User Modeling for the Social Semantic Web

Till Plumbaum[1], Songxuan Wu[2], Ernesto William De Luca[1], and Sahin Albayrak[1]

[1] Technische Universität Berlin, DAI Labor, Berlin, Germany
{till,ernesto.deluca,sahin}@dai-lab.de
[2] Otto-von-Guericke Universität Magdeburg, Magdeburg, Germany
wusongxuan@gmail.com

Abstract. How can we make use of the personal information a single user is spreading all over the Social Web every day? In this paper we investigate what is needed from a user model point of view to support user data sharing and aggregation to enhance personalization and recommendation services. We present a study of 17 social applications to define requirements and attributes for a common user model that allows sharing of user data and analyze what is needed to enhance user model aggregation approaches. As a result, we present a comprehensive user model especially fitted to the needs of the Social Web. Furthermore, we present a WordNet for the user modeling domain as part of the user model to support user model aggregation.

Keywords: User modeling, Social Web, Semantic Web, User Model Aggregation

1 Introduction

Every day, people in the Social Web create 1.5 billion pieces of information on Facebook, over 140 million tweets on Twitter, upload more than 2 million videos on YouTube and around 5 million of images to Flickr[3]. This huge amount of social data attracts researchers who want to use it to learn more about user preferences and interests, and enhance recommendation and personalization systems. What most current system have in common is that they use data from a single application and depend on sufficient user information (user behavior or ratings) to produce good results [1, 2]. By using the distributed personal information a single user produces on a daily base, and by building a holistic model of the user, personalization and recommendation quality can be further enhanced. But, for this holistic model the distributed user data has to be aggregated across applications. This idea is not new, it has existed since the 90's where different research initiatives proposed generic user modeling servers that build a central structure to manage and share user information [9, 10]. These approaches could

[3] http://www.scribbal.com/2011/04/infographic-how-much-daily-content-is-published-to-twitter-facebook-flickr/

not succeed because of their static, predefined user models while application-based user models strongly differ in the information they need to know about a user (as we will show in Section 3). Another reason for the failure was that applications do not want to lose control over their data, thus, a central storage was not wanted. New trends from the Semantic Web can provide a remedy. Instead of having a central server, ontology based user models are proposed to support data aggregation and sharing. Thus, applications can keep their data but use a common "language" to model the information. While semantic technologies help to overcome technical problems, the main questions remain: What user information must a semantic model contain with focus on the Social Web? What requirements must a model fulfill to support data sharing and aggregation?

In this paper, we want to give answer to those questions by analyzing user models from different Social Web applications and draw conclusions about the diversity and type of user information that such a generic user model should have. We therefore discuss existing work and motivate a semantic Social Web User Model (SWUM). Requirements and structure of SWUM will be introduced in Section 3 and is based on the extensive analysis of 17 Social Web applications in Section 3.1 and 3.2. In Section 3.3 we also carefully investigate what is needed to enable an easy, automated, aggregation process. To give a better understanding of the intended use of the SWUM we present a use case in Section 4. The main contributions of this paper are an extensive analysis of requirements of today's Social Web applications regarding stored user data and the introduction of a new Social Web user model that is:

– generally adapted to the needs of Social Web applications and
– that allows an easy data sharing between applications.

2 Related Work

Until the turn of the millennium, most personalization and recommendation research focused on user information available in one application and how to use this information to enhance personalization quality. With the influence of the Social Web, or Web 2.0, and the fact that user information is highly distributed over several applications, research started to explore cross-system personalization approaches. This research can be roughly classified into two major directions [11]:

– A centralized approach with standardized models that aggregate the distributed user information and build the basis for cross-system information transfer.
– A decentralized approach where dedicated software components transfer user information from one application's representation to another.

The work presented in this paper is in alignment with the first direction, the centralized approach. This approach can also be subdivided into two aggregation strategies. The first strategy proposes the use of standardized user models which all involved applications must agree on. The second strategy deals with the mediation of different user model representations using meta-models that connect user data from one application with data from another application, in the same domain, or across domains. The standardization approach involves no computational effort to aggregate data as all data already is in the same format. An effort in this direction is the General User Modeling Ontology (GUMO) created by Heckman et al. [7]. GUMO is a comprehensive user model that intends to cover all aspects of a user's life. The user dimensions covered range from contact information and demographics over abilities, personality right up to special information like mood, nutrition or facial expressions. GUMO is at the time of this writing the most comprehensive generic user modeling ontology. Another approach that came up with the Web 2.0 is the Friend-of-a-Friend (FOAF) ontology. FOAF is a lightweight model that is integrated on the website, the application's user interface, using RDFa. FOAF covers basic user information like contact information, basic demographics and allows to specify some social relations like group membership or "knows" relations to other FOAF profiles. GUMO, which represents the most generic user model, covers only some parts of information that are needed for the Social Web. Especially the *Interest* dimension (in music, books, etc.) and user information like accounts for different Social Web applications, which are crucial, as we show in Section 3, are completely missing. FOAF, which is designed for a Web use, is too simplified. FOAF has a "knows" relationship, which defines a social relation, but the type of the relation remains unclear. Also no user needs and goals can be defined, which is part of many social applications as we will see in Section 3.

The second strategy is to build meta-models that allow defining how application-dependent user data corresponds to user data from another application. This has the advantage that applications are not forced to adopt a predefined generic user model and can rely on their own model. In [13], the authors present an aggregation ontology which gives applications the possibility to define a model, which describes how information in different profiles is related and how data can be aggregated. Furthermore, the ontology not only allows to define relations between data in different application models but also to define the overlap, the similarity, of the modeled information. So it is possible to define that the field "interests" in one application and the field "music interests" in another, is related but only to a certain degree as "music interests" is only subset of "interests". In [16], van der Sluijs et al. present the Generic User model Component (GUC) which builds a central component where all applications have to subscribe to and describe their user model via a schema defining the data structure of the user models for different applications. The authors also suggest the possibility to use different matching and merging techniques to map input schemas and create a merged schema as the union of the input schemata and to construct combined ontologies of the application schemata. While the meta-model approach seems to be a more practical one but to achieve a semantic and syntactic interoperability, the big disadvantage is that is needs a lot of effort to connect all the different user models. This work currently has to be done manually or semi-manually and must be repeated for every new application user model.

To summarize: both strategies and the presented related work have shortcomings. Because of the big differences, regarding the covered user information and representation forms in different applications, the development of a commonly accepted ontology, covering all aspects of user modeling for all domains seems not feasible. The meta-model approach, without automatic aggregation mechanisms, is only applicable in small settings where only a few applications are connected and not for the Social Web. We therefore propose a middle way: We need a new 'common' user model that combines aspects of the presented approaches and focuses on a special domain, the Social Web. Also, the user model should support automatic aggregation by defining a structure that allows finding relations between different user model concepts and allows for a flexible extension of the model.

3 Requirements for a Social Web User Model

To define a user model for the domain of the Social Web, we first have to understand the demands of social web applications on user models. Therefore, we did an extensive survey of the modeled user information of 17 well-known Social Web applications. The list of analyzed applications is shown in Table 1. The applications were chosen because of their size and level of awareness (number of users, global distribution). To be able to consider local differences, we also included applications that are strong in only one or two regions (Orkut in South America, Lokalisten and StudiVZ in Germany). We also selected Social Web applications from different kinds of domains, photo- and video-sharing platforms, short-message services, social networks, etc. To decide if the user information stored by an application is of importance, we picked at least two Social Web applications from the same domain.

Table 1. List of 17 social applications that we analyzed for the requirements analysis

Facebook	`http://www.facebook.com`	Myspace	`http://www.myspace.com`
Windows Live	`http://home.live.com`	YouTube	`http://www.youtube.com`
Flickr	`http://www.flickr.com`	Yahoo	`http://de.yahoo.com`
Picasa Web	`http://picasa.google.com`	StudiVZ	`http://www.studivz.net`
Digg	`http://www.digg.com`	Yelp	`http://www.yelp.com`
Lokalisten	`http://www.lokalisten.de`	Orkut	`http://orkut.com`
Identi.ca	`http://identi.ca`	LinkedIn	`http://www.linkedin.com`
Vimeo	`http://www.vimeo.com`	Xing	`http://www.xing.com`
LastFM	`http://www.last.fm`		

For each evaluated application, we collected the type of information and the internal attribute name. Table 2 shows the type of user information and where the information was found on the Web page. The internal attribute names, used by each application are particularly important as they are later used to define and name the attributes of the Social Web User Model (SWUM).

Table 2. Evaluation example for Yahoo: User information, attribute name and where the information was found on the Web page.

IU Name	Source code ID	Found on
Name	name	Registration Page
Firstname	firstname	Registration Page
Surname	secondname	Registration Page
Gender	gender	Registration Page
Birthday	birthdaygroup	Registration Page
Country	country	Registration Page
Postal Code	postalcode	Registration Page
Yahoo! ID and Email	yahooid	Registration Page

To be able to create our SWUM, we first have to decide which type of information, which user model dimensions, should be part of the model and which attributes in the different dimensions should be supported.

3.1 User Model Dimensions

After collecting all the information, the first step is to determine the user model dimensions that our user model has to cover. As shown in GUMO, a lot of dimensions exist, but not all of them are required in the context of the Social Web. Several dimension are mentioned and discussed in the literature. We present a consolidated taxonomy that bases on [17, 6, 8, 9, 3] and builds the basis for the selection of needed dimensions for our model:

– *Personal Characteristics* (or Demographics) range from basic information like gender or age to more social ones like relationship status.
– *Interests and Preferences* in an adaptive system usually describe the users interest in certain items. Items can be e.g. products, news or documents.
– *Needs and Goals*: When using computer systems, users usually have a goal they want to achieve. Such goals can be to satisfy an information need or to buy a product. The plan to reach such goals is for example to support users by changing navigation paths or reducing the amount of information to a more relevant subset.
– *Mental and Physical State* describe individual characteristics of a user like physical limitations (ability to see, ability to walk, heartbeat, blood pressure, etc.) or mental states (under pressure, cognitive load).
– *Knowledge and Background* describe the users knowledge about a topic or system. It is used in educational systems to adapt the learning material to the knowledge of a student, display personalized help texts or tailor descriptions to the technical background of a user. The knowledge and background is a

long-term attribute on the one hand but can differ and change from session to session depending on the topic. Knowledge and background about certain topics can increase or decrease over time [3].
- *User Behavior*: The observation and analysis of user behavior is usually a preliminary stage to infer information for one of the previous mentioned dimensions. It can also serve for direct adaptation like using interaction history to adapt the user interface to common usage patterns of the user.
- *Context*: In computer science context generally refers to "any information that can be used to characterize the situation of an entity" [4], but the discussion about what context actually is, is still ongoing[5]. In the area of user modeling, the term context focuses on the users environment (e.g. Location or Time, or devices the user interacts with) and human characteristics. Human characteristics describe Social Context, Personal Context and overlap with the *Mental and Physical State* dimension).
- *Individual Traits* refer to a broad range of user features that define the user as an individual. Such features can be user characteristics like introvert or extrovert or cognitive style and learning style.

Based on this user taxonomy, we checked all 17 applications if they cover these dimensions. Fig. 1 shows that social applications only cover some dimensions. All of the applications maintain *Personal Characteristics* and most of them also use *Interests and Preferences* information. Not used at all are the dimensions *Individual Traits* and *Mental and Physical State* which are more used in educational systems than in Social Web applications [3].

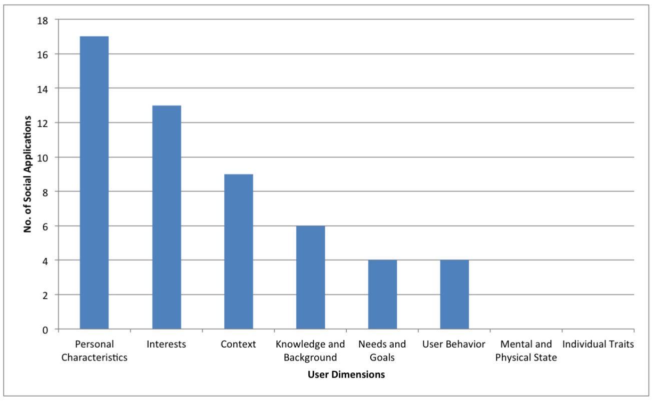


Fig. 1. Number of applications storing user information in the different user dimension categories.

The usage of *Knowledge and Background* and *Context* depends on the focus of the social application. Social business applications, like LinkedIn or Xing, support the *Knowledge and Background* dimension as users can enter their college degree, areas of profession, etc. The support for the dimension *User Behavior* is not easy to work out, as user behavior usually is an implicit feature and not displayed on the user profile page of an application. It can be assumed, though, that almost all applications track user behavior on their site. A positive exception is "Google Dashboard" [4] where a user gets an easy overview of the stored personal information e.g. previous search behavior. The *User Behavior* dimension, although it is an important piece of adaptation and personalization, is to complex to be part of a generic Social Web User Model. For this purpose we recommend a specialized approach with an extra user behavior ontology as presented in [14, 12]. *Context* is an important area as the latest research shows and of importance for a Social Web User Model [15]. However, not all forms of context can be considered as a part of a Social Web User Model. The analysis showed that the Social Context and Location is of importance and therefore those subdimensions of context are part of SWUM. The importance of the context Time also seems of interest, but did not show up in our analysis.

From this analysis it follows that a main requirement for Social Web user model is, that it has to cover the user dimensions *Personal Characteristics, Interests, Knowledge and Behavior, Needs and Goals and Context (Social Context, Location)*. Accordingly, these dimensions are part of our SWUM.

3.2 User Model Attributes

After we selected the dimensions to be covered, we have to define the attributes that the user model should support.

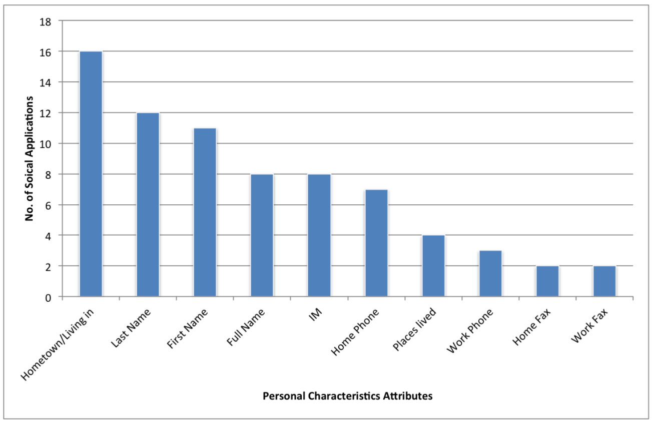


Fig. 2. Attributes of the *Personal Characteristic* dimension and how often they occur in the different applications.

[4] https://www.google.com/dashboard

The procedure for the attribute selection is similar to the procedure used to select the dimensions. We checked the different attributes of the different applications. Fig. 2 gives an example for the*Personal Characteristic* dimension. It shows an excerpt of the attributes and how often they occur in the analyzed social applications. In this way, we selected a set of attributes for each dimension. An example for the *Personal Characteristic* dimension is shown in Fig. 3. The *Personal Characteristic* is divided into two main concepts namely Demographics and Contact Information. The concept Location is a helper concept to model locations and link certain information, e.g. places lived, to it.

Contact Information

- First name: string
- Middle name: string
- Last name: string
- Full name: string
- Nickname: string
- Username: string
- Maiden name: string
- Living in: List of *Locations*
- Places lived: List of *Locations*
- Current City: *Location*
- Hometown: *Location*
- Work Phone: int
- Home Phone: int
- Mobile Phone: int
- Home Fax: int
- Work Fax: int
- Personal Email: string
- Work Email: string
- Personal Homepage: string
- Work Homepage: string
- IM: string

Demographics

- Gender: string
  - Female: bool
  - Male: bool
- Birthday: date
  - Day: int
  - Month: int
  - Year: int
- Birthplace: *Location*
- Language: string
- Other Languages: string
- Family status: string
- Education: *Education*
- Employment: *Employment*
- Employment History: List of *Employments*

Location

- Country: string
- State: string
- City: string
- Street: string
- House number: int
- Postal code: int

Fig. 3. SWUM attributes for *Personal Characteristic* dimension.

3.3 A User Model Word Net

An important outcome of the attribute distribution analysis was that often similar information is stored by most applications, but in differently named attributes, e.g. name (Yahoo) and real_name (LastFM) or homepage (LastFM) and website (Flickr). This problem of attribute name heterogeneity complicates a possible aggregation using a Meta-Model strategy. To cover that problem, we decided to extend our model with a WordNet like lexicon called User Model Word Net (UMWN). WordNet defines word sense relations between words. If a word represents a user attribute, the relatedness between different attributes can be acquired easily. However, many user attributes are not defined in WordNet. Moreover, many terms in WordNet are useless for user profile aggregation.

Hence, the standard WordNet does not help, thus, we designed a reduced WordNet, specialized to serve the user profile aggregation and initially based on the attribute distribution of our analysis. The decision to use a WordNet based structure comes from the fact, that WordNet has a flexible and well-defined lexicon schema, which is publicly known and accepted. The user model terms can be linked to each other accurately by using the properties defined in WordNet. An example is depicted in Fig. 4 where the word sense relations for name and date are shown.

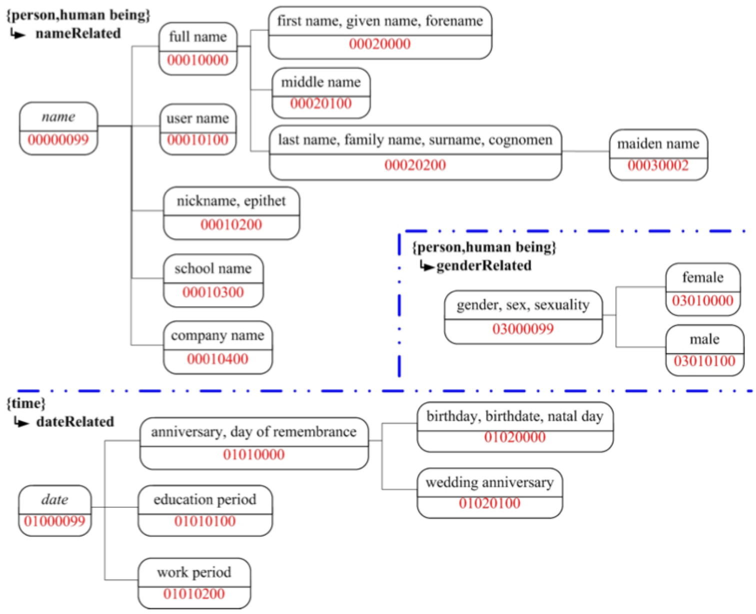
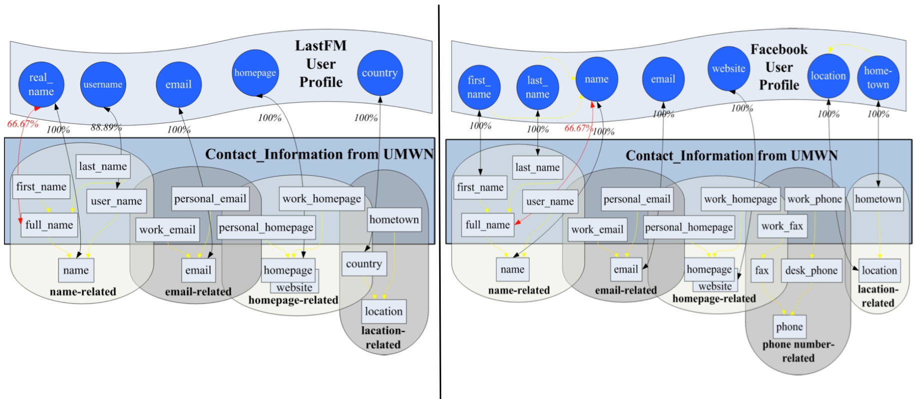


Fig. 4. User Model WordNet relations.

The UMWN is an important step for an automatized aggregation of different user models. It defines different types of word relations. The "Name" concept describes the relations between different types of name attributes that can occur in a user model. The concept "full name" consists of different subclasses like "first name", which has several synonyms ("given name" or "forename"). UMWN is stored in RDF(s)/OWL. Using ontology structures has the advantage that such a model is not static and can be easily extended. Our UMWN is extensible, towards not only to the individuals, but also to the schema of UMWN. Because of the highly distributed and heterogeneous user information in different user models, extensibility is an important feature. The UMWN contains currently ca. 520 syn sets where around 200 are unique in the User Model WordNet and not part of the common WordNet. It also contains over 100 antonyms and homonyms and 200 meronyms.

4 Use Case: Profile Aggregation with the SWUM

To outline the intended usage and functionality of the SWUM (which includes the UMWN) we want to exemplary explain the steps needed to aggregate a Facebook user model and a LastFM user model. The aggregation is a two-step process which we want to explain by the example of the website/homepage attribute shown in Fig. 5. First step is to connect the LastFM attributes to the SWUM (see Fig. 5a). The LastFM user model has the attribute "homepage" which can be directly linked to the SWUM, with a concept match of 100%. The Facebook profile (Fig. 5b) contains the attribute "website" which is also part of our SWUM and thus, the attribute can also be linked to the SWUM without any extra effort.



a) Aggregation of the LastFM profile with the UMWN

b) Aggregation of the Facebook profile with the UMWN

Fig. 5. First step of the aggregation process. Figure a) shows how the attributes of LastFM and the SWUM/UMWN are connected. Figure b) depicts the connections of the Facebook profile.

The second step is then to directly connect the LastFM and Facebook user model as shown in 6. Based on the previously shown aggregation, connecting both models is straightforward. Revisiting the homepage/website example, these attributes can be directly linked because of the UMWN. The UMWN defines a synonym relation between the concepts "homepage" and "website", thus the LastFM and Facebook attribute can be directly linked with a match of 100%.

The aggregation of attributes that are not part of the SWUM can be done not only using the attribute name but also using the attribute content. So could an

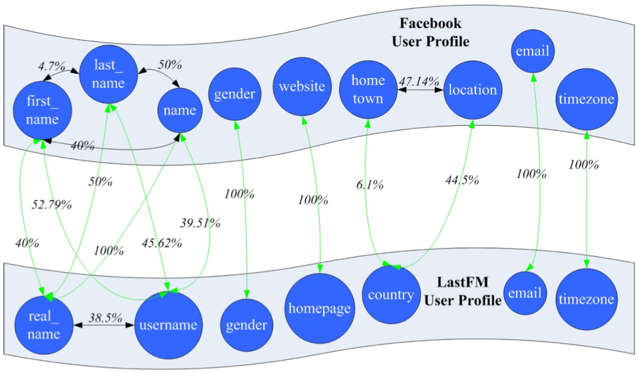


Fig. 6. Aggregated LastFm and Facbook profiles.

analysis show that the LastFM attribute "real_name" often contains the users' full name and thus a connection with the SWUM/UMWN attribute "full_name" can be done. Or the missing attributes can be added to the SWUM which is easy to do as it is a flexible RDF/OWL structure.

5 Conclusion and Outlook

In this paper we wanted to answer the question what are the requirements of the Social Web for a user model to profit from the available distributed user information. We present a new user model, the Social Web User Model (SWUM) that is fitted to the needs of the Social Web. We therefore conducted an extensive analysis of 17 social applications and to specify requirements, which dimensions and attributes are needed, for a Social Web user model. Based on this analysis we defined the dimensions a Social Web user model must cover and explained how the decision process was conducted. The analysis showed, that a Social Web user model only needs to cover certain dimensions of the user, namely *Personal Characteristics, Interests, Knowledge and Behavior, Needs and Goals and Context (Social Context, Location)*. We also presented the procedure to define the attributes of such a Social Web user model. To cover the problem of attribute heterogeneity throughout different social applications, we also equipped our model with a reduced WordNet that is especially tailored to the area of user modeling, the User Model Word Net (UMWN). The complete SWUM and UMWN model is based on RDF/OWL and thus easy to extend and reuse.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Trans. on Knowledge and Data Engineering 17, 734–749 (2005)

2. Bilenko, M., White, R.W.: Mining the search trails of surfing crowds: identifying relevant websites from user activity. In: Proceedings of the 17th international conference on World Wide Web. pp. 51–60. WWW '08, ACM (2008)
3. Brusilovsky, P., Millan, E.: User models for adaptive hypermedia and adaptive educational systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization, chap. 1, pp. 3–53. Springer-Verlag, Berlin Heidelberg New York (2007)
4. Dey, A.K.: Understanding and using context. Personal and Ubiquitous Computing 5, 4–7 (2001)
5. Dourish, P.: What we talk about when we talk about context. Personal Ubiquitous Comput. 8, 19–30 (February 2004)
6. Heckmann, D.: Ubiquitous User Modeling. Akademische Verlagsgesellschaft Aka GmbH, Berlin (2006)
7. Heckmann, D., Schwarzkopf, E., Mori, J., Dengler, D., Krner, A.: The user model and context ontology gumo revisited for future web 2.0 extensions. In: Proceedings of the Int. Workshop on Contexts and Ontologies: Representation and Reasoning. CEUR Workshop Proceedings, vol. 298. CEUR-WS.org (2007)
8. Jameson, A.: Modelling both the context and the user. Personal and Ubiquitous Computing 5, 29–33 (2001)
9. Kobsa, A.: Generic user modeling systems. User Modeling and User-Adapted Interaction 11, 49–63 (March 2001)
10. Korth, A., Plumbaum, T.: A framework for ubiquitous user modeling. In: Information Reuse and Integration, 2007. IRI 2007. IEEE International Conference on Information Reuse and Integration. pp. 291–297 (August 2007)
11. Kuflik, T.: Semantically-enhanced user models mediation: Research agenda. In: 5th international workshop on ubiquitous user modeling (January 2008)
12. Plumbaum, T., Lommatzsch, A., Luca, E.W.D., Albayrak, S.: Serum: Collecting semantic user behavior for improved news recommendations. In: UMAP 2011, Poster and Demo Session, Girona, Spain (2011)
13. Plumbaum, T., Schulz, K., Kurze, M., Albayrak, S.: My personal user interface: A semantic user-centric approach to manage and share user information. In: HCI International 2011 (2011)
14. Plumbaum, T., Stelter, T., Korth, A.: Semantic web usage mining: Using semantics to understand user intentions. In: UMAP '09: Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization. pp. 391–396. Springer-Verlag, Berlin, Heidelberg (2009)
15. Said, A., Berkovsky, S., De Luca, E.W.: Putting things in context: Challenge on context-aware movie recommendation. In: Proceedings of the Workshop on Context-Aware Movie Recommendation. pp. 2–6. CAMRa '10, ACM, New York, NY, USA (2010)
16. van der Sluijs, K., Houben, G.J.: A generic component for exchanging user models between web-based systems. Int. J. Continuing Education and Liflong Learning Vol. 16 Nos. 1/2, 64–76 (2006)
17. Sosnovsky, S., Dicheva, D.: Ontological technologies for user modelling. Int. J. Metadata Semant. Ontologies 5(1), 32–71 (2010)