In: Josef Fink: User Modeling Servers – Requirements, Design, and Evaluation.
Amsterdam, Netherlands: IOS Press (Infix), 2004

Foreword

Quite a few generic user modeling systems have already been developed during the past fifteen years (see Kobsa [2001] for a historic overview). The aim of these 'user modeling shell systems' and more recently 'user modeling servers' is to facilitate the provision of user modeling services to application systems by supplying core functionalities for such services.

The development of user modeling servers to date has been largely shaped by requirements that reflect academic interests, such as application and domain independence, expressiveness of the underlying representation language, and powerful inferential capabilities. The way of looking at the problem of generic services, the search for solutions as well as the appraisal of solutions have all so far been strongly influenced by artificial intelligence research. Results from neighboring disciplines, such as databases and distributed systems, have been largely ignored. Empirical evaluations of the practical applicability of developed user modeling servers (e.g., in terms of performance and scalability) were hardly ever carried out.

The work of Josef Fink is different: the author takes an interdisciplinary and application-oriented approach, puts his work into the context of both academic tradition and budding commercial software, and is not shy of thoroughly evaluating the user modeling server that he developed and implemented.

The author proposes a single contact point for applications to obtain user modeling services, and his solution is therefore 'centralized' in this sense. Unlike previous research, his centralized approach however neither entails the physical centralization of user-related information, nor of the processing of this information. As far as user information is concerned, the Lightweight Directory Access Protocol (LDAP) that the author employs as a representational basis affords the option of replicating information about users temporarily if needed and synchronizing it automatically (this seems to be very useful for mobile and for highly fail safe applications). Regarding the processing of information, his underlying CORBA architecture allows the server to be dynamically distributed across platforms if so desired.

The author's choice of an LDAP-based representation for assumptions and information about users offers significant advantages over the two traditional approaches in academia and industry, namely knowledge representation systems and database systems. These advantages lie particularly in international vendor-independent standardization, demonstrated performance and scalability, dynamic and transparent management of distributed information, built-in replication and synchronization, a rich number of pre-defined user-related information types, and the extensibility of the core representation language for new information types and for data types with associated semantics.

In the last part of his work, the author describes a very interesting simulation experiment to test the scalability of his user modeling server to real-world application scenarios. The design of this study sets yet another milestone in user modeling research. Fink simulates users who submit requests to a web server, which cause queries about these users' interests to be submitted to the user modeling server and also new user interests to be added. His simulation becomes even more valuable due to the fact that the parameters that describe users' web navigation behavior were not extrapolated from web server data (where only a

skewed fragment of this behavior can be observed), but were mostly derived from a large client-side analysis of users' web navigation that was conducted by a well-known Internet usability research company.

The results of this experiment look very convincing: when tested on an entry-level PC with up to 12,500 user models and with a load that is higher than that of 75% of all German web servers, the mean response time was only 50ms per request (100ms with two sigma added). Distributing the server to four platforms brought an improvement by 60%. In an industrial server cluster with a 20-fold increase in workload and 8 million user models, the mean response time was 35ms. Industrial deployments of this server therefore seem possible by all means (as a matter of fact, one profiling application across most German Top 100 web sites with a total workload of several billion page impressions per month is already underway).

The work described in this book constitutes an excellent synthesis of scientific and industrial concerns. Fink defines meaningful requirements on user modeling servers, gives an overview of existing systems, pinpoints their deficiencies, develops a very novel architecture for user modeling servers, implements it, and tests its utility both within an application project and in empirically founded performance experiments. The unusual breadth of his approach (which rests on research in data bases, distributed systems, human-computer interaction, user modeling, statistics, and e-commerce) and his very convincing solutions make this book a worthwhile reading both for researchers and for industrial practitioners.

Alfred Kobsa

University of California, Irvine

Kobsa, A. (2001). Generic User Modeling Systems. *User Modeling and User-Adapted Interaction* 11(1-2): 49-63. http://www.ics.uci.edu/~kobsa/papers/2001-UMUAI-kobsa.pdf