Knowledge Management System Building Blocks

Georg Hüttenegger
Vienna University of Technology, Austria
georg@ist.org

Abstract: This paper describes three building blocks of a technological Knowledge Management (KM) system that provides all relevant and practical means of supporting KM and thus differentiates itself from existing KM tools in goal and approach, as they usually deal with a limited range only. The three blocks described within this paper are: a virtual information pool, which utilizes Enterprise Application Integration (EAI), a single and central user interface providing ubiquitous access, and mechanisms to enrich the available data, essentially based on Artificial Intelligence and Data Mining techniques.

1. Introduction

Many experts define Knowledge Management Systems as IT systems that support KM activities and hence are generally concerned with enhancing competitiveness by the appropriate use of all available knowledge (for further definitions of KM see Sveiby [Sveiby 2001]). There are two main issues to be addressed by IT: To provide users with access to as much relevant information in the most practical way possible and to capture as much information as feasible and add this to the available information. There are already a number of systems and technologies on hand, which provide a great deal of functionality in a wide range of (KM) fields. For further information on existing systems/tools, see Alavi and Leidner ([Alavi 1999] or [Alavi 2001]) as well as Marwick [Marwick 2001].

This paper proposes building blocks for an integrated system that provides as much functionality as possible within current IT constraints. The concept focuses on larger organizations that currently use diverse systems and beyond that is probably more suited for non-routine tasks such as project management, software development, or in general, development of new products. Although a resulting system would certainly address similar issues faced by smaller organizations, complexity and expenditure could prove as a limiting factor. Based on a paper describing the general architecture of the proposed KM system (presented at the ECKM 2003 [Hüttenegger 2003]), this paper provides detailed descriptions of the three major building blocks the solution substantially consists of.

It is of utmost importance to bear in mind that no proposition is being made to introduce a KM system or even parts of it, which can be seen as a general solution for a company striving for KM. Instead, each organization should carefully review their KM needs and not before preliminary results are available, concern themselves with technological issues. This line of argument finds support by current research activities, stating, why purely IT driven KM initiatives have failed (for example McDermott [McDermott 1999] or Malhotra [Malhotra 2002]).

2. Requirements for a KM System

Before describing the actual solution itself, requirements of such an IT system, reasoning on analysis of goals and needs of KM, shall be presented (primarily based on two KM research reports published by KPMG: [KPMG Consulting 1998] and [KPMG Consulting 1999]).

The primary requirements identified are Capturing, Integrating & Connecting (technologically), Transforming, Storing, Flexible & Adaptable Permission System, Connecting People, Data Mining / Knowledge Discovery in Databases / Machine Learning, Handling & Capturing Context, Links, and History, Automation, Searching / Finding and Retrieving, and finally Embedding into Normal Workflow.
Figure 1: shows how these requirements relate to one another. In addition to the requirements listed, importance of a few other issues shall be stressed: Providing Transparency, Providing Traceability/"History", and Non-Functional Requirements (mainly performance, stability, usability).

These requirements call for an integrated system and as a result reduce learning times and ease information access at the same time. The building blocks presented form the heart of such a system.
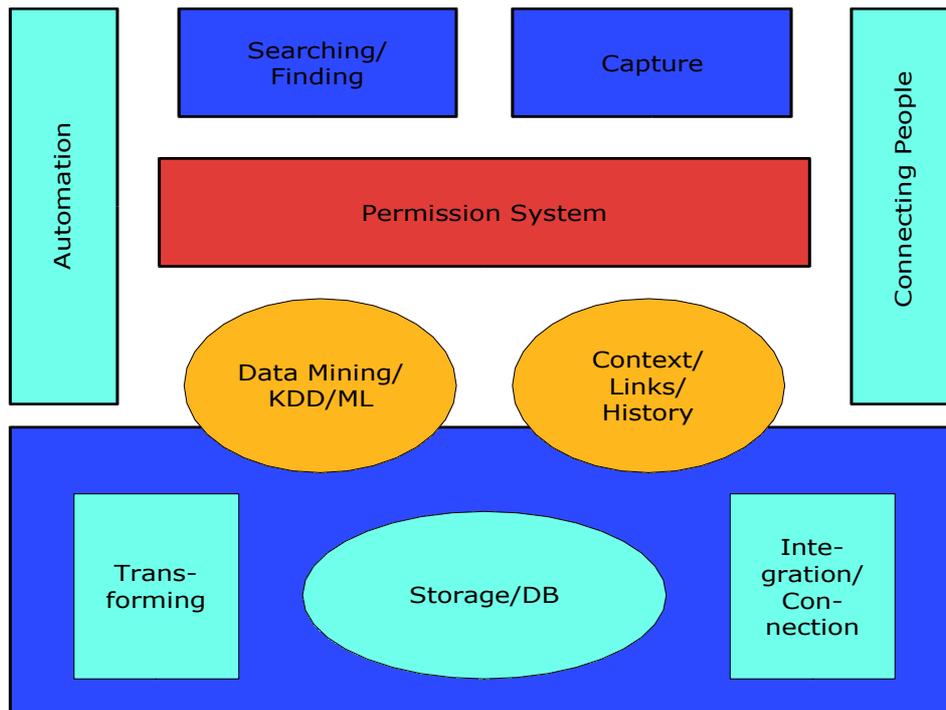
Figure 1: Graphical depiction of system requirements

3. Virtual information pool and technical foundation

The virtual information pool is a knowledge base that incorporates all relevant information while not necessarily storing it, thus forming a virtual pool of information, and allows all relevant data available to the organization to be found and accessed.

As the needs of different organizations are very diverse, there can be no exact definition what "relevant data" actually is. However, some types of data will be common to virtually all organizations, including:

1. All sorts of documents (often found in Document Management Systems, databases, or file systems).
2. Project related data such as memberships, history, or lessons learned.
3. All types of communication data such as e-mails, chats, forums, etc.
4. A skills database.
5. Information on coordination, especially information related to meetings.

The crucial task (from a purely IT point of view) is keeping the information up-to-date with as little manual input as possible, whereas the second most important task is to create an organizational culture and precise policies that foster sharing and helping (in practical terms: to encourage employees to help others instead of just optimizing their own team and/or project).

3.1 Providing ambiguity

This is another major issue faced by the knowledge base, and there are a number of ways of achieving this. One method is to allow users to add their own bi-directional links between stored pieces of information. However, one must bear in mind that many pieces of information will "belong" to multiple locations. This is quite natural for human beings but is not possible with simple file system hierarchies. However, users are already accustomed to this simple concept and will probably need to be "taught" how to work with this particular "feature". [Note: The fact that a single information item can belong in multiple locations often confuses people, e.g. when electing to delete it. The same holds true for shared folders where people perhaps just want to "remove" a piece of information from their personal view of the folder's content.]

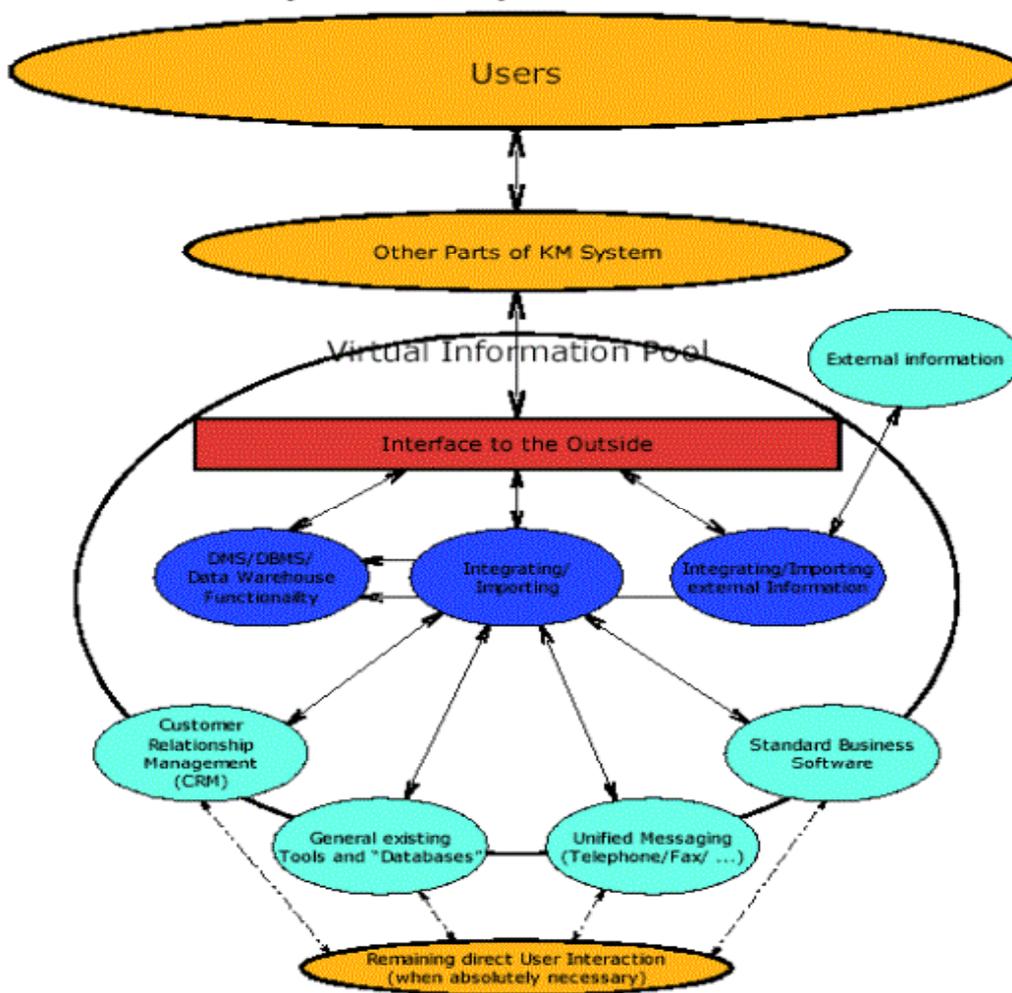### 3.2   Overview of the virtual information pool

Figure 2: Depiction of the virtual information pool

Figure 2 shows the functions and features of the virtual information pool as well as the most important connections within. By integrating/importing external information, the pool is able to include high quality external databases, for example. To allow transparent access through the interface even to sources outside an organization such integration needs careful planning: One has to consider permission, potential cost, and performance issues and solve all of them in a satisfactory way.

Integrating and importing all kinds of information and data forms the heart of the virtual information pool. Especially relevant sources are: standard business software (like human resources, enterprise resource planning, etc.), more specifically customer relationship management, unified messaging (with information about phone calls, faxes, etc.), and general existing 'tools' and 'databases'. Not shown in this picture (as it will

appear in Figure 3) is the CSCW/Groupware aspect that not only provides a user interface but is also needed as a data source.

General DMS/DBMS and especially Data Warehouse functionality are managing/using/enriching all the available information to fulfill the promise of a consistent interface.

On the bottom of Figure 2, one sees "Remaining direct User Interaction". Generally, the approach calls for a complete replacement of such alternative ways of user interaction, on the other hand, as this will hardly ever be possible, at least in a first step, it is included as another data/information source.

3.3   Technical foundation and realization

Since I am focusing on larger organizations, a multitude of systems providing relevant features for the building blocks will be in use in

each individual case. If necessary functions are not yet available, the necessity and possible financial implications of relevant systems should be discussed separately from the building blocks themselves. (One should consider the cost of basic functions separately as they have important merits of their own.) Important base systems include: A directory service of some kind (to hold user relevant information / user configuration) and a Single Sign-On based on the directory services. [Note: While not necessary from a technical point of view, a Single Sign-On function will automatically increase the user friendliness of the system, by eliminating the need to log-on to each of the numerous systems individually.] The pool itself clearly needs some data storage, in which any information not available through two-way integration needs to be stored. Additionally, all functions required to accomplish sophisticated operations on the organization's data store need to be available efficiently through the virtual information pool (thus disguising in which system the data is actually stored). This is another major issue for the pool: to hide the source of the data as far as possible (this in turn allows legacy systems to be replaced without the need to change functions above the virtual information pool themselves).

Storing data is not the primary purpose of the pool (data should only be stored here if absolutely necessary). Rather, the aspects of integration and adding value to information are paramount (e.g. through integration with EAI). Therefore, this concept does not call for an additional Groupware or Document Management System (DMS), but instead recommends the integration and use of existing systems. Whilst this will inevitably pose some difficult challenges (e.g. some common Groupware systems do not even allocate unique keys to documents), re-inventing the complex functionality of current systems is neither feasible nor desirable. There is a far greater selection of DMS available and therefore selecting an appropriate solution for both legacy systems and the new information pool should be easier to achieve.

Unfortunately, a number of existing applications (e.g. Oracle Forms Applications) do not support this approach since they do not provide adequate interfaces (APIs). Integrating such legacy systems simply by extracting their data is rather awkward and leaves any functionality behind. Therefore, the primary objective must be to have adequate interfaces

to all critical applications. Otherwise, they should be migrated and all new systems that are acquired or developed in the future should fit into the new architecture.

It is important in the design stage to plan for a phased implementation of the resulting solution, starting with a major component and then gradually adding significant components in each phase. The actual first phase will ultimately depend upon the individual customer requirements and situation: one approach would be to start with the integration of the more accessible systems and thus demonstrate the realizable benefit; a second approach might be to introduce the building blocks at the same time as another large system is deployed and demonstrate the synergy benefits.

3.4   Conclusion

All this information together forms the knowledge base or virtual information pool—the single most important foundation for a complete and holistic KM system. The solution proposed places two major requirements on this pool: Firstly, the provision of ubiquitous access to the pool and, secondly, the automatic updating and enriching of the information it contains.

4.   Providing ubiquitous access

The need for a virtual information pool together with the importance of a Single Sign-On as an essential additional system component has already been adequately discussed.

Another issue of similar importance is the provision of access to the system with full functionality for a variety of communication devices.
Since most organizations will need separate areas within a resulting KM system, a powerful and flexible permissions system is of utmost importance. An eminently suitable basis would be a directory service. Nevertheless, the chosen permissions system has to allow for groups that may not share their documents (e.g. who are working for different – or even competing – customers), parts of the system should be accessible via Extranets (so that partners can use it), possibly even providing content for the company website.

At the same time, as much information as possible must be available to every authorized user. [Note: This can be especially challenging when access restrictions on documents alone may not be secure. Imagine employees

searching for documents containing information about "layoffs" and finding out that there are several recent documents containing this keyword created by top management]. Finally, yet equally important, the management of such controls must be kept as simple as possible. Important concepts here are access control lists, user groups, and administration rights for parts of the system (so that responsibility can be split and distributed).

Establishing the need for a powerful permissions system, the "window" to the proposed solution: "The single, central User Interface (UI)" has to fulfill some complex requirements:

- Seamless integration of existing and newly introduced systems and their functions.

- Handling different access devices while taking into consideration any limitations imposed by the device itself (also to be addressed when entering/integrating information; sometimes particularly demanding with limited and/or mobile devices like smart phones).

- Providing powerful search mechanisms (full text retrieval and other capabilities supported by the integrated engine such as searching for all available attributes including context, document similarity, etc.).

- Providing hierarchical access to the information: This can be done through taxonomies, maps, directories, etc. For example, taxonomies could be generated automatically and therefore always up-to-date.

- Providing context/links/history for the resulting pieces of information (when entered/updated; connected to which project; stored/modified by whom; show documents with similar content; etc.).

- Providing ways to add new information: This should be integrated into the normal workflow and actually save, instead of generate work

- Integrating Computer Supported Cooperative Work (CSCW) functionality

(Groupware features): These systems often have elaborate clients that cannot be easily integrated into the interface; however, it is essential that key functionality is available through the common interface. This may involve integration at multiple levels: a bi-directional connection for the data itself; full integration of the selected key functionality and access to the remaining functionality via the "native" interface).

- Visualization: This is another powerful mean of providing users with a view of information in the virtual pool. This involves a range of issues such as how much information is presented, in what form, which additional navigational means are offered (star diagrams of the relations between information items are a good example).

- Optional: Off-line access is often another important issue. At the same time, monetary constraints and limited benefits can be arguments against this generally rather expensive functionality.

One of the primary ideas behind the central UI is to provide access to as much functionality as possible from one single GUI. The main benefits are: Greater efficiency as a result of fewer and more consistent GUIs (ideally with complete integration of their functionality), more data available to the knowledge base, and possible savings due to synergies and re-use of available data when building new functionality. [At the same time, it becomes more difficult to use third party software, as it will not integrate nearly as well as individual solutions.]

This central UI is also a viable way to realize a quick win. If the resulting KM system replaces (or completely integrates the functionality of) at least two existing systems, the users will have at least one system less to deal with. Furthermore, users will appreciate the benefits of a complete solution. From a technical point of view, complete integration also enriches the knowledge base.

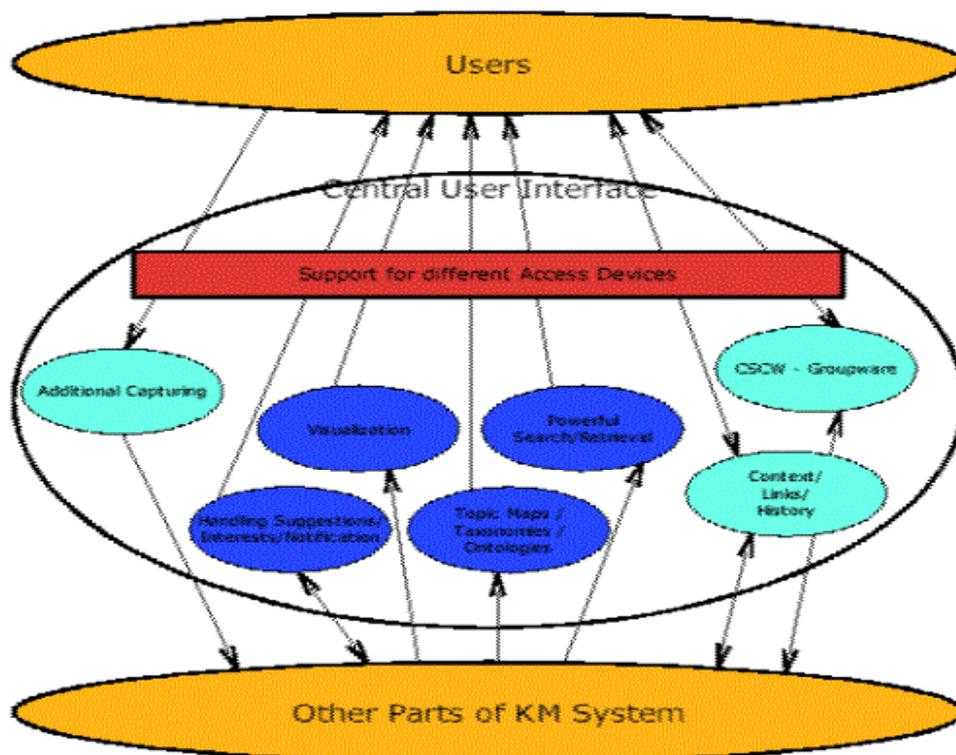## 4.1 Overview of the Central User Interface

Figure 3: Depiction of the Central User Interface

Figure 3 shows all the technologies and functions behind the user interface as well as how they work together to provide this second building block. Through the UI functions, all the functionality of the whole KM system is made available to the end user. Necessary adaptations for different access devices are an integral part of the system thus consistently offering the subset of functions that makes sense for limited devices (like PDAs, mobile phones, bandwidth-limited connections etc.).

This figure also shows three conceptually different sets of functions:

- Capturing:

  As a rather broad function, there are no other examples of this set. It represents the aspect of acquiring additional information. Whilst using implicitly available information this aspect should minimize user input.

- Presenting/Delivering:

  This aspect includes: Visualization, Powerful Search/Retrieval, Handling Suggestions / Interests / Notification, and Topic Maps/Taxonomies/Ontologies. While a certain amount of interaction with the user and other parts of the resulting system is part of these functions, their primary issue is to deliver information to the user. To accommodate different types of users, multiple powerful mechanisms are offered and are accompanied by e.g. push mechanisms, in managing interest profiles together with suggesting possibly relevant new documents to the affected user.

- Interaction:

  Finally, interaction is the most important factor for both the CSCW/Groupware functionality and for Context/Links/History. These functions are similar to the previous ones as they deliver and present information. At the same time, they are providing means to communicate, capture current and provide past context. Summing up, they are essentially bi-directional, which is not the focus of the described functions so far.

On the bottom of Figure 3, one can see "Other parts of KM System" together with the most important connections to the various functions. Please keep in mind that the arrows are indicating the important directions of data and information flows and are not meant to be exact descriptions.

4.2   Conclusion

This section discussed the necessity and the most important benefits of the single and central UI, the two most important reasons are:

- To build one consistent UI at the top level of the virtual information pool, and
- To be able to integrate all functionality/systems where this is not possible at lower levels or at least only to a certain degree.

This interface has to provide diverse access possibilities, such as powerful searching, hierarchical access, and access for limited and off-line devices. This interface is also a good chance to realize a quick win by replacing and integrating at least two previously used systems in the first phase of the implementation of the resulting KM system.

5. Automatic updating / enriching the knowledge base

This section discusses the third major building block of the resulting KM system architecture. The necessity of a shared knowledge pool and a central and single UI has already been discussed. However, these could also be part of a "simple" EAI or CRM project, although taxonomies are already a step further and their importance for the UI has already been described. What distinguishes KM systems from information stores are the more sophisticated means of analyzing and enriching the available information, thereby providing insights not easily gained by other resources. Existing KM tools tend to address mainly one or several of these aspects and consequently, although they can rightfully claim to be KM software, do not attend to all the issues a system has to deal with in successfully supporting KM.

5.1 Overview of the Automation Aspect

Utilizing the integrated knowledge base provided by the virtual information pool, automation is the key to go beyond a "mere" Information Management System (IMS). Situated between the knowledge base (thus being able to work on all available information) and the user interface, it provides functions and features that separate it from simpler and smaller solutions.

The important issues and technologies as well as the most relevant connections between them are depicted in Figure 4. The central user interface – as the only part interacting with the end users – shields them from the automation aspects, whereas the virtual information pool provides diverse data and information necessary for all the automation functions.

Let us now turn to the individual points of the automation aspect:

- Yellow Pages / Skills Database:

  Using the wealth of information available through the virtual information pool, this subsystem updates the skills database automatically. Whilst this may result in a lower quality than would have been achieved through human intervention/input, the information is always up-to-date and the updating process requires little or no manual input. Relevant information includes: Project membership, written documents, accessed information, hours booked on certain sub-projects, discussions via e-mail, forum, chat, etc. [Note: Whilst this type of automatic update is by no means perfect, manual updates can also be unreliable and with the information available in the integrated system, an automatically generated informed guess might be quite acceptable.]

- General Data Mining:

  Expert users can make direct use of the data mining functionality, thus allowing direct exploration of possible correlations together with new insights etc. This positions it nearest to the user interface, as they need close integration.

- Base for Notifications, Interests, Push, Intelligent Hierarchies etc.:

  This issue represents all the automation items below and is relevant mostly from a conceptual point of view. It is up to an actual implementation whether this is a dividing line or not.

- Knowledge Maps / Clustering:

  This is the foundation of providing access to possibly related documents through various means. Users are offered or may search for similar documents and can browse in the available knowledge maps (e.g. being based on generated and automatically updated taxonomies).

- Classifying Messages:

  Especially interesting when information is not addressed personally and has to reach a correct recipient, taking e.g. holidays and substitutes into account. This can be used for customer messages but equally for internal ones (e.g. hotline etc.).

- AI/ML/DM/KDD:

Perhaps the heart of the automation aspect is the utilization of the possibilities of various forms of AI/ML/DM/KDD. This is not only based on the knowledge pool directly but also on the following two points. These technologies allow the resulting solution to provide information not integrated or captured in a number of ways, automated summaries or providing "links" to similar documents being just two of the many possibilities.

- Speech Mining/Image Similarity/Analyzing Video:

  While these technologies are not yet fully mature they help to integrate speech, images and videos as resources into the resulting solution accessible to a certain degree. Especially speech mining has already proven its practical applicability for more than specialized requirements.

- Monitoring Usage/Constant Re-evaluation:

  This aspect is especially important, as users will seldom provide the system with a rating of the information they accessed. Therefore, it is necessary to assess the usefulness of information automatically,

thus showing the potentially most relevant information while at the same time reducing the assessed "value" of no longer relevant information. Re-evaluation can also be important to manage the volume of stored data by providing hints what could possibly be removed from online access (e.g. to an archive).

Not directly depicted in the following figure are the following important functions:

- Handling interests/profiles:

  The resulting system automatically notifies users about pertinent new information items. To enable this, all new documents must be classified to the user's possible interests. Additionally, the system proposes new interests based on access data and other information (e.g. new role, joining new project, etc.).

- Collecting context and history for the information pieces:

  Automatically including all available information by the system when users create a new item/entry. This also holds for modifications to existing information.
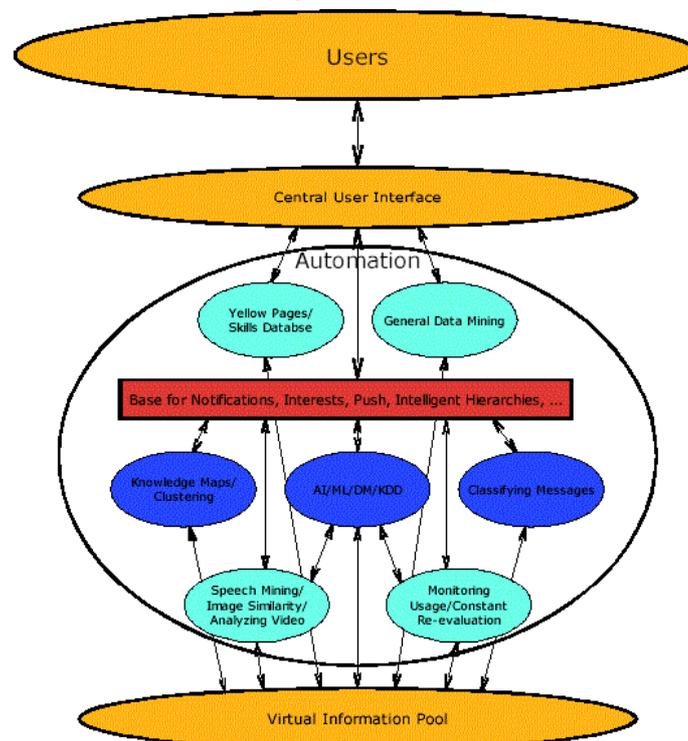

Figure 4: Depiction of the Automation Aspect

5.2   Conclusion

This section describes the most demanding building block of the resulting KM system: Automation and Enrichment of the information

within the system. Generally, these processes automatically capture and generate additional information and provide it to the users, in addition re-evaluating information is of equal

importance. Although this is difficult without manual intervention, it is necessary to offer a working procedure to address this issue and prevent the KM system from becoming "crowded" and useless. Whilst automatically generated information might not provide the desired results, there are already a number of successful examples applying these underlying technologies. [Note: Most KM tools rely on one or more of these technologies.]

6. The big picture

Having described the three major building blocks, namely the virtual information pool, the single and central user interface, and automation and enrichment, a short overview of how these building blocks form a KM system shall round up the big picture:
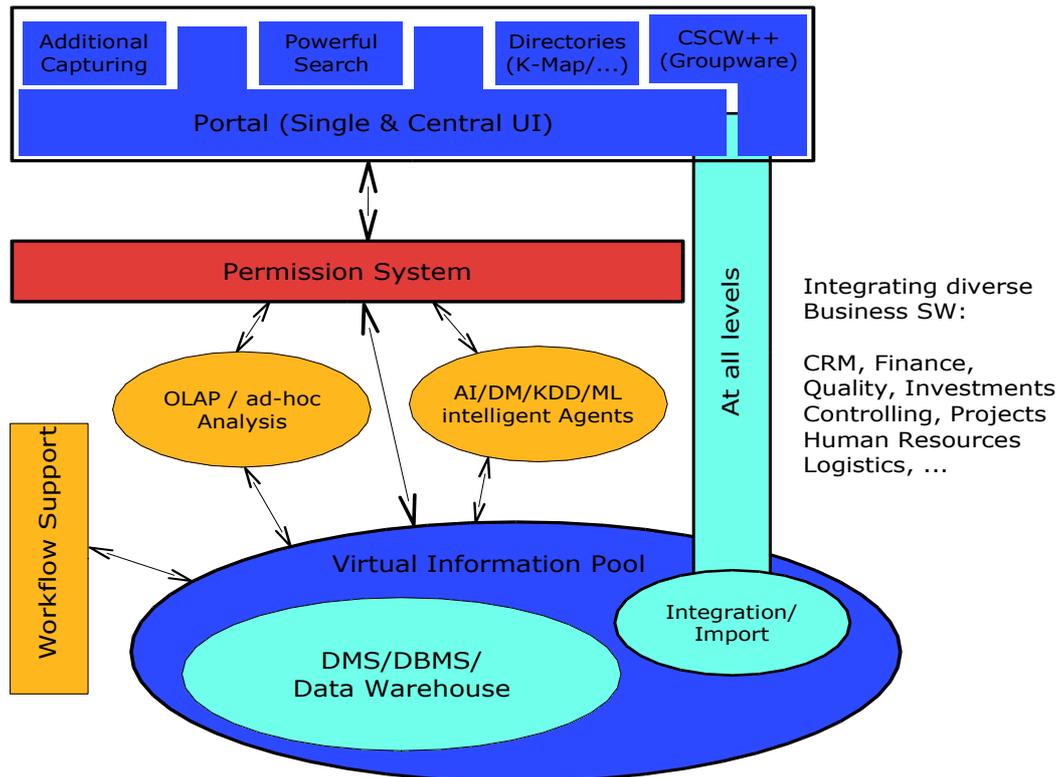
6.1 Putting it together

Figure 5: Overview of the System Architecture

Figure 5 illustrates the resulting system architecture described above and the virtual information pool is immediately recognizable as the center of the solution. The single and central user interface is on top and some of the more important examples of automation and enrichment of the knowledge base are located in the middle. There is also a need to integrate diverse business software (not discussed). These systems, or more precisely, their data, form important parts of the knowledge base, and their relevance becomes clear when human resources data or ERP systems are taken into consideration.

6.2 Further aspects

Lacking space, many interesting issues, such as, why certain technologies have been included in a building block while others are

missing, have not been addressed in this paper. As a matter of equal importance, any actual initiative striving to realize the proposed building blocks will have to approach the necessary design, implementation, and introductory issues with the greatest care.

7. Conclusion: Summary/Further Research

This paper describes three major building blocks of a technological KM system that supports all the major needs of KM. This differentiates the approach from descriptions of "smaller" KM systems (such as the one described by Fagrell et al. [Fagrell 2000]), or KM tools designed to address a narrow set of requirements, or normal integration projects (e.g. CRM projects using EAI technologies).

7.1 Summary

To conclude this paper, consider the issues that the building blocks and a resulting system should address:

- Supporting people in their daily work,

- Making them more efficient and faster,

- Helping them achieve, in terms of quality and – potentially – quantity, better output.

This is the result of:

- Enabling members to find all relevant information in the organization quickly and easily (both by searching and by advanced directories/maps).

- Supplying a list of all available experts in a particular field.

- Providing information about the context and history of earlier and probably similar projects and documents (as examples of how things might be done).

- Enabling communication at all levels (internal experts, project teams, etc.).

- Managing employee interest profiles, suggesting updates, to include potential new interests and actively informing users of new, relevant information.

- Automatically classifying documents, especially incoming ones, and identifying and retrieving similar documents, to provide additional contextual information.

Together, the building blocks address the need to include technological aspects in any well-planned and complete KM initiative.

7.2 Further Research

While IT is definitely not necessary for KM in general, this paper has once again demonstrated the prospects of integrated support IT can provide. This is a major motivation to continue investigating the described building blocks and, given a broader view, a resulting system, together with analyzing, how effective they can be for real world KM problems. From a more general point of view, technology already offers many features of high relevance for KM. Sadly enough, when it comes to capturing tacit knowledge, it is still far from providing a satisfactory solution. [Note: One can be quite certain that this will not be achieved in the near future; indeed, doubt may be indicated, that it ever will be possible.] This remains an area of current research (see Brown et al. [Brown 2001]). Imagine the possibilities, when speech recognition, video analysis, or image similarity recognition become as useful as OCR is today – this may give you a small indication of what might eventually be possible.

References

Alavi, Maryam and Leidner, Dorothy E. "Knowledge Management Systems: Issues, Challenges, and Benefits", Communications of AIS Volume 1, Article 7 (1999).

Alavi, Maryam and Leidner, Dorothy E. "Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues", MIS Quarterly, Vol. 25, (2001) pp. 107–136.

Brown, Eric, Srinivasan, Savitha, Coden, Anni, Ponceleon, Dulce, Cooper, James, Amir, Arnon, and Pieper, Jan "Towards speech as a knowledge resource", Proceedings of the Tenth International Conference on Conference on Information and Knowledge Management, Atlanta, ACM Press (2001) pp. 526–528.

Davenport, T.H. and Prusak, L. "Working Knowledge: How Organizations Manage What They Know", Boston Harvard Business School Press, USA (1998).

Essers, Juup and Schreinemakers, Jos "The Conception of Knowledge and Information in Knowledge Management", in: J. Schreinemakers (ed.) Knowledge Management: Organization, Competence and Methodology – Advances in Knowledge Management Vol. 1. Proc. 4th Int. ISMICK Symposium Würzburg (1996) pp. 93–104.

Fagrell, Henrik, Forsberg, Kerstin, and Sanneblad, Johann "FieldWise: A Mobile Knowledge Management Architecture", Proceedings of the ACM 2000 Conference on Computer supported cooperative work, Philadelphia, ACM Press (2000) pp. 211–220.

Hahn, Jungpil and Subramani, Mani R. "A Framework of Knowledge Management Systems: Issues and Challenges for Theory and Practice", Proceedings of the 21st international conference on Information systems, Brisbane, Association for Information Systems (2000) pp. 302–312.

Hoffmann, Marcel, Loser, Kai-Uwe, Walter, Thomas, Herrmann, Thomas "A Design Process for Embedding Knowledge Management in Everyday Work", Proceedings of the international conference on Supporting group work, Phoenix, ACM Press (1999) pp. 296–305.

Hüttenegger, Georg "A Knowledge Management System Architecture" Proceedings of the 4th European Conference on Knowledge Management (ECKM 2003), Oxford, MCIL 2003, pp. 485–494.

Johanessen, Jon-Arild, Olaisen, Johann, and Olsen, Bjørn "Mismanagement of tacit knowledge: Knowledge management, the danger of information technology, and what to do about it", (2000) [online] http://program.forskningsradet.no/skikt/johannessen.php3, visited November 3, 2003.

KPMG Consulting "Knowledge Management: Research Report 1998", (1998) [online] http://www.brint.com/papers/submit/knowmgmt.pdf, visited November 3, 2003.

KPMG Consulting "Knowledge Management: Research Report 2000", (1999) [online] http://www.kpmg.nl/Docs/Knowledge_Advisory_Services/KPMG%20KM%20Research%20Report%202000.pdf, visited November 3, 2003.

Malhotra, Y. "Why Knowledge Management Systems Fail. Enablers and Constraints of Knowledge Management in Human Enterprises", in: Handbook on Knowledge Management 1: Knowledge Matters, Springer, Heidelberg, Germany (2002) pp. 577–599.

Marwick, Alan D. "Knowledge management technology", IBM Systems Journal Vol. 40, No. 4 (2001) pp. 814--830, [online] http://www.research.ibm.com/journal/sj/404/marwick.pdf, visited November 3, 2003.

McDermott, R. "Why Information Technology Inspired But Cannot Deliver Knowledge Management", California Management Review, Vol. 41, No. 4, (1999) pp. 103–117.

Sveiby, Karl-Erik "What is Knowledge Management?", (2001) http://www.sveiby.com/articles/KnowledgeManagement.html, visited November 3, 2003.