

User Modeling and Adaptive Semantic Web

Editor(s): Krzysztof Janowicz, Pennsylvania State University, USA and Pascal Hitzler, Wright State University, USA

Solicited review(s): Michel Dumontier, Carleton University, Canada and Jie Tang, Tsinghua University Beijing, China

Open review(s):

Lora Aroyo ^a, and Geert-Jan Houben ^b

^a Computer Science, VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, the Netherlands

^b Web Information Systems, TU Delft, P.O. Box 5031, 2600 GA Delft, the Netherlands

Abstract. Historically, personalization and adaptation have been important factors for the success of the Web and therefore they have been important topics in Web research. Many research efforts in the field of adaptive hypermedia and adaptive Web-based systems have resulted in solutions for user-adapted access to Web content, often in terms of systems that provide an adaptive hypermedia structure of content pages and hyperlinks. With pages and links that depend on the user, it is feasible to offer a high degree of personalization. Next to research into engineering and realizing adaptation, research into user modeling has been crucial for the success of adaptation. To apply the right adaptation it is necessary to know the user and her relevant properties and the research field of user modeling has focused on theories and techniques for eliciting knowledge about the user. Naturally, the research fields of adaptive web-based systems and user modeling have always lived in close harmony. In order to create a similar success with personalization and adaptation in relation to the Semantic Web, adaptation and user modeling have to be redefined, with consequences for the research into these topics. In particular, the nature of user modeling changes significantly with the extended distribution and openness that we encounter on the Web of Data, with implications from problems studied in Web science. Promising research shows how Semantic Web-based solutions can aid in the representation of user properties for sharing and linking of user models. In this vision paper we outline the evolution of user modeling and adaptation in connection to the Semantic Web and list research questions and challenges for the relevant research fields.

Keywords: User modeling, adaptation, personalization, linked open data, web science

1. Introduction

In this paper we outline the evolution we see and anticipate for the research fields of *adaptation* and *user modeling* in connection with the Semantic Web.

Adaptation of content access to the user is by definition a process in which a description of the content and a description of the user are combined to decide whether and how to present the content to the user. On the Web this adaptation has typically been designed and realized in closed and scoped applications. The user modeling to represent relevant user descriptions has been characterized by the same assumptions. When we relate this to the Web of Data, new conditions and assumptions come into play. At the Semantic Web we have witnessed already the use of seman-

tics in the integration of content, for the purpose of integrating and linking content between applications, but similarly, and perhaps more importantly, user modeling can and should also be aligned with the conditions and requirements of the new Web. Where the distributed and open nature of data has brought major advances for linking content, the same characteristics of distribution and openness find their way into user modeling.

In this vision we reflect on what is happening and sketch challenges and questions for the relevant research fields. In Section 2 we consider the concepts of adaptation and user modeling in their original context of the classical Web, before we turn to user modeling in relation to the Semantic Web in Section 3 and the corresponding challenges in Section 4.

2. Adaptation and User Modeling

From the conception of the Web its hypertext-based nature triggered researchers to find ways to improve the nodes-and-links structure that gave the Web its success. With the observation that a single fixed hyperlink structure would not fit all users, the research field of adaptive hypermedia [5] investigated how the hyperlink structure could be made adaptive to the user, i.e. how the pages and hyperlinks could be made fitting for each single user.

This adaptation aims to present the best possible hyperlink structure to a user depending on relevant properties of the user, e.g., background, context, or goal. Educational applications have always been a good example for researchers to showcase adaptation solutions [3]. In educational applications often the users are students or learners for whom the acquired knowledge is a basis for adaptation. Imagine a teacher wants to present the student with relevant material to study a certain subject, then the teacher will create a structure of pages and hyperlinks that will make the student go through the material in a manner that satisfies the teacher's intentions and pedagogic principles. That structure will present the student with pages and links depending on the knowledge the student acquired before and during her study of the material. For this purpose, the hyperlink structure will include with pages and links pre-conditions that reflect the teacher's assumptions about the student's knowledge at that moment in the browsing.

In the research field of adaptive hypermedia [5] this approach has been studied in several other domains as well, e.g. e-commerce and tourism, and this has led to the development of systems that support the design and execution of adaptation in hypermedia-based information delivery. Due to the nature of the first trials and the technological hurdles that had to be taken, most research concentrated on systems with a well-defined and limited scope, to make it feasible for the system to "know" the user and how to respond to that. Obviously, the design of adaptation asks for a detailed understanding of the user's knowledge at the time and of the influence that the knowledge should have on the content to be presented. In "closed" applications and systems the design and execution of the adaptation proved to be already challenging enough for researchers to extensively investigate design and usage [21]. Later, the scoping was relaxed when the same approaches were being used at the Web [7,8].

Approaches for adaptation cannot be meaningfully applied without a thorough understanding of the user. That is why the field of user modeling [11] has concentrated on theory and techniques for the elicitation of user knowledge into user models that could serve as the basis for effective user-adaptation. Applying intelligent techniques to calculate relevant properties of the user for adaptation, for example in recommendation or teaching scenarios, researchers established theories and tooling for an accurate and relevant description of the user on the basis of the user's actions in the application.

In linking these two research fields [18], the use of an explicit user model is the classical approach for adaptive systems. Following the reference model from [13], a general view on adaptive systems is that a system contains a description of the domain content, i.e. a *domain model*, a description of the user, i.e. a *user model*, and a way to combine those two to adapt the content for a presentation fitting the user, i.e. an *adaptation model and engine*.

In many cases the user model overlays the domain model, meaning that the user knowledge is expressed as an overlay over the domain content. A good example from the educational scenario would be that the student's knowledge is expressed as a value attached to each domain subject reflecting the degree to which the student has learned the topic. The specific elements in a user model depend of course on the application, but aspects that we often see are history, background, preferences, knowledge level, goals and tasks, context of work, meta-cognitive skills, personality traits, affective states, and attitudes.

3. User Modeling in a Web of Linked Data

With the content moving towards the Semantic Web, it is now interesting to see how the advances in the Semantic Web and the cloud of Linked Open Data [20] impact adaptation and in particular user modeling.

Where in the traditional adaptation approaches the adaptation was often confined to a single closed application, it is natural to try to share and integrate content data to profit from the investment in content made in multiple applications. In the evolution from adaptive hypermedia to adaptive Web-based systems this trend was already visible. Also, open hypermedia systems [4] show a similar approach where the linking is separated from the content data. Similarly, the open corpus adaptive hypermedia systems [6] show

how metadata-based approaches concentrate on content reuse and integration, with obvious connections to semantic techniques and languages. So, when it comes to integrating and linking data, adaptive applications do not differ from other applications and can equally well benefit from results obtained in Semantic Web research. Therefore, we concentrate in this vision paper further on the user modeling aspect.

3.1. Linking User Knowledge

Semantic integration can of course also create benefits for the user knowledge. When adaptive applications have the opportunity to share user knowledge, for example in an educational setting the student's knowledge in a particular subject domain, then with richer and more relevant user knowledge from across application boundaries the applications can provide better adaptation.

This trend in semantic integration of user knowledge aligns with the trend in Web 2.0 and social networking where people share personal information.

Both trends show two aspects that are relevant for linking user model knowledge: the identification of the users and the representation of their properties.

3.2. User Identification

When applications want to share information for and about users, they require mechanisms for the identification of users.

Identity-based protocols as OpenID¹ can be used for users to link their different identities on the Web. Systems can use authentication mechanisms, from basic http authentication to open protocols for secure API authorization like OAuth². The Google Friend Connect³ API exemplifies the use of OpenID (e.g. Yahoo) and OAuth to integrate existing login systems, registered users, and existing data with new social data and activities. It is based on open standards and allows users to control and share their data with different sites. The integration of social flows and data is realized via the OpenSocial⁴ standard specification. The Facebook Platform uses the OAuth 2.0⁵ protocol for authentication and authorization in web applications (both desktop and mobile). The

Facebook Connect extension makes it possible for users to "connect" their Facebook identity to any site by using trusted authentication and to also reuse, among others, their basic profile information and friends list around the Web.

Research like [10] presents an approach to enable interoperability of user-adaptive systems in a ubiquitous environment. It is centered around a semantics-based dialogue for exchanging user model and context data with focus on the user data clarification and negotiation tasks. Further, [9] looks at a framework for user identification for cross-system personalization. It exploits a set of identification properties that are combined using an identification algorithm.

3.3. User Property Representation

Contrary to the identification of users, for the representation of user properties there are hardly any standardized and generic solutions available. This is not a surprise of course, given the traditionally closed environment in which user model knowledge is created and used.

This representation issue can best be explained with an example. If for example in an educational setting a student's knowledge needs to be represented, then one often sees something like a value such as "well-learned" that is associated as the "degree of learning" with a subject like "Programming" or a value of "80%" for the "knowledge level" of a subject "Java". From these examples it is easy to see that for interoperability we need to align (a) the knowledge about the domain, e.g. about the domain concepts such as "Programming" or "Java", and (b) the knowledge *about* the knowledge about the domain, e.g. the "degree of learning" or "knowledge level" and their corresponding values..

Important aspects of the representation of user properties are

(1) to represent *uniquely the object* of the interest or preferences, e.g. "interested in *Java*" or "likes *Brad Pitt*",

(2) to provide a *shared vocabulary* to express different *user activities* which translate into user properties, e.g. "*like*", "*read*", "*view*", "*favor*",

(3) to have a *shared scale(s)* to interpret the user property, e.g. "rate this video with 5 stars" and "rate this book with 2 stars" – it is handy to know that the first value was in a 10 point scale and the second in a 5 point scale, which makes them almost identical in terms of their value for the aggregated user interest,

¹ <http://openid.net>

² <http://oauth.net>

³ <http://code.google.com/apis/friendconnect/>

⁴ <http://www.opensocial.org/>

⁵ <http://tools.ietf.org/html/draft-ietf-oauth-v2-10>

(4) to represent the *notion of certainty and accuracy* of the collected and aggregated user properties, e.g. indicating the source or the context of the collected information and specifying how trustworthy or reliable the source is – if the user had bought a book on Amazon about Java programming or watched a video on TED about it, these could be pretty reliable sources for her interest in this topic, while if she just browsed through several web pages it might be questionable whether she actually read anything.

A good example for most of those aspects can be found in current extensions of FOAF⁶, e.g. the Weighted Interest Vocabulary⁷ for identifying context and source of the collected information, or e-FOAF⁸ for defining temporal properties for the interest value. Additionally, a format like Activity Streams⁹ is used for syndicating social activities on the Web and providing a shared vocabulary to express user activities across applications. This format has already been adopted by Facebook, MySpace, Windows Live, Google Buzz, BBC, Opera, Gowalla, among others. The base schema defines a set of Verbs, e.g. “mark as Favorite”, “post”, “tag”, a set of Object Types, e.g. “article”, “bookmark”, “comment”, a set of Activity Context Elements, e.g. for location and mood, and Event Verbs, e.g. “positive RSVP”.

e-FOAF allows for temporal reasoning in the user interest value calculation over time. Knowing when a specific piece of evidence for the user interest has occurred (`e-foaf:interest_appear_time`) or when the interest value was updated (e.g. `e-foaf:interest_value_updatetime`) helps to increase the accuracy in the calculation of the user interest as an aggregation (`foaf:cumulative_interest_value`) of multiple pieces of evidence around the Web. Additionally, properties as `e-foaf:retained_interest_value`, help express decay or other time-related aspects of the interest values. With properties like `e-foaf:interest_longest_duration` and `e-foaf:interest_cumulative_duration` the strength of the interest can be varied in order to reflect the intensity of the evidence in terms of calculating the cumulative user interest value.

⁶ <http://www.foaf-project.org/>

⁷ <http://xmlns.notu.be/wi/#spec/20091224.html>

⁸ <http://wiki.larkc.eu/e-foaf:interest>

⁹ <http://activitystrea.ms/>

Semantic Web-based user model standards that are used widely in educational settings are IMS LIP¹⁰ and IEEE PAPI¹¹.

3.4. Sharing Adaptation Functionality

For the sake of completeness we mention that following [13]’s reference model, after *domain knowledge* and *user knowledge* linking and integration, the integration of *adaptation functionality* can also be improved but this is an extremely challenging problem given the proprietary nature of many of the currently available solutions and systems. Projects like [16] show first steps in the integration of adaptation functionality, where also semantic technologies are used albeit mainly for domain and user knowledge.

With the advances in the Web of Data for linking domain and content knowledge and for linking user model knowledge as we have just described in this section, we see however the emergence of a new paradigm for adaptation, where adaptive applications are connected to a cloud of Linked (content) Data as well as a cloud of Linked User Data. While this Linked User Data is technically part of the new Web of Data as well, the *distributed and open* nature puts a whole new perspective on user modeling, and opens a whole new array of innovations.

4. Distributed and Open User Modeling

For sharing user model knowledge, experience from the Semantic Web can provide concrete solutions, as for example the research from [2,3,10,19] shows. The main benefit is that each application does not have to build up its user model knowledge alone, which specially in the beginning can be a problem when little knowledge is available: the so-called *cold start*. Also, with more knowledge available to construct a model the chance that the model *accurately* describes the user increases naturally.

This sharing of user knowledge is not just a matter for applications that like to adapt (as we discussed in the previous section), but practically the same ambition and techniques show in the Social Web when a user wants to share her own personal profiles between social networking systems, as for example [13] shows. In [15] it is shown how FOAF can be used to link social networks. Where existing social network-

¹⁰ <http://www.imsglobal.org/profiles/>

¹¹ <http://www.ieee.org>

ing services are highly centralized, as are existing personalized services, the trend towards distribution is clearly visible and helps also to increase the control users have over their own data.

Another example is provided by the NoTube project¹² where a strategy for aggregating user data from various Social Web applications is provided. This work is based on a concrete use case of reusing activity streams to determine a user's interests, and then generating television programme recommendations from these interests. A key component to realize this is the NoTube BeanCounter [1]. The main design rationale is to provide a flexible and extensible architecture that exposes robust, scalable and reliable services to handle different kinds of responses of different social application platforms. A set of APIs allows for modeling the targeted responses in order to gather them, represent them with a set of suitable RDF vocabularies, and integrate them with other pulled information in a fully transparent way – with the help of service-specific adaptor *tubelets* (e.g. a twitter activity stream tubelet) and application server *modelets* (e.g. for movies or songs) which allow for the selection of data source and RDF vocabulary and the generation of RDF-ized user data, integrated with data coming from other adaptors.

Observing the promising results in user modeling for adaptive and social applications with the aid of Semantic Web-based solutions, we now outline research questions and challenges for the new paradigm of *Distributed and Open User Modeling* as a main ingredient for the Adaptive Semantic Web.

4.1. User Identification

The major question in user identification is: *How do we identify a person (or a person's appearance)?*

In the conventional adaptive solutions, closed and with restricted scope, identification mechanisms are often proprietary or pragmatic, and usually these are also not fit for application at Web-scale. The new assumptions and requirements imply a number of research questions:

- How can a person identify herself to an application?
- How can a person manage her identities (across multiple applications)?
- How can applications find a user (identity) in other applications?

- How are trust and privacy provided in mechanisms for user identification?
- How do users behave in systems with shared user identification and what are the social and legal consequences?

The above questions do include technical challenges, but also constitute interesting problems in Web Science. Considering the specific Semantic Web angle we see that standard identification mechanisms and efficient corresponding indexing mechanisms need to be proposed.

4.2. User Knowledge Alignment

After users being identified, the main question with respect to user knowledge is: *How can user model knowledge be shared?*

Answering his question on the Web of Data brings up several research questions related to the representation of user properties:

- How can the objects of user properties uniquely be represented?
- How can a shared vocabulary be constructed for expressing user properties?
- How can shared scales be constructed to interpret values for user properties?
- How can notions of certainty and accuracy be attached to user properties?

The main challenge here for the user modeling research field is to derive from the vast amount of user modeling theories and experience [18], those properties that are typically and effectively used to model user properties and then turn to the area of the Semantic Web to create a standard vocabulary for those properties. Such a vocabulary could borrow from SKOS, RDF/OWL etc., as we see in some of the examples we mentioned before.

Besides its role for sharing user model knowledge, such a vocabulary-based approach would also be the ideal stepping-stone for Web Scientists to study user modeling on the Web of Data, and thus to analyze how user modeling performs under the new conditions of distribution and linking. As part of this study, openness and scrutability need to be investigated as well:

- How is an open approach to user knowledge perceived by the users?
- How can users be given the opportunity to inspect and correct the user knowledge an application maintains about them?

¹² <http://notube.tv>

The role of the user in the elicitation and verification of user knowledge can also be extended, with the aid of semantic techniques, following [12].

Thus, we see the first examples of investigations into the new paradigm for user model knowledge.

5. Conclusion

In this paper we have considered how user modeling evolves from its original environment in connection to a closed and scoped adaptive system to a distributed and open existence in the Semantic Web. We have identified some promising research approaches that show how the Semantic Web can contribute with solutions for user identification and user knowledge representation. On the basis of that experience, we have formulated for the new paradigm a number of relevant engineering and scientific questions for inclusion in the research agenda.

References

- [1] C. van Aart, L. Aroyo, D. Brickley, V. Buser, L. Miller, M. Minno, M. Mostarda, D. Palmisano, Y. Raimond, G. Schreiber, and R. Siebes, The NoTube Beancounter: Aggregating User Data for Television Programme Recommendation. In Proceedings of SDoW 2009, workshop in conjunction with ISWC 2009 (2009).
- [2] F. Abel, D. Heckmann, E. Herder, J. Hidders, G.J. Houben, D. Krause, E. Leonardi and K. van der Sluijs, A Framework for Flexible User Profile Mashups. In Proceedings of APWEB 2.0 2009, workshop in conjunction with UMAP 2009 (2009).
- [3] L. Aroyo, P. Dolog, G.J. Houben, M. Kravcik, A. Naeve, M. Nilsson, and F. Wild, Interoperability in personalized adaptive learning. *J. Educational Technology & Society*, 9 (2) (2006) 4–18.
- [4] C. Bailey, W. Hall, D. Millar and M. Weal, Towards open adaptive hypermedia. In Proceedings of AH 2002, Adaptive Hypermedia and Adaptive Web-Based Systems (2002), 36-46.
- [5] P. Brusilovsky, Adaptive hypermedia. *User Modeling and User Adapted Interaction*, 11 (1/2) (2001) 87-110.
- [6] P. Brusilovsky and N. Henze, Open Corpus Adaptive Educational Hypermedia, in: *The Adaptive Web*. Springer Lecture Notes in Computer Science 4321, 2007, 671-696.
- [7] P. Brusilovsky, A. Kobsa and W. Nejdl (Eds.), *The Adaptive Web*. Springer Lecture Notes in Computer Science 4321, 2007.
- [8] P. Brusilovsky and M.T. Maybury, From adaptive hypermedia to the adaptive web. *Commun. ACM* 45 (5) (2002) 30–33.
- [9] F. Carmagnola and F. Cena, User identification for cross-system personalisation. *Information Sciences* 179 (1-2) (2009) 16-32.
- [10] F. Cena and L. Aroyo, A Semantics-Based Dialogue for Interoperability of User-Adaptive Systems in a Ubiquitous Environment. In Proceedings of UM 2007 (2007), 309-313.
- [11] C. Conati, K.F. McCoy and G. Paliouras (Eds.), *User Modeling (UM 2007)*. Springer Lecture Notes in Computer Science 4511, 2007.
- [12] R. Denaux, V. Dimitrova and L. Aroyo, Integrating open user modeling and learning content management for the semantic web. In Proceedings of UM 2005, *User Modeling (2005)* 9–18.
- [13] P. De Bra, G.J. Houben and H. Wu, AHAM: A Dexter-Based Reference Model for Adaptive Hypermedia. In Proceedings of ACM Hypertext 1999 (1999), 147-156.
- [14] R. Ghosh and M. Dekhil, Mashups for semantic user profiles. In Proceedings of WWW 2008 (2008), 1229-1230.
- [15] J. Golbeck and M. Rothstein, Linking Social Networks on the Web with FOAF: A Semantic Web Case Study. In Proceedings of AAAI 2008 (2008), 1138-1143.
- [16] Grapple project: www.grapple-project.org
- [17] D. Heckmann, T. Schwartz, B. Brandherm, M. Schmitz and M. von Wilamowitz-Moellendorff, GUMO - The General User Model Ontology. In Proceedings of UM 2005, *User Modeling (2005)* 428-432.
- [18] G.J. Houben, G.I. McCalla, F. Pianesi and M. Zancanaro (Eds.), *User Modeling, Adaptation, and Personalization (UMAP 2009)*. Springer Lecture Notes in Computer Science 5535, 2009.
- [19] T. Kuflik, Semantically-Enhanced User Models Mediation: Research Agenda. In Proceedings of UbiqUM'2008, workshop in conjunction with IUI 2008 (2008).
- [20] Linked Data: www.linkeddata.org
- [21] W. Nejdl, J. Kay, P. Pu and E. Herder (Eds.), *Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2008)*. Springer Lecture Notes in Computer Science 5149, 2008.