# An Architecture for Life-long User Modelling

Desmond Elliott, Frank Hopfgartner, Teerapong Leelanupab, Yashar Moshfeghi, and Joemon M. Jose

Department of Computing Science
University of Glasgow
Glasgow, G12 8QQ
United Kingdom
{delliott, hopfgarf, kimm, yashar, jj}@dcs.gla.ac.uk

**Abstract.** In this paper, we propose a unified architecture for the creation of life-long user profiles. Our architecture combines different steps required for a user profile, including feature extraction and representation, reasoning, recommendation and presentation. We discuss various issues that arise in the context of life-long profiling.

## 1 Introduction

The explosion of desktop applications, web applications, and computing devices that are capable of capturing different aspects of our daily lives has resulted in an almost continuous stream of digital life data. Harnessing these individual streams of life data represents an interesting research problem in the areas of life-long user modelling and the exploitation of the data captured in these models.

An example of a computing device that is capable of capturing an aspect of our daily lives is the portable music player. Portable music players allow people to listen to their collection of music on-the-go. *Last.fm*[1], a web-based music service, allows people to store a history of their personal music listening. The personal history of each user is used to infer relationships between the musicians that different people listen to, with the aim of helping them find new musicians they may want to hear, or concerts they may want to attend. One of the shortcomings of the portable music player and *Last.fm* data stream is the lack of context provided by the separation of this aspect of daily life from other aspects such as current or planned location and events. It would be beneficial to collect the data generated by as many devices and applications as possible into a single stream of data, otherwise known as a life-long, to produce a more complete user model and to allow systems to operate on data with a surrounding context.

If the heterogeneous data streams that are capturing aspects of a person's life could be collected into a single stream, that person may be able to benefit from relationships between the different aspects. For example, if you were to regularly listen to *The London Philharmonic Orchestra* on your portable music player and

[1] http://www.last.fm/

your mobile telephone sent a GPS location notification to your life-log, then it would be possible to receive a notification that *The Dresden Philharmonic Orchestra* were scheduled to perform while you were in Dresden for a meeting. This would not be automatically possible where the different data streams of digital data are separated.

Some existing approaches to capturing heterogenous streams into a life-log include MyLifeBits[2] and friendfeed[3]. MyLifeBits is a research project into a lifetime story of everything inspired by the Memex personal information system. MyLifeBits stores content and metadata for many different types of data and describes them using multiple overlapping common attributes. friendfeed is a web-service which collects data using the RSS or Atom publishing protocol standards from blogging services, bookmarking services, photo storage services, status update services, music services, and many other life data collection platforms.

In this paper, we propose an architecture for standardised life-long user modelling for collecting, storing, and processing the data captured by a diverse set of life data capture mediums into a life-long user model. The architecture collects the stream of data generated by people during their daily activities and uses the evolving set of concepts in domain-specific ontologies to extract relationships between the different data produced by the different mediums. These raw data and the linking concepts between them are exploited by collaborative filtering and content-based recommendation algorithms to help individuals receive information pertinent to their on-going daily activities.

The remainder of this paper is structured as follows. In Section 2, we provide a literature survey on related work. In Section 3 we propose the architecture of a life-long user modelling system, and in Section 4 discuss issues arising from life-long user modelling systems.

## 2 Background

In this section we provide an overview of research related to collecting data from users, building user profiles from collected data data, identifying relationships between different aspects of the user profiles, and providing recommendations based on user interactions and profiles.

### 2.1 Data Capture

The large volume of life data produced by individuals may not always be relevant to them in a given situation. This realisation leads to the need to gain an understanding of their interests to adapt the information presentation in a relevant manner. In information retrieval, there are rarely structured data representations to work with, however, relevance feedback techniques are used to

[2] http://research.microsoft.com/en-us/projects/mylifebits/
[3] http://www.friendfeed.com/

observe the performed actions to gain an understanding of which actions are most likely representative of interests. Traditional relevance feedback techniques require users to issue explicit feedback [3, 22]. Kelly and Teevan's overview of implicit feedback for query expansion and user profiling in information retrieval advocates the use of implicit feedback for identifying individual's preferences [15]. Exploiting implicit feedback on a short term basis has been studied intensively in the text retrieval [28], image retrieval [25] and video retrieval [26] domain. These results suggest that implicit indicators of relevance could also be useful for life-long user modelling.

### 2.2 User Profiling

A life-long user model is likely to be affected by changes in interests over time. The Ostensive Model, [5] is a model of developing information needs where the user's needs are described as a set of evolving information objects that exemplify the intentionality of their need. The Ostensive Model reformulates a temporal dimension to the notion of probabilistic relevance using ostensive relevance profiles. One of the most robust profiles, which has been successfully tested in various domains [25, 13, 11], is to decay the degree of relevance of features extracted from objects implicitly affected by users' interests through them over the time of search session. In this decay technique, there is an assumption that recent interactions with information objects is more important than previous interactions as they better exemplify the current information need. We believe a user model constructed using such a decaying relevance of features is suitable for a life-long user model.

A potential application of a life-long user modelling system is to retrieve or recommend a set of information objects to maximise the satisfaction of the user in different situations. To effectively construct a user profile, various information objects recommended in a life-long profiling system can be ranked in accordance to the Probability Ranking Principle [20]. This principle proposes that documents relevant to the user's information needs should be ranked in decreasing order of estimated probability of relevance made by a user about an information object, which is influenced by the objects previously examined through user's interaction recorded in life-log repository. Fuhr [6] proposed an extension to the principle which relaxes some strong assumptions that limit the case of accepting only one query, and not the interaction between a user and a system. In his proposal, the principle is framed within a situation-based framework where the information objects are ranked in the optimum ordering presented in each situation. We believe that applying the situation-based interpretation of this principle for a life-long user model can assist to recommend information relevant to user's information needs in many different situations.

### 2.3 Reasoning

A challenge in user modelling is to identify *events* of users' interests by analysing the information chunks from different sources stored in the users' profiles. Westermann and Jain [27] suggest the fusion of various types of data such as, but not limited to, physical locations, time points, or entities can help with identifying events. A reasonable approach to map the relationship between these data types is to rely on ontologies. Gruber [9] defines an ontology as a "content specific agreement" on vocabulary usage and sharing of knowledge. Tsinaraki et al. [24] argue that in the multimedia domain, ontologies can be used for annotation, indexing, query specification, retrieval and knowledge extraction. An early approach of using ontologies for user profiling is the SmartPush [14] project, where professional editors were asked to enrich information with semantic metadata. These metadata were then used to filter relevant information. Even though their approach is promising, it requires too much manual input, which questions its scalability. Gauch et al. [7] create an ontology-based user profile based on users' browsing behaviour. Their personalised retrieval system outperforms an unpersonalised baseline system, indicating the effectiveness of such profiles. Adopting this approach for creating life-long user profiles will remain the main research challenge in the next few years.

### 2.4 Recommendation

One potential application of creating user profiles is to identify users' interests and hence to recommend relevant items to them. Recommender systems can help users to overcome their information overload problem. Over the last decade, much research has been performed to improve recommender systems in both industry and academia. These applications based their recommendations on the users' likes and dislikes. Examples of recommender systems can be seen in Amazon.com for books, CDs and other products [16], MovieLens for movies [17], and Jester for recommending jokes [8].

We can classify the recommendation approaches into two main categories:

1. **Content-based** The items will be recommended to the user based on the information gathered from the user itself. This information could be the previous rating given by the users for other items or could be based on the user interaction to the system. The root of content-based approach can be traced to information retrieval [2], [21], and information filtering [4]. Due to early and successful improvements in information retrieval and the importance of several text-based applications, most content-based systems were used to recommended textual items such as text documents or news articles [1].
2. **Collaborative filtering** The most successful and widely used recommender systems are based on collaborative filtering. In this approach, items are recommended based on the information gathered from other users who share the same interest. The main problem with this approach is that users tend not to provide judgements about every action that can result in an entry in their profile. This results in very sparse profiles. In order to overcome sparsity issue, it has been proposed to use semantic information to improve recommendation performances such as approaches of Mobasher [18] and Moshfeghi [19]. Other approaches rely on context to recommend items [26].

## 3 Architecture & Issues

In this section we outline the architecture of our proposed system along with its components. Figure 1 shows the proposed architecture, which consists of four main components: a LLUM API, a Life-Log repository, a modelling component, and a recommendation component. The following subsections describe each component in detail.

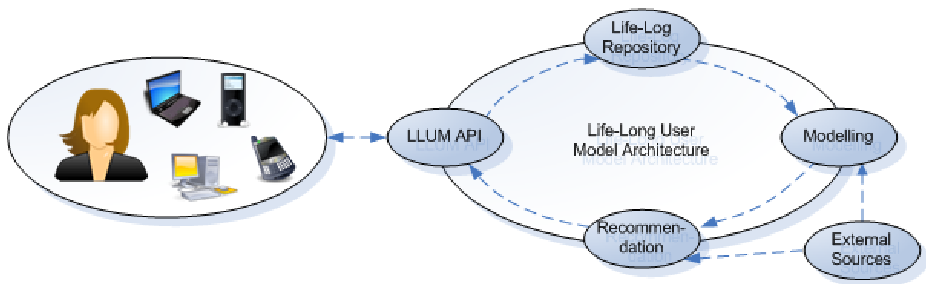


**Fig. 1.** High-level architecture of life-long user modelling

*LLUM API* The architecture of the Life-Long User Model (LLUM) provides an interface to support inputs of continuous life data streams obtained from an individual's devices and outputs of recommendations presented in an appropriate manner with respect to temporal and contextual awareness. This interface, referred to as the LLUM Application Programming Interface, allows our architecture to support many applications and devices. The LLUM architecture can collect the data generated by users in daily life activities and planned events in order to model the life data into a user profile with the surrounding context.

*Life-Log Repository* Data capturing from the incoming data stream reflects different aspects of users' personal interests. We propose a personal repository which can be automatically updated using any device. It would hence contain personal information such as GPS coordinates, music songs you listened to, or feedback provided on recommendations. One challenge for this component is to filter important from unimportant information in the repository, for example by identifying events.

*Modelling* Jain [12] suggests to model user activities into distinct events, which can be defined by temporal, spatial, experiential, casual, informational and structural aspects. For example, an event could be that you enjoy listening to classical music on your way to work using your portable music player. Recently, the use of ontologies has been proposed to combine these aspects to related events. Pursuing this idea, we propose to use ontologies to exploit the repository accordingly.

However, further research is still required for a positive identification of events. An important key for modelling user interests can be the use of external sources. If a user, for instance, listens to audio tracks composed by *Haydn* and *Beethoven*, a user model can be enriched with the information that both are classical composers. Another challenge is to keep these events in a user profile. We propose to store such profiles on online profile servers that can interact with other people's profiles and/or external sources. This inter-linking, aiming at recommending personalised information, will be described in the following component.

*Recommendation* Once a user's interests have been identified and modelled accordingly, this knowledge can be used to recommend other information that might be of a user's interest. For example, if a profile reveals a general interest in classical music and information about the current location of a user, location-sensitive recommendations classical performances occurring during the user's stay in the city could be suggested. Knowledge of the location of friends sharing a similar taste in music can lead to recommending that both could attend the concert together. This inter-linking, however, requires a careful investigation with respect to privacy issues since you might not always want to tell others where you are right now and what your interests are.

## 4 Discussion

In this paper, we introduced an architecture for life-long user modelling. The architecture provides an API for various applications and devices to capture user activities. This can lead to a permanent, personalised data stream which can be stored in an online repository. Since this leads to a large and rapidly growing volume of data, scalability becomes an urgent challenge. Issues such as which information is of importance and which is not needs to be considerd.

Given a repository of personal activity data, we propose to model users based on their events. Identifying these events is another research challenge. The main problem is that insufficient information may be available to positively identify such events, or the information available may be vague or ambiguous. Therefore, we propose the use of ontologies to assist with disambiguation and hence to create ontology-based user profiles. Even though it is now possible to handle ontologies that model large scale datasets consisting of a billion relationships between concepts [23], an important challenge is still to find the correct representation for each concept in a user profile. While it is a simple task to map captured GPS coordinates with an according ontology such as GeoNames, it is much more challenging to identify a user's interest, by exploiting their interactions with retrieval systems such as a personalised news video recommendation system, for example [10].

Another important challenge in user profiling is that users have multiple interests that can change over time. These issues are intensified when we are considering a life-long period. We assume that more recent events are of higher importance to a user and should be modelled accordingly. Nevertheless, dependent on the context, older events can become important again. We believe that new research directions are needed to understand the importance of captured events in the profile.

Another challenge is to identify those factors in the profile which should be employed for recommendation. We propose that both external sources and friends' profiles can be used to recommend information of interest. Considering these additional sources, privacy becomes a serious issue. It needs to be clarified to which degree each individual is willing to provide information that can be used by others. Users should be able to control, modify, and delete all details that are captured and used in a life-long profiling system.

In summary, the architecture we have described in this paper provides an overview of a life-long user modelling system. We have described the main components of such a system and some of the issues arising from attempting to create life-long user models.

## Acknowledgments

This research was supported by the European Commission under contracts FP6-027122-SALERO, FP6-033715-MIAUCE and the Royal Thai Government. It is the view of the authors but not necessarily the view of the community.

## References

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, June 2005.

[2] R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

[3] N. J. Belkin, C. Cool, K. B. Ng, and S. Park. Using relevance feedback and ranking in interactive searching. In *TREC-4 Proceedings of Fourth Text Retrieval Conference*, pages 181–209, 1996.

[4] D. Billsus, C. A. Brunk, C. Evans, and B. G. M. Pazzani. Adaptive interfaces for ubiquitous web access. *Commun. ACM*, 45(5):34–38, 2002.

[5] I. Campbell and C. J. van Rijsbergen. The ostensive model of developing information needs. In *Proc. of CoLIS-96, 2nd Int. Conf. on Conceptions of Library Science*, pages 251–268, 1996.

[6] N. Fuhr. A probability ranking principle for interactive information retrieval. *Inf. Retr.*, 11(3):251–265, 2008.

[7] S. Gauch, J. Chaffee, and A. Pretschner. Ontology-based personalized search and browsing. *Web Intelli. and Agent Sys.*, 1(3-4):219–234, 2003.

[8] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151, 2001.

[9] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, 43(5-6):907–928, 1995.

[10] F. Hopfgartner and J. M. Jose. On user modelling for personalised news video recommendation. In *UMAP'09 - First and Seventeenth International Conference on User Modeling, Adaptation, and Personalization, Trento, Italy*, 6 2009.

[11] F. Hopfgartner and J. M. Jose. *Toward an Adaptive Video Retrieval System*, chapter 6. CRC Press: Boca Raton, Florida, vol. 2 edition, 2 2009.
[12] R. Jain. Eventweb: Developing a human-centered computing system. *Computer*, 41(2):42–50, Feb. 2008.
[13] H. Joho, R. D. Birbeck, and J. M. Jose. An ostensive browsing and searching on the web. In *CIR*, 2007.
[14] S. Jokela, R. Sulonen, and M. Turpeinen. Agents in delivering personalized content based on semantic metadata. In *In Proc. 1999 AAAI Spring Symposium Workshop on Intelligent Agents in Cyberspace*, pages 84–93, 1999.
[15] D. Kelly and J. Teevan. Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum*, 37(2):18–28, 2003.
[16] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.
[17] B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan, and J. Riedl. Movielens unplugged: experiences with an occasionally connected recommender system. *IUI Conference*, pages 263–266, 2003.
[18] B. Mobasher, X. Jin, and Y. Zhou. Semantically Enhanced Collaborative Filtering on the Web. *Lecture Notes in Computer Science*, pages 57–76, 2004.
[19] Y. Moshfeghi, D. Agarwal, B. Piwowarski, and J. M. Jose. Movie Recommender: Semantically Enriched Unified Relevance Model for Rating Prediction in Collaborative Filtering. *ECIR'09*, pages 54–65, 4 2009.
[20] S. E. Robertson. The probability ranking principle in ir. *Readings in information retrieval*, pages 281–286, 1997.
[21] G. Salton, editor. *Automatic text processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1988.
[22] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Readings in information retrieval*, pages 355–364, 1997.
[23] S. Schenk, C. Saathoff, A. Baumesberger, F. Jochum, A. Kleinen, S. Staab, and A. Scherp. SemaPlorer - Interactive Semantic Exploration of Data and Media based on a Federated Cloud Infrastructure. In *Billion Triples Challenge at the 7th International Semantic Web Conference 2008*, 2008.
[24] C. Tsinaraki, P. Polydoros, F. Kazasis, and S. Christodoulakis. Ontology-based semantic indexing for mpeg-7 and tv-anytime audiovisual content. *Multimedia Tools Appl.*, 26(3):299–325, 2005.
[25] J. Urban and J. M. Jose. Ego: A personalized multimedia management and retrieval tool: Research articles. *Int. J. Intell. Syst.*, 21(7):725–745, 2006.
[26] D. Vallet, F. Hopfgartner, M. Halvey, and J. M. Jose. Community based feedback techniques to improve video search. *Signal, Image and Video Processing: Special Issue on Multimedia Semantics, Adaptation & Personalization*, 2(4):289–306, 2008.
[27] U. Westermann and R. Jain. Toward a common event model for multimedia applications. *IEEE MultiMedia*, 14(1):19–29, 2007.
[28] R. W. White, J. M. Jose, and I. Ruthven. An implicit feedback approach for interactive information retrieval. *Inf. Process. Manage.*, 42(1):166–190, 2006.