Creating corporate knowledge with the PADDLE system

Klaus Tochtermann[*+] and Andreas Kussmaul[*]
[*]Research Institute for Applied Knowledge Processing (FAW)
PO Box 2060, 89081 Ulm, Germany, E-Mail: tochterm | kussmaul@faw.uni-ulm.de
[+]Center for knowledge-based applications and systems (KNOW-Center)
Graz, Austria, E-Mail: ktochter@iicm.edu

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

Abstract: The integration of dispersed knowledge available in an organization is often referred to as corporate knowledge. The corporate knowledge of an organization can be built up, shared and maintained most efficiently if modern Internet technologies are applied by knowledge workers. As different knowledge workers have different perspectives on corporate knowledge, customization concepts are needed to tailor knowledge objects according to different needs. Additionally, the acceptance of the corporate knowledge depends mainly on the quality of the knowledge stored therein. Against this background this paper presents concepts to create corporate knowledge, to customize it according to different requirements and to ensure a high level of quality. These concepts and their implementation are discussed in the context of the PADDLE system.

1. Introduction

The crucial insight that has made knowledge management important is that the value of an organization does not depend only on the material assets of the organization (like buildings and machinery), but also on the often very dispersed knowledge “inside the heads”. The coherent integration of such dispersed knowledge in an organization is often referred to as corporate knowledge. Corporate knowledge may relate to problem solving expertise, project experiences, human resources management, lessons learned, and design issues etc. Some companies already openly acknowledge the importance of corporate knowledge by producing not just "financial balance sheets" but also "knowledge balance sheets": up to 80% of the value of an organization may be comprised of corporate knowledge rather than material assets (Murray et al. 1999). Once this crucial fact is accepted it is obvious that organizations must make sure that corporate knowledge is nurtured, protected, archived and increased as much as possible. Since this is a very complex process, systematic strategies are required. For example, according to (Skyrme et al. 1997), the development of the intellectual assets of an organization consists of six steps: Business strategy to determine which role knowledge has for strategic decisions in a company; analysis of competitors which includes an analysis of the knowledge the competitors have; knowledge classification to structure the corporate knowledge according to the knowledge of an organization’s employees; knowledge assessment to assess the value of the knowledge in an organization and to assess the costs for maintaining and extending this knowledge; investments to close gaps in an organization’s knowledge; and portfolio management to make transparent which knowledge is used at present, which knowledge will be used in the future and which knowledge is obsolete.

The complete spectrum of knowledge management activities is difficult to automate (Borghoff et al. 1997) but Information technology as an enabling technology can support specific facets of knowledge management. Corporate knowledge is one such specific facet. It can only be built up and maintained if the knowledge workers (i.e. key people responsible for the total organizational knowledge-creation process at the corporate level) make available their knowledge to their colleagues in an organization-wide knowledge base. Such a corporate knowledge base will only be used, if it can be tailored to different requirements and if a certain degree of quality is assured. This paper presents the approach taken in the context of the PADDLE system (Personal ADaptable Digital Library Environment) to support knowledge workers in extending a corporate knowledge base with their own knowledge. Beyond this we also show how knowledge workers can build up and maintain

their personal knowledge space. Finally, our approach also includes a quality assurance component to ensure that only high-quality and reliable knowledge is added to the knowledge base. The remainder of the paper is structured as follows: Section 2 briefly introduces the PADDLE system. Section 3 introduces our approach to build up a corporate knowledge base which can be tailored to personal knowledge spaces. Concepts for the quality assurance component and a prototype implementation are addressed in Section 4. The paper is concluded in Section 5 with a brief look at future research directions.

2. PADDLE at a glance

Originally, the PADDLE prototype system was developed as a digital library environment for knowledge workers. Currently, the system is extended by specific knowledge management components. PADDLE offers its users a broad variety of customization and personalization features for the information resources accessible by the system (Hicks et al. 1999; Tochtermann et al. 1999). The primary characteristic of PADDLE is that the underlying approach for customization and personalization is metadata based (metadata is data about data and well known from digital catalogs). The use of metadata to support customization was motivated by its ability to exist and be maintained completely independent of the data to which it refers. This allows users or user groups to apply their personal adaptations to information resources without affecting the information resource itself.

Currently two different remote data systems are connected to PADDLE. Together they provide over 2,000 Microsoft Office documents and about 100,000 HTML documents. At a glance, the following possibilities exist to adapt the information resources of almost any Internet based remote system.

1. *Personalization of metadata:* Many information resources on the Internet have metadata which further describe the resources. PADDLE allows its users to adapt existing metadata to their personal needs. This includes the ability to hide and rename metadata fields as well as to add new metadata fields with new values (e.g. a new metadata field “geographical relationship“ with the value “sun50.faw.uni-ulm.de“ can be added to describe that a resource is located on the machine named sun50.faw.uni-ulm.de). Also, the values of existing metadata fields can be changed.
2. *Personalization of search forms:* Users can design personal search forms to enable them to include personalized metadata fields and values in the search process.
3. *Personal working spaces:* Users can store their personalizations in personal working spaces. The working spaces can be made accessible not only for single users but also for user groups.

A detailed technical description of the PADDLE system can be found in (Hicks et al. 1999).

3. The PADDLE environment for creating corporate knowledge

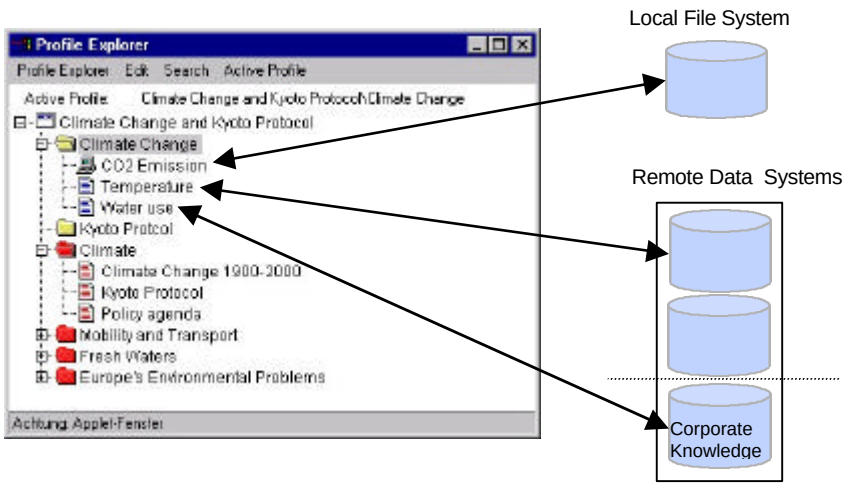
To build up a corporate knowledge base a shift in the paradigm of using Internet information systems must take place. Today, knowledge workers are mostly “passive“ users of a system, that is, they can access, download and read resources but they can hardly add new ones or adapt existing resources to their own or a group’s needs. Relevant studies in knowledge and information management have revealed that value-added services for knowledge management should include specialized knowledge spaces which serve specific needs of specific knowledge workers (e.g., to extend corporate knowledge) (Schatz et al. 1998). While adapting resources is already supported by the PADDLE system, a component for adding new resources to create a corporate knowledge base has not yet been completely developed. A corporate knowledge base can be built up in a totally unstructured way, i.e., by capturing information in different layouts and formats and recording all of the practices of an organization. Even though such an approach will be very inexpensive, it can also generate a lot of irrelevant and unstructured information. Later on the need arises to filter irrelevant from relevant information resources, to unify the layout according to the thematic areas to which an information resource belongs, to structure the information resources, etc. (Buckingham Shum 1997). In PADDLE a different approach is taken. The idea in PADDLE is to provide an environment which supports a systematic and structured way for building a corporate knowledge base. Even though an organization has to spend more time on the development in this environment, it will pay off soon as no re-organization or optimization of the corporate knowledge will be required later on.

The approach for building corporate knowledge with PADDLE covers three aspects. Firstly, an organization

can differentiate between different types of knowledge objects they want to make available in their corporate knowledge base. The types for knowledge objects may include reports, product descriptions, meeting minutes, project reports, work practices etc. Secondly, an organization has to categorize the user groups working with the knowledge base. This is of particular importance as different knowledge workers need different views on the knowledge base. In this context the personalization and customization concepts of PADDLE 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.

In the core system of PADDLE two remote data systems are integrated. Each these systems contains information from professional content providers. With the possibility to add new knowledge objects to the system the question about where to store this knowledge arose. One approach is to put new knowledge objects in context with already existing resources. The following drawbacks restrained us from pursuing this idea: putting new knowledge in context with already existing resources requires write access to the remote data sources - this is very much against the PADDLE philosophy which says that every remote data system should be treated as a black box (Tochtermann et al. 1999). In addition, an update of the remote data systems by the content providers would have caused many problems in keeping the knowledge space 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 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 put their knowledge objects in context with already existing resources we provide the concept of profiles. A profile is a structured collection of resources and knowledge objects which have something in common (e.g., they address the same topic, they are needed to perform a specific task etc.). The PADDLE system distinguishes between public and private profiles. Public profiles can be accessed and used by every knowledge worker of an organization. However, write access to these profiles is only granted to authorized knowledge workers. Knowledge workers can create and maintain their private profiles which are not accessible for other knowledge workers in an organization. Private profiles allow knowledge workers to create their personnel knowledge space in which they can compile not only resources from the remote data system and knowledge objects from an organization's knowledge base but also knowledge objects of any type from their local file system. The following figure depicts this idea. On the left the profile explorer is displayed. It contains all private profiles (e.g., "Climate Change" and "Kyoto Protocol"; light gray in the upper part) of a knowledge worker and all public profiles (e.g., "Climate", "Mobility and Transport" etc.; dark gray in the lower part) of the system. Unlike public profiles, private profile can contain knowledge objects from the local file (e.g., "CO2 Emission" in the private profile "Climate Change"). To easily distinguish between knowledge objects from a



remote data system and those from a local file system different icons are used.

Figure 1: Public and private profiles.

Note, that all customization features provided by the PADDLE system (Tochtermann et al. 1999) can be applied to knowledge objects in profiles. Additionally, the PADDLE concept of working space has been extended to offer a very flexible environment for customizations. 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 very well the basic requirements of the knowledge workers, we encountered another challenge primarily concerning the public profiles. Knowledge objects can be very rich in content (e.g., a product description). Since knowledge workers often only need some parts of a knowledge object to perform their specific task, this bears the risk that knowledge workers get overloaded by too much information and knowledge they do not need. One solution to tackle this problem is to split a knowledge object into several "smaller" knowledge objects according to the different groups of knowledge workers. As a result, the number of knowledge objects would increase dramatically. Also, the togetherness of these "small" knowledge objects has to be provided, which in turn requires further structuring concepts. The solution we have chosen is based on the PADDLE approach to support personalization and customization. All knowledge objects (except for those in the remote data systems) are represented in XML. Different XSL style sheets are used to adapt a knowledge object "on the fly" to the specific needs of a group of knowledge workers. The following figure depicts this idea. The PADDLE middleware component provides different XSL style sheets for different groups of knowledge workers. Whenever a knowledge worker of a certain group of knowledge workers accesses a knowledge object, the XSL style sheets provided for this group customizes the knowledge object "on the fly" according to specific needs of this group. With this approach we can assure that knowledge workers are always accommodated with those parts of a knowledge object which are relevant for their tasks.

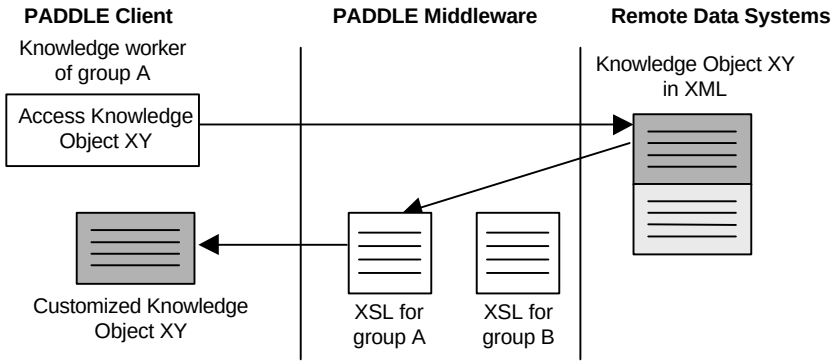


Figure 2: Customization of Knowledge Objects using XML and XSL

Figure 3 shows that different layouts and contents of a knowledge object can be chosen depending on the applied customization. The two windows in the lower part display the same document, however, with different customizations applied to it. Obviously the layout is different. In addition, the customization on the right does not provide the knowledge worker with the metadata which exist for this document (c.f. document on the left).

Figure 3: Customized knowledge objects

Technically, Java servlets of the PADDLE middleware are used for the customization. Using pre-defined XSL style sheets, the servlets generate different HTML documents on the basis of existing XML documents.

4. Quality Assurance

The level of quality of the knowledge in a corporate knowledge base directly influences the level of acceptance to use the corporate knowledge. Therefore, before knowledge can be made available widely, it should undergo a quality assurance process.

In our approach the quality assurance process is part of a linear workflow which serves the purpose to disseminate, review, edit and release knowledge objects to a predefined group of users. The first step of the quality assurance process ensures that the knowledge objects can be searched for effectively. This formal check can be passed if all metadata of a knowledge object are defined according to a pre-defined data type definition. In addition, the plausibility of the defined metadata is checked, that is if the value of a metadata field belongs to a pre-defined data type. While the first step of quality assurance can be done automatically, the second step of quality assurance, the content check, requires intellectual assessments of experts and, thus, cannot be carried out automatically. This check of the quality of the content of a knowledge object ensures that the resources meet the quality standards of an organization.

As to the content, we differentiate between two quality certificates (restricted and public) of the knowledge objects. The quality certificate is defined by the users who want to add knowledge objects to the corporate knowledge base. The certificate determines if a knowledge object has to go through both steps or only the first step of the quality assurance process. The certificate also determines how widely a knowledge object is made available. Knowledge objects with a restricted quality certificate have to pass the first step of the quality assurance process only; this ensures that they have well-defined metadata and, thus, can be searched for effectively. Since no check against the quality of the content is carried out, such knowledge objects are made available only in defined areas of the knowledge base to which selected knowledge workers have access. The public quality certificate requires that a knowledge object goes through both steps of the quality assurance process. Once they pass these steps they are made available to all users of the corporate knowledge. The differentiation between these two types of certificates has proven valuable for sharing of knowledge objects which exist in a premature or preliminary version only. In organizations, such knowledge objects are of great importance for well-defined groups of users (e.g., strategic planning group) which have to rely on latest information about new trends, new legislation etc.

Finally, users can assign priorities to the knowledge objects before they enter the quality assurance process. Priorities are of particular importance for the second step of the quality assurance process as they provide the experts a rough order in which they should assess the quality of the content of an information resource. The following figure depicts the main steps of the quality assurance process.

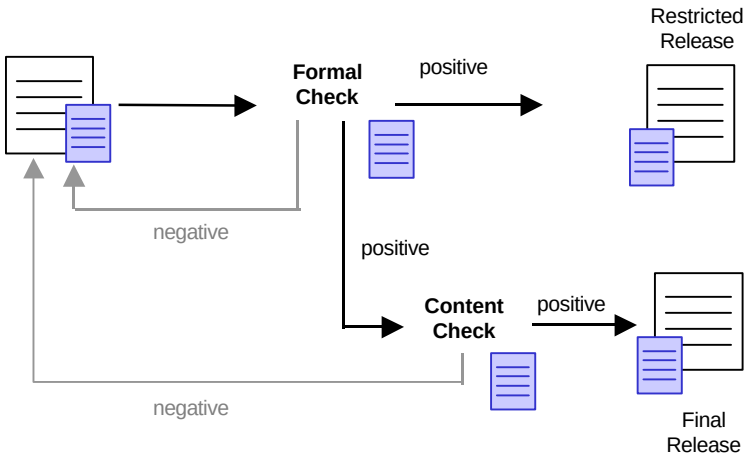


Figure 4: Quality assurance process

Technically, the quality assurance component capitalizes on XML and DTD defined for different types of

knowledge objects. The formal check verifies if the metadata of knowledge objects are specified according to a pre-defined DTD. For example, this includes if all relevant metadata fields have meaningful values. Any complaints why a knowledge object did not pass the checks are recorded in a report (shaded rectangle in figure 4). This report is a separate document but linked to the respective knowledge object.

5. Outlook

Currently, the knowledge management component of PADDLE focuses on explicit knowledge but not tacit knowledge. While explicit knowledge can be captured and codified, tacit knowledge is knowledge which cannot be easily articulated because we are not aware of it. The idea to capture tacit knowledge with the PADDLE system is based on the observation that profiles and customizations of knowledge objects represent some of a knowledge worker's tacit knowledge (Stenmark 1999). For example, the way in which knowledge objects are put in context with each other, how knowledge objects are structured in a profile, how and in which working space they are customized etc. reflects tacit knowledge of a knowledge worker which we want to exploit for automatically structuring the corporate knowledge base. Also, the quality assurance component needs further improvements. The idea is to develop a strategy that can include several experts in specific fields in the quality assurance process of one knowledge object. This is revealed to be of great importance particularly for knowledge objects which cover a broad range of topics. Finally, the retrieval of knowledge objects will be improved through the integration of a GIS-based geographical search component (Tochtermann et al. 2000).

6. Literature

Borghoff, U.M. and Pareschi, R. (1997). Information Technology for Knowledge Management. Journal of Universal Computer Science Vol. 3, No. 8, Springer Pub. Co., pp. 835-842.

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.

Hicks, D.L., Tochtermann, K., Eich, S., and Rose, Th. (1999). Using Metadata to Support Customization. Proceedings of the Third IEEE Metadata Conference, Bethesda, Maryland (USA), http://computer.org/conferen/proceed/meta/1999/papers/19/dhicks.html

Murray, P. and Myers , A. (1999). The Facts about Knowledge. Studie der Cranfield School of Management. http://www.info-strategy.com/knowsur1/

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

Skyrme, D., Amidon, D.M. (1997). Creating the Knowledge-Based Business, Business Intelligence Ltd. Wimbledon 1997.

Stenmark, D. (1999). Using Intranet Agents to Capture Tacit Knowledge. Proceedings of WebNet 1999 - World Conference on the WWW and Internet, Hawaii 1999.

Tochtermann, K. Schwartz, S. (2000). Geographical Search Components for Information Portals. Proceedings of WebNet 2000 - World Conference on the WWW and Internet, Texas 2000.

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 1999.

Acknowledgements

The research was conducted in the PADDLE project (Personlizable ADaptable Digital Library Environment) funded by the program "Verteilte Verarbeitung und Vermittlung Digitaler Dokumente" of the Deutsche Forschungsgemeinschaft (DFG).