Requirements for Personal Information Agents in the Semantic Web

Wolfgang Woerndl

Technische Universität München, Munich, Germany
woerndl@in.tum.de
http://www11.in.tum.de/persons/woerndl/

Abstract. In this paper we first present some requirements for personal information agents in the Semantic Web. Then we outline our PINA project which tries to combine elements of identity management and information agents in the Semantic Web. The fundamental idea is to store references to Semantic Web annotations in identity servers as part of user profiles.

1 Introduction

Personalization of information offerings appears to be a promising concept to help people finding relevant information in the world wide network of information sources. Thus, a variety of systems have already been developed using user data they have collected or information users have explicitly made available. These systems offer personalized Web pages or make recommendations based on user profiles. User profiles thereby contain information such as demographic data (e.g. age, gender, Email addresses), specified interests or past transactions (e.g. bought books).

In addition, work on building the *Semantic Web* is recently gaining more and more attention (see [1] for a recent overview). The goal of the Semantic Web activities is to make available the meaning of information to computers. Thereby, software agents or other programs analyze and evaluate semantic meta data of information items to improve services. An important building block in this regard are ontologies that are shared between information providers.

A combination of personalization and Semantic Web could be beneficial because additional semantic information[1] to data sources could be used to improve customization of search results or other filtering services.

However, personalization also raises issues of privacy and trust. Firstly, any personalization application potentially poses privacy problems, because users have to provide information about themselves and want to know how their information is being used. Secondly, there is also the problem of trust. In the existing Web, it is more or less up to the user to (manually) decide whether

[1] For example, meta data to information sources, (semantic) annotations to Web pages or ontological classification of information items. In the following, the term "annotations" is used to describe all kinds of meta data or semantic information.

J.-B. Stefani, I. Demeure, and D. Hagimont (Eds.): DAIS 2003, LNCS 2893, pp. 260–265, 2003.

information, e.g. search engine results, might be trustworthy or not. In the Semantic Web, this will not be the case, because agents have to determine the trustworthiness of information.

In this paper, we present requirements for personal information agents in the Semantic Web. We will also briefy introduce the Personal Information Agents (PINA) project which tries to combine elements of identity management and Semantic Web agents.

2 Requirements for Personal Information Agents

In this chapter, we outline some requirements for personal information agents in the Semantic Web in various research areas.

2.1 Agents in the Semantic Web

According to James Hendler, the ideal internet agent is as follows: "A good internet agent needs these same capabilities. It must be communicative: able to understand your goals, preferences and constraints. It must be capable: able to take options rather than simply provide advice. It must be autonomous; able to act without the user being in control the whole time. And it should be adaptive; able to learn from experience about both its tasks and about its users preferences." ([2])

Apart from the agent requirements such as autonomy, personal information agents have to take into account the semantic character of information. More precisely, the following aspects have to be considered:

- Segmentation of search and customization to take advantage of the agent paradigm
- Interoperability: handle multiple ontologies, for example different user profile models, and integrate mapping mechanisms between ontologies
- Heterogeneity: possibility to query different information sources
- Balance the trade-off between expressibility and (computational) complexity

2.2 Identity Management and User Profile Modelling

The basic idea of *identity management* is to separate user profiles and identities from the services that are using them. An identity management system would allow people to define different identities, roles, associate personal data to it, and decide whom to give the data to and when to act anonymously [3].

A generic user model and storage in an integrated repository is needed. User profiles should be stored in a non-redundant manner [4]. Information about users should be reusable for different applications and domains so that users do not have to enter their information such as Email addresses or interests again and again. Thereby, an ontology-based approach might be useful [5]. So far there is no widely accepted "user profile ontology" and a mapping of different

personalization ontologies is required in real world scenarios. Another important aspect with regard to user modelling and identity management is to consider different roles and identities of users, for example "work" or "private" identites.

When combining Semantic Web and personalization technologies, one of the most important questions is: what is the relationship between Semantic Web annotations and user profiles? Are annotations part of a user's profile? Or maybe references to annotations? If annotations are not part of a user's profile at all, it is very difficult for agents to make any inferences about the trustworthiness of information sources.

2.3 Privacy and Trust

Privacy is "the claim of individuals, groups or institutions to determine for themselves, when, how and to what extent information about them is communicated to others." ([6]). The aspect of control for the user is essential. User need to know how, why and what part of their profile is being accessed. It is not reasonable to build user-adaptive systems without considering privacy. In addition, we can identify two aspects of privacy in our scenario. Firstly, privacy of users who provide semantic annotations to information sources. Secondly, privacy of users who access information sources or search for relevant pieces of data. Current systems often adress only one aspect and neglect the other.

The two most important features of privacy preserving identity management are authentification and authorization. *Authentification* provide means for users to securely assert their identity without necessarily revealing their true identity. An important feature is to define and control different pseudonyms which is the scope of the Liberty Alliance project (see www.projectliberty.org). The goal of this project is to define an open standard for federated identity management that allows users to link elements of their identity between accounts without centrally storing all their personal information. However, their focus is the management of identities and authentification in the WWW to provide a so called "single sign-on" (SSO) service and the Liberty Alliance cannot easily applied to agent-based information services.

Authorization is about controlling the access of services to user profile attributes. More precisely, requirements in our scenario are [3]:

- Flexible access right control system, e.g. through rules and negotiation
- Possibility to use a pseudonym instead of real identity
- Purpose binding of data accesses
- Possibility for the user to monitor access rights and accesses and revoke granted access rights if necessary
- Control whether user data may be distributed to other services (and users)
- Integration of cryptographic techniques for anonymous data transfers

Possible solutions to *trust* in the Semantic Web include the signing of information items by persons or institutions. Agents can then evaluate the digital signatures to proof the trustworthiness of annotations before presenting personalization results to users. But trust has to be adressed in combination with privacy. Therefore, the top layer in the Semantic Web layer cake by Tim Berners Lee

(available at http://www.w3.org/2002/Talks/01-sweb/slide12-0.html, for example) should be named "Trust & Privacy" not just "Trust" because all efforts to improve trust and build a "Web of Trust" potentially decrease the privacy of users. In other words, there is a trade-off between trust and privacy in any personalization system that has to be taken into account when designing the application.

3 The Personal Information Agent (PINA) Project

The goal of the Personal Information Agent (PINA) project is to bring together identity and user profile management on the one hand, and Semantic Web technologies and agent technologies on the other hand. The purpose is to support semantic personalization of information sources and improve adaption of information to user profiles. This is especially done with respect to user privacy. A fundamental idea is to store references to Semantic Web annotions as part of user profiles.

The components of our architecture are depicted in Fig. 1.

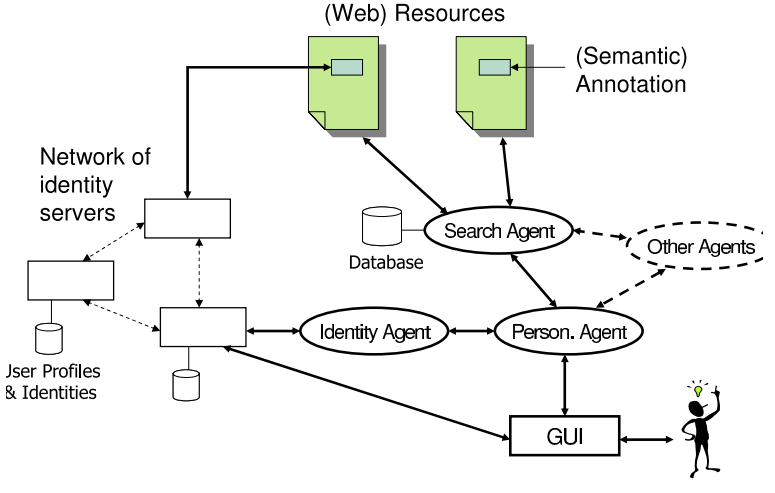

Fig. 1. PINA components.

The basic building blocks are as follows:

- (Semantic) Web resources with additional annotations[2]
- A network of identity management servers to store references to semantic annotation and handle the user profile management
- Information agents to provide personalization and other services
- Appropriate user interfaces and/or client-side tools

[2] The annotations could be stored in designated annotation servers separate form the information items.

In [3] we describe an identity management infrastructure which separates the identity management from the service applications. This is done in the domain of community support systems but can be used for other services that use personal information. Agents can access user profiles via an agent-based interface (FIPA). Different user identities are also part of the framework and can be used by personalization agents. We thereby cover the identity network part in Fig. 1. We are currently implementing the link to Semantic Web annotations.

Mechanisms to derive trust in the Semantic Web can then be designed using the binding of Semantic Web annotations to user identitites. A user can define and control different pseudonyms to mark Semantic Web annotations. The real identity of the user does not have to be disclosed. For example, a user can provide annotations under a pseudonym "wolfgangw" or "foo23". Agents then derive the trustworthiness of annotations by using these pseudonyms instead of real identities of users. The authenticity of pseudonyms is proven by the identity management network.

In [7,3] we also explain a concept for authorization in distributed management of user profiles. Thereby authorization is done by combining privacy enhancing technologies with access control. User profile agents negotiate access right to user information with service agents using privacy policies of services and preferences and access rules of users. Our approach is well suited for the agent scenario because the solution adheres to the agent paradigm of autonomy of components and also uses a message based protocol for the negotiation.

4 Conclusion

In this paper we have presented some requirements for personal identity management in the Semantic Web. We also outlined the architecture of the PINA project. The main idea is to combine identity management and (personal) information agents in the Semantic Web. The briefly presented solution stores a reference to a Semantic Web annotation in a distributed user profile management network. This approach allows for personalization services that exploit different identities of users and other identity management features. PINA also allows the realization of privacy mechanisms as summarized in [7] and [3].

Next steps in PINA include implementation of more components to test the effectiveness of the system in delivering customized information to users. In addition to identity agents and ontology agents or servers, we envision filter and personalization agents and a component that provides an appropriate user interface. Another point is whether the Liberty Alliance specification can be used (or adopted) to handle the authentification using different user identities between (personal) agents. For far, we use an easier authentification schema that is provided by our identity management infrastructure.

References

1. Fensel, D.; Hendler, J.; Lieberman, H., Wahlster, W. (Eds.): Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press (2003)
2. Hendler, J.: Is There an Intelligent Agent in Your Future? Nature (1999)
3. Koch, M., Woerndl, W.: Community Support and Identity Management. Proc. Europ. Conference on Computer-Supported Cooperative Work (ECSCW2001), Bonn, Germany (2001) 319–338
4. Kobsa, A.: Generic User Modeling Systems. User Modeling and User-Adapted Interaction 11, Kluwer (2001) 49–63
5. Razmerita, L., Angehrn, A., Maedche, A.: Ontology-Based User Modeling for Knowledge Management Systems. Proc. User Modeling 2003, Lecture Notes in Artificial Intelligence (LNAI) 2702, Springer (2003) 213–217
6. Westin, A.: Privacy and Freedom. New York (1967)
7. Woerndl, W., Koch, M.: Privacy in Distributed User Profile Management. The Twelfth International World Wide Web Conference (WWW2003), Budapest, Hungary (2003)