Towards Support for Personalization in Distributed Digital Library Settings

David L. Hicks
Department of Computer Science
Aalborg University Esbjerg
Niels Bohrs Vej 8
6700 Esbjerg, Denmark
Phone: ++45 7912 7632
Fax: ++45 7545 3643
Email: hicks@cs.aue.auc.dk

Klaus Tochtermann
Know-Center
Inffeldgasse 16c
8010 Graz
Austria
Phone: ++43 316 873 5671
Fax: ++43 316 873 5688
Email: ktochter@know-center.at

1 Introduction

The world of information today is an increasingly diverse and distributed one. Both the amount and variety of information available in digital form continue to increase at a rapid rate. Evolving network infrastructures enable large information repositories to be queried and accessed from virtually anywhere. These trends in the availability and management of information have important implications for knowledge workers. In order to work effectively, knowledge workers must increasingly be prepared to look to digital sources, including those available over the network, for more and more of their information needs. Remotely located information items managed by external systems distributed across networks represent an important part of the overall information needs of knowledge workers. For example, an employee in the process of preparing an environmental impact report might need to gather information from several sources, interacting with a geographic information server at one location to obtain maps and related geographic data, a government environmental information server at another site to obtain the required statistical data, and an image server at yet another site to obtain satellite imagery for the report.

The ability for knowledge workers to personalize their information space is an important capability, one that can facilitate their ability to perform complex tasks [8]. For example, the user described in the previous example would probably find it convenient to be able to attach annotations to a satellite image obtained from an image server in order to communicate personal observations to a coworker during the process of preparing the report. The importance of user customization and personalization capabilities has been noted in the literature. Nuernberg suggests the need to support the easy and fast personalization of information accessed by users of web client applications, in order for the information to be used more effectively [4]. Additionally, they point out that the new digital processes that will characterize future information systems (e.g., agents, user profiling, and other automated personalization mechanisms) will likely require even further personalization and customization functionality than available in existing systems. Marshall notes the need for supporting personal annotation for the holdings contained in digital libraries, citing the importance of providing a digital analogue to this familiar and convenient form of marking up and working with paper-based documents [3]. Roescheisen reports that the process of a user personalizing an information space adds value to it [6].

One possible approach for supporting the personalization of distributed information would be for the systems that knowledge workers interact with to support the personalization process. For example, the creation of the annotation described in the previous example could be supported by the information system that manages the satellite image to which the annotation refers. This type of approach might be feasible for a system with a localized and limited user base. Tracking personalization information for a widely used network based information system, however, is a much different task. These systems have a potentially large number of distributed users. Supporting personalization with a centralized approach in this type of environment would rapidly become difficult as the number of users grows large. Additionally, personalization functionality is beyond the original design scope of most current network based information systems. Few have either the incentive or resources to support the personalization process [5, 1].

This paper describes an approach for providing personalization support targeted for knowledge workers that must interact with information from diverse and distributed sources. The next section introduces the approach with a brief overview of the architecture upon which it is based. A prototype implementation is briefly described in the next section. The following section discusses ongoing and future research plans within the context of the topic of the workshop.

2 The PADDLE Personalization Architecture

The Personal ADaptable Digital Library Environment (PADDLE) architecture was designed to create a personalization environment for knowledge workers, especially those with diverse information needs. Personalization in this context refers to the ability of a user or group of users to customize or modify information objects in a way that reflects personal preferences, and facilitates their ability to perform a task. As described earlier, the information world of today is an increasingly distributed and heterogeneous one. This often requires knowledge workers to interact with a variety of different systems in order to obtain the information they require. A primary goal of the PADDLE architecture is to support personalization for all of the information objects with which knowledge workers interact, regardless of where the information is stored or by what system it is managed. This goal significantly shaped the architecture and lead to two of the primary characteristics of its approach for supporting personalization: that it is *decentralized* and that it is *metadata based*.

The approach is decentralized in that the information required to represent personalizations for individual users is not centrally stored within information repositories. As described earlier, network based information repositories can have a large if not unlimited user base. A strategy that centralizes personalization functionality at the information repository would be increasingly difficult to realize as the number of users increases. The PADDLE architecture instead uses an approach that captures personalization information locally (with respect to the user) as users interact with information items and then maintains it in a decentralized way.

The PADDLE approach is metadata based in that metadata serves as the mechanism for capturing and maintaining personalizations that are made to information items. In its most basic form, metadata is simply data about data. The most common use of metadata is as a mechanism for describing information resources. For example, the metadata descriptions contained in digital catalogue systems describe information resources in a way that enables users to determine if a particular resource is likely to be relevant for their task at hand. The descriptions need to be general enough to be appropriate for the variety of users of the digital catalogue system. The role of metadata in the PADDLE architecture is a somewhat unconventional one. Instead of being used to describe information resources in a general way, such as the descriptions contained in a digital catalogue, metadata is used at a much finer level of granularity. It serves as the basis for creating *individualized* descriptions (or personalizations) of information items.

An overview of the PADDLE architecture is illustrated in Figure 1. The shaded part of the figure represents a user's local computing environment. Client applications are the tools that are used by knowledge workers to access information. Example client applications include a web browser, a database front end, or any tool used for information access. The information resources illustrated in Figure 1 are the artifacts such as documents, images, etc. that are accessed by knowledge workers. They can be located anywhere on the network. The primary functional component of the architecture is the Customization Metadata Manager (CMDM). As illustrated in Figure 1, the CMDM is positioned between client applications and the information items they access. It is a server process that performs a range of functions in response to client application requests. The most important functionality provided by the CMDM is the creation of metadata to capture personalizations made to information items.

Also shown in Figure 1 is the customization metadata store. This facility provides persistence for personalizations that have been defined for information items. Personalizations stored within the customization metadata store are automatically applied to information items as they are accessed. The personalization metadata store is structured into contexts, which provide a mechanism to partition personalization information according to individual users or user groups. Contexts can be arranged hierarchically, providing a layering mechanism to support multiple levels of customization, such as individual, departmental, or company/organization wide personalizations. Note, the information items themselves are not stored in the customization metadata store, it only contains personalizations.

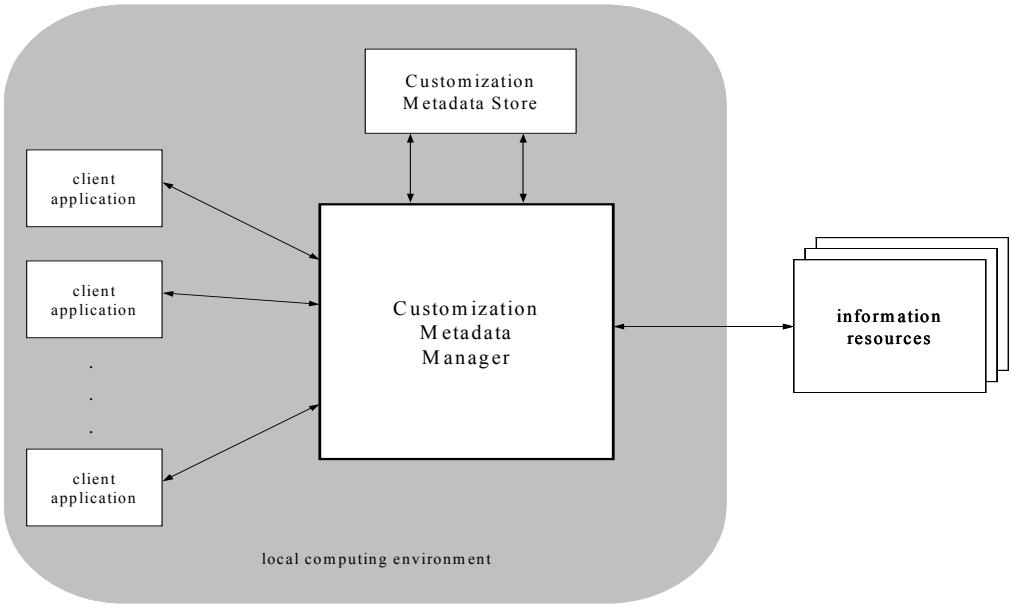

Figure 1. The PADDLE Customization Architecture

An example usage scenario is helpful to demonstrate the interactions between the various components of the architecture. Consider an image browsing tool being used by a knowledge worker to access a remotely located satellite image. The browser tool might be a client application in the environment shown in Figure 1. The system where the satellite image is actually located would correspond to an information resource in Figure 1 that is being accessed remotely. In order to access an image, the browser tool can issue a request for the CMDM to retrieve it. The CMDM would contact the appropriate remote information system to retrieve the image, and then check its customization metadata store to determine if any personalizations have been defined for the image by the current user. If no personalizations have been defined for it, the image would simply be passed along directly to the browser for display to the user. If personalizations have been defined for the image, the CMDM would apply them before passing the image along to the browser. While examining and working with the image, a user might decide to somehow personalize it, such as by adding an annotation, or perhaps changing an existing one. The browser tool could support such personalizations by requesting the CMDM to create customization metadata records to capture it. The records are stored in the customization metadata store and will be automatically applied the next time the image is accessed by this particular user.

3 Prototype Implementation

A prototype environment for supporting personalization has been designed and implemented based on the PADDLE architecture. Currently it supports only a small subset of the kinds of information personalization operations envisioned within the architecture. Specifically, the prototype supports the personalization of information items contained within an environmental database system that describes the state of the environment in Germany [2]. The information within the database is well structured into records consisting of a series of fields (approx. 40). The prototype system provides the user with the capability to personalize information at the field level.

Figure 2 illustrates the personalization component interface of a database browser for the environmental database that has been integrated into the prototypic personalization environment. The browser is a client application within the architecture (Figure 1). Using the interface of the browser, users can specify an alternate value for a field value of a database record, delete one of the fields from a record, or add a new field to a record. This enables them to personalize records from the environmental database to enhance their ability to work with the information. For example, a user could personalize the value of a field to something more meaningful for them in order for the corresponding record to located easier in the future.

In Figure 2, the "Semantic Relationship" field of the displayed database record has been updated for that purpose. Alternatively, a user might wish to organize a subset of the records of the database according to some new dimension. This could be facilitated by creating a new field for the records and assigning appropriate values. In Figure 2, the "Temporal Relationship" field has been added for this purpose.

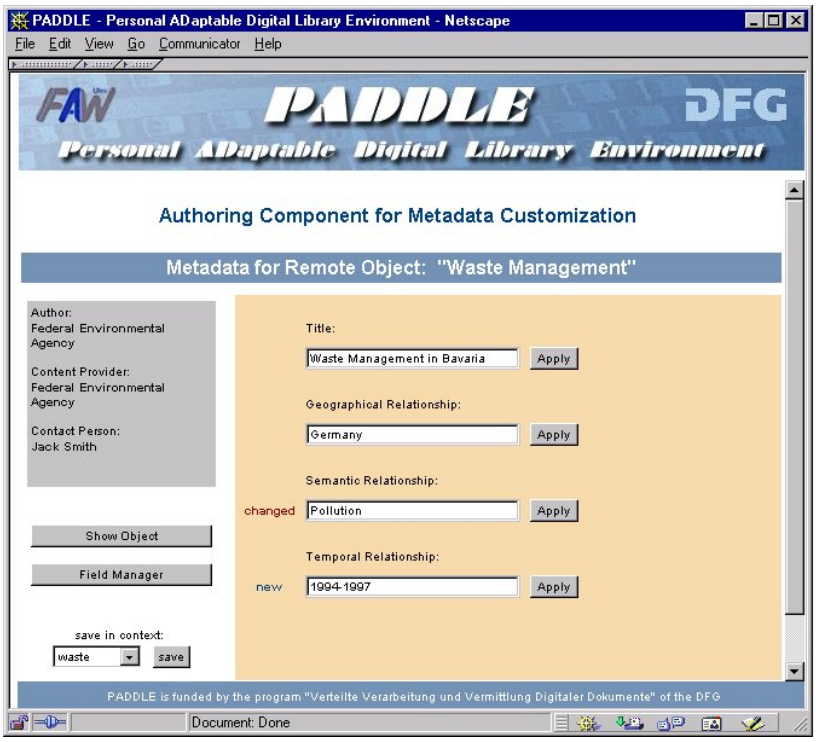

Figure 2. Personalization component of database browser.

An additional important capability that the prototype system provides is the ability to utilize personalization information when performing searches. When performing a search operation for information from the database, users can specify the domain of the search to be the original (non-personalized) information contained in the database, the personalized information, or a combination of both. They can also define their own search forms to accommodate the inclusion of new fields they might have defined for records of the database.

4 Discussion

The prototype system has provided an initial glimpse of the feasibility of the PADDLE approach for supporting personalization. So far, the approach appears to be a valid one. Maintaining personalization information in a decentralized way addresses the scalability problem inherent in the centralized approach. Also, keeping this information local to the user who actually made the personalizations enables it to more easily be used for other important purposes such as searching.

The personalizations supported by the current prototype, however, are quite limited compared to those envisioned within the PADDLE environment. Specifically, the prototype supports the personalization of the well-structured information records from a particular database system. The same approach might be expected to also work for other types of structured information. However, the information artifacts with which knowledge workers interact are not always so well structured. In order to truly gauge the feasibility of the PADDLE approach for a general digital library setting, less structured types of information also need to be considered. Nothing in the PADDLE architecture limits it to structured information types. Enhancing the customization metadata manager to expand the information types it supports is clearly one

of the next steps needed to enable a more thorough and realistic evaluation of the approach for supporting personalization.

An important area to be examined is the investigation of automated personalizations within the PADDLE architecture. Efforts so far have focused on supporting user initiated or explicit personalizations. Implicit personalizations, those made automatically based on usage patterns, user profiles, and other information that is gathered or otherwise known about users, are an important category for supporting knowledge work. It is important to know how well the PADDLE approach for representing and managing personalization information can accommodate implicit personalizations. For example, can the personalization information contained in the customization metadata store (Figure 1) be processed or mined in a way that facilitates the creation of implicit personalizations?

A specific area to be investigated within the implicit category of personalization concerns tacit knowledge, knowledge that cannot be easily articulated because the user may not even be aware of it. For example, the particular information items a user has chosen to personalize, when considered as a whole, might represent a form of tacit knowledge. Tacit knowledge might also be present in the way in which personalizations have been structured, e.g., organized into contexts within the customization metadata store (Figure 1) [7]. It might be possible to extract this kind of knowledge by processing personalization information and then utilize it to make useful suggestions to the user.

References

[1] Hicks, D. L. and Tochtermann, K. 1998. "Environmental Digital Library Systems". In Proceedings of the First GI (German Computer Science Society) Workshop on the Use of Hypermedia in Environmental Applications, Ulm, Germany, May 15-16, 1998.

[2] Hicks, D.L., Tochtermann, K., Eich, S., and Rose, T. 1999. Using Metadata to Support Customization. In Proceedings of the Third IEEE Metadata Conference, Bethesda, Maryland (USA).

[3] Marshall, C. 1997. "Annotation: from paper books to the digital library". In Proceedings of ACM Digital Libraries'97, Philadelphia, Pennsylvania, USA, July 23-26, 1997. pp. 131-140.

[4] Nuernberg, P. J., Furuta, R., Leggett, J. J., Marshall, C. C., and Shipman, F. M. 1995. "Digital Libraries: Issues and Architectures". In Proceedings of ACM Digital Libraries '95, Austin, Texas, USA, June 11-13, 1995. pp. 147-153.

[5] Phelps, T. and Wilensky, R. 1997. "Multivalent Annotations". In "Research and Advanced Technology for Digital Libraries", in Proceedings of the First European Conference on Digital Libraries, ECDL'97, Pisa, Italy, September 1997, Springer, pp. 287-303.

[6] Roescheisen, M., Winograd, T., and Paepcke, A. 1995. "Content ratings and other third-party value-added information – defining an enabling platform". D-Lib Magazine, August 1995.

[7] Stenmark, D. 1999. Using Intranet Agents to Capture Tacit Knowledge. Proceedings of WebNet 1999 – World Conference on the WWW and Internet, Hawaii, 1999.

[8] Van House, N. A. 1995. "User Needs Assessment and Evaluation for the UC Berkeley Electronic Environmental Library Project: a Preliminary Report". In Proceedings of ACM Digital Libraries'95, Austin, Texas, USA, June 11-13, 1995. pp. 71-76.