Journal of Universal Computer Science, vol. 7, no. 7 (2001), 550-565
submitted: 1/5/01, accepted: 7/6/01, appeared: 28/7/01 © Springer Pub. Co.

Personal Digital Libraries and Knowledge Management

David Hicks
(Aalborg University Esbjerg, Esbjerg, Denmark
hicks@cs.aue.auc.dk)

Klaus Tochtermann
(Know-Center, Austria
ktochter@know-center.at)

Abstract: The efficient management of knowledge has become imperative for almost all types of organizations. Many approaches exist for dealing with knowledge management at a corporate level. But there is also a need to support knowledge management also at an individual level, a level which takes the specific needs, experiences and skills of knowledge workers into account. While largely unexplored within the field of knowledge management, in the field of digital libraries advanced personalization and customization concepts exist. Within this context, this paper examines these concepts and how they can be exploited to address the challenges which are typical for knowledge management. As the paper will show, many synergies exist, if knowledge management at an individual level is dealt with in combination with personal digital libraries.

Keywords: Personalization, Information Systems, Knowledge Management, Digital Libraries

Categories: H.1, H.2, H.3, H.4, J.3

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. Advances in storage technologies enable enormous repositories containing a variety of information to be compiled and maintained online. Evolving network infrastructures make it possible for large information repositories to be queried and accessed from virtually anywhere. These trends in the availability and management of information have important implications for both individual knowledge workers and the organizations within which they work.

At the *individual level*, these trends have produced an environment in which 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 the information with which they work is an important capability, one that can facilitate their ability to perform complex tasks [Van House et al. 1995]. 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 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 [Nürnberg et al. 1997]. 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 [Marshall 1997]. Roescheisen reports that the process of a user personalizing an information space adds value to it [Roescheisen et al. 1995]. Though it is an important capability for enabling knowledge workers to work more efficiently and effectively with information, personalization can be a difficult capability to support, especially in today's increasingly diverse and distributed information environment.

At the *organization level*, the recent trends noted above in the information landscape of today have produced an environment in which knowledge management has become a critically important capability for many types of corporate organizations. The primary objective of knowledge management is to leverage the knowledge that resides within an organization to achieve the organization's goals more efficiently and cost effectively. The formalized knowledge or organizational memory within a company is the key to maximizing the return on the intellectual assets and investments of an organization [Tochtermann 2000]. This is reflected in the fact that many companies now include the category of intellectual assets on their balance sheets, in addition to more traditional material assets such as labor and capital, when measuring the value of the organization. For many companies the percentage of overall worth represented by intellectual assets is already substantial, and continues to rise [Murray et al. 1999].

The knowledge management process within an organization involves more than just the initial creation of a corporate or organizational memory [Skyrme 1997]. To more completely support the information needs of organizations, a more integrated view is needed, one that includes support for: the retrieval of information from the organizational memory so users can more easily locate relevant information items, the visualization of that information to assist users in contextualizing and visualizing complex information structures, and knowledge transfer techniques that facilitate and promote the usage of corporate knowledge bases for learning and educational purposes to help employees stay current in today's rapid pace of innovation. Only through a more integrated approach can organizations derive the maximum potential benefit from their investment in establishing an organizational memory [Tochtermann 2000].

Though important and necessary in today's business climate, the knowledge management process presents many difficulties. An intrinsic problem of knowledge management is the complex nature of the interrelated knowledge contained in corporate memories. In addition, once established and adopted by the employees of an organization, organizational memories tend to grow rapidly with large amounts of new knowledge being added on a continual basis. This can become overwhelming to the users of the knowledge base as its size grows larger, and can also reduce the quality of the knowledge base if knowledge objects are not screened or evaluated before being added to the knowledge base.

This paper examines the relationship between the support for personalization of information by knowledge workers and the knowledge management process within organizations. Specifically, it considers the beneficial interactions and synergies that are possible between supporting personalization in a personal digital library setting and the development and use of a corporate knowledge base. In Section 2 a strategy for supporting the personalization process in digital libraries is presented. The next section examines how this strategy can be built upon and extended to offer support for the development, use, and extension of a corporate knowledge base. Section 4 examines issues surrounding the maintenance of quality within the knowledge base. The related literature is considered in Section 5. The paper is concluded in Section 6.

2 A Personal Digital Library Environment

As described in the previous section, the characteristics of today's information landscape present a challenging environment in which to support the personalization process. The diverse and distributed nature of the information with which knowledge workers must interact pose particular difficulties. The straightforward approach for supporting personalization in which the system that owns or manages an information object supports the personalization of the object is not necessarily a feasible one. For instance, in the example from the previous section, a user created an annotation for a satellite image being used to prepare a report. One strategy to support this capability would be for it to be provided by the system that manages the satellite image. This type of approach might be possible 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 very 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. An additional problem is that 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 [Phelps et al. 1997; Hicks et al. 1998].

This section presents a strategy for supporting the personalization of diverse and distributed information objects. It begins with a description of a personal digital library architecture intended to support knowledge workers. A prototype system based upon the architecture is then briefly described.

2.1 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 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 use of metadata to support personalization was motivated by one of its primary characteristics: that it can exist and be maintained completely independent of the data to which it refers [Tochtermann 1997]. For example, consider a digital catalog system. The information described by the metadata contained in the catalogue might exist locally (with respect to the catalogue), it could be managed by a remote system, or indeed it might not even be available in digital form. In any case, it is possible for the metadata descriptions contained in the catalogue to be defined and maintained separately from the information they refer to. The approach described here exploits this characteristic of metadata to enable the customization of data objects in a way that places no restrictions on where the objects are stored or by which system they are managed. Write or update access to the data objects being customized is not assumed or required [Hicks et al. 1999].

The most common use of metadata is as a mechanism for describing information resources in a general and broadly applicable way. 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 instead 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.

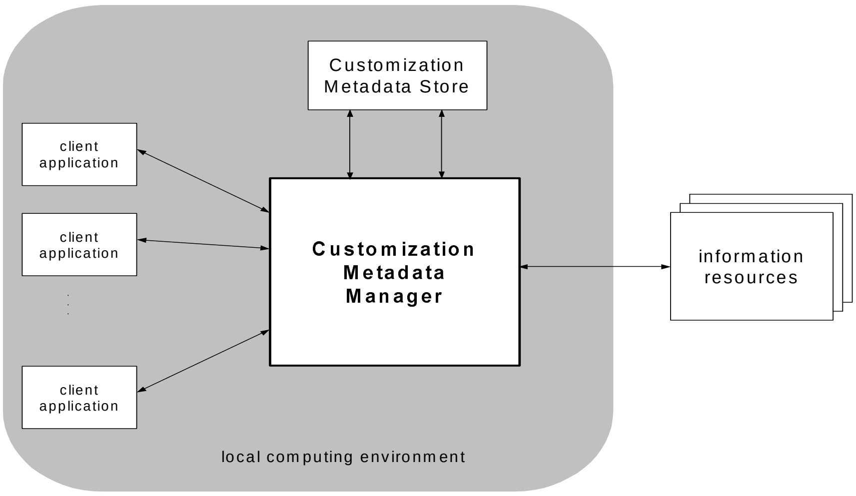

Figure 1: Overview of the PADDLE Customization Architecture

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. Note, the information items themselves are not stored in the customization metadata store, it only contains personalizations. 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 criteria. It is also possible for contexts to be shared by a group of users, to support collaborative activities.

Contexts can be arranged hierarchically, providing a layering mechanism for personalizations. 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, 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.

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 provider 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 them. 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.

2.2 A Prototype Implementation

A prototype personal digital library 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 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 [Tochtermann et al. 1999].

A client application has been implemented and integrated into the 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 one contains a collection of over 2,000 Mircosoft Office documents, the second one consists of over 100,000 HTML documents, and the third one is the electronic thesis archive of a University. Each of these information systems provide 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.

As an example, consider a user working with an HTML document representing an environmental report from one of the information systems. In the prototype environment, the user would have the ability to personalize the metadata description for the report by, for instance, changing a generic value which has been specified for the subject field from “environmental report” to something more specific, such as “coastal wetlands report”. This more precise value might reflect more specific or detailed knowledge the user posses concerning this specific information resource, and personalizing it will facilitate working with the resource. Alternatively, the user might wish to define a new metadata field for the descriptions of the documents in the information system to enable them to be classified according to some new criteria, such as their relevance for a particular task being performed. Finally, the user might wish to hide or delete one or more of the fields in the document descriptions if they are not relevant for the task at hand. This would enable them to focus and work more effectively with those fields that are relevant.

Note that all of these personalizations would be performed within the particular context specified by the user. This prevents the personalizations made by one user from interfering with those of another. Of course, if the current context is a group context and not an individual one, each of the group members would see the effects of the customizations being performed.

In addition to supporting personalizations, the prototype system also enables the definition of personalized search forms. Users can design search forms that allow them to include personalized fields when searching for information. This enables them to exploit personalization information when performing a search. For instance, the user described in the previous example could search for resources by specifying a value for one of the fields that has been customized, such as “coastal wetlands report” for the subject field. Alternatively, a value can be specified for one of the newly created fields.

In order to help users organize and keep track of the documents with which they work, the prototype system provides the working space mechanism. The working space provides a way to group together and organize a set of related information items. For example, a user could create a working space to help keep track of the information items needed to perform a specific task. The items could be organized within the workspace according to topic area, or whatever criteria is relevant to the task at hand.

3 Integrating Knowledge Management into Personal Digital Libraries

The PADDLE architecture as well as the prototype implementation were both originally designed to establish a personal digital library environment for knowledge workers. The capabilities it provides, however, can also be used to support various aspects of the knowledge management process. This section examines ways in which the functionality provided within the architecture has been built upon and extended to support a specific facet of knowledge management, the creation, use, and extension of a corporate knowledge base by knowledge workers.

To build up a corporate knowledge base a shift in the paradigm of using Internet information systems must take place. Today, knowledge workers are primarily "passive" users of information systems, that is, they access, download, and read information resources, but they do not add new ones or tailor existing resources according to their own or to a group of user's needs. As noted in the previous section, most current Internet information systems do not support this functionality. However, relevant studies in knowledge and information management have revealed these capabilities to be important. They have shown that value-added services for knowledge management should include the support of specialized knowledge spaces that serve the needs of specific knowledge workers. For example, knowledge spaces can be used to provide the capability to extend corporate knowledge [Schatz et al. 1999]. While adapting resources is a capability that is already supported by the PADDLE architecture and prototype, extensions are required to develop a component for adding new resources to create a corporate knowledge base.

A corporate knowledge base can be built up in an informal and unstructured way, i.e., by capturing information in different layouts and formats and recording all of the practices of an organization. Though such a free form approach would be inexpensive to implement, it can also generate a lot of irrelevant and unstructured information. To maintain the usability of the corporate knowledge base, the need will soon arise to perform a number of normalization operations such as filtering the irrelevant from the relevant information resources, unifying the layout according to the thematic areas to which an information resource belongs, structuring the information resources, etc. [Buckingham Shum 1997]. In the PADDLE system, a different approach is taken. The idea in PADDLE is to provide an environment which supports a systematic and structured way of building a corporate knowledge base. Even though this approach involves more time for the initial development of a knowledge base, it will soon pay off as no re-organization or optimization of the corporate knowledge will be required later.

The approach for building corporate knowledge with PADDLE involves three activities. Firstly, an organization must determine the different types of knowledge objects they want to make available in their corporate knowledge base. The types of knowledge objects may include reports, product descriptions, meeting minutes, project reports, work practices, etc. Secondly, an organization has to categorize the user groups that will be working with the knowledge base. This is of particular importance as different knowledge workers need different views of the knowledge base. In this context the personalization and customization concepts of the PADDLE approach play an important role. Thirdly, for each type of knowledge and each group of users, templates are required which support the knowledge workers in preparing in a coherent way the knowledge they want to add to the knowledge base.

As mentioned earlier, three remote data systems containing information from professional content providers have been integrated into the current PADDLE prototype environment. Introducing the possibility to add new knowledge objects to the system prompts the question concerning where to store this new knowledge. One approach is to store new knowledge objects along with existing resources. The following drawbacks dissuaded us from pursuing this strategy. Firstly, storing new knowledge along with existing resources requires write access to remote data sources. This is quite counter to one of the fundamental PADDLE philosophies that states remote data systems should be treated as a black box [Tochtermann et al. 1999]. Secondly, an update of a remote data system by the content providers could cause many problems in keeping the knowledge base consistent. For example, an update may change or even delete the context of knowledge objects added by the knowledge workers. The resulting situation might be one of chaos where dangling knowledge objects have to be assigned to new contexts. Therefore, we have chosen a different approach: all new knowledge objects are stored in a separate database which is completely under control of the PADDLE system and completely independent from the other remote data systems.

In order to allow users to logically put their knowledge objects into the context of existing resources we provide the concept of profiles. Profiles build upon the workspace concept described earlier. A profile is a structured collection of resources and knowledge objects which have something in common. For example, objects in a profile may address the same topic or they may be the ones needed to perform a specific task. The PADDLE system distinguishes between two types of profiles: public and private. Public profiles can be accessed and used by all knowledge workers of an organization. However, write access to these profiles is only granted to authorized persons. Individual knowledge workers can create and maintain private profiles which are not accessible to the others within the organization. Private profiles allow knowledge workers to create a personal knowledge space in which they can compile resources from remote data systems, knowledge objects from an organization's knowledge base, and also knowledge objects from local sources, such as their local filesystem.

The profile mechanism is depicted in Figure 2. On the left side of the figure a screen dump from the profile explorer is displayed. The profile explorer is a graphical tool for defining and working with profiles. In this case, a number of profiles can be seen in the screen display. In the upper part of the screen the user's private profiles (e.g., "Climate Change" and "Kyoto Protocol") are shown, indicated by a light gray icon. The public profiles (e.g., "Climate", "Mobility and Transport", etc.) are shown below the private ones, and are indicated with a darker gray icon. Unlike public profiles, private profiles can contain knowledge objects from local sources, such as the local file system. For example, the private profile "Climate Change" contains the "CO2 Emission" information object. Note that a different icon is used to distinguish those knowledge objects that come from local sources.

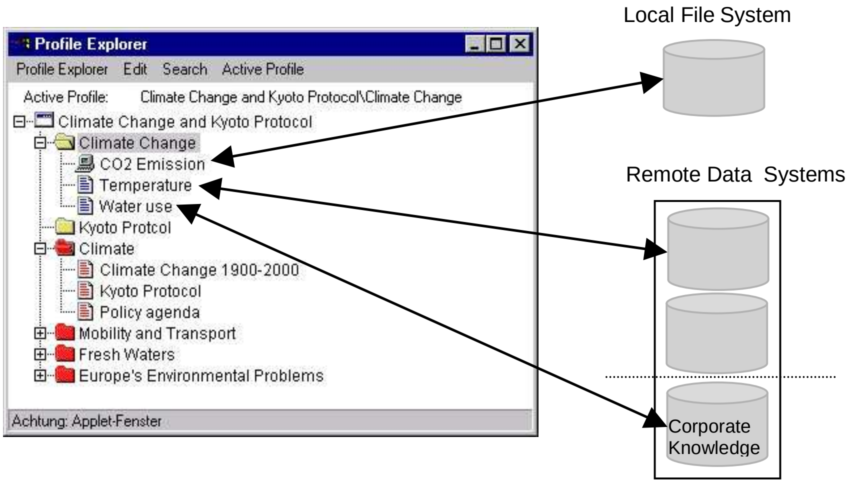

Figure 2: Public and private profiles

All customization features provided by the PADDLE system as described earlier can be applied to the knowledge objects in profiles. Additionally, the PADDLE concept of working space has been extended to offer a very flexible environment for personalizations. Each customization of knowledge objects in profiles is valid in the context of only one working space. This makes it possible to apply different customizations of the same knowledge objects in different working spaces.

Even though the differentiation between private and public profiles match up well with the basic requirements of knowledge workers, we encountered another challenge primarily concerning public profiles. Knowledge objects can sometimes be very rich in content (e.g., a product description). When working with rich content knowledge objects, often only a subset of the complete content of the object is required for a specific task. Sometimes the presence of the additional, irrelevant content can hinder a knowledge worker's ability to perform the task. They become overloaded or distracted by the presence of so much information, especially the irrelevant information they do not need.

One strategy to address this problem is to divide a knowledge object into several smaller ones. However, there are difficulties with this approach. Firstly, it would be difficult to choose the correct granularity into which to divide knowledge objects. Different knowledge workers as well as different groups of knowledge workers have different needs, so choosing a single, correct granularity level would be difficult.

Secondly, when knowledge objects are divided into smaller ones, the overall number of knowledge objects would increase dramatically, and this could create storage concerns. Thirdly, the process of dividing knowledge objects into smaller ones would destroy the knowledge that is inherent in their relationship as subcomponents of the same object. A mechanism would be required to preserve the knowledge that is implicit in their original connectedness.

The solution we have chosen is based on the PADDLE approach to supporting personalization and customization. All knowledge objects (except for those located at remote data systems) are represented in an XML format. Different XSL style sheets are then used to adapt a knowledge object "on the fly" to the specific needs of a knowledge worker or group of knowledge workers. Figure 3 depicts this concept. The PADDLE middleware component provides different XSL style sheets for different groups of knowledge workers. Whenever a knowledge worker or a certain group of knowledge workers accesses a knowledge object, the XSL style sheets provided for this group are used to customize the knowledge object "on the fly" according to specific needs of the group. With this approach we can assure that knowledge workers are always accommodated with those parts of a knowledge object which are most relevant for their task at hand.

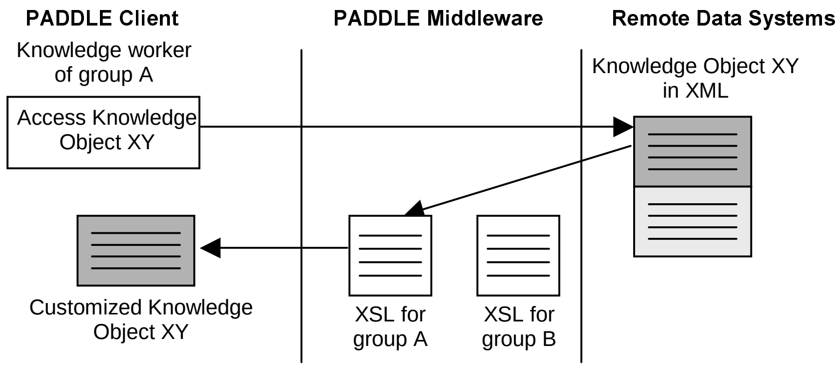

Figure 3: Customization of Knowledge Objects using XML and XSL

Figure 4 illustrates how the XML/XSL strategy for supporting customization works in practice. The two windows shown in the lower part of the figure illustrate the same document being displayed in different ways, according to the personalizations defined by different user groups. There are obvious differences in the formatting of the document such as font, font size, text color, etc. In addition, there are differences in how the content of the document is being displayed. In the display on the left, the metadata fields for the document are being displayed at the top followed by the actual document content. In the display on the right, only the document contents are displayed, accommodating a user group performing a task for which the metadata fields are not relevant.

At the technical level, Java servlets of the PADDLE middleware components are used for performing these customizations. Using pre-defined XSL style sheets, the servlets generate different HTML documents on the basis of existing XML documents.

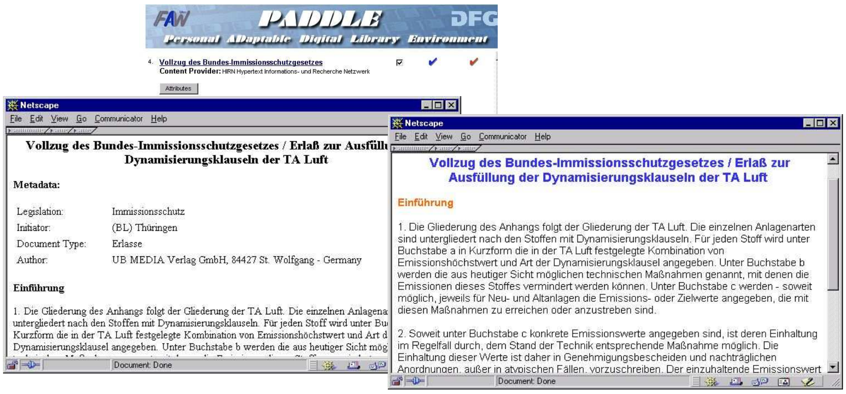

Figure 4: Customized Knowledge Objects

4 Maintaining Quality

The level of quality of the knowledge in a corporate knowledge base directly influences its acceptance and use by members of the organization. In order for people to use the knowledge base and exploit it as a tool for the tasks they perform, they must be confident in the knowledge it contains. To maintain an acceptable quality level, it is imperative that new knowledge go through a quality assurance process before being made widely available.

In the PADDLE approach, quality assurance is part of a linear overall workflow process that serves to edit, review, release, and disseminate knowledge objects to a predefined group of users. The first step of the quality assurance portion of the process ensures that the knowledge objects can be searched for effectively. It is a formal check to make sure that all of the metadata fields of a knowledge object are assigned valid values according to a predefined data type definition. In the second step of the quality assurance process, the plausibility of the values of metadata fields are checked to ensure they are reasonable. While the first part of the quality assurance process can be done automatically, the second part, the check to make sure the values are reasonable, requires the intellectual assessments of experts in the relevant field, and thus cannot be done automatically. This check regarding the quality of the content of the knowledge object provides an opportunity to ensure that resources being added to the knowledge base meet the quality standards of the organization.

When assessing the quality of the content of knowledge objects, two levels of quality certificates are defined: restricted and public. The quality certificate is defined by the user wishing to add knowledge to the corporate knowledge base. The certificate chosen for a knowledge object determines if it has to go through both steps of the quality assurance process, or only the first one. The certificate also determines how widely a knowledge object is made available within the knowledge base. Knowledge objects with a restricted quality certificate only have to pass the first step of the quality assurance process. This ensures that they have well-defined metadata and can thus be searched for effectively. Since no check against the reasonableness of actual metadata values is carried out, such knowledge objects are made available only within restricted areas of the knowledge base to which selected knowledge workers have access.

The public quality certificate requires that a knowledge object go through both steps of the quality assurance process. Once an object with a public quality certificate passes both of these steps, it is made available to all users of the corporate knowledge base. The differentiation between these two types of certificates has proven valuable for the sharing of knowledge objects which exist in a premature or preliminary version only. In some organizations, such knowledge objects are of great importance for certain groups of users (e.g., a strategic planning group) which have to rely on the latest information about new trends, new legislation, etc.

Finally, users can assign priorities to the knowledge objects before they enter the quality assurance process, indicating how urgent it is for the knowledge object to be inspected and released into the knowledge base. Priorities are of particular importance for the second step of the quality assurance process as they provide the experts involved a way to determine the order in which they should assess the quality of the content of a potential new knowledge object. Figure 5 depicts the main steps involved in the quality assurance process.

At the technical level, the quality assurance component capitalizes on the XML and DTD defined for different types of knowledge objects. The formal check (step 1 of the process) verifies if the metadata provided for knowledge objects adheres to the predefined DTD. For example, this includes if all relevant metadata fields have meaningful values. If a knowledge object did not pass one of the checks involved in the process, the reasons are recorded in a report (represented as the shaded rectangle in Figure 5).

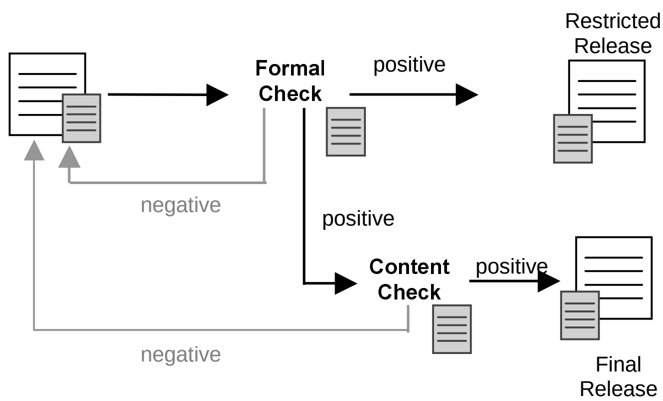

Figure 5: Quality assurance process

5 Related Work

To date, much digital library research has focused on infrastructure issues such as the technologies necessary to create large information repositories and information retrieval and indexing techniques [Goh et al. 2000]. However, support for personalization and customization is a topic that is starting to be considered. The Stanford Digital Library Infobus [Paepcke et al. 1999] provides an infrastructure which is similar to the one described here. Much like the CMDM in the PADDLE approach, the Infobus is designed to pull all components of a distributed digital library together. Library services built into the Infobus provide the necessary support functions, including query 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 CDMD. Also, the Stanford Digital Library supports hierarchical metadata models while at present the PADDLE approach allows only flat models.

The Patron-Augmented Digital Library project seeks to develop a digital library to support digital scholarship [Goh et al. 2000]. 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. The publishing phase of the Synchrony project is intended to provide control over the addition of new materials to the library. This shares many similarities with the quality control mechanisms described here.

6 Conclusion

This paper has presented and integrated research we have conducted in knowledge management and digital libraries during the past four years. The focus has not been on the presentation of new concepts that have not been presented elsewhere. Instead, the emphasis has been on the synthesis of results from these two areas, and the main contribution has been placing them into the same context for the first time. This opens up new perspectives which can provide a starting point for further research in the fields of knowledge management and digital libraries. And indeed, this paper showed that many synergy effects exist between knowledge management and digital libraries. Sophisticated knowledge management tools and digital libraries are both based on many of the same technologies, which facilitates the integration of concepts developed in each research area. The paper also showed that efficient knowledge management requires more than just a consideration of the organizational perspective. A lot of its success depends on how well individual knowledge workers are supported in performing their specific and personal tasks. These tasks often required knowledge that is specially customized and personalized according to a knowledge worker's needs, skills, and experiences. The highest leverage effect can be achieved when a personal digital library system provides the personalized information and a personal knowledge management tool supports the creation, storage, and use of the personal knowledge of a knowledge worker. It is the combination of the two which can lead to a tremendous increase in efficiency for almost all types of knowledge intensive tasks.

References

[Buckingham 1997] Buckingham Shum, S. 1997. Negotiating the Construction and Reconstruction of Organisational Memories. Journal of Universal Computer Science Vol. 3, No. 8, Springer Pub. Co., pp. 899-928. http://www.jucs.org/jucs_3_8/negotiating_the_construction_and

[Goh et al. 2000] 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.

[Hicks et al. 1998] 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.

[Hicks et al. 1999] 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).

[Marshall et al. 1997] 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.

[Murray et al. 1999] Murray, P. and Myers, A. 1999. The Facts about Knowledge. Study of the Cranfield School of Management. http://www.info-strategy.com/knowsurl1/.

[Nürnberg et al. 1995] Nürnberg, 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.

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

[Phelps 1997] 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.

[Roescheisen et al. 1995] 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.

[Schatz et al. 1999] Schatz, B., and Hsinchum, Ch. 1999. Digital Libraries: Technical Advances and Social Impacts; IEEE Computer, Special Issue "Digital Libraries", Vol. 32, No. 2, pp. 45-50.

[Skyrme et al. 1997] Skyrme, D. and Amidon, D.M. 1997. Creating the Knowledge-Based Business, Business Intelligence Ltd. Wimbledon 1997.

[Tochtermann et al. 1997] Tochtermann, K., Riekert, W.F., Wiest, G., Seggelke, J., Mohaupt-Jahr, B. 1997. Using Semantic, Geographical, and Temporal Relationships to Enhance Search and Retrieval in Digital Catalogs; LNCS 1324 Springer, Proceedings of the 1st European Conference on Research and Advanced Technology for Digital Libraries, Pisa, September 1997, pp. 73-86.

[Tochtermann et al. 1999] 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.

[Tochtermann 2000] Tochtermann, K. 2000. Four Dimensions of Knowledge Management. In Proceedings of WebNet 2000 - World Conference on the WWW and Internet (AACE), San Antonio, Texas, US, 2000.

[Van House 1995] 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.