Supporting Forgetting and Semantic Enrichment of e-Memories through Annotation

Reza Rawassizadeh
Institute of Software
Technology and Interactive Systems
Vienna University of Technology
Favoritenstrasse 9-11/188
Vienna, Austria
rrawassizadeh@acm.org

Martin Tomitsch
Design Lab, Faculty of Architecture,
Design and Planning
The University of Sydney
148 City Road
2006 NSW Australia
martin.tomitsch@sydney.edu.au

Elaheh Momeni
Research Group Multimedia
Information Systems
University of Vienna
Liebiggasse 4/3-4, 1010
Vienna, Austria
elaheh.momeni.roochi@univie.ac.at

A Min Tjoa
Institute of Software
Technology and Interactive Systems
Vienna University of Technology
Favoritenstrasse 9-11/188
Vienna, Austria
amin@ifs.tuwien.ac.at

Katarzyna Wac
Institute of Services Science
University of Geneva
Rue de Drize 7, Battelle A
CH-1227 Geneva, Switzerland,
katarzyna.wac@unige.ch

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
The metaphor of life-logging (personal e-memory) promises a complimentary assistance to the biological human memory. However life-log information, which is collected from raw context sensors data, is weakly coupled with real life events. Further forgetting, an inherent feature of human memory, is not supported in life-logs. To address these limitations of life-logs we propose two novel approaches, which are: (I) supporting digital forgetting, and (II) facilitating reminiscence from users' e-memories through semantic enrichment. Forgetting is achieved based on the information expiration timestamp and personal preferences. Reminiscence is achieved by annotating life-log information with meaningful human readable tags and based on its social access-scope.

Keywords
Life-log, personal e-memory, forgetting, annotation, semantics

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Human Factors

Introduction

Life-logs (called also "personal e-memory") are tools or systems which digitally record users' life events by sensing and collecting contextual information and digital readable activities of the user. Context sensors are the core component of life-logs and they are responsible for sensing and collecting information for the life-log dataset. These datasets host large amounts of contextual information about the owners and their activities. In addition, dataset size increases continuously. On one hand information retrieval (or in more human centric terms, a recall on the personal e-memory) is a major challenge. Since the data comes to the life-log dataset in different formats, from heterogeneous resources, basic issues that are being solved in RDBMS[1] are critical challenges in the context of PIM[2] and life-log systems [5]. On the other hand, forgetting is a necessary feature of any (biological) memory system [10], which should be taken into account, while designing electronic memories for human assistance [8].

A possible approach for assigning semantics and enriching life-log datasets based on described requirements is the use of annotation. We suggest employing annotation to (1) mark selected information objects for forgetting, and (2) assigning semantics to raw sensor information.

Annotation can be stored embedded in the information objects or separated from the target information objects [4]. In this work we chose to perform and maintain annotation separately because of the portability of the annotation dataset, flexibility of the tool and the heterogonous characteristic of the dataset, which makes it harder to embed annotation in it. The result of this research provides set of data files, which contain annotations and stored separately from the life-log dataset, but they have a spatio-temporal link with the life-log dataset.

In this paper, first we describe the theories and principles of which will be used in this research. Next we define the system and our implementation approaches. At the end we conclude the paper.

Related Works

To our knowledge no life-log tools that support forgetting exist. The research therefore builds on other efforts that enrich personal digital data storage, to facilitate reminiscences. Horvitz et al. [3] annotate personal datasets through memory landmarks. They employed a calendar crawler to construct predictive models which identify memory landmarks. The process of identification uses Bayesian learning methods to perform supervised training and infer landmarks from users' calendars. Kim et al. [6] proposed a method to annotate life-log information by fusing sensor data from a set of heterogeneous sensors. Aziwa et al. [1] tackle the issue of life-log information retrieval by identifying keyframes and summarizing them. As method of summarization they suggest Spatio-Temporal sampling for keyframes extraction and detecting senses in a conversation. This approach was used for video, audio, accelerometer and RFID sensors of a life-log application. In general all these efforts try to assist users in better reminding their past stored information.

Theories and Principles

It is important to clarify what type of memory is the focus of this research. Kroener et al. categorize electronic

[1] Relational Data Base Management System

[2] Personal Information Management

memories into: object memory, community memory, application memory and personal memory [7]. In this research we focus on personal memory, which is also another interpretation of the term life-log. Furthermore the following theories and principles constitute the foundation of this work:

Spatio-Temporal Aspect: We are living in a spatio-temporal world. Meanings, all of our life events except dreams happen in a specific location and at a specific date-time. Location and date-time are the most important factors in contextual information. The process of reminiscence is also relying heavily on date-time and location [10]. Based on the currently available technologies it is not always possible to sense the location, because location sensors such as GPS are not functional in every environment. For instance GPS cannot (yet) work indoors. However most of the operating systems have date-time, which is accessible since the target device has not been turned off. This means that date-time is a necessary field for any life-log record and all digital information objects should be stored with a timestamp. It leads us to conclude that, first date-time and then the location, are the most important links between different information objects which came from different and heterogeneous sensors.

Sharing: In our research [9], we have described sharing life-log benefits users in: health and medical studies, sousveillance (unlike surveillance it is bidirectional and not unidirectional), software personalization, learning social patterns and behavior, matchmaking and historical studies. There we proposed three scopes for sharing *Private*, *Friend* and *Public*. A private information object will never be shared and only the owner can receive access to that information. Public information is open for public access (read) in the target domain, and Friend information is restricted only to a specific community access. Here the process of automatic annotation will be done based on these three access limitation layers. In simple term, private information can use private source, public information can use public information, and Friend information can use community information to get annotated. This will be discussed in more detail in the System Definition and Implementation section.

Meta-data Annotation: Annotation is attaching extra information to a piece of information [4]. Annotation can be used for different purposes such as indexing, bookmarking, assigning meta-data, etc. Here by annotation we mean tagging and thus meta-data creation which is data about the data. Meta-data information objects will be extracted from other resources and they will be collected based on date-time and (if possible) location.

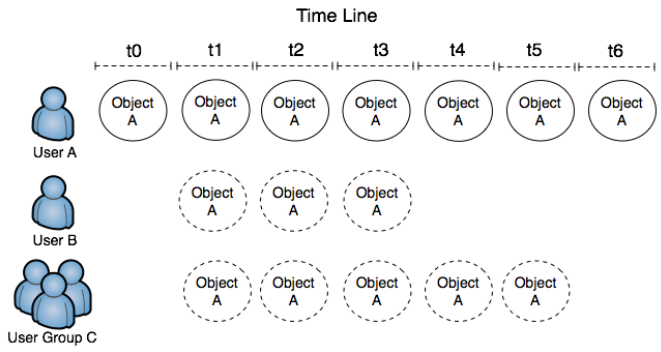

Figure 1: Access Expiration Definition for a Shareable Information Object.

Forgetting: Scharter [10] explained that one of the features of biological memory is forgetting. It is notable that the biological memory does not delete any information permanently, only the link to that information will be

removed while the information will remain in the memory permanently. We use a similar pattern to stimulate digital forgetting. This means that information objects will not be deleted, but they will not appear in the query result, when users issue a query over the dataset. In this context, we have defined two types of forgetting: *Personal Forgetting* and *Social Forgetting*. When users mark an information object to share, they should define the expiration date-time for the shared information object. This means that any shared information will be removed from the target community when the expiration date-time is reached. Figure 1, shows that User A created a data object at time t0, then shares this data object at time t1 with User B and a group of Users (User Group C). User B's access to data object A will be expired at time t3 and Users in Group C cannot access this object after time t5.

Personal forgetting should be done manually. This means that a user will mark which information should be forgotten by specifying the location and/or the date-time. Figure 2 shows how a user can select a specific area on the map or/and specific date-time to forget selected information objects. After the user selects an area on the map or a date-time period, she can click on the forget button and information objects from the selected location or date-time period will not appear in the query result anymore. However they will stay in the dataset.

System Definition and Implementation

In this section first we describe the implementation of forgetting followed by a description of the implementation of the semantic enrichment approach.

Forgetting

As has been described before, there are two types of forgetting, personal and social. In order to perform forgetting in a personal scope we use a GUI as shown in Figure 2, there the user can decide what to forget based on location or date-time. Afterward when a query engine is searching the dataset, it will check the annotation file and bypass forgotten records. In order to perform the forgetting in a social community, we are developing a Facebook application, which can remove the shared information object based on the given expiration date-time. Available social networking systems are not capable of hosting life-log information [9], but this prototype will be designed to implement forgetting for the Facebook, by defining expiration date-time for users' posts and remove them afterward. In the future we will extend it for a life-log based social networking system.

Semantic Annotation

Semantic annotation can be done either manually or automatically. Due to the fact that life-log dataset structure is always growing in size, manual annotation is not feasible and cumbersome. In some special occasions persuasive approaches such as games [2], could be used to motivate users annotating their life-log information, but this is beyond our research. We perform automatic annotation by aggregating information from different information sources. We store annotation in the JSON format based on their spatial and temporal dimension (if existing) in three scopes:

Personal (Private) annotating, means collecting the private information of the user such as users' Google calendar information for enriching personal information. For the temporal dimension we use a Google Apps Script to read and export Google calendar content for the specific date to the annotation file.

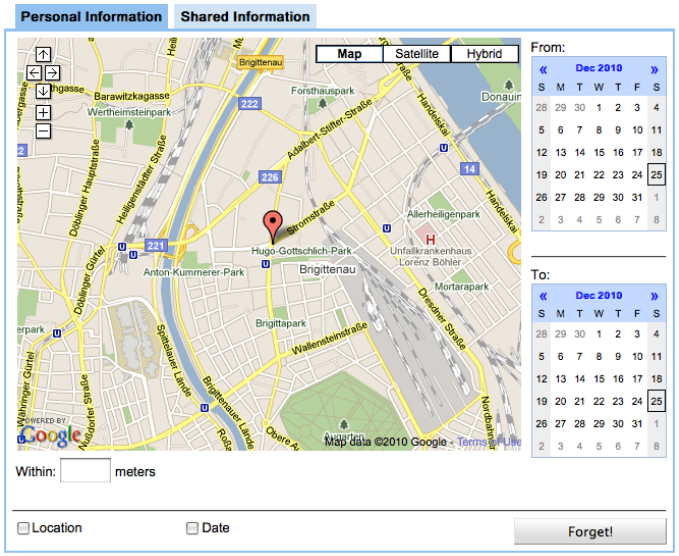

Figure 2: Spatio-temporal based forgetting. Users can specify location and date period, then based on these two factors, all information can be marked as forgotten in the dataset.

Public annotating, means using public information available on the Web for the enrichment process. For spatial dimension our system used LinkedGeoData[3], which uses the information collected by the OpenStreetMap[4] project and makes it available as an RDF knowledge base. It interlinks this data with other knowledge bases in the Linking Open Data initiative (for example to Wikipedia resources) that improves the discoverability of the annotated resource. As we already described it is not always possible to sense the location, however by using LinkedGeoData REST/Linked Data services we can find out information about points of interest (which are not sensed by sensor) in a circular area around the specific location of a person. The main advantage of using this service is that it not just finds popular locations, it also finds public unpopular locations, for instance restaurants or post offices. Moreover, the service enables limited search for specific classes of locations based on the LinkedGeoData ontology (For instance, lgdo:FastFood, lgdo: Restaurant). For the temporal dimension we used a Web crawler (from yahoo pipe) to extract news of a specific date and the current resident country of the user from news portals such as Google News, Yahoo News, Wikinews, and NY Times.

Community (Friend) annotating, means collecting communication information of the user from her social media systems to enrich her personal information. For the temporal dimension we developed a Facebook crawler to extract users' wall content, likes and comments of the friend in order to enrich friend-based information. Then we stored them in the annotation data files with their time-stamp. Yet in the spatial dimension we did not provide any location-related information for the friend scope annotation.

Figure 3 shows the overview of the semantic annotation framework based on the two layers. Process layer contains three enrichment components (based on the three described scopes) for collecting information from external services. Data layer, which contains three annotation files, are collected by process layer components.

[3] . http://linkedgeodata.org/

[4] . http://www.openstreetmap.org/

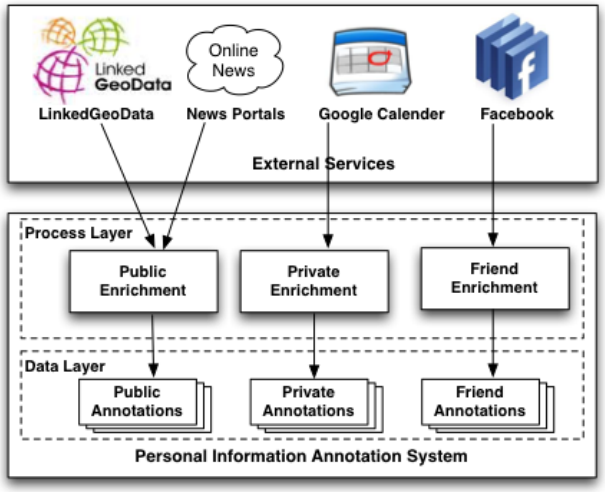

Figure 3: Overview of the semantic annotation framework

Conclusion and Future Work

Forgetting is one of the major requirements of personal e-memories. In this paper, we propose a forgetting mechanism, which supports social and personal forgetting. Additionally we introduced information annotation based on the information access level definition. We hypothesize that this form of annotation benefits users by using life-log, in order to recall and reminiscence. As future work, we plan to finalize the development of the social forgetting tools and to evaluate the provided approach in a group of users and not just single users.

References

[1] Aizawa, K., Tancharoen, D., Kawasaki, S., Yamasaki, T., Efficient retrieval of life log based on context and content. In Proceedings of *the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 22-31, 2004.

[2] Fogg,B.J, Persuasive Technology: Using Computers to Change What we Think and Do. *Morgan Kaufmann*, 2003.

[3] Horvitz, E., Dumais, S., Koch, P., Learning Predictive Models of Memory Landmarks. In *Proceedings of the CogSci 2004: 26th Annual Meeting of the Cognitive Science Society,* 2004.

[4] Hunter, J., Collaborative Semantic Tagging and Annotation Systems. *Annual Review of Information Science and Technology* 43(1):187-239, 2009.

[5] Jones, W. Personal information management. *Annual review of information science and technology*, 41(1):110-111, 2007.

[6] Kim, I.J., Ahn, S.C., Ko, H., Kim., H.G., Automatic Lifelog Media Annotation based on Heterogeneous Sensor Fusion. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. MFI 2008*, pages 703-708, 2008.

[7] Kroner, A., Schneider, M., Mori, J., A Framework for Ubiquitous Content Sharing. *Pervasive Computing* 8(4):58-65, 2009.

[8] Mayer-Schoenberger, V. Delete: The Virtue of Forgetting in the Digital Age. *Princeton University Press*, NJ, USA, 2009.

[9] Rawassizadeh, R., Toward Sharing Life-log Information with Society, In *Behaviour and Information Technology Journal* (to Appear).

[10] Schacter, D.L. How the Mind Forgets and Remember:The Seven Sins of Memory. *Mariner Press*, USA, 2002.