Important Aspect of Knowledge Management

Hermann Maurer

Institute for Information Processing and Computer Supported New Media
Graz University of Technology
Inffeldgasse 16c
A-8010 Graz, Austria
hmaurer@iicm.edu
http://www.iicm.edu/maurer

This paper is dedicated to my colleague and friend Thomas Ottmann on the occasion of his 60-th birthday. Happy birthday, Thomas, and lots of further great ideas and success stories!

Abstract. In this paper it is explained what Knowledge Management (KM) is and why it will play an important role in the future. This implies that KM is indeed more than just the sophisticated use of information systems or distributed databases for complex tasks. With the introduction of the Maurer-Tochtermann model for KM and the description of the basic elements of new techniques, it can be demonstrated quite readily that KM is indeed something fundamentally new. At the end of this paper a system developed in Europe called Hyperwave will be mentioned briefly as perhaps being the leading example of this new type of software, i.e. software for KM systems.

1 Introduction

For quite some time, the term "knowledge management" (KM) has been increasingly used in connection with the efficient exchange of structured information within large organizations, and for describing the use of heterogeneous sources of knowledge. Since information systems, databases, and networked systems, such as many internet or intranet applications, have similar objectives, the question arises whether KM can really be considered something new, or whether it is just "old wine in new bottles." In this paper, it will be argued that KM is in fact leading toward different objectives, and that important new concepts and technologies have already been developed in this area. The term KM is vague, since the underlying term "knowledge" itself is vague: even though we all know (!) what the word "knowledge" means, it is nonetheless difficult to describe it. The same problem occurs when trying to describe such words as data, information and knowledge.

If you ask "When was Archduke Johann of Austria made the imperial vice regent of the German empire?", the answer is "1848," a simple piece of data. And yet, if you put "Archduke Johann of Austria" together with "imperial vice regent of the German empire," then suddenly we know (!) something that we did not know before. Thus, given that new knowledge is created by proper linking and structuring of data, information and existing knowledge, it could be argued that the definition for knowledge is “structured information.” Even though this definition is not totally satisfactory, (since we know(!) that “knowledge” is more than that), it will be used as the working hypotheses for this paper. Since we already are having problems with trying to define “knowledge,” how can we expect that defining terms such as “knowledge management” or “knowledge management systems” will be any easier?

With this in mind, it certainly comes as no surprise that there is currently a whole slew of different definitions for KM, e.g. [1] - [4], [8], [19].

In order to avoid further difficulties with definitions that we also encounter when trying to define terms such as “intelligence,” “life,” “consciousness,” etc., it makes sense to steer away from precise definitions and proceed in a more pragmatic fashion. An approach that goes in this direction will be introduced next. It will be shown beyond a doubt that KM is in fact a new field of know(!)ledge, one in which information technology plays an essential part.

For the sake of completeness, it must be pointed out that the separation between KM and other disciplines is quite subtle: KM is viewed differently by the different groups. For example, book [3] treats many areas of IT as branches of KM that this author would have more likely handled separately. There are also positions that go to the other extreme, in which the definition of KM moves away from information technology, towards that of organizational science [4]. The author of this paper does not by any means ignore the organizational aspects that play a role in any good KM system, but rather concentrates on those parts having to do with information technology and computer science.

2 Pragmatic Starting Points for Knowledge Management

The saying “if only our employees knew what our employees know, then we would be the best organization in the world” articulates very well the most essential aspect of KM: a group of people always has more knowledge than any of the individuals, and the same knowledge will even be perceived in different ways by different people.

The challenge presented here is to come up with a way to capture the knowledge (or parts of it) that resides in peoples heads; to somehow archive it in electronic form; and to make it easily accessible to other people. The advantages for the organization that manages to do this are considerable. It would eliminate duplication of work, cooperation between employees would be made easier, knowledge would not be lost when employees leave, the training of new employees would be made substantially easier, etc. Of course, the capturing of this knowledge must occur without being a significant burden to the employees and the system has to be able to deliver the knowledge quickly, when it is required, even before the employees are aware that they are in need of additional knowledge.

It will be shown that there are indeed methods that exist which are able to implement the above-mentioned features to some extent. KM of this type will be referred to as "KM for Organizations" in the following discussions. It should be noted that even very simple solutions can help quite a lot. For example, it may not be necessary for every employee to know what every other employee knows, but employees that are interested in a certain topic should at least know whether there are any other employees that could help them with that topic. However, knowledge management of this sort is less ambitious. This type of KM is already being used in some organizations as so-called "Knowledge Domains" [5] or "Yellow Pages".

On the other hand, a much more innovative use of "KM for Organizations" could be envisioned, which would be important not only for organizations, but for all of society as well. One in which a modern version of the internet and omnipresent computers could be used to make large parts of the collective knowledge of mankind readily available to everyone, almost as an expansion of the human mind. Reflections on how and why something like this might not be all that unrealistic in the future are found in [6].

There is another more pragmatic approach to KM, that also finds its basis in a saying, but which comes from a different direction. In organizations that work with large amounts of data that come from heterogeneous sources there is the saying, "If only we were able to find a way to automatically link new data with existing data such that the data that belongs together would be recognized and classified as such, and that related information would also be linked together, then the problem of archiving information would finally be solved." This problem will be referred to as "KM for Archives," for short.

The following two examples will help to illustrate this situation. In the Journal of Universal Computer Science, J.UCS, (see [13] and [14]), the papers are not just classified by numerous attributes and are therefore able to be located using several different criterion, they are also designed to have "links into the future" added to them. The term "Link into the future" can best be explained using an example: say there is a contribution A, written in 1995 that is used in a new paper B, written in 2002: then B usually will quote A, this being a "link into the past". It is, however, possible in digital libraries [15] that paper A also contains a reference to B, i.e. a "link into the future", a feature clearly imposible in printed publications. Such "links into the future" are quite easy to create in digital libraries when literature references have a fixed format. However, generating these links is also possible beyond the scope of any one digital library using various means. The most obvious is the use of so-called "citation indexes." Another way is the use of a system that is able to recognize "similarities" or "logical connections" between documents. These methods will be described in more detail in section 3.

A further example is the electronic version of the Brockhaus Encyclopedia [16]. For each entry, there is a "knowledge web" that graphically displays other entries that have a logical connection with the current one. In the Brockhaus model, the recognition of these connections occurs through the previously mentioned "similarity concept," as well as through corresponding metadata [17]: every entry is classified under one or more categories. In this manner, references between contributions such as Bush (the politician) and Bush (the vegetation) can be eliminated. In the year 2003, a Brockhaus model for an electronic knowledge web containing the complete ontology [12] of the German language is being created for the first time!

It will be shown in the following that "KM for Archives" is included under the definition of "KM for Organizations," as one of its important branches. This is why the following discussion can be focused on the latter.

3 A Model for Knowledge Management in Organizations

In addition to the previously mentioned methods that described the main elements of a knowledge management system, the communication model by Maurer-Tochtermann in Figure 1 clearly illustrates in which way KM goes beyond (distributed) information systems and databases.

Figure 1 shows a group of people exchanging knowledge with one another, whereby a large portion of this knowledge is exchanged over a networked computer system (this being asynchronous, i.e. time-delayed, as well). Each of the arrows 1 through 7 has the following particular meaning attached to it:

Arrow 1 shows that people exchange knowledge directly with one another, e.g. during coffee breaks, on business trips, over the telephone, etc. It is this type of exchange that the organizational side of knowledge management is mainly concerned with and will not be handled further in this paper.

Arrows 2 through 4 symbolize the different ways that information is put into the KM system, while arrows 5 and 6 indicate that there are at least two ways for knowledge to be delivered from the KM system to the users. More precisely, arrow 2 symbolizes the explicit input of information in the system and arrow 3 symbolizes the implicit input. It should be noted that the gathering of this information and knowledge occurs as a by-product of other activities that are carried out, i.e. new knowledge is created within the KM system without additional work being required. Arrow 4 signifies that a KM system is also able to "systematically" generate information by observing users. Arrow 5 symbolizes the traditional querying of information in an information system or database. Users enter explicit queries and receive the corresponding answers from the KM system, which can be anything from fragments of information to large coherent documents, such as manuals, books or teaching modules. Perhaps even "more interesting" is arrow 6, which symbolizes that a KM system can act on its own to deliver knowledge to the user, even without an explicit request. Finally, arrow 7 symbolizes that a KM system is able to generate new knowledge out of existing knowledge.

Figure 1 illustrates quite clearly the differences between classical information systems (databases) and KM systems. Taking away the arrows 3, 4, 6 and 7 away, then what is left is the exact model of a classical information system. Thus, the question of whether KM systems are really different from classical information

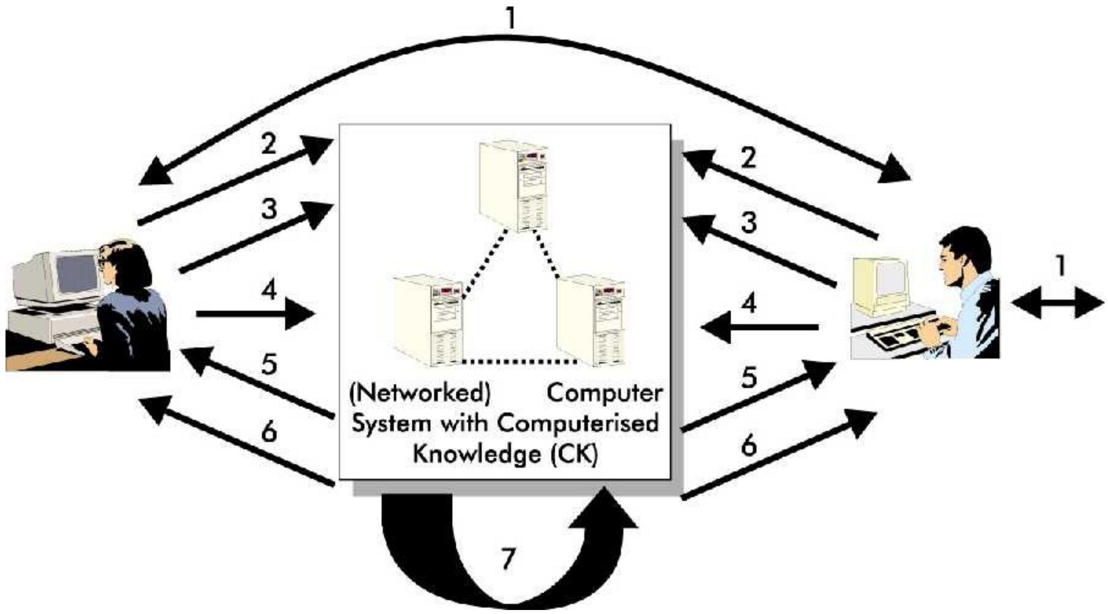

Fig. 1.

systems depends on whether the behavior symbolized by arrows 3, 4, 6 and 7 is actually feasible. The answer to this question will be answered affirmatively in the following sections.

KM for Organizations, depicted in Figure 1, can also stand for KM for Archives, if arrows 1 and 3 are ignored. This confirms the previous statement that attention can be focused on KM for Organizations since KM for Archives is included within it by definition.

4 A Few Knowledge Management Techniques

In this section, it will be shown how the actions depicted by the arrows 3, 4, 6 and 7 in Figure 1 can be realized in a modern KM system.

Arrow 3 in Figure 1 symbolizes the implicit input of information. In other words, the increase of the amount of knowledge in KM systems occurs as a by-product of other non-related actions carried out by people. The KM system could be designed to react on even the simplest of actions, such as the announcement of an event that is printed and sent out by e-mail to a distribution list. The KM system would record the event in the event calendar, and automatically move it to the Events of the past year folder after the date of the event had passed. At the years end, it would also add it to the "Events since XXXX" folder. All information that is put into a web server (from telephone lists to the structure and activities of the company and its departments) would also make its way into the KM system. There is hardly any organization today that can stay competitive without being IS0 9000 certified. This brings its own benefits with it. The extensive amount information that has to be constantly maintained presents a powerful and natural source for the KM system. The information collected on current projects with their goals, team members, tools, milestones, problems, protocols, documentation, etc. carries with it extensive knowledge about what is going on within the organization. Even completed projects suddenly prove to be very useful, since much can be learnt about current and future trends from positive and negative experiences in earlier projects. Workflow procedures in an organization are also components that are quite often kept in electronic form and therefore are logical candidates to be automatically integrated into the KM system. The same goes for year-end reports, product lists with descriptions and prices, manuals and other internal reports, and so on, including everything up to the integration of existing information systems, for which so-called knowledge portals [20] are needed. E-mail (with the appropriate mechanisms for authorization) should be kept centrally in the KM system as well, rather than managed by the individual users. The reason why bringing this information together is so important is clear: only when an adequate store of information is available can the information in a KM system be linked together automatically, resulting in structured information, i.e. knowledge.

Arrow 4, which symbolizes the creation of systemic knowledge from the input information, is currently the weakest part of KM systems. The basic idea is that general rules can be derived from entries that come from certain sources (e.g., from specific employees or databases), which would then be able to be used in similar situations. Rules like these are often closely connected with the activities symbolized by arrow 5 (explicit data search). For example, a good KM system would recognize that some searches are carried out over and over again by one or more users. It would correspondingly design shortcuts and generate automatic bookmarks; or it would realize that almost all persons who search for X, for example, also search for Y. Thus, the next time a user searches for X, information about Y will be offered automatically as well, in the sense of arrow 6.

The power of the arrows 6 and 7 can perhaps be better illustrated when a few techniques are explained that are of key importance to a good KM system.

One of the most important concepts in a KM system is the idea propagated by Maurer [9] of the active document. In an exaggerated sense, the following could be envisioned: users can ask any question (typically by typing it in, but perhaps in the future by simply asking it aloud) about a document that they have called up and the KM system will provide the answer!

Although it is clearly impossible for a KM system to be able to answer every question, there are two methods that come very close to delivering the intended result. First, certain questions can be converted into database queries and answered in this way. Secondly, questions asked for the first time can be answered by experts (perhaps even with a delayed response). Afterwards, the use of a trick prevents that the same question has to be answered a second time by an expert: the KM system saves questions and the corresponding answers. If another question is asked about the same document, it will be checked whether this question is equivalent to a previous question. In this case, the question can be answered by the KM system without having to call on an expert. This behavior becomes most practical when there are many persons accessing the same document (say on an intranet or internet server), and where answers stay the same over an extended period. Of course, there is the (huge) problem of determining whether question X is equivalent to question Y.

This problem can be tackled in a number of ways. Question X and Y can be checked for similarities through a comparison of the words (with the support of thesauruses, semantic networks or ontologies if necessary). If it is suspected that question X is equivalent to question Y, then the KM system will ask whether question Y is an acceptable substitute for question X, in which case the question can be answered by the KM system. If question X does not appear to be equivalent to any of the previously asked questions, then and only then will the question be forwarded to an expert. A refinement of this method also works sometimes using approaches from artificial intelligence, such as "case-based reasoning." However such methods are, up to now, still of lesser practical importance.

Of course, it would be more elegant if it could firmly be proven whether two texts X and Y are equivalent. However, this is not possible with the current methods or even those of the foreseeable future, except under very restricted conditions, e.g. if only certain syntactical forms are allowed and the choice of subject is limited (see [9]).

However, another quite simple approach to solving the problem of active documents exists, which can be termed "localized FAQs." More precisely, when users have questions, they mark the corresponding section of the document and type in their question. The question will then be answered by an expert, either during given "consultation hours," or otherwise at a later time. An icon will be inserted before the marked section of the document, which will signal to later readers that "a question was asked about this section and was answered by an expert." Subsequent users can check the Question and Answer dialog before asking further questions about the same section. In large-scale experiments it has been shown that there are seldom more than a few questions asked about the same section of a document (and if more do occur, the author of the document is well-advised to revise it since it is obviously not clear enough). This list of questions quickly brings the document to a "stable" condition. In one company of 150,000 employees that required everyone to read extensive materials, it turned out that after the first 600 employees had read the information only .03% of those remaining had new questions! That means that for the first 600 employees, experts needed to be available (in this case, around the clock), but for the remaining 149,400 employees (i.e. for 99.6%) only 45 new questions arose, and these were able to be answered without causing many problems with a time-delayed response.

The above discussion makes it obvious that a good KM system has to include the concept of "active documents."

As was just discussed concerning the comparison of two questions, X and Y, a problem that is often encontered is to figure out whether X and Y are "equivalent" or "similar." There are a number of ways to do this, the most common of which being improvements on the approach of using the frequency of the most important words as a measure for similarity. This approach is able to determine with high probability whether two texts are similar. These methods not only allow for the automatic classification of texts (in the sense of arrow 7 in figure 1), but also for active notification when there is a suspicion that two documents are similar (in the sense of arrow 6 in figure 1). A few concrete examples will help illustrate the practical application of similarity tests such as these.

Example 1: Consider an organization that has branches world-wide. When a new project is started at a location A, a description of it is documented in accordance with ISO 9000. The KM system takes this document, translates it from the local language into English and then compares the resulting (poor) translation with the English descriptions of all the current projects in the organization. If a similarity is found with a project at a location B, then both A and B will be notified that a duplication of work might be happening. The KM system may be wrong in some cases, i.e. the similarities in A and B are already known or irrelevant, but even if just a few cases are found in which duplication can be avoided, it is still very significant financially.

Example 2: In an organization that has a large internal research department, an employee adds a new publication A to the digital library. Soon afterwards, the employee receives an e-mail notification because an analysis of similarities between documents shows that there are two similar contributions, B and C, in the library as well. At the same time, the authors of B and C are notified that a similar paper was recently added to the digital library. Just as in example 1, duplication can be avoided and cooperation is increased. Ideally, publication A would not only be compared to documents in the organization's internal digital library, but also against those in other digital libraries accessible over the internet (for example, as in J.UCS [13], [14]).

Example 3: Through the regular review of non-private e-mails, it is determined that two groups, A and B, seem to be working on similar problems. Both groups are notified accordingly.

Example 4: In a discussion forum, a topic is brought up that had already been thoroughly covered a few months ago. The KM system recognizes this and curtails an unnecessary discussion with a link to the previous entries.

All together, it is quite clear that the testing for similarity between documents is one of the most powerful of the set of tools that a KM system needs to have. Computed links can be graphically displayed quite well through a knowledge web, such as in Brockhaus Multimedial [16]. An example of this is Figure 2, which displays an automatically computed knowledge web for the word space probe ("Raumsonde" in German).

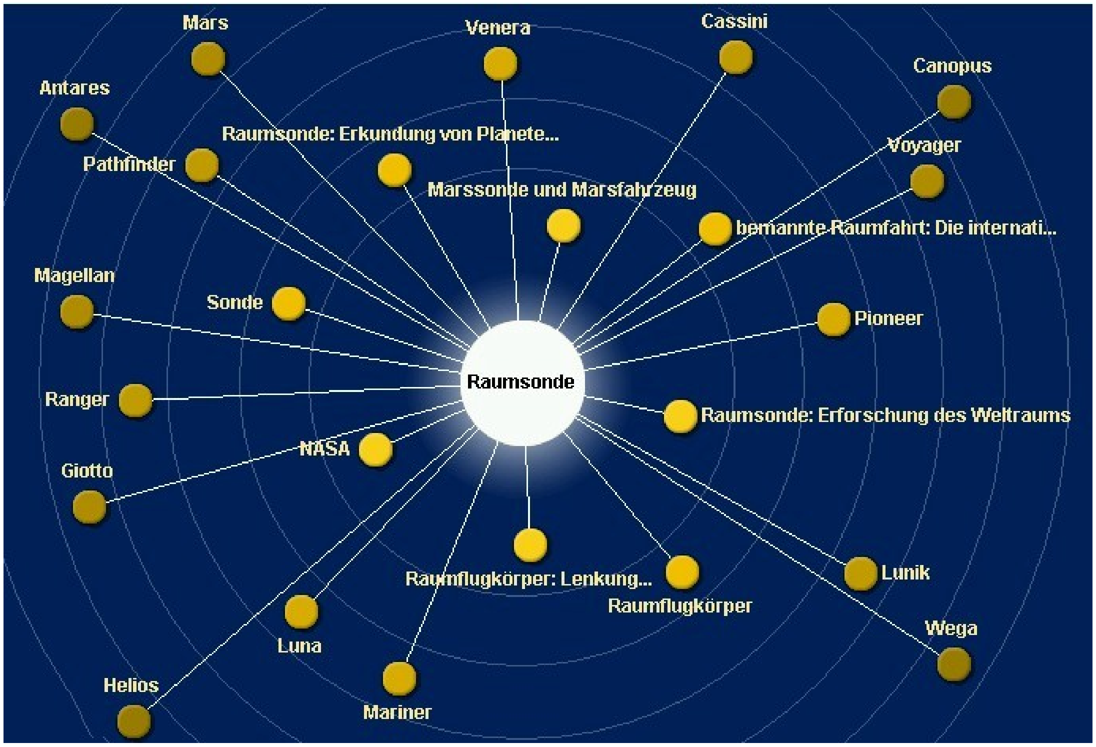

Fig. 2.

A more complex variation on recognizing similarities is the recognition of logical connections or links. However, the methods for doing this are not yet generic enough for general use; they must instead be adapted to the particular requirements individually. For this reason, one example that follows will have to suffice to explain this variation.

Let us assume we have a very large store of data, for instance, all publicly accessible WWW servers and databases, including all reports available in electronic form from newspapers and news agencies. The task would be to determine which people are closely connected to a specified person X.

If a KM system finds the text, "X visited Nassau on October 15, 2002," then the information "October 15, 2002" will be entered in the "X database." In an analysis of an entry found elsewhere, "Maurer was on North Eluthera from the 10th to the 20th of October, 2000," the following occurs: (a) the word "Maurer" will be recognized as a proper name by syntax- analysis techniques, and (b) the entry will be compared with every entry in the X database. Since the KM system is able to identify certain geographical relationships, it knows (!) that North Eluthera belongs to the Bahamas and that Nassau is the capital of the Bahamas. Since October 15, 2000 falls between the 10th and 20th of October 2000, it is conceivable that Maurer ran into person X. In the "links database," it would be saved as (X, Maurer, 1), meaning that there is one circumstance in which Maurer and person X could have met. If the KM system later finds that "X met with person Z" and that "Maurer went to school with person Z," then the set (X, Maurer, 1) would be replaced by (X, Maurer, 2). Should the counter reach a critical size, e.g. it reaches (X, Maurer, 100), the system alarm is sounded. This now means that it is highly probable that person X and Maurer are connected in some way. The "sounding of the alarm" is another typical sort of action that is symbolized by the arrow 6 in Figure 1, while the process of calculating such connection is one of the actions symbolized by arrow 7.

It should be clear that the creation of the relationships described in this example assumes a rather complex system. However, a look at the past shows that it performs quite well in e.g. tracing person, with terrorism a currently very hot topic.

There are much less complex applications that also have to do with logical connections or links, for instance, in the area of E-commerce: similarities between customers A and B concerning e.g. their tastes in literature can be found out through their purchases. Now, once this has been established, if customer A buys a book written by author X, whose books A has not previously read, and soon afterwards buys another book by author X, then it can be assumed that customer A likes author X. Hence author X can be recommended (most often successfully) to customer B.

Research in the area of the recognition of logical connections is still very much going on, but it is already clear that this will be an additional tool that is of great importance to KM systems.

5 Summary

In this paper, it was shown that knowledge management systems (KM systems) differ in a number of decisive aspects from classical information systems. The methods described have already been realized in a few existing systems. Hyperwave[7], [18] is probably the most fully developed of these, since it offers active documents, automatic classification, recognition of similarities and many other essential components of a KM system.

References

1. Woods, E., Sheina, M.: Knowledge Management - Applications, Markets and Technologies. Ovum Report (1998)
2. Studer, R., Abecker, A., Decker, S.: Informatik-Methoden fr das Wissensmanagement Angewandte Informatik und Formale Beschreibungsverfahren, Teubner-Texte zur Informatik, vol. 29 (1999).

3. Karagiannis, D., Telesko, R.: Wissensmanagement Konzepte der Knstlichen Intelligenz und des Softcomputing, Lehrbcher Wirtschaftsinformatik, Oldenbourg Verlag (2001).
4. Davenport, T., Prusak, L.: Working Knowledge: How Organizations Manage What They Know. Harvard Business School Press, Boston, (1998).
5. Helic, D., Maurer, H., Scerbakov, N.: Knowledge Domains: A Global Structuring Mechanism for Learning Resources in WBT Systems. Proceedings of WEBNET 2001, November 2001, AACE, Charlottesville, USA (2001) 509-514
6. Maurer, H.: Die (Informatik) Welt in 100 Jahren. Informatik Spektrum, Springer Verlag (2001) 65-70
7. Maurer, H. (ed.): HyperWave: The Next Generation Web Solution. Addison-Wesley Longman, London (1996)
8. Sivan, Y.: The PIE of Knowledge Infrastructure: To Manage Knowledge We Need Key Building Blocks. WebNet Journal 1,1 (1999) 15-17
9. Heinrich, E., Maurer, H.: Active Documents: Concept, Implementation and Applications. Journal of Universal Computer Science 6, 12 (2000) 1197- 1202.
10. Tochtermann, K., Maurer,H.: Umweltinformatik und Wissensmanagement. Tagungsband des 14. Internationalen GI-Symposiums Informatik fr den Umweltschutz, Bonn, Metropolis Verlag (2000) 462-475
11. Meersman, R., Tari, Z., Stevens, S. (eds.): Database Semantics. Kluwer Academic Publishers, USA, (1999).
12. Uschold, M., Gruninger, M.: Ontologies: principles, methods and applications. The Knowledge Engineering Review 11,2 (1996) 93-136
13. www.jucs.org
14. Krottmaier, H., Maurer, H.: Transclusions in the 21st Century. Journal of Universal Computer Science 7,12 (2001) 1125-1136
15. Endres, A., Fellner, D.: Digitale Bibliotheken. dpunkt Verlag Heidelberg, Germany (2000)
16. Der Brockhaus Multimedial 2002 Premium. DVD, Brockhaus Verlag, Mannheim (2001)
17. Weibel, S.: The state of the Dublin core metadata initiative D-Lib Magazine 5, 2 (1999) 300
18. www.Hyperwave.de
19. Petkoff, B.: Wissensmanagement. Addison Wesley (1998)
20. Hyperwave: Hyperwave Information Portal White paper (1999) http://www.hyperwave.de/publish/downloads/Portalwhitepaper.pdf