International ACM Workshop on Contextualized Attention Metadata: Collecting, Managing and Exploiting Rich Usage Information (CAMA 2006)

Martin Wolpers, Jehad Najjar, Erik Duval
Ariadne Foundation and Katholieke Universiteit Leuven, Belgium
{martin.wolpers, jehad.najjar,erik.duval}@cs.kuleuven.be

1 Introduction

Providing flexible access to relevant information in an effective and efficient way remains a challenge. Approaches based on attention metadata present novel opportunities to deal with this challenge. Attention metadata indicate what end users do with information objects. Applications such as authoring tools, search tools, repositories, email applications, blogs and wikis can generate attention metadata. Collecting, managing, merging, analyzing and exchanging distributed attention metadata enables advanced tools to exploit data about user experiences in heterogeneous contexts for more flexible user information interaction. For instance, users of one application can be recommended relevant material, based on their attention given to information that they have interacted with in other applications. More concretely, this approach could allow a user recommendation service to take into account whether the end user is processing her email or authoring a scientific paper while she is listening to the music.

CAMA 2006 was held in conjunction with the 15th ACM Conference on Information and Knowledge Management (CIKM 2006) in Arlington, USA, on November 6-11, 2006. Contextualized attention metadata is a cross-disciplinary research topic that combines diverse fields like databases, information retrieval, knowledge representation, knowledge management, user modelling, recommender systems, learning technologies, etc. The main objective of the workshop was to bring together researchers, practitioners, and developers to discuss recent advances and identify future challenges.

2 Topics of Interest

The main topics of the workshop included:

- Attention Metadata collection frameworks
- Management of large volumes of attention metadata
- Real-Time stream based attention metadata processing
- Methods and algorithms for analysis of attention metadata
- Log Analysis in databases and web applications
- Information mining in attention metadata streams
- Information extraction in attention metadata streams
- Privacy and Security
- Personal Information Management
- Representation of attention metadata
- Attention metadata services
- Applications of attention metadata streams
- Merging of attention metadata streams

The workshop attracted 16 international researchers from various communities, ranging from database and information retrieval to knowledge management and eLearning. The workshop was held in an highly interactive format with continuous discussions in between and during presentations. Participants found the workshop very productive and successful. More details about the workshop can be found online at http://ariadne.cs.kuleuven.be/cama2006/.

The workshop program consisted of six paper presentations and two keynote speakers. The three short and three long papers were organized in two tracks dealing with the two main issues of attention metadata, namely capture and usage of metadata focussed on attention.

In order to have a highly interactive environment for the full day workshop, we started the first session by asking the workshop participants to contribute their thoughts. They were asked to state their definition of attention metadata and how they envision using it in their contexts? The summary of their answers yields a number of definitions of attention metadata:

Attention metadata describes the behaviour with which users deal with information. It can be collected by observing the user's activities. It possibly allows deriving conclusions on the user's interests and goals by relating activity to content. Furthermore, by relating attention metadata of several users, behavioural patterns might be observed.

Steve Gillmor, our first keynote speaker, elaborated on current developments with GestureBank (http://www.gesturebank.com/), a non-profit organization that aims to open the usage of attention metadata to a wider public, by providing a solution for secure and reliable storage of private attention metadata. Currently, attention metadata in GestureBank is captured with the AttentionTrust (http://attentiontrust.org/) recorder. Enabling the user to stay in control of her data is their main concern.

David Wiley (Utah State University), our second keynote speaker, discussed how attention metadata can support web-based reading in working groups. The approach is based on observing how group members read web content. The observations are used to suggest further reading by providing personalized RSS feeds that include the suggestions.

3 Paper Presentations

The first session of the workshop dealt with the use of contextualized attention metadata. The applications ranged from usage metrics to ranking and decision making.

Joe Pagano (Library of Congress, USA) presented *Benefits and challenges of developing a public sector metrics program using commercial tools*. The paper discusses how the Library of Congress developed a metrics program using commercially available tools. Individual commercial tools for (web) metrics analysis are discussed in relation to their usability in the Library of Congress. The focus lies on organizational and technical issues, e.g. the integration and relation of data from different tools that use different data formats and implementation challenges in different information system environments.

Xavier Ochoa (Escuela Superior Politécnica del Litoral, Ecuador) discussed the *Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects.* The paper proposes and details the use of Contextual Attention Metadata, gathered from different tools used in the lifecycle of a Learning Object, to create ranking and recommending metrics. Four types of metrics are detailed: Link Analysis Ranking, Similarity Recommendation, Personalized Ranking and

Contextual Recommendation. While designed for Learning Objects, the authors show that these metrics could also be applied to rank and recommend other types of reusable components like software libraries.

David Archer (Portland State University, USA) discussed *Capturing and ReUsing Human attention in Corporate Decision Making.* This context involves gathering, organizing, conflict-solving and using fine-grained information from diverse sources and formats to evaluate scenarios and choose courses of action. Though the pertinent information is readily available it is typically not integrated and ready for use. Tools for organizing this information are scarce, in part because the desired structure is often not pre-defined or available in an existing ontology. The authors introduce a construct called a manifestation, along with the notion of collections of manifestations, to support decision making based on the requirements and constraints of a corporate decision making environment while focussing on the entity in question.

The second session focussed on the capturing contextualized attention metadata from various perspectives.

Dimitrii Zagorodnov (University of Tromso, Norway) presented *WAIFR: Web browsing attention recorder based on a state-transition model.* This paper presents a model for a software component that monitors user interaction with a computer system and distributes summaries of that interaction to remote repositories. The model is illustrated with a prototype, implemented as a Web browser extension, for recording temporal aspects of Web browsing, such as time spent reading a page or keeping it open. To give the user control over privacy exposure and network resource usage that are inherent in remote monitoring, the design allows the user to specify how detailed the summaries are and whom they are sent to.

Bernhard Schandel (University of Vienna, Austria) presented *The SemDAV Project: Metadata Management for Unstructured Content.* In this paper, the authors point out why the file system has significant drawbacks for efficient management of metadata. To overcome these issues, their approach extends the well-known WebDAV protocol enabling metadata to be retrieved, stored, managed and used in searching and browsing situations.

The last presentation was given by Stefania Costache (L3S Research Center, Germany) on *Application Independent Metadata Generation.* The paper deals with the capture and use of attention metadata to personalize the usage of information stored on the user's desktop. To achieve this, contextual information across heterogeneous media types, file formats, and applications should be annotated and linked. The paper presents a light weight system which monitors the file structure and automatically generates metadata based on user activities. The authors illustrate the utility of extracted metadata by showing how it can be leveraged to enhance conventional full-text desktop search.

4 Discussion

The workshop was a successful event fostering discussions about the various issues involved in attention metadata. Participating researchers from various communities contributed their expertise and research findings to advance the field of attention metadata further.

Lively discussions continued throughout the whole workshop, using the paper presentations as topic feeds rather than separating papers from each other. All participants actively commented on the presentations by reflecting on their own work and by relating to the other presentations.

The discussion ranged from the issues of collecting attention metadata, contextualizing attention metadata, privacy and security to possible uses of attention metadata. The participants agreed that tools are still missing to collect contextualized attention metadata from heterogeneous sources. Nevertheless, research in the diverse communities already shows that the storage of such data is not the real issue anymore. Instead, research focuses on the observation and exchange of contextualized attention and how this data is then analyzed and used. RSS as the "transport medium" is a promising way of sharing attention metadata, because the data provider can control content and access to it through appropriate tools. All participants agreed that contextualized attention metadata enables user profiles that capture the user's behavioural patterns more precisely. Therefore, the participants expect that patterns describing the user behaviour will be used for personalization. Furthermore, such user profiles will be dynamic in the sense that they are able to incorporate new real-time observations about the user.

Privacy and security issues require that the user stays in control of her data. Attention metadata provided allows for a highly detailed and personal user profile. As such, this profile is of high value to all sorts of interested parties, like for instance for advertisement purposes. In order to prevent the abuse of these personal data, technologies need to empower users to stay in control of their data. Users need to be able to control which data they provide to whom and when. With GestureBank, Steve Gillmor presented one solution that allows users to fully control their data while, at the same time, enabling a non-profit business model that supports the possibly commercial interests of attention metadata providers and parties interested in their data.

5 Conclusion

The workshop enabled participants to get acquainted with and discuss attention metadata related research approaches stemming from various communities. As a result, the participants conclude that the number of available sources for contextualized attention metadata is rapidly growing, spanning from the collection of simple web browsing-based attention and behaviour to mouse gestures and monitoring work on the user's desktop.

The participants identified the two following research issues where further work is necessary. The use of contextualized attention metadata has not yet been researched to its full extend. Very preliminary work emerges that shows highly encouraging results. Commercial and industrial uptake is not yet there. Emerging fields are for example advanced usage statistics, behavioural pattern mining and finally advanced personalisation to meet the requests of the single user.

Furthermore, significant advances are necessary to ensure privacy and security.

We want to thank the organizers of the CIKM conference for their support. Special thanks go the members of program committee, who did a great job on reviewing the submissions. Finally, we also thank the authors and the participants for moving this research field forward.

6 Future Workshops

CAMA2006 was the first initiative that brought together researchers interested in attention metadata to discuss uses, challenges and future directions. Next year, we will organize CAMA2007 to continue the fruitful discussions on contextualized attention metadata and its applications. The workshop was supported by ACM CIKM organizers and the European Network of Excellence PROLEARN.