User Modeling in Information Portals

George D. Magoulas
University of London, UK

INTRODUCTION

The concept of information portal spans over various domains such as document collections, enterprise information portals, digital libraries, subject gateways, Web directories, and government portals (Tatnall, 2005).

Users seeking for content through an information portal increasingly look for more intelligent services and support in order to avoid disorientation and develop a holistic understanding of how all the information fits together that will help them to better formulate their search goals and information needs. One of the key tools in offering more intelligent services to the users of information portals is personalization technologies (Lacher, Koch, & Woerndl, 2001; Riecken, 2000). Personalization aims to tailor information and services to each individual user's characteristics, usage behavior, and/or usage environment (Brusilovsky, 2001). Nevertheless, to provide effective personalization, an understanding of the individual user and their cognitive characteristics, goals, and domain knowledge is needed (Benyon & Höök, 1997; Manber, Patel, & Robinson, 2000). This understanding about users can be achieved through a user modeling process by means of a user-guided approach, in which user models are created on the basis of information provided by each user (Fink, Kobsa, & Nill, 1997) or an automatic approach, in which the process of creating a user model is hidden from the user (Brusilovsky & Schwarz, 1997).

This article provides a background on existing approaches for developing user models. It identifies the basic types of information that need to be stored in a user model and discusses tools for automated user modeling. Lastly, it discusses future trends in user modeling for Web portals.

BACKGROUND

Adopting an appropriate approach to user model development and deployment is important for achieving personalization. In the 70s, user modeling was performed by the main application and often it was not possible to separate the user-modeling component from other system components. In the 80s, distinctive components were introduced to carry user-modeling tasks, and later on the concept of reusable user modeling components was proposed (Finin, 1989). Taking inspiration from the field of expert systems, user models were developed as shells in order to support complex reasoning processes about the user and to be usable in a wide range of domains (Kobsa, 1990). In the middle 90s, the advent of the World Wide Web and the development of Web-based applications led to client-server architectures for Web personalization and allowed the deployment of user modeling servers (Kobsa, 2001). However, user-modeling servers in many cases are developed as domain dependent and are not considered flexible enough as their user model representation is closely interlinked with other data processing modules (Fink & Kobsa, 2000).

One way to introduce flexibility is to construct a user model automatically, minimizing the user's involvement in the modeling process. Thus, an automatic approach has been proposed to create user models by observing users in an unobtrusively way, and collecting information even when users are not willing to give feedback of their actions, or their preferences change over time (Montaner, Lopez, & de la Rosa, 2003; Semeraro, Ferilli, Fanizzi, & Abbattista, 2001). This is based on the idea that a typical user exhibits patterns when accessing a Web–based system such as an information portal and the set of interactions containing those patterns can be stored on a database. Intelligent computational techniques can then be applied to recognize regularities in user trails such as particular skills, aptitudes, and preferences for processing information and constructing knowledge from information (Zukerman, Albrecht, Nicholson, 1999).

In order to automatically create user models for information portals, the following issues need to be examined in detail: (1) what information should a user model contain and (2) what techniques can be used to automatically model the user. These questions are answered in sections next.

WHAT INFORMATION CAN BE INCLUDED IN A USER MODEL?

There are no standards for developing use models, only guidelines about what a user model can represent (Kobsa, 2001). Among a wide range of user-related data that can be stored in a user model, we consider nine elements for user modeling in information portals:

1. **Personal Information:** Gender, age, language, culture, etc. Some of these factors affect the perception of the interface layout. For example, gender differences affect access in the sense that males and females have

different requirements with respect to navigation support (Czerwinski, Tan, & Robertson, 2002) and interface features as they exhibit significant differences in their browsing and information management behavior (Large et al., 2002). The preferences of males and females also differentiate remarkably in terms of attitudes, information seeking strategies (Vaughan, 1993; Zoe & DiMartino, 2000), and media preferences (Parush & Bermanb, 2004).

2. **Information Processing Preferences:** These refer to a user's information processing habits and have an impact on user's skills and abilities such as preferred modes of perceiving and processing information and problem solving (Chen, Magoulas, & Macredie, 2004; Magoulas, Papanikolaou, & Grigoriadou, 2003). They can be used to personalize the navigation support, the presentation, and organization of the content and search results (Magoulas, Chen, & Dimakopoulos, 2004).
3. **Hardware Specifications:** It concerns the hardware used to access the information space and affects personalized services in terms of screen layout and bandwidth limitations (Cohen, Herscovici, Petruschka, Maarek, & Soffer, 2002).
4. **Physical Context:** This dimension captures the physical environment from where the user is accessing the portal (office, home etc.) and can be used to infer the goals of that user and adapt the content accordingly (Maamar, AlKhatib, Mostéfaoui, Lahkim, & Mansoor, 2004).
5. **User History:** This dimension captures user past interactions with the portal and can be used to personalize any kind of service under the assumption that a user is going to behave in an immediate future in the same way it has behaved in the immediate past. Among other data may include pages visited that contain pointers to specific keywords or browsing habits (Sugiyama, Hatano, & Yoshikawa, 2004).
6. **Content Preferences and Interests:** These are usually provided in the form of keywords or topics of interest for that user and can be used to filter the content (Middleton, De Roure, & Shadbolt, 2001; Tanudjaja & Mui, 2002).
7. **Motivation:** It indicates the reason for which that user is searching information in a particular session (Sellen, Murphy, & Shaw, 2002). For example, it is not the same to search for information about China as a tourist searching for information about his or her destination or as a manager preparing a business report.
8. **System Experience:** It indicates the prior knowledge a user has about an information space (e.g., level of computer skills, experience with other Web portals). This information can be used to personalize the navigation, the search results, or provide intelligent help. For example, system experience may depend on users' familiarity with the features and functionalities of a library portal (Stelmaszewska, Blandford, & Buchanan, 2005) or with her familiarity with some functionalities of an educational portal (Mitchell, Chen, & Macredie, 2005).
9. **Background Knowledge:** This dimension relates to the existing level of understanding of a particular user on the domain knowledge. Note that the level of expertise of a user can vary with the domain and influences the navigation behavior leading to disorientation problems (Last, O'Donnell, & Kelly, 2001).

WHAT TECHNIQUES CAN BE USED FOR AUTOMATIC USER MODELING?

A variety of techniques have been proposed to build sophisticated user models such as probabilistic Web mining and soft computing methods.

Probabilistic methods (Zukerman & Albrecht, 2001) such as Markov models, Bayesian classifiers, and Bayesian networks can be used to capture the transitions of a user between the different states of a portal. For example, they can be used for modeling user's navigation behavior from low-level information provided by temporal sequences of navigation actions and tracking of user's navigation behavior in an information portal, as well as for predicting users' interests of a particular type of content by analyzing the pages that they have previously visited.

Web mining is a special kind of data mining that deals with the task of extracting implicit, previously unknown, but potentially useful information from Web data (Pal, Talwar, & Mitra, 2002). Data collected from a portal can be distributed, heterogeneous, and high dimensional so Web mining methods analyze data logs looking for trends, patterns, and relationships, without knowledge of the actual meaning of the stored data (Erinaki & Vazirgiannis, 2003; Pierrakos, Paliouras, Papatheodorou, & Spyropoulos, 2003). For example, they can be used for extracting structured relations from unstructured text collections in information portals, or for finding unexpected information such as new services and products in an enterprise information portal.

Soft computing techniques have been used successfully for representing imprecise knowledge about the user and creating user models (Frias-Martinez, Magoulas, Chen, & Macredie, 2005). Fuzzy logic, one of the most popular soft computing methods, facilitates creating user models in environments such as an information portal, where, usually, users are not wiling to give feedback on their actions, and as a result, the degree of uncertainty is very high. Nevertheless, the process of applying fuzzy logic-based techniques involves making several informed decision for creating a user model. For example, in user modeling the concept of