Personalizing Information Spaces: A Metadata Based Approach[Ψ]

David L. Hicks
Department of Computer Science
Aalborg University Esbjerg
Niels Bohrs Vej 8
6700 Esbjerg, Denmark
Email: hicks@cs.aue.auc.dk

Klaus Tochtermann
Know-Center
Inffeldgasse 16c
8010 Graz
Austria
Email: ktochter@know-center.at

Abstract

The amount and variety of information available today in digital form continue to increase at a rapid rate. To keep pace with these trends, knowledge workers must be empowered with enhanced information management capabilities in order to continue to work effectively. Enabling knowledge workers to personalize the information items with which they work is an important capability, one that can facilitate the ability to perform complex tasks. This paper examines a metadata based approach to supporting the personalization process for knowledge workers, especially those that must interact with diverse and distributed information objects. An architecture for supporting the personalization process is described along with a prototype personalization environment that is based upon it. Two of the primary characteristics of the approach are that it is metadata based and decentralized. The advantages of the approach are discussed, along with an examination of the challenges that it presents for ongoing and future research.

Keywords: Personalization, Metadata, Customization, Information Systems

1. Introduction

Both the quantity and variety of information available in digital form today are increasing at a staggering pace. The world of information is becoming an increasingly diverse and distributed one. 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 [16]. 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 [17]. 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. Nürnberg, et. al. have suggested a 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

[Ψ]This paper is an expanded version of a workshop paper that appeared in [5].

https://doi.org/10.23106/dcmi.952106610

more effectively [8]. 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 [6]. Roescheisen reports that the process of a user personalizing an information space adds value to it [12].

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 instance, 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 develop the technology necessary to support the personalization process [10, 3].

This paper describes an approach for providing personalization support for information spaces targeted for knowledge workers that must interact with information from diverse and distributed sources. The two primary characteristics of the approach are that it is decentralized and that it is metadata based. The next section introduces the approach with a brief overview of the architecture upon which it is based. A prototype implementation that is based on this architecture is described in Section 3. The following section discusses the use of metadata in the personalization architecture along with issues that it generates that will be examined in ongoing and future research. Section 5 of the paper examines related research and is followed by a conclusion.

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 and distributed information needs. Personalization in this context refers to the ability of a user or a 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 has 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 difficult to realize. The PADDLE architecture instead uses an approach that captures personalization information locally as users interact with and personalize 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. When metadata is used this way, the descriptions encoded within it need to be general enough to be appropriate for the variety of users that will consult the digital catalogue system [4].

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

https://doi.org/10.23106/dcmi.952106610

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. These individualized descriptions can be used for a variety of purposes. Examples include tracking a personalization that a user would like to make to an information item or recording user observations concerning an information item, such as how relevant it was for a specific task being performed.

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 shown in Figure 1 represent the information objects used by knowledge workers as they perform their tasks. Examples include documents, images, database records, or any information object with which knowledge workers interact.

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 and management of metadata to capture personalizations made to information items.

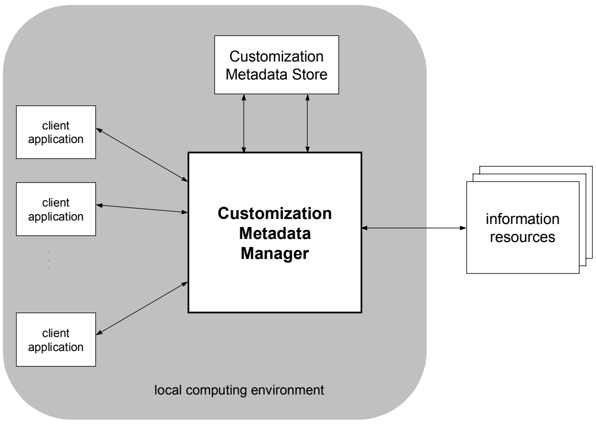

Figure 1. The PADDLE Customization Architecture

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 by client applications within the personalization environment. Note, the information items themselves are not stored in the customization metadata store, it only contains personalizations. The actual information items continue to reside in the information system where they were originally located.

The customization metadata store is structured into contexts, which are collections of related personalizations. Contexts provide a mechanism to partition the customization metadata store according to individual users or user groups. Each user can define personalizations within their own private context, preventing the personalizations made by one user from overlapping or interfering with those of another. When necessary, a user can define more than one context in order to organize their personalizations according to the multiple tasks they are working on, or some other criterion. It is also possible for contexts to be shared by a group of users, to support collaborative activities.

As illustrated in Figure 2, contexts can be related to one another hierarchically, providing a layering mechanism for personalizations [11]. This enables separately defined but related changes to a set of objects to be combined. When arranged this way, multiple levels or scopes of customization can be supported. For example, an organization may wish to define a corporate wide context that contains a set of customizations for information items that should be seen by all users within the organization. A particular department within the organization might wish to extend it with a set of customizations appropriate for members of the department. These could be organized into a departmental context. Finally, an individual member of the department might wish to further personalize information items through the creation of a private context. These contexts could be related hierarchically so that when a user accesses an information item, such as Object B in Figure 2, any personalizations defined for it in the corporate context are first applied, then any defined within the departmental context are applied, and finally those from the individual context are applied. Note that when Object A is accessed in the scenario presented in Figure 2, the corporate view of the object would be presented to the user since no departmental or individual level personalizations have been defined for it.

https://doi.org/10.23106/dcmi.952106610

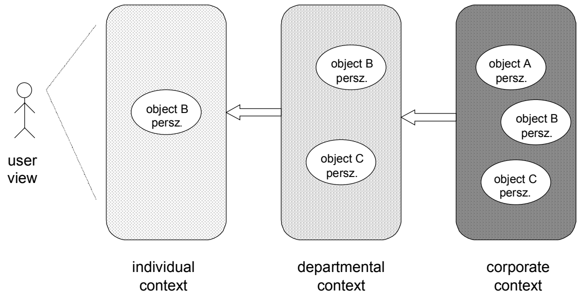

Figure 2. Personalization contexts related hierarchically.

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. If integrated into the PADDLE environment, the browser tool could be a client application in the arrangement 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 the image. 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 within the current context. 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 within this particular context.

3. Prototype Implementation

A prototype personalization environment has been constructed based on the PADDLE architecture. The two main elements of the architecture, the customization metadata manager and the customization metadata store, have been implemented as individual software components. The CMDM has been implemented in Java and is based upon the Netscape Fasttrack Web server. A Microsoft Access database currently provides the functionality of the customization metadata store. The software components communicate using a range of standard protocols. Communication between the CMDM and the metadata store takes place using Java RMI, to facilitate distribution. Communication between the CMDM and external or remote information systems is flexibly defined using abstract Java classes so that a range of different protocols can be accommodated. Further details of the base implementation can be found in [14].

A client application has been implemented and integrated into the personalization environment that enables users to access information objects from remote sources. The client application interacts with the CMDM to enable users to view information objects as well as perform customizations on those objects. Currently three different information systems have been integrated into the prototype environment, each of which contains information from professional content providers. The first information system contains a collection of over 2,000 Microsoft Office documents, the second one consists of over 100,000 HTML documents, and the third one is the electronic theses archive of a German University. Each of these information systems provides metadata descriptions of the resources they contain. The prototype environment currently supports the personalization of these metadata descriptions by users of the client application. The types of personalizations permitted on the metadata descriptions include: the ability to change the value of a metadata field, the ability to hide or delete a metadata field, and the ability to define a new metadata field and specify a value for it.

Figure 3 illustrates the personalization component of the client application that enables the user to personalize metadata descriptions of information resources. As an example personalization, a user might decide to change the value of a field to something more meaningful for them in order for the corresponding record to be located easier in the future. In Figure 3, the "Semantic Relationship" field of the displayed resource description 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 done by creating a new field for the resource descriptions and assigning appropriate values. In Figure 3, the "Temporal Relationship" field has been added for this purpose. Using the new field the information items examined can be classified by the user according to the time period to which they pertain.

https://doi.org/10.23106/dcmi.952106610

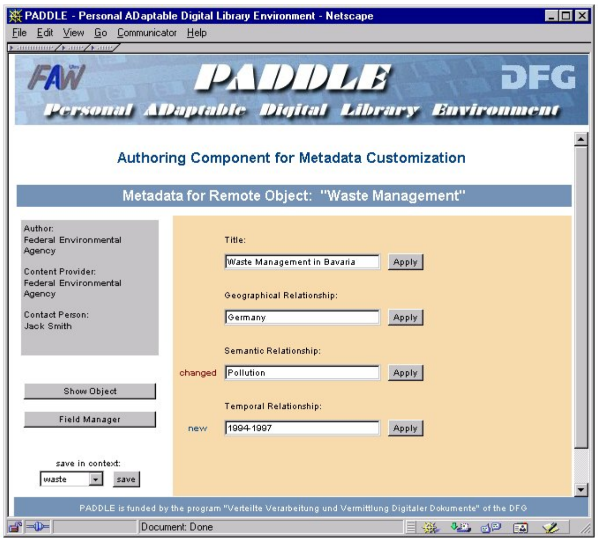

Figure 3. Personalization Component of Client Application.

In order to help users organize and keep track of the documents with which they work, the prototype system provides an information space profile mechanism. A profile provides a way to logically group together and organize a set of distributed but related information items. It also provides a convenient way to access these information items. For example, a user might create a profile to help keep track of the information items needed to perform a specific task. The items could be organized within the profile according to their topic area, how relevant they were for the task, or whatever criteria are appropriate to the task at hand.

Figure 4 illustrates the profile explorer. The profile explorer is a client application that enables users to graphically create, maintain, and interact with profiles. The folder icons in the screen display correspond to profiles that have been created. The items listed hierarchically below folder icons represent the information items contained in the profile. The user can graphically manipulate these items while organizing them into profiles. The information items themselves can be accessed by the corresponding icon.

4. Discussion

As described earlier, three information systems have been integrated into the prototype environment. Each of the systems provides metadata based digital catalogue type descriptions of the information objects they contain. The prototype system currently supports the personalization of these information object descriptions. Initial experiences with the prototype indicate that the ability to personalize these descriptions is a quite useful one, reflecting how frequently the descriptions are consulted by users. When working this way, of course, users are not actually personalizing the information objects themselves; they are instead creating personalizations of the metadata descriptions of the objects. The metadata that is created and stored in the personalization metadata store (Figure 1) to track and maintain these personalizations is essentially metadata about metadata. This has the interesting effect of elevating in status the digital catalogue descriptions provided by these information systems. The entries these catalogues contain start to become information objects in their own right as they themselves begin to acquire their own metadata descriptions as personalizations are defined for them.

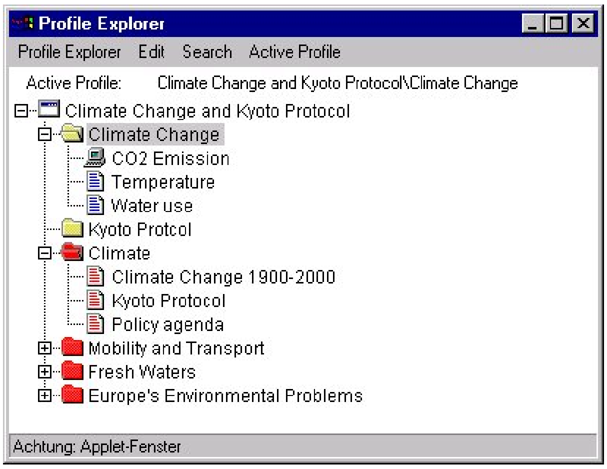

Figure 4. The Profile Explorer Tool.

An interesting and important question concerns how well the metadata approach presented here can be extended to support the personalization of additional types of information objects. Currently the prototype supports the personalization of the well structured digital catalogue entries provided by the information systems that have been integrated into the environment. 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. For example, consider the actual information items such as documents, images, graphs, data sets, spreadsheets, etc. that are typically provided by information systems. Often these items are not as well structured as the entries of the digital catalogues that describe them. In other cases, the items may be structured but little may be known

https://doi.org/10.23106/dcmi.952106610

about the internal details of the structure. Either way, personalization support is also important for these items. Nothing in the PADDLE architecture limits it to supporting only structured information types, but clearly strategies will be required in order to accommodate less structured types of information or information objects for which the details of the structure are not known.

The decentralized aspect of the PADDLE approach for supporting personalizations results in the metadata that is created to track personalization operations being stored locally (with respect to the user). It is contained in the personalization metadata store (Figure 1). Personalization metadata can be viewed as an information resource in its own right. Storing this information locally offers the best chance for it to be exploited in support of the user. For example, it can be used to enhance a user's ability to search for information items, especially those objects that have been accessed previously and for which personalizations have been defined [15]. In [14] a description is provided of how the prototype system has been extended to support this enhanced searching capability. Additionally, since personalization metadata by definition is used to record individual user preferences, encoded within it is a substantial amount of information about users. With the proper data mining techniques, it should be possible to automatically extract useful data about the user that can serve as input to agents and other types of automatic recommender systems [5, 13].

There are potential disadvantages, however, with a decentralized approach for supporting personalization. One of the more difficult ones is the maintenance of consistency between information objects and the personalizations that have been defined for them. When an information resource stored in a remote information system is changed or deleted, inconsistencies can arise between the resource and the metadata created to track personalizations that have been defined for it. So far this has not been a major concern for the prototype system. The three information systems that have been integrated each contain relatively static data. New information items are occasionally added, but existing ones are very rarely changed, so inconsistencies do not often arise. As systems containing more dynamic information are integrated into the prototype, however, consistency preserving strategies will be required. A notification strategy might be possible where information systems send notices to registered individuals when particular information objects are updated. Alternatively an automated consistency checking process could run in a background mode searching for potential inconsistencies and notifying the user when they are found.

Another important consideration when working with metadata, especially when multiple client applications and information systems are involved, is the definition of standards. This is necessary in order for applications to interact, interoperate, and coexist with each other. For the research described here, this issue primarily concerns the standardization of personalization metadata. Standards such as the Dublin Core [1] have been defined for regularizing the descriptions of information items such as those contained in a digital catalogue system. Is the definition of a similar standard possible for the metadata records created to track personalization information? If so, it could significantly ease the interchange of this information between different personalization systems, and also facilitate interoperability between them. Several factors will likely influence the development of such a standard such as the type of the information object being personalized and the particular client application that is used to access and define personalizations for it. As mentioned earlier, the prototype system currently supports personalizations for only well structured information objects – digital catalogue descriptions – so standardization of the personalization metadata has not yet become a major issue. Once support for less structured information types is integrated into the environment, however, this will become a significant issue.

5. Related Work

Metadata has been used in some existing systems to support various forms of information customization. For example, the HyperWave system provides a predefined set of attributes that, on an individual object basis, can be used to maintain customization or descriptive metadata values for data objects [7]. The research described here differs from these systems in the intended set of personalizations supported. The PADDLE system is intended to support a variety of client applications interacting with a range of information systems. It allows client applications to define any required amount of metadata fields to support whatever personalization operations necessary. Predefining the allowed set of metadata fields permitted would reduce the range of personalizations possible, and also restrict the support it could provide to new client applications.

Support for personalization and customization is starting to be examined in the digital libraries area. The Stanford Digital Library Infobus provides an infrastructure which is similar to the one described here [9]. Much like the CMDM in the PADDLE approach, the Infobus is designed to pull the

https://doi.org/10.23106/dcmi.952106610

components of a distributed information setting together. Digital library services built into the Infobus provide the necessary support functions, including query and translation and metadata facilities. The main difference, however, is that the Infobus models documents stored in the remote information sources as objects while in the PADDLE approach, no abstraction of the original documents exists in the CMDM.

The Patron-Augmented Digital Library project seeks to develop a digital library to support digital scholarship [2]. Four phases have been identified in the digital scholarship process: acquiring, structuring, authoring, and then publishing information. The Synchrony prototype digital library system has been developed to support each of these phases. While in the PADDLE approach to supporting personalization, the type of customizations allowed is essentially open ended, in Synchrony, emphasis is placed on supporting a subset of customizations identified as especially relevant to digital scholarship, such as creating annotations.

6. Conclusion

This paper has described an architecture for supporting the information personalization process. The architecture is based upon an unconventional use of metadata in which it serves as the basis for tracking very fine grained individualized descriptions (or personalizations) of information items. The architecture has been specifically designed to help knowledge workers cope with the increasingly diverse and distributed information world of today in which they must interact with a variety of different sources to get the information they need.

A prototype system has been developed based on the architecture. An initial client application supports interactions with the three information systems that have so far been integrated into the personalization environment. Currently the prototype allows the personalization of information contained in the digital catalogue descriptions that these information systems provide for the information resources they contain.

Initial experiences indicate the approach is a valid one for supporting the personalization of structured information types such as digital catalogue descriptions. The architecture has the advantage of maintaining personalization information (metadata) locally where it can most effectively be exploited to assist the user in tasks such as searching for information. Additionally, this information can serve to inform agents and other automated systems of individual user preferences.

Future research will focus on expanding support for personalization within the architecture and prototype to include less structured types of information. Another important area to be examined is the definition of a standard for personalization metadata.

References

[1] Dublin Core Metadata Initiative, URL http://dublincore.org

[2] Goh, D. and Leggett, J. 2000. Patron Augmented Digital Libraries. In Proceedings of the Fifth ACM Conference on Digital Libraries. June 2-7, 2000, San Antonio, TX, USA. pp. 153-163.

[3] 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.

[4] 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).

[5] Hicks, D. and Tochtermann, K. Towards Support for Personalization in Distributed Digital Library Settings. In Proceedings of the Joint DELOS/NSF Workshop on Personalisation and Recommender Systems in Digital Libraries, Dublin, Ireland, June 18-20, 2001.

[6] 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.

[7] Mauer, H. 1996. HyperWave - The Next Generation Web Solution. Addison Wesley, http://hyperg.iicm.tugraz.ac.at/hgbook.

[8] 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.

[9] Paepcke, A., Wang Baldonado, M. Q., Chang, C. C. K., Cousins, S., and Garcia-Molina, H. 1999. Using Distributed Objects to build the Stanford Digital Libraries Infobus. IEEE Computer, Vol. 32, No. 2, pp. 80-87.

[10] 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.

[11] Prevelakis, V. 1990. Versioning issues for hypertext systems. In Object Management, D. Tsichritzis, Ed. Centre Universitaire d'Informatique, Universite de Geneva, Switzerland, pp. 89-105.

[12] 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.

[13] Stenmark, D. 1999. Using Intranet Agents to Capture Tacit Knowledge. Proceedings of WebNet 1999 –

https://doi.org/10.23106/dcmi.952106610

World Conference on the WWW and Internet, Hawaii, 1999.
[14] Tochtermann, K., Hicks, D., and Kussmaul, A. 1999. Support for Customization and Personalization on the Web. Proceedings of WebNet 1999 - World Conference on the WWW and Internet, Hawaii, USA, 1999.
[15] Tochtermann, K., Hicks, D. and Kussmaul, A. Creating corporate knowledge with the PADDLE system. 2000. In Proceedings of WebNet 2000 - World Conference on the WWW and Internet, San Antonio, Texas, USA, 2000.
[16] Tochtermann, K. and Maurer, H. Knowledge Management and Environmental Informatics. Journal of Universal Computer Science, Vol. 6, No. 5 (2000), pp. 517-536.
[17] 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.

Acknowledgements

This research was conducted as part of the PADDLE project (Personal ADaptable Digital Library Environment) funded by the program "Verteilte Verarbeitung und Vermittlung Digitaler Dokumente" of the Deutsche Forschungsgemeinschaft (DFG).

https://doi.org/10.23106/dcmi.952106610