2008 Third International Workshop on Semantic Media Adaptation and Personalization

Enhancing User Privacy in Adaptive Web Sites with Client-Side User Profiles

C. Kolias[1], V. Kolias[2], I. Anagnostopoulos[1], G. Kambourakis[1], E. Kayafas[2]
[1] *University of the Aegean, Department of Information and Communication Systems Engineering*
[2] *National Technical University of Athens, School of Electrical and Computer Engineering*
[1] *{kkolias, janag, gkamb@aegean.gr}*, [2] *vkolias@medialab.ntua.gr*, [2] *kayafas@cs.ntua.gr*

Abstract

Web personalization is an elegant and flexible process of making a web site responsive to the unique needs of each individual user. Data that reflects user preferences and likings, comprising therefore a user profile, are gathered to an adaptive web site in a non transparent manner. This situation however raises serious privacy concerns to the end user. When browsing a web site, users are not aware of several important privacy parameters i.e., which behavior will be monitored and logged, how it will be processed, how long it will be kept, and with whom it will be shared in the long run. In this paper we propose an abstract architecture that enhances user privacy during interaction with adaptive web sites. This architecture enables users to create and update their personal privacy preferences for the adaptive web sites they visit by holding their (user) profiles in the client side instead of the server side. By doing so users will be able to self-confine the personalization experience the adaptive sites offer, thus enhancing privacy.

1. Introduction

Adaptive hypermedia applications made their first steps in the early 1990s and have gain popularity after 1996 with the emergence of the World Wide Web (WWW). Nowadays, as web sites become increasingly complex and informationally massive, the need for "personalized" or "user-adaptive" web sites has become a necessity. Adaptive hypermedia applications are considered as hypermedia systems which adapt their content, structure and/or presentation of the networked hypermedia objects, to each individual user's characteristics, usage behavior and/or usage environment [1]. Based on this definition, adaptive web sites are regarded as adaptive hypermedia systems that can be accessed from the web i.e., by a common web browser.

As stated in [2], adaptive hypermedia can be used within the context of various spheres of action including educational hypermedia applications, on-line information systems, on-line help systems, information retrieval hypermedia, institutional hypermedia, and systems for managing personalized views in information spaces. Today, most research efforts and industrial implementations mainly focus on e-commerce and e-learning fields which have benefited by personalization to a great extent, e.g. by improving the service provided and consequently the user experience. For example, a personalized educational web site automatically adapts its methodologies and contents according to the knowledge level of the student, his general interests and his performance, in order to enhance the educational process. In contrast, a conventional e-learning application will try to find a "golden path" of teaching method for all students and will present the same information to them without discrimination. Likewise, an adaptive e-commerce application will significantly assist the user in the shopping process by recommending products that possibly meet his preferences, automatically redirecting him to product categories that reflect his likings etc.

Generally, adaptation can be distinguished into two types: (a) adaptive presentation and (b) adaptive navigation support. The former refers to actions such as modifying textual and multimedia content of the web site, while the latter to actions such as link hiding, sorting, annotation, direct guidance and hypertext map adaptation [2].

In contrast to adaptable websites where the user has manual control over its appearance and structure, in adaptive websites the personalization process is usually adjusted automatically by the website. This however happens not in an opt-in or voluntary basis but without the user being aware of internal site mechanics. In other words, the user is not in control of how the site behaves. This transparent to the web page visitor process is based on dynamically build user models which

978-0-7695-3444-2/08 $25.00 © 2008 IEEE
DOI 10.1109/SMAP.2008.32

IEEE computer society

represent users' behaviour to the system. These models are build by acquiring specific user data every time a user visits such an adaptive web site. According to [1] user models are classified into the following categories:

- **User data**: information about user's preferences and other personal information (for example, name, sex etc)
- **Usage data**: information about the user's behavior on various web pages. For example, what pages he visited and how long he remained in each one of them.
- **Environment data**: information about the user's network location characteristics revealed by the client machine. For example, IP address, memory, screen resolution, browser type)

User modeling is an active research field of Artificial Intelligence (A.I.). There are many approaches for creating accurate user models but the common denominator of all is that their accuracy depends on the amount of user data gathered. As a result, adaptive web sites require a large amount of information in order to better personalize their content. However, gathering personal information, some of which may be sensitive, raises serious privacy concerns.

Recent studies prove that up to 87% of Internet users are highly concerned about their digital privacy and up to 41% would leave a site that requires registration information [3]. Adaptive web sites intentionally gather user information in a concealed manner, in order not to disturb the interaction with the website. Therefore users are not aware of what kind of personal information is monitored, as well as for how long and under which conditions the logged data will be stored. This situation however might result in undesirable situations. For example, in an e-commerce site, user behavior can be related to real information like a name, an address and a credit card. Additionally users feel uncomfortably knowing that their personal information will be stored on a remote location, left to the adaptive service provider's will to share it with third parties of unknown intentions (e.g. spamming). In other words, users are unaware of and unable to take control over their own user profiles created on each adaptive web site. All these factors raise serious ethical dilemmas on behalf of the user and potential law conflicts on behalf of the service provider.

In this paper we propose a high level architecture that will enable users to take control over their user profiles constructed by adaptive web sites. Users will be able to construct their own privacy preferences by specifying which information they consider as sensitive as well as to create and store each adaptive web site's user profile on their (client) machine. Under these circumstances, user profiles will be shared across different web sites according to the privacy settings specified by the user.

The rest of the paper is organized as follows: Next section addresses previous work on the topic. The proposed architecture is described in Section 3, while Section 4 elaborates on the advantages and disadvantages of our solution. Finally Section 5 concludes the paper and gives some directions for further research.

2. Related Work

So far many approaches have been proposed and adopted as possible solutions to the need of increased user privacy in adaptive web sites. A premature solution followed by many web sites was to commit to a general privacy policy (which would be publicly accessible in the site), where the purpose of using user data and the type of data itself were analytically described. A user could then inspect those policies and evaluate them. The problem with this practice is that sometimes the text of the privacy policy document can be quite lengthy, generic or fuzzy, and may contain technical terms. As a proven fact users are not willing to invest time for actions that are not immediately relevant with their purpose, and in many cases, they cannot fully understand.

To overcome the aforementioned problem, the Worldwide Web Consortium (W3C) [4] recommended A P3P Preference Exchange Language (APPEL) [5] and the Platform for Privacy Preferences (P3P) [6] standards. P3P is a language based on XML, which provides a platform for the service providers (running adaptive sites) to express their privacy policies. APPEL is a similar language which enables users to express privacy preferences. Consequently, tools like the AT&T Privacy Bird [7] have been created which evaluate if the web site's privacy policy meets to the user's privacy preferences. If not, then the utility advices the user to leave the site. In both of the situations described above the user cannot continue interacting with the site if he is not willing to make privacy compromises. At any case the site will not scale down its services to meet his privacy preferences.

Other approaches on the topic can be categorized in pseudonymous profiles generation, task-based and client-side profiles personalization according to [8].

Solutions that fall in the first category attempt to safeguard user privacy by allowing users to access web sites using an alias. The idea is based on the assumption that most personalization actions on web sites do not require the user's real name and other identification information. In this way a malicious entity who gains access to the user's profile will not be able to relate the information contained in it with a real person. This approach has proven problematic in situations where a real name and other sensitive personal information are needed in order to complete a transaction, such as in e-commerce sites.

Task-based personalization attempts to reduce privacy risks by focusing on presenting content related with specific tasks, such as searching for a specific category of books. By doing so, personalization occurs only after placing requests for specific tasks and it is based on data created by the website itself, rather than data gathered from previous activity of the user. The advantages of this approach are that very little if no data is stored in the web sites and its implementation is generally simple. On the downside the personalization services are generally of lower quality, comparing to other approaches.

In client-side personalization the user profile - usually in the form of cookies - is stored on the client's machine rather on the web server per se. This increases the feeling of safety but in most of the cases the user has no direct control over the data stored and being transferred and cannot prevent the transfer of data he considers sensitive to web sites that he does not fully trust. Moreover, this category of implementations is usually application specific and does not promote the collaboration and exchange of user models among different web sites needed for faster and more complete user adaptation.

3. The proposed Architecture

The proposed architecture closely resembles most client-side personalization approaches. An important differentiation however is that according to our approach the service provider does not construct the user's profile alone but both the client and the service provider must collaborate and exchange a series of messages, for a personalization action to be completed. Additionally, the client may sacrifice a better personalization experience to gain increased privacy.

3.1 Architecture components description

The architecture distinguishes the following seven core components, namely: (a) the client, (b) the user modeling agent, (c) the adaptive service provider, (d) the user profiles, (e) the user data request document, (f) the privacy preferences document and (g) the user data response document.

The client is the physical location where the user modeling agent resides and user information, which is organized in various documents, is kept. It is the entity that makes requests for acquiring web pages from adaptive web sites and receives personalized content through the user modeling agent.

The user modeling agent is an application installed on the client side that aims to assist the user in the personalization experience of the various adaptive sites. The user modeling agent is responsible for the following tasks: it presents information to the user in human readable form, constructs, organizes and stores various user data in user profiles, receives and interpret requests on user profile data and produces responses by evaluating user's privacy preferences. The user modeling agent's role is assistive and it generally acts as an intermediate.

The service provider is the personalization component of adaptive web sites. Unlike most modern practices our architecture assumes that the service provider will not store any user data, although it does not enforce any technical mechanism that restricts him from doing so. Service provider may monitor user behavior if the user specifically wishes so, but the data gathered must be sent to the client's repository for later uses and must be deleted from the web server after each session. The characteristics, exact description and purpose of the information gathered in such cases should always be publicly available for users and web site developers. This practice might help achieving inter-site personalization and construction of general user models without the need for semantic web representation as discussed in [9].

Profiles are sets of user information organized in one or multiple xml files. Users may have multiple profiles stored on their machine. Some of this information is populated by the users, some by user agents and other by service providers. Typically user data is filled manually by the users themselves, environment data by their agents and usage data by the service provider of each site. Any information populated by the service provider requires user authorization first. Typically, usage data are included in the profile document in the form of xml references. Any information included in user profile is accessible by the user at any time. The language used to describe user information follows a standard format.

User data requests are special types of queries produced by service providers and sent to the user modeling agent. Every time service providers need information in order to generate a personalized web page, they request data that reside at the user profile on the client side. Queries of such kind follow a specific syntax. Service providers can also query profiles for information submitted by other service providers, e.g. concerning usage data of other adaptive sites. This is possible because data contained in the user models follow a specific syntax and the interpretation of those data is well known. By specifying rules on the privacy policy document users may deny to provide answers to part or the whole of these requests. The service provider then adjusts the generation of the web page according to the data returned. Any information included into a user profile is accessible by the user at any time.

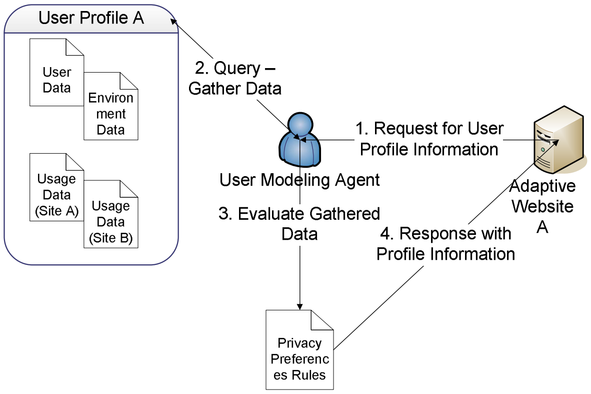

Fig. 1- An overview of the architecture components

The privacy preferences document is a special type of document in which the user specifies the privacy constraints that should be applied in the information sent toward the service provider. With the privacy rules defined in this document the user is certain that any kind of sensitive information will not be transmitted to a specific service provider. Initially, general privacy rules can be defined by the user in the privacy preferences document but the rules contained in it will be dynamically updated as he visits new sites.

User data responses are a special type of documents produced by the user modeling agent and sent toward the service provider. The data that constitutes the document is gathered from the user profile based on the provider's requests and privacy preference documents. An overview of the architecture components is depicted in Figure 1.

3.2 Architecture components interaction

In this section we describe how the previously introduced components cooperate in order for a personalization process to be accomplished. There are three steps in this process: (a) the initialization phase, (b) the negotiation phase and (c) the personalization phase. The architecture components interaction process is presented in Figure 2.

During the initialization phase the user modeling agent firstly prompts the user to create a new user profile, by filling a set of personal information or alternatively by selecting an existing one. At the same time the agent scans the client for environmental characteristics and updates the user profile accordingly. Then the client may visit the adaptive website. At this point a new usage data entry is created in the user profile by the adaptive website, if one does not exist already. Before the client starts to interact with the site, the service provider requests if he is allowed to monitor the user behavior for this session.

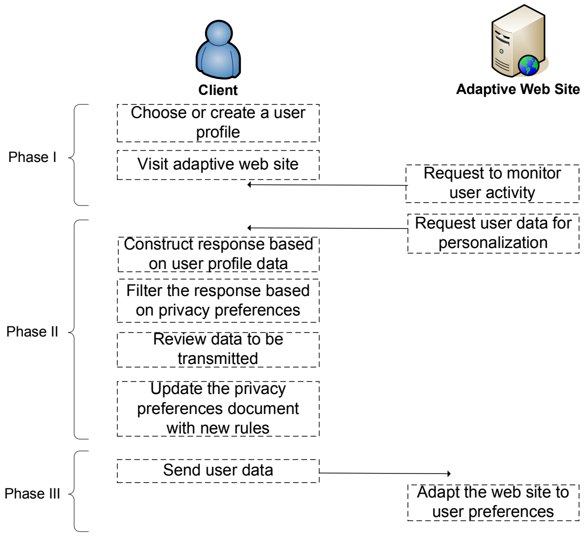
Client
Adaptive Web Site
Phase I
Choose or create a user profile
Visit adaptive web site
Request to monitor user activity
Phase II
Request user data for personalization
Construct response based on user profile data
Filter the response based on privacy preferences
Review data to be transmitted
Update the privacy preferences document with new rules
Phase III
Send user data
Adapt the web site to user preferences

Fig. 2 - The architecture components interaction

The negotiation phase is a multi-step process that results in the transmitting of user information to the service provider. Initially the service provider constructs a request (following a specific syntax) of certain user data for a corresponding personalization activity. The requested data may be user data, environmental data, usage data of its site (gathered from previous interactions with the system), or usage data from any other service provider it considers relevant. During the process the user agent gathers the corresponding data contained in the user profile to construct a response. Next, the user agent filters the response according to the rules contained in the privacy preferences document(s). Before sending the response toward the service provider the user agent presents the profile information to be sent to the user in order to make last minute changes or to deny sending his profile information for the current session. Finally, the privacy preferences document is updated with new rules (if any).

During the personalization phase, the service provider receives the personalization information the client sent, which is either all the information it requested for, or a subset of them. Then it adjusts the personalization experience and generates the final web page.

3.3 A real life scenario

In this section a simple, real life scenario is presented in order to examine the operation of the proposed system in more detail.

A user wants to visit an online bookstore. Before he starts interacting, he chooses the profile he usually adopts when dealing with e-commerce sites. This profile contains confidential information such as credit card information, real name and shipping address, as well as usage data from various sites (coming from previous interactions with them). Upon connecting to the site the user agent informs the user that the bookstore site may monitor his behaviour, gather usage data and update the user profile lying in the client machine at the end of the session.

For the personalization activity to be executed the web site needs to know the user's real name and a set of other information relevant to a typical user behaviour on the site, for example the favourite book category of the user. As a result the welcome page of the bookstore recognises the user by his name and makes some book recommendations according to user preferences. The corresponding request data document is presented in Table 1.

The user agent gathers the data requested from the information contained in the selected user profile. During this process, it filters the information to be sent according to the rules contained in the privacy preferences document. The user happens to have a high level of trust for this site so the rules are permissive. The user agent presents the information to be sent to the user but he does not wish to conceal any of it. So personalization action completes with success. The transaction completes and the user leaves the site.

Some time later the user decides to visit a new auction website for the first time. In order to personalize the content of its first page the particular website requires user personal data such as the real name, native language and relative usage data (from other sites) like the user's last purchases. The corresponding request data document is given in Table 2.

The user agent creates the response document. Since it is the first time that the user visits this website, there is no privacy rule in the privacy preference document, therefore none of this information is rejected. The user agent presents the data to the user, who decides to disclose his real name, in order not to relate it with the purchases he made in the past on various other websites. A new privacy rule is created and the privacy preferences document is updated. Table 3 presents the resulting privacy preferences document. The information is sent to the auction website, which adjusts its appearance accordingly. Thus, it welcomes the user with a general welcome message (since no real name was provided), translates part of its webpage and recommends some products based on user's past purchases record coming from other sites.

```
<Request-Data>
        <Purpose>
                <Type>
                        Recommendation
                </Type>
                <Type>
                        Personal Message
                </Type>
                <Requesting_Page>
                http://www.bookstore.com/welcome.html
                </Requesting_Page>
        </Purpose>
        <Sender>
                http://www.bookstore.com/
        </Sender>
        <Data>
                <First_Name/>
                <Last_Name/>
                <Most_Searched_Category/>
                <Last_10_Purchased/>
        </Data>
</Request-Data>
```

Table 1. The data request document of the bookstore site

3.4 Security assumptions

Since user data is held on the client rather the server, the architecture assumes that the user is able to store data safely on his machine and that the user modelling agent provides mechanisms for accomplishing that, for instance through means of encryption. Furthermore, since an exchange of messages which potentially contain sensitive data is expected to be conducted during personalization activities, the use of Secure Socket Layer (SSL) is presupposed. Client devices must support SSL and X.509 certificates for guarantying message confidentiality and integrity when in transit.

```
<Request-Data>
        <Purpose>
                <Type>
                        Recommendation
                </Type>
                <Type>
                        Personal Message
                </Type>
                <Type>
                        Translation
                </Type>
                <Requesting_Page>
                http://www.auctionwebsite1.com/welcome.html
                </Requesting_Page>
        </Purpose>
        <Sender>
                http://www.bookstore.com/
        </Sender>
        <Data>
                <First_Name/>
                <Last_Name/>
                <Native_Language/>
                <Last_10_Purchased/>
        </Data>
        <Affiliates>
                <Affiliate>
                        <URL>

        http://www.bookstore.com
                        </URL>
                        <Data>
                                <Last_10_Pruchased/>
                        </Data>
                </Affiliate>
                <Affiliate>
                        <URL>

        http://www.website1.com
                        </URL>
                        <Data>
                                <Last_10_Pruchased/>
                        </Data>
                </Affiliate>
        </Affiliates>
</Request-Data>
```

Table 2. The data request document of the auctions site

```
<Privacy-Preferences>
        <Deny_All>
                <Service>
                        http://www.website1.com
                </Service>
                <Service>
                        http://www.website2.com
                </Service>
        </Deny_All>
        <Deny>
                <Data>
                        <First_Name/>
                        <Last_Name/>
                        <Service>
                        http://www.auctionwebsite1.com
                        </Service>
                </Data>
        </Deny>
</Privacy-Preferences>
```

Table 3. The final privacy preferences document

4. Evaluation

The end user requirements that must be fulfilled by the system as summarized in [10] - [16]) are:

- Purpose specification
- Openness
- Simple and appropriate controls
- Limited data retention
- Pseudonymous interaction
- Decentralized control

As it will be explained further down our architecture corresponds to all aspects mentioned above.

Purpose specification is achieved in two ways. First, as part of the architecture i.e., every site is bound to have a publicly accessible document in which it will be described what kind of personalization activity the site is capable of, what kind of information it requires for each activity and in what way this will benefit the user. Second, every request executed by the adaptive service specifies what will be the personalization action i.e., the purpose of the request.

Openness is fulfilled since every time a user visits an adaptive site, a request is created by the service provider for allowing him to monitor user behavior. The user is able to control if data is collected about him or not. Moreover, since the user profiles are stored on his side in simple XML documents, he is able to view and manually modify the information contained in each one of them, at any time.

The architecture specifies the user modeling agent as one of its basic components. Its role is to communicate and negotiate with the service provider, automate procedures and present the information exchanged in a human-readable manner. The existence of this component must guarantee that the user will have a straightforward and smooth control over the system.

The architecture model assumes that the service provider will not store personal information that has been disclosed by the user. Also, data gathered by the adaptive sites must be instantly deleted after the end of the session. A mechanism for assuring that these rules will be adhered is not assumed by the architecture. Since the personalization process will always be done on-the-fly at every time the user visits the site, storing user information would be a meaningless procedure and would constitute suspicious behavior from the part of the service provider. A similar behavior might be regulated by international privacy laws. Moreover, the user always holds the right for denying disclosure of any information that he considers sensitive, if he does not trust the service provider. In this way data retention is always limited.

Users can create and host multiple user profiles. This is done deliberately because: (a) multiple physical entities might use the user's device to access a web site, (b) a user might act on behalf of another user (and not according to his own preferences) at a given time, (c) a user might wish to interact with a specific site with a different identity. A user might use any user profile from the ones stored in his device. This feature constitutes a mechanism for providing pseudonymous user interaction with the adaptive web sites.

The overall architecture of the system is distributed. Data are stored and processed on his side and integrated user profiles are built by the interaction with various adaptive sites.

Apart from the attributes mentioned earlier more advantages can be recognized from the adoption of this approach. These are summarized as:

- It is simple to implement and maintain in comparison to other approaches on the field.
- It allows portability of the user profiles, since all information is stored and organized in XML files.
- It is lightweight since no real process of data is performed on the client side apart from query execution and filtering.
- It allows inter-site collaboration for the creation of robust user profiles in sorter time.
- It allows the scaling of the personalization effect according to the user privacy preferences.

On the downside the architecture:

- Requires increased user interaction at least when users visit sites for the first time and/or privacy preference rules do not exist.
- Service providers must not store usage data they gather about clients, but no technical mechanism is provided to enforce that.
- Clients take the responsibility for storing and managing their data in a secure way.
- Service providers must comply with the standard in order other sites can use their usage data.

5. Conclusions and Future work

In this paper we proposed an architecture that can flexibly adjust and narrow down the personalization experience in order to preserve the user's privacy. We believe that although personalization in web sites is an important aspect, the user should be given the chance to trade it for greater privacy. We strongly support inter-site collaboration for achieving rapid personalization and we demonstrated that through our architecture this is not whittled but the contrary it is enforced.

Nowadays, because of the pluralism of the products, and services the adaptive sites provide, it may be prove unrealistic to expect all the service providers to conform to a uniform way of data representation and description. Resource Description Framework (RDF) could prove a valuable asset for the representation of the products and services of heterogeneous environments and might help to the construction of generic user profiles on the client side. This in turn, is expected to favour inter-site user profile exchange and optimize the way user agents and service providers communicate.

References

[1]. Kobsa A., Koenemann J., Pohl W. Personalised hypermedia presentation techniques for improving online customer relationships. The Knowledge Engineering Review, Vol. 16(2), 111–155, Cambridge University Press, 2001

[2]. Brusilovsky P., Adaptive hypermedia. User Modeling and User-Adapted Interaction, 2001

[3]. Kobsa A., Personalized Hypermedia and International Privacy, 2002

[4]. W3C, http://www.w3.org/

[5]. W3C, A P3P Preference Exchange Language 1.0 (APPEL1.0), http://www.w3.org/TR/P3P-preferences/

[6]. W3C, The Platform for Privacy Preferences 1.0 (P3P1.0) , http://www.w3.org/TR/P3P/

[7]. AT&T Privacy Bird, http://www.privacybird.org/

[8]. Lorrie Faith Cranor, 'I Didn't Buy it for Myself' Privacy and Ecommerce Personalization, 2003

[9]. Dickinson I., Reynolds D., Banks D., Cayzer S., Vora P., User Profiling with Privacy: A Framework for Adaptive Information Agents

[10]. Brar Ajay, Judy Kay, Privacy and Security in Ubiquitous Personalized Applications

[11]. Adams A., Multimedia Information Changes the Whole Privacy Ball Fame, Proceedings of Computer, Freedom and Privacy. Toronto, Canada, 2000

[12]. Australian Privacy Act, Information Privacy Principles under the Privacy Act 1988

[13]. Belotti V., Sellen A., Design for Privacy in Ubiquitous Computing Environments, Proceedings of the Third European Conference on Computer Cooperative Work. Milan, Italy, 1993

[14]. Hong J.I., Landay I. and J., An Architecture for Privacy-Sensitive Ubiquitous Computing, Proceedings of the 2nd International Conference on Mobile Systems, Applications and Services. Boston, Massachusetts, USA, 2004

[15]. Hong J.I., Boriello G., Landay J. A., Mc Donald D. W., Schilit B. N., Tygar J.D., Privacy and Security in the Location-Enhanced World Wide Web. Proceedings of Fifth International Conference on Ubiquitous Computing, Seattle WA, USA, 2003

[16]. Langheinrich M., A Privacy Awareness System for Ubiquitous Computing Environments. Ubicomp 2002

[17]. Schreck J., Security and Privacy in User Modeling. Kluwer Academic Publishers, 2003