggregated configuration. This ratio becomes 0.7 out of 4 satisfied objective functions when we aggregate them. The number of satisfied objective functions is thus equivalent from about 18.5% to 17.5%. In the S_3 preference set, the number of satisfied objective functions is increased from about 16.8% to 33%, and it is increased from 18.8% to 26% in the S_4 preference set case.

For the "ReasonFrance" website, the behavior of executions with or without any aggregation is quite similar. The results are improved for only the first preference set S_1 when using aggregation, allowing us to obtain up to 60% of terminated executions. A good solution is returned in less than 5 seconds with aggregation and some terminated executions provide a good solution in less than 1 second. In best cases, this time is about 750 milliseconds. In all preference sets (except S_2), the aggregation allows us to really improve the number of satisfied objective functions even if no good solution is found.

D. NFB website

The last studied website is presented in Section II. It corresponds to the publications Web page of the National Federation of the Blind (NFB, nfb.org/publications, may 2014).

1) Execution results: We applied the same preference sets, again with a population size of 250. Table XIII presents figures on all of these executions.

Good adaptations usually correspond to aggregated configurations. Without aggregation, in the most complex problems (S_2 and S_4), the algorithm gives a few solutions (S_2) or none (S_4). In two other cases, with or without aggregation, all executions terminate before 10 seconds.

The S_1 preference set gives quite good adaptations in less than 200 milliseconds. The aggregation configuration for the S_1 preference set only achieves negligible improvement. In this second configuration, the execution time falls to 100 milliseconds. Furthermore, the average number of generations and the maximal number of generations is about twofold lower. On the same website, the S_3 preference set gives good results. Without any aggregation, a good adaptation is computed on average in about 260 milliseconds versus 130 milliseconds with aggregation. The aggregated configuration thus brings an improvement, while also reducing the average number of generations and the maximal number of generations by about twofold. In the aggregated configuration or in the non-aggregated configuration, all executions lead to a good adaptation in less than 400 milliseconds.

Solving the S_2 preference set is highly facilitated by aggregating objective functions. When only six percent of the

TABLE XI. "UBOLEA" ALL PREFERENCES - EXPERIMENT FIGURES.

Variables		search space
27		8×10^{122}

Prefs. set	S_1	S_{1aggr}	S_2	S_{2aggr}	S_3	S_{3aggr}	S_4	S_{4aggr}
Preferences	21	21	48	48	22	22	49	49
Objective functions	21	2	48	4	22	3	49	5
Exec. time (s)								
Average	9.91	1.70	10.00	10.00	10.00	10.00	10.00	10.00
Standard deviation	0.45	3.26	0.00	0.00	0.00	0.00	0.00	0.00
Min	6.74	0.34	10.00	10.00	10.00	10.00	10.00	10.00
Max	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Generations								
Average	323	77	176	375	250	417	171	351
Standard deviation	24.0	147.6	7.1	2.8	24.3	7.6	6.3	5.9
Min	242	16	161	368	210	365	161	310
Max	364	459	208	381	346	425	196	358
Non cpltd. exec.								
Percentage	94	14	100	100	100	100	100	100
Satisfied obj. fct (avg.)	7.4	1.0	8.9	0.7	3.7	1.0	9.2	1.3
Standard deviation	4.2	0.0	1.6	0.6	1.4	0.3	1.4	0.6
Min	1	1	6	0	2	0	6	0
Max	19	1	13	2	7	2	12	3

TABLE XII. "REASONFRANCE" ALL PREFERENCES - EXPERIMENT FIGURES.

Variables		search space
59		3×10^{266}

Prefs. set	S_1	S_{1aggr}	S_2	S_{2aggr}	S_3	S_{3aggr}	S_4	S_{4aggr}
Preferences	41	41	100	100	42	42	101	101
Objective functions	41	2	100	4	42	3	101	5
Exec. time (s)								
Average	10.00	4.51	10.00	10.00	10.00	10.00	10.00	10.00
Standard deviation	0.00	4.49	0.00	0.00	0.00	0.00	0.00	0.00
Min	10.00	0.74	10.00	10.00	10.00	10.00	10.00	10.00
Max	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Generations								
Average	271	175	198	317	312	348	193	293
Standard deviation	18.6	173.9	11.1	1.6	15.8	5.9	14.0	1.8
Min	230	29	171	313	279	307	167	289
Max	318	397	225	321	360	356	229	297
Non cpltd. exec.								
Percentage	100	40	100	100	100	100	100	100
Satisfied obj. fct (avg.)	4.7	1.0	16.6	0.0	8.0	1.3	16.8	0.9
Standard deviation	1.6	0.2	2.3	0.1	3.2	0.7	2.3	0.4
Min	1	0	12	0	2	0	13	0
Max	9	1	22	1	17	2	23	2

executions without aggregation lead to good adaptations in nearly 10 seconds, the aggregation configuration achieves 70% of terminated executions in less than 4 seconds. In addition, the average number of generations is 156 in the aggregated configuration, while it is 325 generation without. Solving the last preference set (S_4) is also highly facilitated when using aggregation. Aggregation allows us to increase by up to 60% the number of terminated executions before 10 s, compared to 0% without any aggregation.

2) Visual examples of computation: We randomly chose solutions returned by our implementation of NSGA-II algorithm and we manually applied them on the NFB website. We recall that the original appearance of this website is shown in Figure 1. We hereafter discuss the displayed results.

The first preference set S_1 includes preferences that guarantee minimal brightness contrast between the text and its direct background. The minimal required contrast is slightly over the contrast recommended by the Web Content Accessibility Guidelines (WCAG 2.0). Figure 11a is the application of one computed adaptation. In the left menu, in top navigation bar and the content, the contrast is not especially high, but it satisfies the minimal requested contrast. The first part of the menu is somewhat light green and the second part yellow. These two colors are far from the original colors. The same

TABLE XIII. "NFB" ALL PREFERENCES - EXPERIMENT FIGURES.

Variables		search space
9		4×10^{40}

Prefs. set	S_1	S_{1aggr}	S_2	S_{2aggr}	S_3	S_{3aggr}	S_4	S_{4aggr}
Preferences	5	5	14	14	6	6	15	15
Objective functions	5	2	14	4	6	3	15	5
Exec. time (s)								
Average	0.16	0.10	9.73	3.49	0.26	0.13	10.00	4.66
Standard deviation	0.05	0.02	1.17	4.37	0.05	0.02	0.00	4.61
Min	0.07	0.05	2.54	0.21	0.13	0.07	10.00	0.30
Max	0.30	0.14	10.00	10.00	0.39	0.20	10.00	10.00
Generations								
Average	8	5	325	156	13	7	325	184
Standard deviation	2.3	1.1	37.3	192.7	2.5	1.2	13.9	181.3
Min	4	3	112	11	7	4	298	15
Max	16	8	385	447	20	11	367	424
Non cpltd. exec.								
Percentage	0	0	94	30	0	0	100	40
Satisfied obj. fct (avg.)	–	–	8.3	2.1	–	–	7.9	2.9
Standard deviation	–	–	1.8	0.4	–	–	1.6	0.8
Min	–	–	5	1	–	–	5	2
Max	–	–	12	3	–	–	13	4

applies for the color of the top navigation bar and the color of the main content. The computed colors only satisfy the minimal brightness contrast constraint.

When we add to the minimal brightness contrast constraint and the color closeness constraint, we reduce the possible colors for each object. Figure 11b shows the application of an adaptation computed using the S_2 preference set. If we consider the left menu, the new colors of the background and of the text allow us to have the minimal required contrast. Moreover, the background colors are not too far from the original colors chosen by the designer. We have a problem with the top navigation menu. This is normal regarding our implementation of the distance between two colors. We will explain this effect in the application of the S_4 preference set.

In Figure 11c, instead of adding the color closeness constraint, we add the uniform background brightness constraint. This constraint allows us to have near light emission for all backgrounds on the page to avoid dazzle. Like in the first example (Figure 11a), colors are randomly chosen during the computation process. However, the minimal required contrast is respected, and all backgrounds are dark. Thus, the uniform background brightness constraint is also satisfied.

With all preferences (S_4 preference set), the result is not totally satisfactory. The minimal required contrast is respected. The near background brightness is also satisfied. Unfortunately, the closeness to the original colors is not very good. As already noted in the S_2 example, some colors are far from the original colors. This is especially the case for grey-based colors, including black and white. The color difference is revealed by comparison of the hue of the two colors. The hue of the black, the white and all grey colors is actually equal to 0 (there is no hue). Colors have hues from 0 to 360 degrees, thus, colors with a degree of around 0 are considered close to all grey scale colors. The white background of the main content can thus to be transformed into a dark colored background. In a future work, we will study how to change the implementation of our color distance function to avoid such effects.

E. Threats to validity

1) Construct validity: We have considered several configurations to apply the NSGA-II algorithm, especially for analyzing the effect of the population size and the potential benefit or drawback of aggregated objective functions. During the experiment, we collected information, including the execution time, the number of generations, and the number of satisfied objective functions (aggregated or not). The execution time tells us whether all objective functions could be satisfied in less than 10 seconds and indicates if the approach can be used in practice. The number of generations allows us to analyze the effect of the population size regarding the elapsed time. For the same time, a larger population can correspond to fewer generations (this refers to the complexity of NSGA-II). For the non-terminated executions, we collect the number of satisfied objective functions to get a quality indication. For each preference set, we compare the non-aggregated implementation with the aggregated one. The case without aggregation is easy to interpret because it easily shows quality results for each preference. However, the aggregated configuration allows us to reduce the number of objective functions, which plays an important role in the complexity of the algorithm. Unfortunately, with aggregation, it is harder to know the impact of changes on each preference. Furthermore, in aggregation, we chose to associate two objective functions to a general preference, but we could associate more objective functions. However, with two objective functions, we study configurations, which are very different from the non-aggregated configurations, while keeping a multi-objective approach. Other values could be

(a) S_1 on the NFB website

(b) S_2 on the NFB website

(c) S_3 on the NFB website

(d) S_4 on the NFB website

Figure 11. Adaptations of the NFB website.

chosen for the mutation (versus crossover) probability. The analysis of the results would have been more complex, without giving any substantial benefit.

2) Internal validity: The extraction of variables from the websites was done manually. Two persons participated in gathering this information. All colors were obtained from Web pages using two different software systems: the "gcolor2" on Linux system and the digital colorimeter included in the MAC OS platform (OS X 10.8.5). Even if colors can differ from one screen to another, such software systems, by getting values from the "display manager" functionalities, are not concerned by this issue. The defined model (variables, constraints) simplifies pages by only encoding variables and constraints concerned by the currently used preferences.

3) External validity: We have considered five very different websites in terms of size, architecture and colors used. Websites were also chosen for noticed accessibility problems for visually impaired people. People who reported accessibility problems have a visual impairment (low or high). We focused the study on the contrast problem and brightness problem because they are the most difficult problems to solve today with existing assistive technologies. Several dimensions of the search space and variables included in the constraints were considered with the aim of studying variations in the resolution. We feel that the variability in the results obtained between NSGA-II executions is significant on this kind of input data.

4) Conclusion validity: We have considered several measures, including min, max, standard deviation and average execution time, generation number and number of satisfied objective functions. We considered 60 executions for each configuration. In our previous paper [1], we compared results of NSGA-II with results obtained with an exact algorithm from preference theory. The exact algorithm appeared to be totally unusable (on a standard computer) when we had a search space of 10^5. We also checked if there are cases where our algorithm returns a good solution in the random initialization phase.

VI. Conclusion and Future Work

Widely used assistive technologies now partially meet their objective of helping people with low vision. However, their general purpose often leads to rather simple adaptations that do not provide relevant improvement for specific visual impairment. Moreover, they may also radically change the appearance of websites: part of the website design may be lost, while navigation marks (deliberately placed or not) helping internet users may disappear. As there is a substantial website diversity in terms of architecture, number of objects, colors used or even in font styles, we propose and test a new approach to adapt Web pages. This approach is based on real user needs (modeled by preferences) to find an adaptation that fits for these preferences.

To address scalability issues when we work with real websites, our approach is based on a multi-objective optimization algorithm (NSGA-II). We describe and carry out an experimentation on five very different websites. The chosen websites have been noticed by users with some impairment as websites with accessibility issues. Research has shown that this approach works fine for small and some medium websites with contrast or color distance preferences. We also learn lessons about limitations of the algorithm regarding running time and the number of considered objects (variables). This experiment also highlights the significant benefits provided by the aggregation of objective functions.

As future work, we plan to investigate other resolution approaches by modeling our problem as a CSP (Constraint Satisfaction Problem). The evaluation of our CSP modeling on existing constraint solvers will be compared with the results obtained with NSGA-II. The main and original target is to help people with low vision, thus we will test our approach in real conditions with users with visual impairment. This evaluation will be conducted on frequently visited websites by previously defining quality indicators. Testing websites with visually impaired people may help us to get information on how to facilitate the expression of preferences by the concerned end user. In parallel, we will work on object or structure recognition from Web pages. This phase is important to automate the whole process. Object and structure recognition is especially complex and requires refined heuristics for HTML4 Web pages because they are not explicitly defined in the source code.

ACKNOWLEDGMENT

The authors would like to thank Berger-Levrault, which supported this work with a grant, and Rémi Coletta for his valuable remarks about the way of reporting the evaluation.

REFERENCES

[1] Y. Bonavero, M. Huchard, and M. Meynard, "Web page personalization to improve e-accessibility for visually impaired people," in Proceedings of the Second International Conference on Building and Exploring Web Based Environments (WEB 2014,), 2014, pp. 40–45.

[2] "Visual impairment and blindness, fact sheet n°282," World Health Org., http://www.who.int/mediacentre/factsheets/fs282/en, Oct. 2013, accessed: 2014-11-09.

[3] Web Content Accessibility Guidelines, World Wide Web Consortium, http://www.w3.org/TR/WCAG20/, accessed: 2014-11-09.

[4] User Agent Accessibility Guidelines, World Wide Web Consortium, http://www.w3.org/TR/UAAG20/, accessed: 2014-11-09.

[5] Authoring tools Accessibility Guidelines, World Wide Web Consortium, http://www.w3.org/TR/ATAG20/, accessed: 2014-11-09.

[6] Web Accessibility Initiative - Accessible Rich Internet Applications, World Wide Web Consortium, http://www.w3.org/WAI/intro/aria, accessed: 2014-11-09.

[7] "References on evaluation tools by WAI," http://www.w3.org/WAI/ER/existingtools.html, accessed: 2014-11-09.

[8] M. Y. Ivory, J. Mankoff, and A. Le, "Using automated tools to improve Web site usage by users with diverse abilities," Information Technology and Society, vol. 3, no. 1, 2003, pp. 195–236.

[9] J. T. Richards and V. L. Hanson, "Web accessibility: a broader view," in WWW '04: Proceedings of the 13th international conference on World Wide Web. ACM Press, 2004, pp. 72–79.

[10] C. Domshlak, R. I. Brafman, and S. E. Shimony, "Preference-based configuration of Web page content," in Proceedings of International Joint Conference on Artificial Intelligence (IJCAI), 2001, pp. 1451–1456.

[11] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic personalization based on Web usage mining," Com. ACM, vol. 43, no. 8, 2000, pp. 142–151.

[12] K. Z. Gajos, J. O. Wobbrock, and D. S. Weld, "Automatically generating user interfaces adapted to users' motor and vision capabilities," in Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST), C. Shen, R. J. K. Jacob, and R. Balakrishnan, Eds. ACM, 2007, pp. 231–240.

[13] J. Abascal, A. Aizpurua, I. Cearreta, B. Gamecho, N. Garay-Vitoria, and R. Miñón, "Automatically generating tailored accessible user interfaces for ubiquitous services," in Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS), K. F. McCoy and Y. Yesilada, Eds. ACM, 2011, pp. 187–194.

[14] A. Mereuta, S. Aupetit, and M. Slimane, "Improving Web accessibility for dichromat users through contrast preservation," in Proceedings od the 13th International Conference ICCHP (International Conference on Computers Helping People with Special Needs), ser. Lecture Notes in Computer Science, K. Miesenberger, A. I. Karshmer, P. Penáz, and W. L. Zagler, Eds., vol. 7382. Springer, 2012, pp. 363–370.

[15] L. Balme, A. Demeure, G. Calvary, and J. Coutaz, "Sedan-Bouillon: a plastic Web site," 2005, plastic Services for Mobile Devices (PSMD), Workshop held in conjunction with Interact'05, Rome, 12 Sept. 2005.

[16] M. Macías, J. González, and F. Sánchez, "On adaptability of Web sites for visually handicapped people," in Proceedings of the second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH), ser. Lecture Notes in Computer Science, P. D. Bra, P. Brusilovsky, and R. Conejo, Eds., vol. 2347. Springer, 2002, pp. 264–273.

[17] K. Cullen, L. Kubitschke, T. Boussios, C. Dolphion, and I. Meyer, "Web accessibility in European countries: level of compliance with latest international accessibility specifications, notably WCAG 2.0, and approaches or plans to implement those specifications," http://ec.europa.eu/digital-agenda/en/news/study-report-web-accessibility-european-countries-level-compliance-latest-international, 2009, accessed: 2014-11-09.

[18] S. Kaci, Working with Preferences: Less Is More, ser. Cognitive Technologies. Springer, 2011, iSBN:978-3-642-17279-3.

[19] J. Horn, N. Nafpliotis, and D. Goldberg, "A niched Pareto genetic algorithm for multiobjective optimization," in Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE, 1994, pp. 82–87.

[20] J. Knowles and D. Corne, "The Pareto archived evolution strategy: A new baseline algorithm for Pareto multiobjective optimisation," in Proceedings of the Congress on Evolutionary Computation, vol. 1. IEEE, 1999, pp. 98–105.

[21] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," IEEE Trans. Evolutionary Computation, vol. 3, no. 4, 1999, pp. 257–271.

[22] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Trans. Evolutionary Computation, vol. 6, no. 2, 2002, pp. 182–197.

[23] M. Harman, S. A. Mansouri, and Y. Zhang, "Search-based software engineering: trends, techniques and applications," ACM Comput. Surv., vol. 45, no. 1, Dec. 2012, pp. 11:1–11:61. [Online]. Available: http://doi.acm.org/10.1145/2379776.2379787

[24] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization," Evolutionary computation, vol. 10, no. 3, 2002, pp. 263–282.

[25] "Conversion from RGB to HSV," http://en.wikipedia.org/wiki/HSL_and_HSV, accessed: 2014-11-09.

A Non Intrusive Method for Measuring Visual Attention Designed for the Study and Characterization of Users' Behavior in Serious Games

Zahen Malla Osman, Jérôme Dupire, Alexandre Topol, and Pierre Cubaud
Centre d'Etude et de Recherche en Informatique et Communications
Conservatoire National des Arts et Métiers
Paris, France
Email: {zahen.malla_osman, dupire, topol, cubaud}@cnam.fr

***Abstract*—The study and characterization of user behavior in the context of human computer interaction allows us to improve the design of interactive applications. Using certain sensors to analyze user behavior such as an eye tracking system can be intrusive and uncomfortable. In this paper, we report on an experiment to determine the minimum field of view that permits the user to perform an effective search task in a 3D virtual environment, by analyzing how the user controls the virtual camera. Our study exploits a model based on the use of several novel non intrusive temporal and quantitative measures of visual attention, such as: fixation, gaze, and movement. Seven out of ten measures gave significant results with the same findings.**

Keywords-Field of view; virtual environment; video games; visual attention; virtual camera; eye tracking; non intrusive.

I. INTRODUCTION

The use of non intrusive visual attention measures to analyze users' behavior in 3D virtual environments was originally proposed in our previous work [1]. Here, we extend this work by presenting more details about our study and its implementation in serious games.

We commonly refer to the term "attention" to indicate the presence of important information. Attention is an ability function for selecting an object or stimulus among many others [2]. It describes the cognitive process of selectively concentrating on one aspect of the environment while ignoring others; it has also been defined as the allocation of processing resources [3].

Visual attention is the ability of a vision system, whether human or artificial, to quickly select the most pertinent information from the environment in which it operates [4].

Some diseases, such as Alzheimer's disease, can currently be measured using attention measures, due to the fact that attention varies depending on physiological factors (e.g., fatigue and rest) or psychological factors (e.g., stress and motivation).

The model presented in this paper was built during the design of *Le Village aux Oiseaux. Le Village aux Oiseaux* [5] is a therapeutic game for seniors who suffer from Alzheimer's disease. This game is designed to train the player's attention network in order to decrease cognitive loss due to Alzheimer's disease. The gameplay of *Le Village aux Oiseaux* has been developed as a 3D FPS (First Person Shooter) video game. The player takes the role of a photographer who helps the inhabitants of a small town to prevent the destruction of their town. The player's mission is to take pictures of birds: they aim to fix a target (bird) on the screen and press a button to validate. FPS games, however, are usually designed for young male gamers; certain modifications are needed to make these games accessible for seniors.

In order to provide such training, a major challenge is to know whether our player is focused on the game and to have an approximation of how high his/her attention capabilities are. Having such data allows us to (1) help the player to concentrate by adding special events (e.g., visual and sound effects) in the game, (2) modify the game difficulty to adapt to the player.

Taking into consideration the physiological and the behavioral manifestations such as electro-dermal activity, heart rate, blood pressure, electromyographic activity, and encephalographic activity allows us to analyze user behavior. Currently, these signals are measured by sensors such as ECG, EEG sensors, or eye trackers.

Eye tracking has been used to measure visual attention for many years. It is the process of using sensors to localize the position and the behavior of the eyes. Eye trackers work with micro-cameras that focus their targets on both eyes and record their movements when the user fixes his/her attention on a stimulus. They capture the contrast between the pupil and the retina using a projection of infrared light by analyzing the light reflected by the cornea of the human eye. Thus, from an initial calibration, the eye tracker can analyze eye movements from top to bottom and left to right. An eye tracking system helps us to determine what a person is looking at, what he/she is not looking at, but also what he/she does and does not pay attention to. Through eye tracking systems we can provide many visual attention measures, such as: fixation, gaze, and movement, in order to analyze users' ocular behavior.

A principal means of interacting with 3D VEs (Virtual Environments), in the case of video games, for example, is the use of the virtual camera, which is relatively easy to access and manipulate via game engines. The use of this virtual camera can show interesting results for non invasive study and characterization of user's behavior - especially in the absence of eye tracking systems, which can sometimes be unavailable.

Our work is focused on the FOV (Field of View) effect of the virtual camera for determining the minimum FOV that allows users to perform an effective search task in a 3D VE.

The remainder of this paper is organized as follows: Section II presents related work. Section III describes our experiment that analyzes user behavior in a 3D VE via the virtual camera. Section IV summarizes our paper and provides an outlook for future work.

II. RELATED WORK

Gaming is an increasingly prevalent cultural pastime [6], and today the video game is one of the most popular types of software applications in the world. More than half of all Americans play video games, for example [7][8]. Analyzing user behavior allows us to monitor the cognitive status of users while they play video games. We can, for example, determine phases during which the player is motivated or bored, in order to give information to game designers, allowing them to improve the design of their game by identifying, the areas of interest that attract the player's attention.

Video games can provide a framework for testing many types of attention measures [9], e.g., playing video games, such as Pac Man, can improve the reaction times of older adults [10]. An action game can improve visual/attentional skills, a strategic game can improve the skills of executive control, and a puzzle game can enhance certain spatial skills [9].

Numerous studies, both academic and industry-based, have been conducted to analyze user behavior. *Le Village aux Oiseaux* has been inspired by a study by Green and Bavelier. Their study shows that action games, particularly FPS games, improve the attention network of their players [6].

Attention measures can be applied in serious games that are primarily tools for learning, e.g., we can imagine an artificial intelligence that adapts with the game's environment based on user behavior. According to Guardiola et al., a serious game is a rule-based formal system with a variable and quantifiable outcome, where different outcomes are assigned different values; the player exerts effort in order to influence the outcome and feels attached to that outcome. Specifically, however, a serious game is combined with a defined real life objective [11]. This definition is based on the classic definition of games given by Juul [12].

Games are designed to be pleasurable and interesting activities; games provide a powerful means of inspiring motivation in users. Generally, games are played for the sake of playing them. By contrast, therapeutic activities are designed to maximize their efficiency.

The McGraw-Hill Concise Dictionary of Modern Medicine[1] proposes this definition of therapy: "A general term for any form of management of a particular condition, treatment intended and expected to alleviate a disease or disorder; any technique of recovery, which may be medical, psychiatric, or psychological." The dictionary also proposes a list of more than two hundred different therapies. Some of them are very general (e.g., physical therapy), while others are very specific (e.g., Nicotine replacement therapy). *Le village aux Oiseaux* aims to reduce the effects of Alzheimer's disease by training the patient's attentional network [5].

Early identification of cognitive decline is used in geriatric medicine to identify imminent functional impairment and/or to help delay disease progression [13][14][15], e.g., an adaptive intelligent computer game can be used to make an attention deficit/hyperactivity disorder diagnosis [16].

During everyday interactions, our eyes provide a lot of information that reflects our emotional and mental states. Eye movement data reflect moment-to-moment cognitive processes during task execution [17]. Any given image contains multiple elements and the human eye fixes much more on some elements while others may receive little attention; when we look at an object in space (e.g., a wall with windows and doors), our eyes concentrate much more on some parts of this object (e.g., one of the windows), while the other parts of the object may receive less attention [18].

Studying ocular behavior in the context of human computer interaction (e.g., web browsing or video games [19][20][21][22]), allows us to identify and provide many indicators that can be used to evaluate user attention in order to improve the design of an user interface such as, for example, a digital library [23].

Much research has been conducted towards the study of ocular behavior during playing video games. Players of FPS video games concentrate their attention on one region of the game's environment, paying more attention to the center of the screen around the reticule because they shoot enemies through the reticule. By contrast, the attention area is larger in an adventure game because the player's attention is not constrained by any specific area of the screen [24][25].

There are many types of eye behaviors: fixation, being the moment when the eyes are relatively stationary, taking in or encoding information with a minimum duration of 100 milliseconds [26]; saccade, being the eye movement that occurs between fixations with durations of approximately 150–200 milliseconds [27]; and gaze, being the moment when the eyes look at a display element [28]. When we look at an object in a visual display, we may make many fixations on this object. The number of these fixations shows the importance of the display area. A large number of fixations, however, can also reflect a poorly designed interface [28]. Some researchers use a positive score for each correct fixation in the environment, in order to examine user performance, by asking users to fix on specific objects in the environment among other objects (obstacles) [4].

To study user behavior in a 3D VE, a common approach is to ask users to complete search task in order to know how he/she interacts with the 3D VE [29], e.g., users may have to find objects that have specified numbers displayed on them [21], or to find a maximum number of hidden keys distributed in a 3D VE [30].

Our idea was to use the virtual camera of a 3D VE to examine several visual attention measures, such as: fixation, gaze and movement. The use of the 3D VE's virtual camera provides an indirect method for analyzing the effect of FOV

[1] © 2002 by The McGraw-Hill Companies, Inc.

on user behavior, given that the useful FOV is the total area of the visual field within which individuals can obtain useful information without moving their heads or eyes [31][32].

In order to study and characterize user behavior in a 3D VE through the virtual camera, we selected several visual attention measures employed by Gibbs et al. [19]. The measures selected are expressed by the number of fixations, fixation duration, and gaze duration. We also introduced new measures to give more information about how the user performs an effective search task in our 3D VE. The measures that we added are expressed by: the number of gazes, the number of movements, the movement duration, the sum total duration of all fixations per task, the sum total duration of all gazes per task, the sum total of all movements per task, and the total duration of each task.

The principal motivation of this work is to build a novel method for non invasive study and characterization of user behavior in 3D VE (e.g., *Le village aux Oiseaux)*, in order to show how users interact within 3D VE with a small FOV (Field of View). Our goal was to examine the FOV effect of the virtual camera on user behavior, and to determine the minimum FOV that allows the user to perform an effective search task in a 3D VE [1]. FOV size is very important for rapid extraction and identification of information in a 3D VE. The effective search task, in the context of our experiment, consists of a simple navigation within the 3D VE for the purpose of finding all objects (e.g., hidden buttons distributed around the VE) using: the least possible number of fixations and the shortest fixation duration; the least possible number of gazes and the shortest gaze duration; the least possible number of movements and the shortest movement duration; the shortest sum total duration of all fixations per task; the shortest sum total duration of all gazes per task; the shortest sum total duration of all movements per task; and the shortest total duration of each task. Our results provide information that can be of benefit to game designers, allowing them to improve gameplay, manage the difficulty of game environments, and optimize the distribution of visual resources.

III. EXPERIMENT

Gibbs et al. used an eye tracking system to determine whether ocular behavior differs between newspaper websites and TV-oriented websites. They used several visual attention measures to test ocular behavior, such as: number of fixations, fixation duration, and gaze duration. Within the contest of FPS video games, our research uses these measures employed by Gibbs et al., as well as our own measures to analyze user behavior, using the VE's virtual camera instead an eye tracker. The aim of our experiment is to determine the minimum FOV that permits the user to perform an effective search task in a 3D VE, and to generate knowledge for game designers that can help them manage and adapt the difficulty of a 3D VE according to user behavior [1].

The users in our experiment use a mouse and a keyboard to manipulate the virtual camera of our 3D VE as they would in a FPS video game (e.g., Half Life, Counter Strike). The measures employed in our experiment, consist of various types, such as: fixation, being a short pause in movement, represented quantitatively by the Number of Fixations (NF) and temporally by the Fixation Duration (FD), which vary between 100 and 300 milliseconds; gaze, which is the time spent looking at a display object, represented by the Number of Gazes (NG) and the Gaze Duration (GD), which starts from 300 milliseconds; the movement between two fixations or gazes, represented by the Number of Movements (NM) and the Movement Duration (MD), which starts from 100 milliseconds.

We also added four measures to those specified above: the Sum Total Duration of all Fixations per task (STDF), the Sum Total Duration of all Gazes per task (STDG), the Sum Total Duration of all Movements per task (STDM), and the Total Duration of each task spent by the user to complete the required task (TD).

A total of 14 volunteers (10 male and 4 female) participated in this experiment. Their ages varied between 25 and 42 years, with a mean of 30. All participants are right-handed, healthy, and have normal or corrected-to-normal vision. The experiment was performed on a desktop personal computer (Intel 3 GHz Core 2 Duo, 4 GB RAM) running Windows 7 Professional, with an LCD display with a resolution of 1920×1040 pixels.

A. Procedure

The purpose of the following experiment is to compute visual attention measures and to study the FOV effects on user behavior during a visual search task in a 3D VE. Fig. 1 shows our 3D VE, which is a virtual art gallery similar to the static environment of Lee et al. [21]. We used the Unity3D game engine version 3.5 to create our 3D VE, including all the objects and the buttons. The virtual camera is positioned at the level of the eyes of the user's avatar.

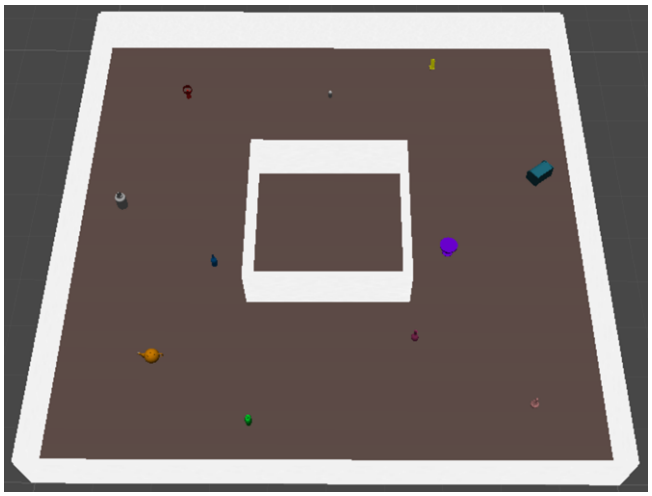


Figure 1. Our 3D virtual environment (the art gallery).

The participants were first invited to complete a short form to provide information including their name, age, gender, and whether or not they often play FPS video games.

Secondly, the participants were asked to perform a free navigation in the 3D VE with a FOV of 80°, simply

navigating in the 3D VE and observing the virtual objects using the mouse and the keyboard to control navigation motion. This step was created as a training phase to learn manipulation of the virtual camera. The participants used the mouse to change the orientation of the virtual camera (yaw and pitch angles) and the keyboard to move the virtual camera. We used an 'AZERTY' format keyboard with the following key mapping: Z: forward, S: back, D: right, Q: left.

Finally, the participants were asked to perform a visual search task to find and validate hidden buttons in the 3D VE. They had to find ten buttons randomly distributed on the surfaces of objects in the 3D VE (each object in our VE contains one hidden button).

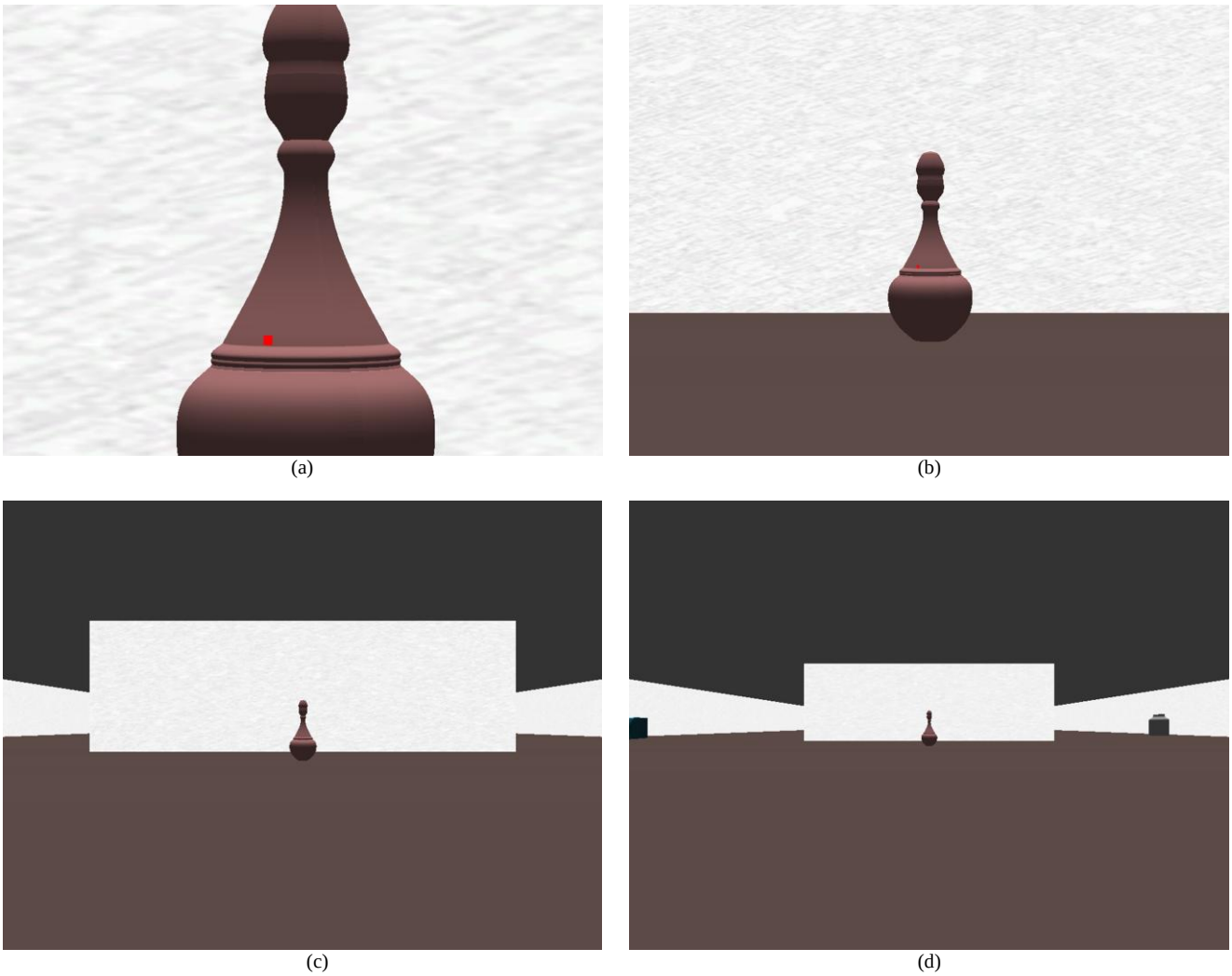
Each participant had to find all the hidden buttons using the reticule area (a rectangle 250×150 pixels situated in the center of the screen), and validate them by pressing the Space key. A number is displayed at the top left of the screen to indicate how many hidden buttons are left.

The participant was asked to repeat the search task six times, knowing that we had changed the positions of all the objects and buttons, as well as the FOV size of the virtual camera before each of the six attempts at the task (10° for the first attempt, 20° for the second, 30° for the third, 50° for the fourth, 80° for the fifth, and 110° for the sixth attempt). The order of the attempts was randomized for each of the participants in order to eliminate the adaptation effect.

The purpose of changing the FOV size (i.e., from 10° to 110°) was to discover how FOV affects user behavior and to determine the minimum FOV that enables the user to perform an effective search task in a FPS type 3D VE, given that the default FOV in a FPS game ranges from 75° to 110°.

Each one of our visual attention measures was computed for each attempt in real time and was recorded for a posteriori analysis.

Fig. 2 shows our 3D VE using four different sizes of FOV. We can see that the navigation with a FOV of 10° (see Fig. 2.a) can be difficult, because the size of FOV is very small compared to the other sizes of FOV (e.g., 30°, 80°, or 110°).



(a) (b) (c) (d)

Figure 2. Four different views of our 3D virtual environment (VE) for the same virtual camera position : (a) VE with FOV of 10°, (b) VE with FOV of 30°, (c) VE with FOV of 80°, (d) VE with FOV of 110°.

B. Results

A one-way Analysis of Variance (ANOVA) was conducted to see whether the FOV of the virtual camera affected user behavior during the search task in our 3D VE. A total of 14 subjects took part in the experiment. We sought to discover whether there is a significant difference between the measures that we obtained by changing the FOV size between 10°, 20°, 30°, 50°, 80°, and 110°. We expressed our measures by way of a natural logarithm and tested the measures' normality using the Shapiro Wilk test [33]. Then, we used the ANOVA test to analyze the variance between all our measures. We note that, for our statistical analysis, we do not take into account the random spatial distribution of objects, nor the random order of the tasks.

Table I shows the means, standard deviations, and analyses of variance of all our measures. Our ANOVA results show a significant difference between certain measures used in our experiments when we changed the FOV; such as: the Number of Fixations (NF), the Number of Gazes (NG), the Number of Movements (NM), the Sum Total Duration of all Fixations (STDF), the Sum Total Duration of all Gazes (STDG), the Sum Total Duration of all Movements (STDM), and the Total Duration of each task (TD). However Fixation Duration (FD), Gaze Duration (GD), and Movement Duration (MD) do not show any significant difference.

- *Impact of FOV on the Total Duration of each task (TD)*: We observed that the TD decreases when the FOV increases (see Fig. 3). The boxplot presents the TD means of all participants in the six sizes of FOV. We found that the TD becomes convergent from a FOV of 30°. We also found that there was not much change in user behavior when he/she used a FOV of 30°, 50°, 80° or 110°; however, a FOV of 10° or 20° shows a lot of change in user behavior. For example, users took a long time to complete the task when they used a FOV of 10° or 20°, while they took less time when they used other FOVs.
- *Impact of FOV on the Number of Fixation (NF)*: We also observe that the NF measure decreases when the FOV increases (see Fig. 4).

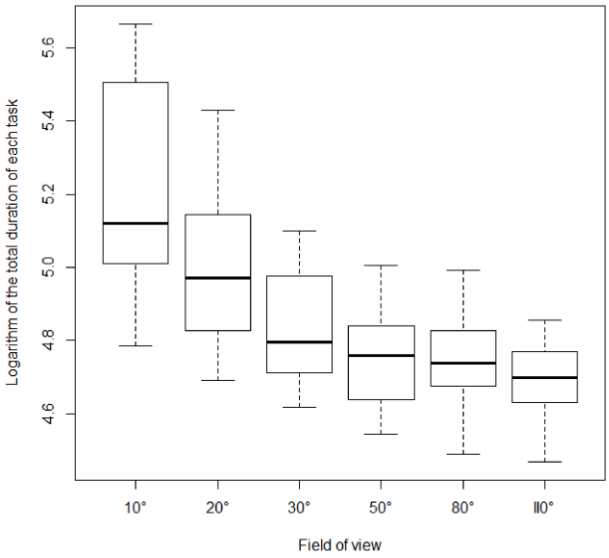


Figure 3. The Total Duration of each task (TD) by Field of View (FOV).

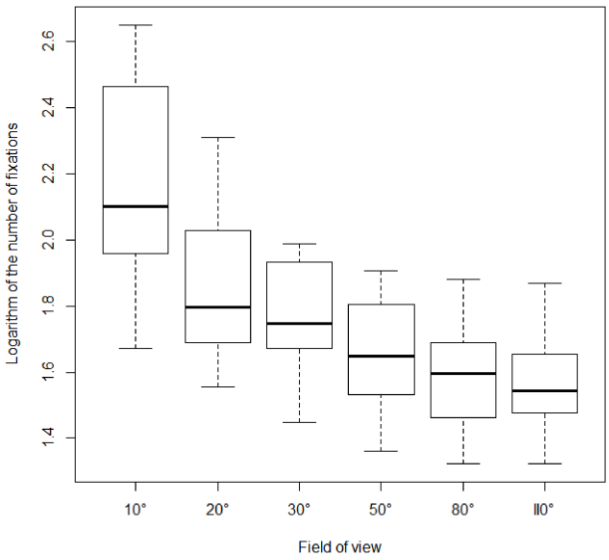


Figure 4. The Number of Fixations (NF) by Field of View (FOV).

TABLE I. MEAN, STANDARD DEVIATIONS AND ANALYSES OF VARIANCE OF THE VISUAL ATTENTION MEASURES IN THE SIX SIZES OF FOV

	10°	**20°**	**30°**	**50°**	**80°**	**110°**	***F***	***p***
NF	2.15 (0.32)	1.84 (0.23)	1.78 (0.16)	1.66 (0.18)	1.60 (0.17)	1.57 (0.14)	14.92	<0.0001 ***
NG	1.93 (0.34)	1.66 (0.27)	1.50 (0.24)	1.45 (0.23)	1.43 (0.23)	1.40 (0.17)	8.99	<0.0001 ***
NM	2.35 (0.32)	2.06 (0.24)	1.96 (0.18)	1.87 (0.19)	1.82 (0.19)	1.79 (0.13)	12.68	<0.0001 ***
FD	2.23 (0.02)	2.23 (0.02)	2.21 (0.02)	2.22 (0.04)	2.22 (0.03)	2.21 (0.03)	1.12	0.358
GD	2.90 (0.09)	2.95 (0.12)	2.93 (0.11)	2.89 (0.10)	2.88 (0.11)	2.90 (0.12)	0.94	0.489
MD	2.44 (0.19)	2.53 (0.25)	2.49 (0.24)	2.51 (0.26)	2.57 (0.26)	2.47 (0.30)	0.41	0.838
STDF	4.37 (0.33)	4.07 (0.21)	3.99 (0.17)	3.89 (0.18)	3.82 (0.15)	3.78 (0.13)	15.81	<0.0001 ***
STDG	4.83 (0.40)	4.61 (0.33)	4.43 (0.32)	4.33 (0.31)	4.31 (0.32)	4.30 (0.27)	5.80	<0.0001 ***
STDM	4.79 (0.22)	4.59 (0.25)	4.45 (0.15)	4.37 (0.13)	4.39 (0.14)	4.26 (0.22)	13.91	<0.0001 ***
TD	5.20 (0.30)	4.99 (0.21)	4.85 (0.16)	4.76 (0.13)	4.75 (0.14)	4.69 (0.11)	14.84	<0.0001 ***

*** p <0.0001, ** p <0.001, * p <0.01, NF: the Number of Fixations, NG: the Number of Gazes, NM: the Number of Movements, FD: the Fixation Duration, GD: the Gaze Duration, MD: the Movement Duration, STDF: the Sum Total Duration of all Fixations, STDG: the Sum Total Duration of all Gazes, STDM: the Sum Total Duration of all Movements, and TD: the Total Duration of each task.

The decrease in the NF measure allows us to note that the users of our application reduce their number of fixations as the FOV increases. This is because a large FOV offers a larger view of the environment and affords users comfortable navigation for performing the required search task.

- *Impact of FOV on the Sum Total Duration of Fixations (STDF)*: We also found that the STDF becomes convergent from a FOV of 30° (see Fig. 5).

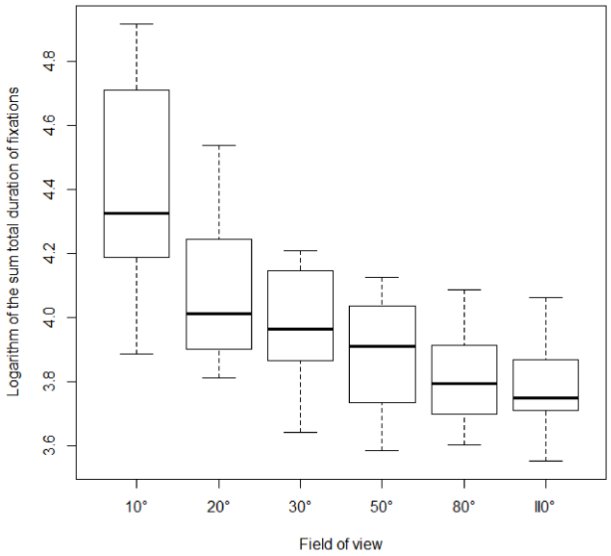


Figure 5. The Sum Total Duration of Fixations (STDF) by Field of View (FOV).

- *Impact of FOV on the Number of Gazes (NG)*: The NG in the fourth task (FOV = 50°) was high compared with the third, fifth, and sixth tasks (respectively: FOV = 30°, 80°, 110°). This is because users had difficulty in finding the hidden buttons in this task (see Fig. 6).

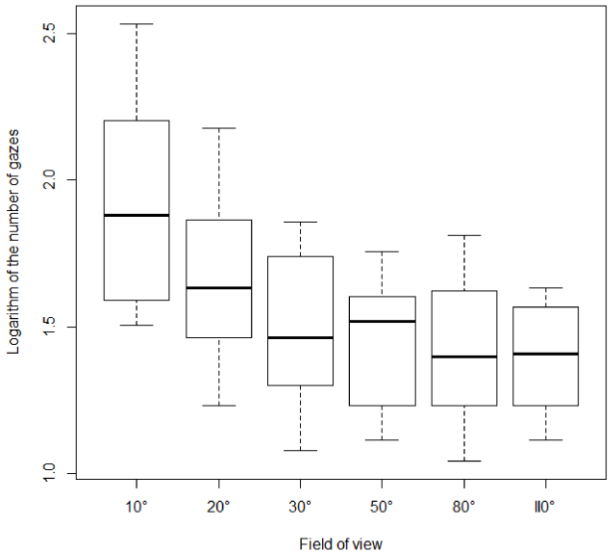


Figure 6. The Number of Gazes (NG) by Field of View (FOV).

- *Impact of FOV on the Sum Total Duration of Gazes (STDG)*: We observed that the STDG in the sixth task (FOV = 110°) was high compared to the fifth task (FOV = 80°), due to the use of a large FOV (see Fig. 7).

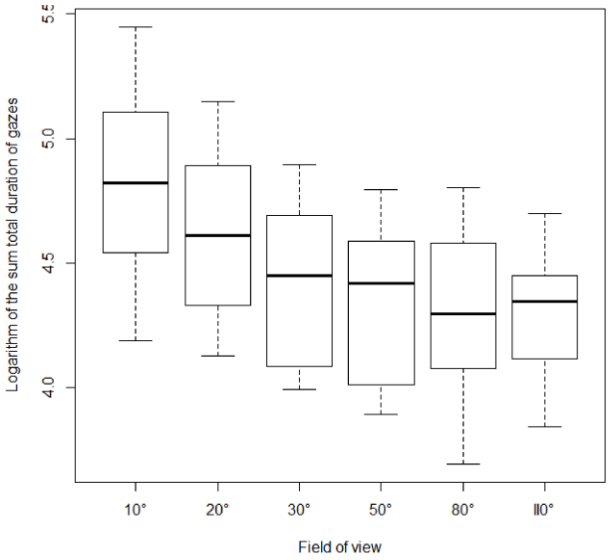


Figure 7. The Sum Total Duration of Gazes (STDG) by Field of View (FOV).

- *Impact of FOV on the Number of Movements (NM)*: We also found that the NM in the fourth task (FOV = 80°) was high compared with the third, fifth, and the sixth tasks (respectively: FOV = 30°, 80°, 110°) (see Fig. 8). We can note also that the users needed fewer movements to accomplish the required task when they used a large FOV offering a large view of the environment. Consequently, users can perform the required search task quickly when they use a large FOV.

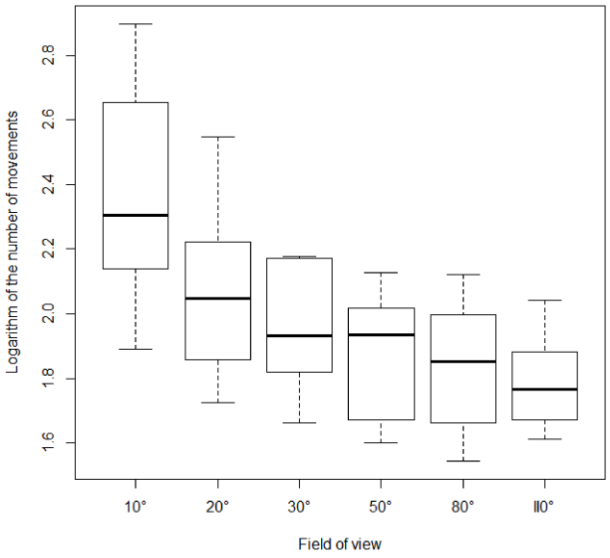


Figure 8. The Number of Movements (NM) by Field of View (FOV).

- *Impact of FOV on the Sum Total Duration of Movements (STDM)*: The boxplot presents the STDM means of all participants in each of the six sizes of FOV. We observed that there is a user in the third task (FOV = 30°) that is out the boxplot. This is because the user had difficulty in manipulating the virtual camera (see Fig. 9).
  As we noted above, we observed also that the STDM measure decreases as the FOV increases. This result allows us to observe that users need less movement time to accomplish the required visual search task in our 3D VE when they use a large FOV.

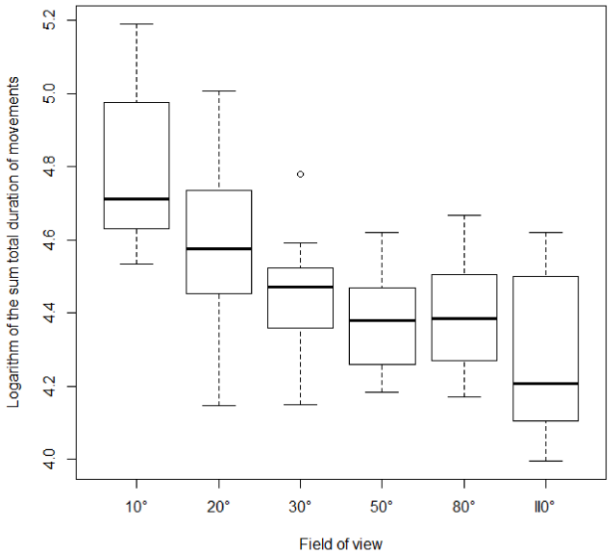


Figure 9. The Sum Total Duration of Movements (STDM) by Field of View (FOV).

- *Difference between the video game players and the non-video game players using the Total Duration of each task (TD)*: After analyzing all our subjects without taking into consideration their video games experience, we divided our subjects into two categories: Video Game Players (VGP) and Non-Video Game Players (NVGP). Fig. 10 shows a comparison between the video game players and the non-video game players using the TD measure. We found that the non-video game players took more time than the video game players to achieve the required visual search task.
  The video game players performed better that the non-video game players in our 3D VE.
  We can also see that the FOV for the video game players became more stable using a size of 50° (see Fig. 10), allowing us to observe that video game players need a minimum FOV of 50° to perform an effective visual search task in our 3D VE. We also notice much less variance in the results from the video game players.

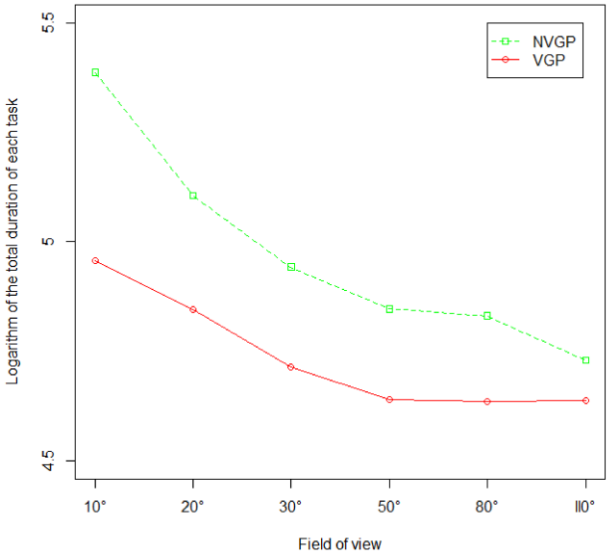


Figure 10. The difference between the Video Game Players (VGP) and the Non-Video Game Players (NVGP) using the Total Duration of each task (TD).

- *The minimum Field of View (FOV) for user search task in 3D Virtual Environment (VE)*: To determine the minimum FOV that allows the user to conduct an effective search task within our 3D VE, we performed another Analysis of Variance (ANOVA) that examined all our visual attention measures between each pair of FOV that we used in our experiment.
  We defined two groups of FOVs according to visual attention measure values: Group 1, with FOVs of 10° and 20°, and Group 2 with FOVs of 30°, 50°, 80°, and 110° (see Fig. 11). The results of this ANOVA do not show a significant difference between a FOV of 10° and a FOV of 20° (p=0.0585), but they do show a significant difference between a FOV of 10° and a FOV of 30° (p=0.015*), 50° (p=0.012*), 80° (p=0.011*), and 110° (p=0.0002***). The ANOVA results also show a significant difference between a FOV of 20° and a FOV of 110° (p=0.005**), but they do not show a significant difference between a FOV of 20° and other FOVs. Finally, this ANOVA shows that there is no significant difference between FOVs of 30°, 50°, 80°, and 110°.
  Our visual attention measures become convergent from a FOV of 30°. We found that our subjects in Group 1 were slower than in Group 2 because they performed the visual search task quicker, with: the least possible number of fixations, the least possible number of gazes, the least possible number of movements, the shortest sum total duration of all fixations per task, the shortest sum total duration of all gazes per task, the shortest sum total of all

movements per task, and the shortest total duration of each task.

We note that users can use a FOV of 30° as a minimum FOV for performing an effective search task in our 3D VE.

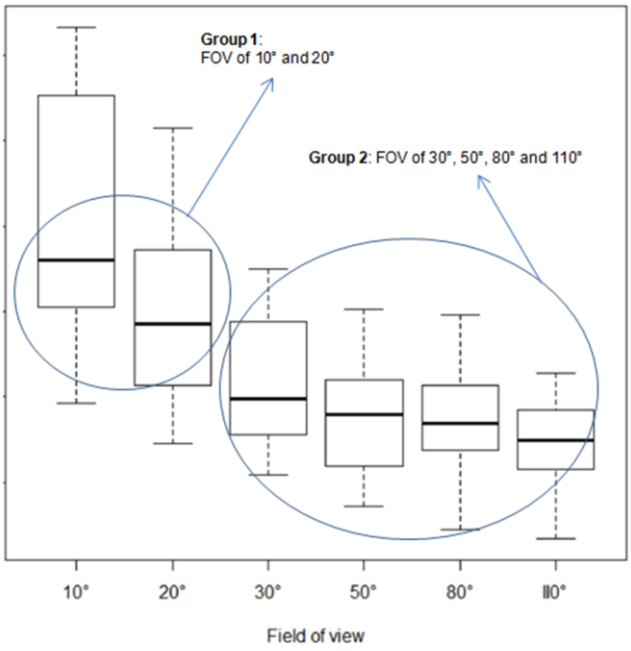


Figure 11. Two groups of Field of View (FOV) following the ANOVA results: Group 1 with FOVs of 10° and 20°, Group 2 with FOVs of 30°, 50°, 80°, and 110°.

C. Discussion

As evidenced by other studies, eye tracking systems have been used to measure visual attention for many years. These systems have provided many visual attention measures, such as fixation, gaze, and movement, for the purpose of analyzing users' ocular behavior. They have also proved useful for evaluating the design of interactive applications.

However, the use of this type of sensor can be intrusive, e.g., users of video games prefer to be comfortable while they are playing, making sensors like eye trackers, EEG, or ECG sensors unsuitable for use with video games.

The use of an indirect method – for monitoring users of video games, for example – can offer advantages for the non-intrusive study of user's behavior and for improving the design of interactive applications, such as navigation in a 3D virtual environment.

Our model uses a novel non intrusive technique for the study and characterization of user behavior in a 3D virtual environment, by analyzing how the user controls the virtual camera. Our model can also help with determining the performance level of users when they perform a visual search task in 3D VE.

Upon analyzing our results, we found that the Field of View (FOV) of the virtual camera in our 3D Virtual environment (VE) can affect user behavior when he/she performs a visual search task within the VE.

The FOV size had a significant effect on user behavior during our experiment. We notice that in Table I there is a significant difference between the results based on most of our measures: the Number of Fixations (NF), the Number of Gazes (NG), the Number of Movements (NM), the Sum Total Duration of all Fixations (STDF), the Sum Total Duration of all Gazes (STDG), the Sum Total Duration of all Movements (STDM), and the Total Duration of each task (TD). This difference between these measures is due to the change in the FOV (i.e., between 10°, 20°, 30°, 50°, 80°, and 110°), where we observe that the FOV affects user behavior during navigation within a 3D VE.

We note that these measures decrease as the FOV increases, e.g., the NF mean value for all the subjects had a natural logarithm of 2.15 when we used a FOV of 10°, and this NF decreased to 1.57 when we used a FOV of 110°. Additionally, the NG mean value for all subjects had a natural logarithm of 1.93 when we used a FOV of 10°, and this NG decreased to 1.40 when we used a FOV of 110°. We observe also that the NM mean value for all our subjects had a natural logarithm of 2.35 when we used a FOV of 10°, and this NM decreased to 1.79 when we used a FOV of 110°. We found also that the STDF decreased as the FOV increased, where the STDF mean value for all the subjects had a natural logarithm of 4.37, and this STDF decreased to 3.78 when we used a FOV of 110°. The STDG mean value for all the subjects had a natural logarithm of 4.83, and this STDG decreased to 4.30 when we used a FOV of 110°.The STDM mean value for all subjects had a natural logarithm of 4.79, and this STDM decreased to 4.26 when we used a FOV of 110°. Finally, the TD mean value for all subjects had a natural logarithm of 5.20, and this TD decreased to 4.69 when we used a FOV of 110°. The decrease is important for determining the FOV within which one can navigate effectively within a 3D VE.

As for Fixation Duration (FD), the Gaze Duration (GD), and the Movement Duration (MD), Table I does not show any significant difference in these visual attention measures. This shows that our users use stable and constant durations when performing the visual search task. We can also note that FOV size does not affect these types of visual attention measures in our 3D VE.

The ANOVA preformed on each pair of FOVs allows us to define two groups of FOVs according to measure values: Group 1, with FOVs of 10° and 20°, and Group 2 with FOVs of 30°, 50°, 80° and 110°, given that there is not a significant difference between a FOV 10° and a FOV 20°; and between a FOV 30°, 50°, 80°, and 110°; but that there is a significant difference between a FOV of 10° and a FOV of 30°, 50°, 80°, and 110°; and between a FOV of 20° and a FOV of 110°.

User behavior in Group 1 was less effective than user behavior in Group 2 because users in Group 2 performed the search task quicker with: the least possible number of fixations, the least possible number of gazes the least possible number of movements, the shortest sum total duration of all fixations per task, the shortest sum total

duration of all gazes per task, the shortest sum total of all movements per task, and the shortest total duration of each task. We found that these measures become convergent from a FOV of 30°. We note that the user can use a FOV of 30° as a minimum FOV for performing the search task in a short time with minimum movement of the virtual camera.

We observe also that the user can perform an effective search task using this FOV of 30° in cases where we did not find much change in user behavior based on the virtual camera when he/she uses a large FOV, such as 80° or 110°. This shows that users in our experiment can perform an effective visual search task better when visual attention measures' values are smaller.

Finally, we showed also in Fig. 10 that the NVGPs spent more time than the VGPs to achieve a visual search task in a 3D VE, and therefore we can deduce that the VGPs perform better on the required task than the NVGPs because the VGPs are accustomed to playing video games. We can also observe that the video game players' ability to perform an effective search task seems to become more reliable and precise from a FOV of 50°.

IV. Conclusion and Future Work

In this paper, we have presented our experiment for determining the minimum field of view that permits the user to perform a search task in a 3D virtual environment using the virtual camera, which is accessible in all game engines.

We used several novel non intrusive visual attention measures to monitor user behavior: the number of fixations, the fixation duration, the number of gazes, the gaze duration, the number of movements, the movement duration, the sum total duration of all fixations, the sum total duration of all gazes, the sum total duration of all movements, and the total duration of each task.

Our results, which are based on the use of a virtual camera of a 3D virtual environment, show differences in user behavior resulting from differences in the field of view.

The participants in our experiment could perform an effective search task better when the visual attention measures' values were smaller.

We have shown that the field of view of the virtual camera affects user behavior during navigation within a 3D virtual environment to complete a visual search task.

Our quantitative and temporal measures were evaluated by changing the field of view of the virtual camera. We found that the user needed less time to achieve his/her visual search task if he/she used a large field of view. We showed that the minimum field of view for performing an effective search task in a 3D virtual environment is 30°. Finally, we showed that video game players perform better in the 3D virtual environment.

The results generated by our experiment could be usefully applied to the design of video games that are based in 3D virtual environments.

For future work, we plan to increase the number of participants in our study and to test our model with a single category of users: video game players, which can give much more specific analysis (the obtained results using video game players had much less variance).

We chose not to modify the color code of our 3D virtual environment in this study, as our 3D virtual environment was inspired by the static environment of Lee et al. [21], which used a dark plain color for the floor. We intend to consult a graphic designer/ergonomist in order to verify all the visual aspects of our 3D virtual environment.

We plan to integrate an additional tool into our application to identify the regions of interest in our 3D virtual environment, using a video game, to provide additional information to game designers on how gameplay can be improved and difficulty managed by modifying the field of view of the virtual camera relative to difficulty level or the player's needs.

Finally, we also plan to test our model with Alzheimer's patients in order to show how our model can be used in the service of cognitive rehabilitation: specifically, for facilitating search tasks and adapting the difficulty of a 3D virtual environment (e.g., *Le Village aux Oiseaux* is a therapeutic game for seniors suffering from Alzheimer's disease).

Acknowledgment

This work has been partly funded by the ANR CONTINT Transread project.

We would like to thank K. Neil and M. Ouattara, and all the people who accepted to test and give feedback on our experiment.

References

[1] Z. M. Osman, J. Dupire, A. Topol, and P. Cubaud, "Non intrusive measures for determining the minimum field of view for user search task in 3D virtual environments," Proc. of the Seventh International Conference on Advances in Computer Human Interactions (ACHI), IARIA, 2014, Barcelona, Spain, pp. 208-213, ISSN: 2308-4138, ISBN: 978-1-61208-325-4.

[2] B. Cadet, "Psychologie cognitive," Press Editions, 1998, ISBN: 2-912-404-09-6.

[3] J. R. Anderson, "Cognitive psychology and its implications," Worth Publishers, 2004, ISBN-10: 1-4292-1948-3.

[4] H. Hugli, T. Jost, and N. Ouerhani, "Model performance for visual attention in real 3D color scenes," Proc. of the First International Work-Conference on the Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach (IWINAC 05), 2005, vol. 3562, pp. 469–478, doi: 10.1007/11499305_48.

[5] S. Mader, S. Natkin and G. Levieux, "How to analyse therapeutic games: the player/game/therapy model," Proc. of the International Conference on Entertainment Computing (ICEC 12), Springer-Verlag, 2012, vol. 7522, pp. 193-206, doi: 10.1007/978-3-642-33542-6_17.

[6] C. S. Green and D. Bavelier, "Action video game modifies visual selective attention," Nature Publishing Group, 2003, vol. 423, pp. 534–537.

[7] A. Lenhart, S. Jones, and A. R. Macgill, "Adults and video games," Pew Internet and American Life Project, 2008, Available: www.pewinternet.org [retrieved: 2014-11-24].

[8] Entertainment Software Association (ESA). Essential facts about the computer and video game industry. ESA EF 2013.

[9] W. R. Boot, A. F. Kramer, D. J. Simons, M. Fabiani, and G. Gratton, "The effects of video game playing on attention, memory, and executive control," Acta Psychologica, 2008, vol. 129, pp. 387–398.

[10] J. E. Clark, A. K. Lanphear, and C. C. Riddick, "The effects of videogame playing on the response selection processing of elderly adults," Journal of Gerontology, 1987, vol. 42, pp. 82–85.

[11] E. Guardiola, S. Natkin, D. Soriano, E. Loaser and P. Vrignaud, "Du jeu utile au jeu sérieux," Hermes, 2012, vol. 62, pp. 87-93.

[12] J. Juul, "The game, the player, the world: looking for a heart of gameness," Utrecht University, 2003, vol. 120, pp. 30-45.

[13] B. Seltzer, P. Zolnouni, M. Nunez, R. Goldman, D. Kumar, J. Ieni and S. Richardson, "Efficacy of donepezil in early-stage Alzheimer disease: a randomized placebo-controlled trial," Arch Neural, 2004, vol. 61(12), pp. 1852-1856, doi: 10.1001/archneur.61.12.1852.

[14] E. Braverman, "Cognitive delcine of aging: important neuroendocrinological predicators of early cognitive decline in a clinical setting Weill Cornell medical center," PATH Medical, Available: www.worldhealth.net [retrieved: 2014-11-24].

[15] H. Jimison, M. Pavel, J. McHanna, and J. Pavel, "Unobtrusive monitoring of computer interactions to detect cognitive status in elders," Information Technology in Biomedicine, IEEE Transactions, 2004, vol. 8(3), pp. 248-252, doi: 10.1109/TITB.2004.835539.

[16] L. Andrade, L. Carvalho, C. Lima, A. Cruz, P. Mattos, C. Franco, A. Soares and B. Grieco, "Supermarket game: an intelligent computer game for attention deficit/hyperactivity disorder diagnosis," Proc. of Mexican International Conference on Artificial Intelligence (MICAI 06), 2006, IEEE, pp. 359-368, doi: 10.1109./MICAI.2006.45.

[17] K. Rayner, "Eye movements in reading and information processing: 20 Years of Research," Psychological Bulletin, 1998, vol. 124(3), pp. 372–422.

[18] A. Yarbus, "Eye movements and vision," Plenum Press, 1967, ISBN-10: 0306302985.

[19] W. J. Gibbs and R. S. Bernas, "Visual attention in newspaper versus TV-oriented news websites," Journal of Usability Studies, 2009, vol. 4, pp. 147–165.

[20] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scottand, and A. M. Wichansky, "Eye tracking in web search tasks: design implications," Proc. of the 2002 Symposium on Eye Tracking Research & Applications (ETRA 02), 2002, ACM, pp. 51–58, doi: 10.1145/507072.507082.

[21] S. Lee, G. J. Kim, and S. Choi, "Real-time tracking of visually attended objects in interactive virtual environments," Proc. of the ACM Symposium on Virtual Reality Software and Technology (VRST 07), 2007, ACM, pp. 29–38, doi: 10.1145/1315184.1315187.

[22] K. Yokoi, K. Watanabe, and T. Kawai, "Dynamic evaluation of distribution of visual attention during playing video game," Proc. of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE 06), 2006, Article No. 96, doi: 10.1145/1178823.1178934.

[23] J. Dupire, A. Topol, and P. Cubaud, "Using game engines for non 3D gaming applications," Proc. of the International Conference on Computer Games (CGAMES 05), 2005, pp. 304–307.

[24] M. S. El-Nasr and S. Yan, "Visual attention in 3D video games," Proc. of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE 06), ACM, 2006, Article No. 22, doi: 101145/1178823.1178849.

[25] A. Kenny, H. Koesling, D. Delaney, S. Mcloone, and T. Ward, "A Preliminary investigation into eye gaze data in a first person shooter game," Proc. of the 19th European Conference Modelling and Simulation (ECMS), 2005, vol. 5, pp. 146–152.

[26] P. M. Fischer, J. W. Richards, E. J. Berman, and D. M. Krugman, "Recall and eye tracking study of adolescents viewing tobacco advertisements," Journal of the American Medical Association, 1989, vol. 261(1), pp. 84–89, doi: 10.1001/jama.261.1.84.

[27] L. J. Muir and I. Richardson, "Perception of sign language and its application to visual communications for deaf people," The Journal of Deaf Studies and Deaf Education, vol. 10(4), 2005, pp. 390–401, doi: 10.1093/deafed/eni037.

[28] R. J. K. Jacob and K. S. Karn, "Eye tracking in human-computer interaction and usability research: ready to deliver the promises," The mind's eye: cognitive and applied aspects of eye movement research, Elsevier, 2003, pp. 573–605.

[29] S. Hughes, "Coordination strategies for assisted viewpoint interaction," Proc. of the IEEE Symposium on 3D User Interfaces (3DUI 06), IEEE, 2006, pp. 17-24.

[30] S. Hillaire, A. Lecuyer, T. Regia-Corte, R. Cozot, J. Royan, and G. Breton, "A real-time visual attention model for predicting gaze point during first-person exploration of virtual environments," Proc. of the ACM Symposium on Virtual Reality Software and Technology (VRST 10), ACM, 2010, pp. 191–198, doi: 10.1145/1889863.1889907.

[31] K. K. Ball, B. L. Beard, D. L. Roenker, R. L. Miller, and D. S. Griggs, "Age and visual search: expanding the useful field of view," Journal of the Optical Society of America, 1988, vol. 5, pp. 2210–2220, doi: 10.1364/JOSAA.5.002210.

[32] K. Murphy and A. Spencer, "Playing video games does not make for better visual attention skills," Journal of Articles in Support of the Null Hypothesis, 2009, vol. 6, pp. 1–20, Available: www.jasnh.com [retrieved: 2014-11-24].

[33] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," Biometrika, 1965, vol. 52, pp. 591–611.

www.iariajournals.org

International Journal On Advances in Intelligent Systems
issn: 1942-2679

International Journal On Advances in Internet Technology
issn: 1942-2652

International Journal On Advances in Life Sciences
issn: 1942-2660

International Journal On Advances in Networks and Services
issn: 1942-2644

International Journal On Advances in Security
issn: 1942-2636

International Journal On Advances in Software
issn: 1942-2628

International Journal On Advances in Systems and Measurements
issn: 1942-261x

International Journal On Advances in Telecommunications
issn: 1942-2601