PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 119
Informatyka Ekonomiczna 18 2010

Witold Abramowicz, Agata Filipowska, Jakub Dzikowski, Paweł Kalisz, Jacek Małyszko, Łukasz Werno
Poznan University of Economics

EGO – WHERE USER MODELING MEETS IDENTITY MANAGEMENT

Summary: Current identity management and user modeling systems suffer from certain limitations. Although, theoretically, they have similar goals: to represent a user and to support him while carrying out certain activities, there is a clear gap between them in means of aspects of users' activity, which they represent. This paper addresses this issue, proposing a system, which aims at incorporating strengths of both types of systems to provide a single, unified method of representation of users in the Web. Description of the system includes functional analysis, system architecture, identity life cycle scheme and more.

Key words: identity management, user modeling, adaptive systems, personalization.

1. Introduction

The Web is nowadays becoming an integral part of both industry and people's everyday life [Mahonen 2006]. As a result, more and more real-world activities are, at least partially, transferred to the Web. These activities include for example shopping [Sangwan, Siguaw, Guan 2009], communication with friends [Preece 2000], fulfilling one's information needs [Abramowicz 2008]. During these activities Internet users often encounter some obstacles.

Information overload [Grise, Gallupe 2000] makes it often impossible for an individual to process the information delivered as a response to a query and find the information he really needs [Ho, Tang 2001]. The information systems could help users in searching the information they need, but first they have to 'understand' users' needs [Brusilovsky, Millán 2007].

Other problems users may encounter in the area of virtual communities [Rheingold 1993]. Users create communities by gathering around specific topics, related to their interests, in online forums, chats, massively multi-player online games, etc. [Nabeth 2006]. A problem, that a user might come across in this situation, is an obligation to create a new user account and perform a registration process each time, when he wants to become a member of a new portal and new community [Grohol 2006]. Additionally,

accounts on these portals are separate entities, and user cannot easily build his single, unified online identity and update it, if needed [Jaquet-Chiffelle 2008].

In the past there have been numerous initiatives aimed at solving such problems. Research on user modeling, adaptive systems and identity management systems are the most obvious examples [Brusilovsky, Millán 2007; Meints 2009, Zwingelberg 2009]. Nevertheless, the current solutions suffer from certain limitations. As section 2 of this paper shows, there is a lack of a single method of representing users, which would be adequate to a broad set of applications, required by the rapidly growing users' involvement in the Web.

The presented paper proposes a system addressing the current limitations. It introduces Ego – a project aiming at creation of a system based on virtual identities, that will enable Internet users to easily present their needs in a broad set of applications. Our approach focuses on giving the control over the identity to the user himself and providing a framework for exploiting information stored within identities on a wide range of applications. Such system will improve the overall accuracy of an information flow within the Web and help users in their everyday Internet activities.

This paper is structured as follows: section 2 presents the current state of the art in the domains of user modeling and identity management. It raises issues of adaptive systems and identity management. Third section presents the proposed approach, including the functional analysis, the Ego system architecture and the identity life cycle scheme. The article concludes with a summary including also directions of the further research.

2. Related work

2.1. User identity in the information society

The notion of identity in the information society is gaining recently much attention. It is expected, that the popularity of the Internet will result in new forms of online identities, merging the real world and life of individuals with the digital ones [Mahonen 2006]. FIDIS project aims at pointing out, that proper identity management is a key in the modern information society development [Rannenberg, Royer, Deuker 2009]. The project, among others, raise a problem of developing, organizing and standardizing concepts connected with identities, such as user modeling and identity management [Hedbom, Van Alsenoy 2009]. We will focus on these two concepts.

2.2. User modeling and adaptive systems

The issues of understanding the user needs and adaptation to them have been topics of interest of various researches for many years, since the introduction of the classic papers on the subject [Goldberg, Nichols, Oki 1992; Rich 1979]. The under-

standing of users' needs requires building user models [Brusilovsky, Millán 2007]. A user model is a source of knowledge about the user, containing assumptions covering all users' activity aspects that can be used by the particular system [Chen, Magoulas 2005]. Information systems that adapt their contents or behavior to the user needs are called adaptive systems [Brusilovsky, Millán 2007].

There is a wide range of applications of adaptive systems, ranging from recommendation systems and information filters, intelligent e-learning, to various systems adjusting interaction patterns to users' preferences [Kay 2001]. One of the most successful applications of this kind of systems is a recommender system embedded in the Amazon.com online shop [Linden, Smith, York 2003]. It builds users' models based on their historical shopping behaviors and on marks, that users have given to in-stock items [Gregory, Jacobi, Benson 2001]. These models are then exploited by the recommender system that as a result provides a personalized offer to each client [Gregory, Jacobi, Benson 2001].

Another example of an adaptive system is the Google search engine. User models are built here upon information about users' search and browsing history for users logged-in to their Google account [Sullivan 2007] or accept anonymous cookie file [Horling 2009]. These models are used to personalize search results in Google Search Engine [Horling 2009; Sullivan 2007].

Both systems share some limitations. Although they gather a lot of information about users, the user models are inaccessible from the outside of the system. Apart from trust and privacy issues, the user models, although potentially very useful for the users, are stored beyond their control and cannot be transferred from one system to another. As a result, every time, when one registers to a new system, the process of learning the user's needs by the new system has to be started from scratch.

2.3. Identity management systems

Another interesting type of systems from the point of view of this paper are the identity management systems. Identity management systems deliver a wide range of functionalities, from implementing authentication, authorization, and accounting to user-controlled context-dependent role and pseudonym management [Meints, Zwingelberg 2009]. These systems work both in corporate environment and, in a wider context, on a scale of the whole Web (Global Identity Management systems) [Meints, Zwingelberg 2009].

W3C [*Requirements*… 2001] has defined requirements for a global Identity Management service. The list of requirements for this service, among others, includes interoperability, extensibility and negotiated privacy and security.

A popular example of a global identity management system is OpenID[1]. OpenID eases registration and logging-in processes by providing a single login and password on many different systems and allows reusing of some basic information

[1] http://www.openid.net.

about the user across these systems [Becker 2006; Recordon]. The OpenID system is based on a decentralized architecture, what means, that there is no central repository of identities. Each user can choose where to store his identity – even at his own local server [Becker 2006].

The classical identity management systems focus on user authentication, and amount of other information about users managed by such systems is limited [Meints, Zwingelberg 2009]. In the case of OpenID, identity contains only basic information about the user, such as his name, pseudonym and the e-mail address [Becker 2006]. Little information makes it impossible to process such identity in order to help user in fulfilling his information needs.

Based on the above analysis, one may note, that the identity management and user modeling are currently heavy researched, but the current systems still suffer from some limitations. There is a need for providing a more universal approach to the identity management, if the Web is to meet growing requirements of more and more active users. The next section brings a proposal of a solution to these issues.

3. Proposed approach

The proposed approach aims at creating a model of identity, which will allow reusing of users' representations to improve addressing and flow of information throughout the Web. We assume that the system should have the following features:

– versatility (extensibility and interoperability),
– decentralization,
– full control of a user over his identity in means of access control issues.

In this section we discuss certain features, which future identity systems should incorporate. We also propose a schema for an identity life cycle.

3.1. Usage scenario

The basic idea of the Ego system is to create a versatile application, which will automatically build universal user model and manage its evolution. Enabling user to create and manage their representation should enhance Internet browsing and allow delivering more user-oriented content. The process starts when a user creates an account and provides some basic data about him or her. From this time on, a special component of the Ego system starts learning user's information needs and building (and updating) his or her model (called in the Ego project user's virtual identity). The creation will be based upon the analysis of user's online activities (for example extended analysis of web pages the user visits).

In a typical scenario (Fig. 1), a user (having logged into the Ego system in a form of a plug-in toggle activation) browses a particular web page. A certain component (scrobbler) retrieves information about the content of the website and

browsing context (date and time, localization of the user, hardware and software platform). Acquired information is sent to identity provider (identity server), which – using the information – updates user's virtual identity. During browsing the Internet by the user, service providers (websites, recommender systems) can query user's virtual identity and personalize presented content utilizing Ego response.

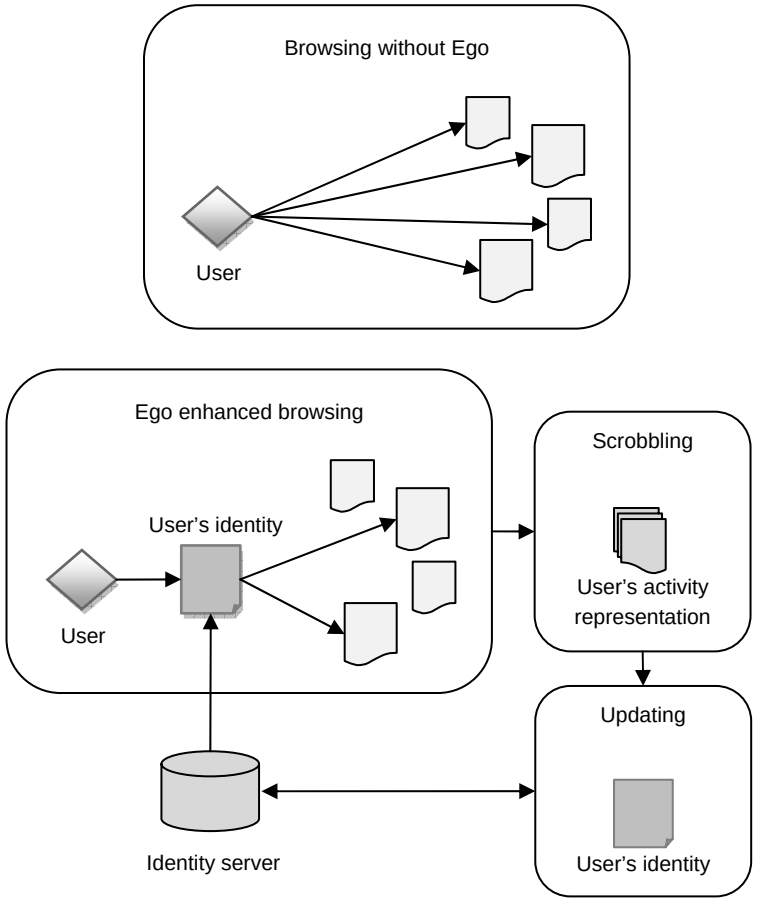

Fig. 1. Browsing with and without Ego system

Source: own elaboration.

3.2. Functional requirements

To address the main objective of the system, it should support the following functional requirements:

1. The system must gather various information about the user from heterogeneous and mostly unstructured sources.

2. The process of information gathering must not be bothersome to the user.

3. The user needs to have control over information stored in his or her identity and over access rights to the information.

4. User's virtual identity should comprehensively represent user's information needs and their evolution.

5. The external applications should be able to query user's virtual identity in order to receive relevant information for the purposes of personalization.

3.3. System architecture

There are two principles we are following as far as system architecture is considered and these are versatility and decentralization. By the versatility we understand the feasibility to be implemented in the scope of a single adaptive system as well at a scale of the whole Internet. The decentralization we understand similarly as in the OpenID case, which means that anyone can use the system or be an Ego provider, without being registered or approved by any central organization. In this section we present the architecture that allows gaining these two features.

The basic view on the architecture of the proposed system is presented in the Fig. 2. The architecture comprises several elements: scrobblers, identity servers, identity catalogues and client services.

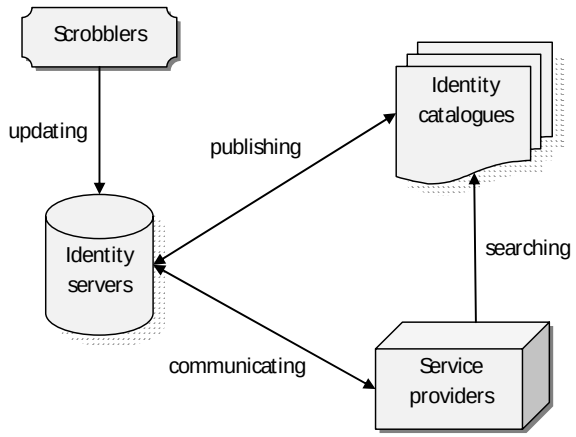

Fig. 2. The components of the Ego architecture and flows between them

Source: own elaboration.

Scrobblers are applications that examine users' activities within the system and provide the identity servers with information about users actions. Scrobblers can be for example:

– plug-ins to a Web browser, which extract contents of web pages read by the user,
– server-side applications, which gather information about resources accessed by users,
– plug-ins to media players, which send information about, for example, songs that were listened by the user.

These programs can analyze resource content, extract the essence of accessed documents and perform various operations (tokenization, lemmatization or stemming, morphological analysis and named entity recognition). For the purpose of user modeling in Ego system, any scrobbler needs to return a fixed representation of current user's activity (surrogate of information resource, the source and the user's identifiers, timestamp etc.).

Acquired information in a form of an RDF document is sent to a chosen **identity server** via the HTTP protocol. The identity server's main functionality is to store and update the user's virtual identity to make it better reflect user activities, interests and knowledge. This issue is described in the identity lifecycle section.

The identity servers provide users with a control panels for managing access rights for information stored within identities. It allows an identity owner to decide, which services, Web agents or websites can access which part of his or her identity information.

According to the decentralization principle, the identity sever can be deployed by virtually anyone. To enable versatility, such a server can work both in small networks, for example in a company environment, and at a scale of a whole Internet. General-purpose user's virtual identity stored on the Internet-level server is the global identity. Every other identity will be referred to as a local identity. The user can have a global identity and simultaneously the local one in his company, which stores only information about his in-work activities. It is possible to connect local identities with the global one. Such assignment implies that information stored in the local identities will be used to update the global identity.

The **identity catalogues** publish identities to other services or users within the Web. The client services can search for certain identities and information, according to permissions assigned by identity owners.

The last link of the Ego identity system are **client services**. Using purpose-built query protocols, and both authentication (that particular service) and authorization (user has agreed to share his data) methods, client services retrieve information about the user from his identity and can exploit it for many different purposes.

3.4. Identity life cycle

During its lifecycle, an identity is subject to many different processes. These are creation, updating, merging, user interference, access control and querying (Fig. 3).

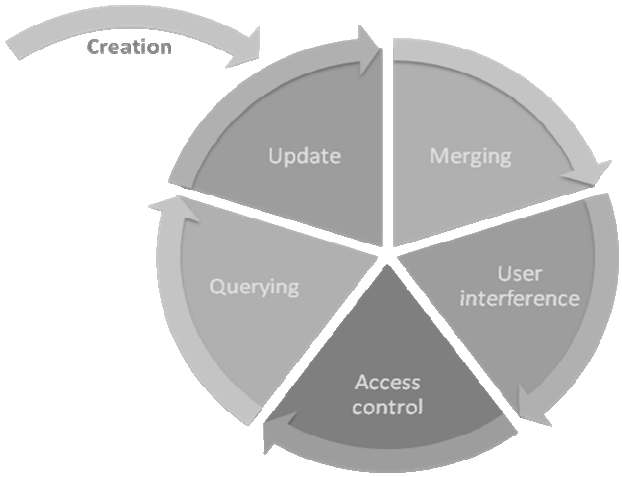

Fig. 3. The identity life cycle

Source: own elaboration.

Creation. To assure the usefulness of an identity from the very beginning of its existence, the system enables using start-up identities. A start-up identity is a predefined user stereotype, which includes expected user characteristics. For example, a newly hired employee in a company can be fitted with a predefined identity based on a role he plays in an organization. The stereotype in this case will be built by merging identities of other users, that play the same role in the organization.

Updating is a process of incorporating into the identity the structured data retrieved by scrobblers from recent activities of the user. It is crucial at this point to establish such method of updating, which assures gradual evolution of an identity, and not its rapid changes. The identity must reflect the current characteristics of a user, but also be aware of its history.

Merging. It is possible for the user to have many identities. Therefore, it is necessary to provide an efficient method of merging several identities, so that querying would concern a broader view of user's information needs and interests. This functionality potentially can be used also for other purposes, for example in already mentioned generating of start-up identities by merging individuals' identities.

User interference. The user should have a possibility to control the information stored in his identity. For example, he or she should be able to determine what kind of information about him or her should be gathered, or even to delete certain information from identity. Still, this interference should be limited to ensure objectivity of information stored in the identity. As a result, user interference should improve his or her virtual identity.

Access control. Another dimension of control of the user over his identity is a possibility to assign permissions to access some parts of the identity by certain services or groups of services. It is also important to make this activity simple and not bothering to the user.

Querying is the most important phase in means of utilization of information stored in identity. Based on permissions assigned, services can send certain queries to the identity. As a result, services can gain knowledge about a user, which can be exploited, for example, for adaptation to user needs.

4. Conclusions and further work

The current state of the Internet development faces many issues that can and need to be solved. There is a need to create a system which combines advantages of identity management and user modeling so as to provide more accurate information about Web user to wide range of adaptive systems.

Ego system will cover requirements of versatility and distribution, and will allow users to improve or adjust their identities. The system will provide new methods of learning about users' interests and pursues and deliver more direct content.

Further work will focus on the modularization of the system and on considering the possibilities of integration with existing solutions, such as OpenID. Additional methods of identity utilization will be considered, such as, for example, it's presentation on social networking sites.

Literature

Abramowicz W., *Filtrowanie informacji*, AE, Poznań 2008.

Becker P., *The case for OpenID*. Retrieved February 19, 2010, from http://blogs.zdnet.com/digitalID/?p=78 (Published 2006, December 4).

Brusilovsky P., Millán E., User models for adaptive hypermedia and adaptive educational systems, [w:] P. Brusilovsky, A. Kobsa, W. Nejdl, *The Adaptive Web*: *methods and strategies of web personalization*, Springer, Berlin 2007, s. 3-53.

Chen S., Magoulas G., *Adaptable and Adaptive Hypermedia Systems*, IRM Press 2005.

Goldberg D., Nichols D., Oki B.T., *Using collaborative filtering to weave an information tapestry*, Communications of the ACM 1992, 35 (12), s. 61-70.

Gregory D., Jacobi J.A., Benson E.A., *Collaborative recommendations using item-to-item similarity mappings*, 2001, United States Patent 6266649.

Grise M., Gallupe R., *Information overload*: *Addressing the productivity paradox in face-to-face electronic meetings*, "Journal of Management Information Systems" 2000 (16).

Grohol J., *Anonymity and online community: Identity matters*, Retrieved February 19, 2010, from http://www.alistapart.com/articles/identitymatters/(2006, April 4).

Hedbom H., Van Alsenoy B., *D19.3: Standardisation report*, FIDIS (Future of Identity in the Information Society) 2009.

Ho J., Tang R. *Towards an optimal resolution to information overload: an infomediary approach*, Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work 2001.

Horling B.K. *Personalized Search for everyone*. Retrieved February 19, 2010, from The Official Google Blog: http://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html (2009, December 4).

Jaquet-Chiffelle D., *D2.13 virtual persons and identities*, FIDIS (Future of Identity in the Information Society) 2008.

Kay J., *User Modeling for Adaptation*, [w:] C. Stephanidis, *User Interfaces for All*: *Concepts Methods, and Tools*, Lawrence Erlbaum Associates 2001.

Kobsa A., Wahlster W., *User Models in Dialog Systems*, [w:] A. Kobsa, W. Wahlster, *User Models in Dialog Systems*, Springer, Berlin 1989.

Linden G., Smith B., York J., *Amazon.com recommendations: Item-to-item collaborative filtering,* IEEE Internet Computing 2003, 7 (1), s. 76-80.

Mahonen P., *The Future Networked Society*, EIFFEL Think Tank 2006.

Meints M., Zwingelberg H., *D3.17: Identity management systems – recent developments*, FIDIS (Future of Identity in the Information Society) 2009.

Nabeth T., D2.2: *Understanding the identity concept in the context of digital social environments*, *FIDIS* (Future of Identity in the Information Society) 2006.

Nabeth T., *D2.3: Models*, FIDIS (Future of Identity in the Information Society) 2005.

Preece J., *Online Communities: Designing Usability and Supporting Sociability*, John Wiley & Sons, New York 2000.

Rannenberg K., Royer D., Deuker A., *The Future of Identity in the Information Society*: *Challenges and Opportunities*, Springer, Heidelberg 2009.

Recordon D., *ACM Workshop on Digital Identity Management*, New York, s. 11-16.

Requirements for a Global Identity Management Service, W3C Workshop on Web Services, San Jose, CA USA 2001.

Rheingold H., *The Virtual Community*: *Homesteading on the Electronic Frontier*, Perseus Books, 1993.

Rich E., *User modeling via stereotypes*, "Cognitive Science" 1979, 3 (4), S. 329-354.

Sangwan S., Siguaw J. A., Guan C., *A comparative study of motivational differences for online shopping*. ACM SIGMIS Database 2009, 40 (4), s. 28-42.

Sullivan D., *Google Search History Expands, Becomes Web History*, Retrieved February 19, 2010, from http://searchengineland.com/google-search-history-expands-becomes-web-history-11016 (2007, April 19).

EGO – NA POGRANICZU MODELOWANIA UŻYTKOWNIKA I ZARZĄDZANIA TOŻSAMOŚCIĄ

Streszczenie: Obecne systemy zarządzania tożsamością oraz systemy modelujące użytkownika, pomimo podobnego celu – reprezentacji użytkownika oraz wspierania go w określonych czynnościach – obejmują całkowicie różne aspekty aktywności użytkownika. W artykule przedstawiono wymagania dla systemu łączącego zalety zarządzania tożsamością oraz modelowania użytkownika, tworzącego ewoluujące w czasie, uniwersalne modele użytkownika, ułatwiające personalizację w sieci. Opisano wymagania funkcjonalne dla takiego systemu, jego architekturę oraz cykl życia wirtualnej tożsamości użytkownika.