iPres 2013 Workshop Proposal

1st International Workshop on Archiving Community Memories

Thomas Risse, Cosmin Cabulea, Dominik Frey, Wim Peters

Abstract

The first International Workshop on Archiving Community Memories aims at promoting and discussing approaches, technologies and issues in preserving Web and Social Web content for building community memories. For this purpose it's intended to bring together communities involved or interested in Web archiving. We also explicitly encourage a closer dialogue between the technical oriented communities with people from Web archival and application communities as well as cultural heritage institutions in general in order to approach the topic from all relevant angles and perspectives.

1 Motivation and Objectives of the Workshop

Given the ever increasing importance of the World Wide Web as a source of information, adequate Web archiving and preservation has become a cultural necessity in preserving knowledge. This is especially the case for non-traditional digital publications, e.g., blogs, micro-blogs, social networks. The challenge with new forms of publications is that there can be a lack of alignment between what institutions see as worth preserving, what the owners see as of current value, and the incentive to preserve together with the rapidness at which decisions have to be made. For ephemeral publications such as the Web, this misalignment often results in irreparable loss. Given the deluge of digital information created and this situation of uncertainty, a first necessary step is to be able to respond quickly, even if in a preliminary fashion, by the timely creation of archives, with minimum overhead enabling more costly preservation actions further down the line.

In addition to the "common" challenges of digital preservation, such as media decay, technological obsolescence, authenticity and integrity issues, web preservation has to deal with the sheer size and ever-increasing growth and change rate of Web data. Hence, selection of content sources becomes a crucial and challenging task for archival organizations. Instead of following a "collect-all" strategy, archival organizations are trying to build community memories that reflect the diversity of information people are interested in.

Beside the creation of Web archives, their usage in applications plays an increasingly important role. Allowing the easy access to information based on different facets and across time is just one aspect. The possibility to look into the past, to understand how things are evolving opens the space for new application scenarios and analysis approaches. In line with the above, contributions to the workshop should focus on, but are not limited to:

- Web and Social Web Harvesting
- Focused & Topical Crawling
- Deep Web Capture
- Social Web Analysis
- Information Extraction
- Video and Image Analysis
- Appraisal and selection of content
- Applications & Use Cases
- Semantic Web Technologies
- Temporal Analytics

- Semi/full-automatic storytelling based on semantic services
- Legal issues
- Visualization of Heterogeneous Social Media Content
- Interactive Techniques for Semantic and Sentiment Analysis
- Interactive Techniques for Exploring Big Data over Time (Timeline)

It is the aim of the workshop to bring together researchers and practitioners involved in the ARCOMEM system development with users and developers of Web archive community and other application areas interested in archiving and preserving Web and Social Web content. Target audience are Web archivists, journalists, broadcasters, parliament archives, political parties archives, social science researchers, HCI researcher and designer etc. The workshop should stimulate the exchange of experiences made and best practices among the participants and help to identify upcoming challenges in Web archiving and the application of Web archives.

2 Format and Structure of the Workshop

The workshop will be a full day workshop grouped into several sessions. The sessions consists of presentations of approaches and technologies developed within the ARCOMEM project and complementing contributions from outside of the project. The workshop allows the submission of 2-4 pages of short and position papers, which will be peer-reviewed by members of the program committee. The paper will be published in electronic proceedings.

The workshop will start with a keynote talk and ends with an open panel discussion. Throughout the workshop, the organizers will develop a list of topics for the discussion at the end of the workshop. The discussion shall summarize and provide feedback on the workshop itself and on the ongoing work in the area of preserving community memories. Furthermore it is planned to have poster sessions in the coffee break to allow discussions about presented topics in smaller groups.

Schedule

- June 1st: Dissemination of Workshop Homepage & 1st CfP
- June 15th & June 30th: 2nd CfP and 3rd CfP
- July 15th: Deadline for submissions
- August 7th: Acceptance Notification
- August 21st: Camera Ready papers

Expected Number of participants: About 30-40

Special Requirements: Room for 50 people equipped with flip chart/black boards, pin boards for posters and beamer.

3 Organizing Committee

Thomas Risse is the deputy managing director of the L3S Research Center in Hannover. He received a PhD in Computer Science from the Darmstadt University of Technology, Germany in 2006. Before he joined the L3S Research Center in 2007 he lead a research group at Fraunhofer IPSI, Darmstadt. He was the technical director of the European project BRICKS about decentralized digital library infrastructures, coordinator of Living Web Archive (LiWA) project and technical director of the ARCOMEM project on Web archiving. Thomas Risse's research interests are Semantic Evolution, Digital Libraries, Web Archiving, and Self-organizing Systems. He was co-organizer of the DLSci Workshop (Digital Library Goes e-Science: Perspectives and Challenges) in conjunction with ECDL

2006. He is co-organizer of the SAME workshop series on Semantic Ambient Media Experience in the years 2008-2012 held in conjunction with ACM Multimedia 2008, AmI 2009 & 2010, CT 2011 and Pervasive 2012.

Cosmin Cabulea is a media researcher specializing on data crunching, visualization, data driven product development and data driven journalism at DW. He has been with Deutsche Welle since 2008, originally with the Department of Media Research and Audience Analysis, so he knows how to unearth and analyse audience needs and interests. He switched positions in 2011 and, now in the Innovations Projects Department, uses the gained knowledge to develop innovative and targeted offers for projects in which DW is involved. Currently he is responsible for the ARCOMEM project. Moreover he is in charge of various tasks related to data driven journalism. He holds a master degree in media studies with minors in psychology and political science from the University of Bonn.

Dominik Frey is a researcher and web archivist at SWR, a German public-service broadcaster (part of the ARD). He manages and develops the company's internal web archive as a research service for journalists. In addition he is in charge of SWR's contributions to the ARCOMEM project. Before joining SWR in 2009, he worked as a web analyst and SEO manager for FOCUS Online. He holds a master degree in sociology with minors in business and computer science from the University of Freiburg.

Dr. Wim Peters is a research fellow in the department of Computer Science at the University of Sheffield. He has been active in the field of computational linguistics for 16 years, and has participated in various EU and national projects covering multilingual thesaurus creation, corpus building and annotation, information extraction in various domains, semantic resource analysis, ontology creation and evaluation, and web archiving.

Some of his previous projects are EuroWordNet (multilingual resource creation), LOIS (legal wordnet building), DALOS (knowledge acquisition from legal texts), NeOn (life cycle of ontology networks) and CLARIN (the creation of a grid-based research infrastructure for the humanities and social sciences). His present responsibilities include the coordination of the ARCOMEM project (www.arcomem.eu).

He is an active member of various ISO/W3C committees on the standardization of terminological lexical and ontological resources, and of several conference and conference workshop program committees. He acts as a reviewer of many international journal and conference submissions.

4   Preliminary Programm Committee

- Bogdan Cautis, Télécom ParisTech, France
- Robert Fischer, Südwestrundfunk, Germany
- Jonathon Hare, Uni Southampton, UK
- Dimitris Koryzis, Hellenic Parliament, Greece
- France Lasfargues, Internet Memory Foundation, France
- Amin Mantrach, Yahoo! Research Barcelona, Spain
- Diana Maynard, Uni Sheffield, UK
- Pierre Senellart, Télécom ParisTech, France
- Yannis Stavrakas, Institute for the Management of Information Systems, Greece