Efficient Information Retrieval: Tools for Knowledge Management

Katarina Stanoevska, Alexis Hombrecher,

Siegfried Handschuh, Beat Schmid

Institute for Media and Communications Management, University of St. Gallen

Müller-Friedberg-Strasse 8, CH-9000 St. Gallen, Switzerland

Tel: +41 71 224 2793, Fax: +41 71 224 2771

E-mail: {Fistname.Lastname@unisg.ch}

Abstract:

Knowledge has become an important resource in many organisations. The success of an organisation depends on its ability to transform personal knowledge of employees into organisational knowledge. This knowledge can then be made widely available to the entire organisation and be reused when needed.

One necessary prerequisite for reusing knowledge, coded and stored in documents, are appropriate classification and retrieval procedures. Classification accompanies the process of knowledge externalisation and retrieval supports the process of knowledge internalisation by enabling the capturing of appropriate coded knowledge.

In this paper we will evaluate currently available retrieval mechanism with respect to their effectiveness in knowledge management. We will then present a comprehensive classification and retrieval technology based on the Q-Technology, which provides support for the automated and intelligent classification and retrieval of knowledge.

Efficient Information Retrieval: Tools for Knowledge Management

. . . who controls the vocabulary, controls the knowledge

George Orwell

1 Introduction

Knowledge has become an important resource in many organisations. The success of an organisation depends greatly on its ability to transform personal knowledge of employees and knowledge within an organisation, not explicitly belonging to an employee, into organisational knowledge. This knowledge can then be made widely available to the entire organisation and be reused when needed.

One common form of permanent storage of organisational knowledge is documents. A document is structured information intended for human perception, that can be interchanged as a unit between users and/or systems [14]. Examples of documents available in organisations, which carry information as a prerequisite for knowledge creation, are handbooks for different tasks summarising process knowledge, project reports and summaries, product descriptions, planning documents and others. Thus documents contain externalised, coded knowledge related to different aspects and topics of an organisation's processes and tasks.

One necessary prerequisite for reusing knowledge coded and stored in documents are appropriate classification and retrieval procedures. Classification accompanies the process of knowledge externalisation and retrieval supports the process of knowledge internalisation by enabling the capturing of appropriate coded knowledge. Thus classification and retrieval are important parts of creating organisational knowledge.

Recent developments in information and communication technologies, and especially their convergence, have opened unprecedented opportunities for efficient classification, management, retrieval and dissemination of documents. Documents are now available on an interactive ubiquitous carrier, which bridges both time and space. In addition, information and communication technology also represents the first carrier, which can mimic human reasoning. This ability complements human capabilities to classify, structure and search for appropriate documents and can be applied for automated classification, retrieval and restructuring of knowledge.

The importance of classification and efficient, as well as qualified, document retrieval has given rise to the development of different approaches and technologies for supporting this task. They are usually created by adopting existing approaches for knowledge representation and focus on different aspects of document classification and retrieval. Some of them provide fast retrieval in large stores of documents. Others focus on a very specific type of meta-description for documents, which limits their applicability to certain application areas. Thus, they differ in how far they support the transformation from personal to organisational knowledge and from organisational to personal knowledge.

In this paper we will present a comprehensive classification and retrieval technology based on the Q-Technology, which provides support for automated and intelligent classification and retrieval of knowledge. This paper is divided into the following sections: Section 2 describes the role and importance of knowledge classification and retrieval in the knowledge creation cycle. Section 3 gives an overview of the most widely used knowledge classification and retrieval mechanisms. In section 4 the Q-Technology is described. Section 5 gives an overview of an implementation of a knowledge medium, which is used for the management of scientific publications.

2 The Role of Information Retrieval and Classification in Knowledge Management

To know, i.e. knowledge is a feature of human beings. We define knowledge as the internal state of an agent following the acquisition and processing of information [9]. An agent can be a human being, storing and processing information in his mind, or an abstract machine, including devices to store and process information [9].

With human knowledge we distinguish between tacit and explicit knowledge [7]. Tacit knowledge is person dependent. It comprises the subjective insights, intuitions, and hunches of individuals. It is knowledge, which is deeply ingrained into the context. Harris defines it as a combination of information, context, and experience [2]. Explicit knowledge, on the other hand, is externalised tacit knowledge, meaning tacit knowledge that has been coded on a carrier. Externalised knowledge is information. This potential knowledge is realised when information is combined with context and experience of humans to new tacit knowledge according to Nonaka [7]. This cyclical knowledge creation process is illustrated in Figure 1.

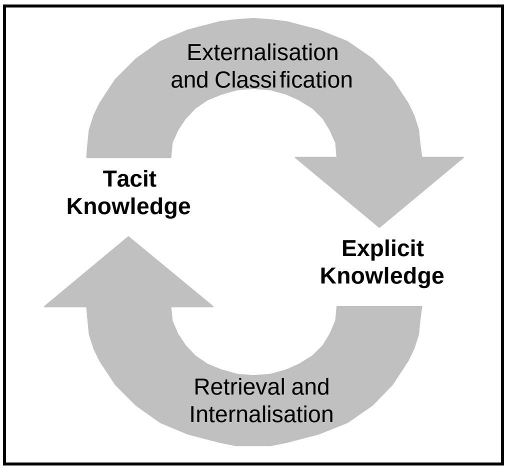

Figure 1: The Tacit-Explicit Knowledge Cycle

The creation and classification of documents is part of the externalisation process in the knowledge cycle. The top arrow in Figure 1 illustrates this. After having created or acquired tacit knowledge, humans will put their ideas on paper. Today, this is often done electronically in the form of digital documents. In the next step, for the purpose of the knowledge cycle, these digital documents need to be classified so they can be retrieved at a later time by anyone interested. In order to be efficient and successful, this should be smoothly integrated into the working process. Efficient classification should be as exact as possible, yet require the least possible effort on the part of the classifier.

Retrieval is part of the internalisation process. The bottom arrow in Figure 1 represents this. When we want information about a certain topic, we ask others who we think might possess that information or we read about it. Hence, retrieval must provide the appropriate information for reuse or creation of new knowledge. In order to achieve this, only relevant answers must be supplied to a user. The retrieval of irrelevant answers, i.e. information overload, must be avoided. Therefore, answers should be as exact and as complete as possible.

In the next section we will look at different concepts for classification and retrieval systems. We will compare the systems, paying particular attention to how well the different systems aid knowledge management using the above two ideas as a criterion.

3 Overview of Existing Approaches

Many different retrieval mechanisms and related classification approaches exist. We would like to give a brief overview of the most widely used and relevant ones. They will then be analysed from a knowledge management point-of-view.

3.1 Full Text Search

3.1.1 Basics

Full text search is probably the best known and most widely used search method. The general idea behind this method is to search through documents looking for key words. The best-known Internet search engines (Lycos, Infoseek, Altavista, etc.) make this very efficient by constructing an index of key words found in documents. On the Internet, so-called software robots perform this indexing, by searching through WWW pages. Constructing an indexing is nothing more then trying to capture the content of a document [8]. Certain rules are applied which eliminate specific words (and, or, the, etc.) from the list of indexed words.

The full text search engines use the previously created indexes and matches them syntactically to the search words when queries are performed. This increases the speed of the system tremendously. As a result, the query returns those pages, which are indexed most often for the given query term or terms. Queries can be somewhat refined by using logic operators (AND, OR, NOT).

3.1.2 Evaluation

The general advantage of the full text search method is that it is very fast due to the automated indexing mechanism, which can be performed using robots. This means that classification of documents is as efficient as it can be since no human intervention is needed. Even though classification may be poor, its automation makes this part of the externalisation process, from the knowledge management point-of-view, very efficient and it integrates smoothly within the working process.

In many cases answers returned by a full text search engine are irrelevant and incomplete. Not only may the results be of poor quality, meaning they are of no real interest to the user, but the number of results may be very large, which means the user needs to navigate once again through a large heap of information. Alternatively, the query can be modified through user evaluation. This is known as relevance feedback in IR [1].

For example, a user may want to search for research papers about databases. The search term entered could be something of the sort "*research AND databases*". This query will very likely not return all the research papers with databases as the topic because many research papers about databases will not have the above mentioned search terms in them. Furthermore, the query is likely to return a lot of results, which are completely irrelevant, since only a syntactical match is guaranteed. The fact that the two words, research and database, can be found within a document does not mean that this paper is actually about that topic.

The poor results often achieved with this type of retrieval and classification mechanism make it a poorer tool with respect to the internalisation process of the knowledge cycle. Since users do not receive exactly the answers they are looking for, the process of retrieval is inefficient. They have to perform extensive selection activities or relevance feedback instead of directly consuming the information. This process is often time intensive.

3.2 Case Based Reasoning

3.2.1 Basics

From a logical point of view CBR is similar to metadata-based retrieval methods (see section 3.3) [12] [13]. Documents are categorised by linking them with attributes. The classification is done indirectly. The content of the document is indirectly described through formal specification of possible problem-areas. The content of the document itself is not categorised. The query is also realised by describing a particular problem area. The answer contains documents, which, based on the meta-description, pertain to a specific case. This type of search mechanism is mostly used in "Help-Desk" environments [13].

In the area of help-desk environments this type of information retrieval mechanism has been very successful since the required work per customer sinks dramatically. The newly documented knowledge is readily available to co-workers working in a similar environment.

3.2.2 Evaluation

CBR tools, due to the categorisation mechanism, provide efficient answers. This means, unlike in the full text search mechanism, queries will return relevant answers. If the classification is done correctly, the query will return all documents, which are relevant for a special case, not leaving any out of the result set that should logically be included. The internalisation process of explicit knowledge is therefore very efficient.

Poor performance, as the amount of data becomes very large, can be a disadvantage of CBR tools. But, a certain amount of information is required within the system before useful results can be retrieved. Only when (almost) all possible types of cases have been entered in the system does the CBR tool have its greatest value. The process of updating the system with new cases is often an external process, meaning the process is not part of the overall help-desk environment. This makes externalisation of cases (documents) very expensive and inefficient. Hence, CBR systems are the mirror image of full text search tools, with respect to the internalisation process, where externalisation is efficient and internalisation inefficient.

3.3 Metadata-based Search

3.3.1 Basics

The metadata-based search and retrieval method is based on meta-descriptions given for documents. Metadata is data about data. Smith [11] describes it as the characterisation of information objects for the purpose of locating, evaluating, and accessing appropriate sets of objects. In database systems this is often referred to as the catalogue or schema of a database. The categorisation is achieved by adding attributes external to the content of the document, such as author or date of creation, as well as attributes describing the content of the document. The attributes have a specific semantic meaning and thus allow for, in contrast to full text retrieval methods, a semantic search. The system used in libraries is based on this principle of meta-descriptions.

Attributes should describe the properties and content of documents as fully as possible. As mentioned, since attributes can be added, aspects of the document, which go beyond its content, can be considered and used for classification. Therefore, attributes can also be used to give qualitative information about documents. For example, an attribute can describe the relevance of a document with respect to a certain topic.

Two different types of metadata-based retrieval methods are possible, depending on whether the key words are connected via clearly defined rules or whether they are freely defined. We refer to the first type as basic and the second as intelligent. The basic type is similar to the system used in libraries, where every text contains a predefined set of attributes (author, title, ISBN number, etc). By intelligent we mean that the retrieval system (the machine) can deduce information from the semantic network, which enables it to produce alternatives to a query with no answers. In this section we will focus on the basic type since it is more widely used and in section 4.3 we will explain the intelligent type in more detail.

3.3.2 Evaluation

These types of systems have a high descriptive quality due to the fact that the amount of attributes describing a document can be very large. This means the system is very detailed. Searching for a research document in a database will return very efficient (good) answers. Consider a categorisation containing at least the two attributes “type of document” and “topic of document”. Now, if a possible attribute value (see section 4.1 on Q-Vocabulary) for ”type of document” is *research paper* and a possible attribute value for “topic of paper” is *databases*, then a query asking for all research papers about databases will return all relevant documents and only those. This makes the internalisation process within the knowledge cycle more efficient than the above mentioned approaches. Sine the system can have an unlimited number of attributes for the classification, the classification can be infinitely fine. Furthermore, metadata-based systems with predefined semantics can provide alternatives when searches return empty sets. This intelligence further enhances the internalisation process of the knowledge cycle.

A disadvantage of metadata-based retrieval systems is that prior to categorising the documents, they have to be read and understood by the person doing the classification. The classification can be subjective, meaning it heavily depends on the person doing the classification. Here we come back to the idea of context. The context is different for every person and hence each person will classify things differently. Of course the types of attributes defined for a system can also be the cause for this (e.g. value of paper → good, bad). One could argue both for and against the fact that the possibility for subjectivity should be eliminated. Thus, building a good model is very difficult.

The externalisation process is not part of the working process. No approaches have been suggested where a smooth integration of the classification process within the knowledge cycle takes place. Therefore, the cost of externalisation is very high making it inefficient.

3.4 Hypertext Systems

3.4.1 Basics

The idea behind hypertext systems is to link informational units in a non-linear manner [4] [5] [3]. This is different from the structures found in a book. Hypertext systems form network-like structures whose connections are expressed in the form of relations. The World Wide Web is an example of a simple hypertext system. By simple we mean that the World Wide Web only uses one connection method, the link, whereas in other systems much more complex relations exist between information units.

In hypertext systems the basic search mechanisms are browsing and serendipity. Browsing is the traversing through the information in hypertext systems by clicking on links. When searching for information in this manner one tends to stumble across information of completely different value (theme). This information may then become more relevant as the information originally searched for. This is referred to as the serendipity effect. This can be either positive or negative. It is a positive effect when not looking for any specific information, when one is not 100% sure about the required information. The danger is that it is very easy to loose track and focus when weeding through the vast amount of information.

3.4.2 Evaluation

As explained above, when a user of a hypertext search mechanism is not certain about the information he/she is looking for, the serendipity effect of browsing can be very useful. In this manner the user can get an overview of the information available in the system. This is clearly an advantage of hypertext systems. But, overall the retrieval of information is not very efficient, as it is very difficult to find information by simply following links (without first querying a search engine). Thus, the internalisation of information within the knowledge cycle is inefficient.

The cost of setting-up hypertext systems is relative expensive. Since the relations (links) between documents can only be automated partially, most of the work has to be done manually. Since this manual process is once again outside the working process of knowledge creation, this externalisation is inefficient with respect to knowledge management.

3.5 Overall Evaluation

At this point it makes sense to briefly compare the different technologies with each other. Once again, the evaluation criterions are efficiency of internalisation and cost of externalisation of information (knowledge) within the knowledge cycle. The two criterions are placed on two axes.

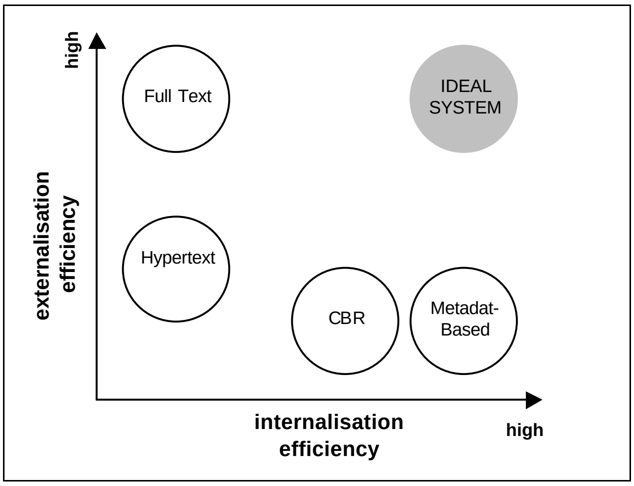

Figure 2: Evaluation of information retrieval systems with respect to each other (*Note:* This figure is not drawn to scale. It should just be used to compare the relative efficiencies of the different systems)

From Figure 2 we can see that most retrieval systems are either efficient in the externalisation or internalisation of knowledge within the knowledge creation cycle. Ideally, a system should be efficient in both processes (ideal system circle). The full text search mechanism is very efficient in the externalisation of knowledge, but very inefficient in the internalisation of knowledge. Metatadata-based systems, on the other hand, are very efficient at the internalisation of information, but inefficient at externalisation. Hypertext systems, in particular the WWW, rank very poorly overall.

An ideal system for the knowledge creation cycle would be one, which is efficient at internalisation while still having low externalisation costs. In the next section we will introduce a technology, which comes close to the ideal system, the Q-Technology.

4 Q-Technology

At the Institute for Media and Communication Management (MCM), at the University of St. Gallen, Switzerland, a generic information retrieval system has been built based on the metadata paradigm described above. In this section we will explain this system and its most relevant features with respect to knowledge management. Unlike most metadata-based systems, the Q-Technology allows for an intelligent information retrieval system. This intelligence will also be explained in detail.

Within the concept of the Q-Technology, the Q-Vocabulary is the means for structuring, categorising and systemising the knowledge, while the underlying Q-Calculus implements the inference mechanism.

4.1 Q-Vocabulary

The Q-Calculus is a formal language for the description and classification of sets of objects, which was developed by Schmid [11]. The basic language constructs offered are:

- *Basic sorts*, which delimit sets of objects or abstract concepts by naming them. In the context of encyclopaedia sorts delimit the domain of discourse.

For example in a research paper about intelligent agents, sorts could be:

Agent, Role

- *Basic scales*, which refer to features of objects and, thus, contain the possible values of classification criteria.

For example:

Type ={Reactive, Proactive}. *ApplicationBranch = {Banking, Manufacturing, Tourism}.*

- *Attributes*, which combine sorts with scales. The scale of an attribute defines partitions on the sort. In other words, it defines a classification structure for the set of objects denoted by the sort. For example: If the scale “Type” is applied to the sort “Agent” by a defined Attribute “Agent.Type”, then the set of agent objects is divided in subsets of proactive and reactive agents. The denoted agents belong either to the one or the other subset.

Agent.Type = Sort: Agent ->Scale: Type

One sort can be the domain of several attributes. The Cartesian product of the attributes of one sort defines the maximal search space delimited by the sort.

Based on the above described basic language constructs more complex, i.e. derived terms can be defined.

- By using logical operators on scale elements to define and name sub-sorts of objects.

For example:

Reactive Agents = Sort: Agent, Attribute: Type = Reactive *Proactive Agents = Sort: Agent, Attribute: Type =Proactive*

- By applying multiplication on sorts to construct complex object sets.

RoleofAgent = Agent X Role

Basic and derived terms are the foundation for further definition of new derived terms. The set of logically related terms referring to a special domain of discourse, i.e. world, forms one Q-Vocabulary. The terms of one Q-Vocabulary form a semantic net. The well-defined relations between the terms can be evaluated by the Q-Inference, thus allowing for complex and intelligent search.

4.2 Classification of Documents

Adding vocabulary terms to a document gives it a meta-description. The vocabulary terms reflect the objects, which are considered in the document:

- *Document X {Sort* → Attribute}

Thus, the definition, i.e. the agreed upon knowledge, becomes an additional part of the document. The Q-Vocabulary terms form a semantic meta-layer over the document, thereby defining a flexible and content oriented search space over the information objects, enabling a content and semantic search.

A new sort can be defined for a specific domain. In the case of documents, a sort might be *publications*, *abstracts*, *definitions*, etc. Publications have certain classification criteria, basic scales, such as:

- *Research Area := {Databases, Knowledge Management, E-Commerce}*
- *Language := {English, German, French}*
- *Audience := {Management, Researcher, General Public}*

Looking more closely at the research area databases, we quickly realise that the field of databases is in itself a whole research area. Thus we can further subdivide Databases as follows:

- *Databases : {Object Oriented, Hierarchical, Relational}*

The process of describing objects in more and more detail can go on and on. The subject of relational databases is once again a whole research area. Thus, we can build complete hierarchies describing objects (Figure 3).

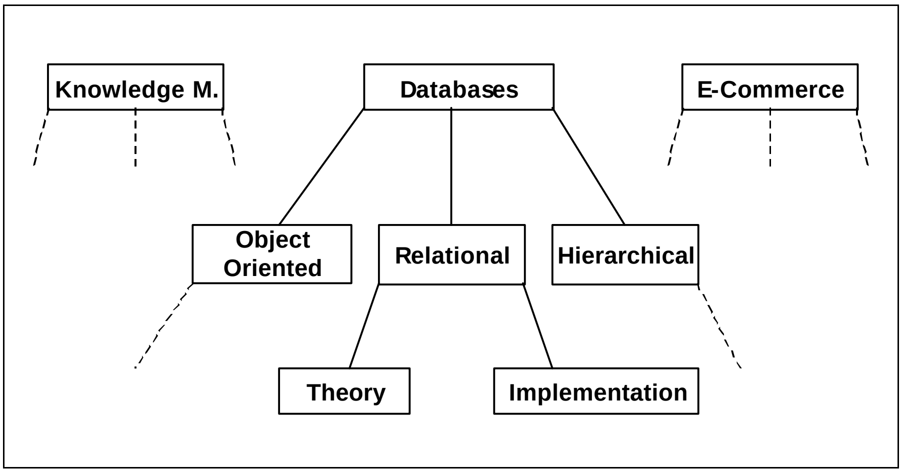

Figure 3: Hierarchical structure of documents

With the Q-Technology we are able to classify a document as precisely as we would like. In classical metadata systems, a user can classify a document by putting it in a category. He may have a list of possible categories as given in Figure 3. Consider a research paper about the theory of relational databases. If choosing to classify the document as belonging to databases, the description is not very precise. But, a query requesting database papers would at least return a positive result. If, on the other hand, the document is classified more precisely as a relational database document, traditional metadata-based systems will not return a positive result when searching only for papers on databases. Here, only a search for papers on relational databases would return a positive result. The Q-Technology solves this problem. When a paper is classified as belonging to relational database theory, a query about database papers will return the relational database theory paper, as should be the case. Within the tree structure, Any single node is logically related to the parent node. It is simply a sub-quantity of the parent node.

Figure 4 illustrates how the Q-Technology places documents, as precisely as possible, in a multidimensional space. Classical metadata systems cannot place a document in a specific square. Here the squares referring to database papers are not subdivided as was the case in the above example. But, further subdivision into hierarchical, object oriented, or relational databases would simply open-up another dimension within the database squares.

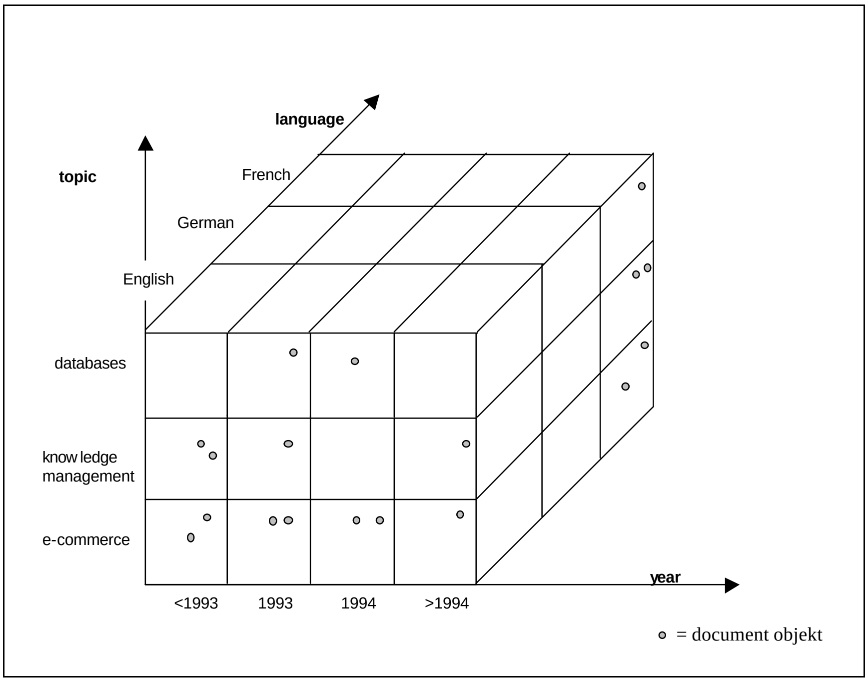

Figure 4: The document objects in the search space

In the next section we will explain how these constructs of the Q-Technology allow for the implementation of intelligent search mechanisms, which can greatly aid in the internalisation process of the knowledge creation cycle.

4.3 Intelligent Search

The semantic data model of the Q-Vocabulary represents relationships between sorts. This kind of semantic structuring allows for associations, which one does not get from conventional catalogue systems. Through the processing of the semantic structures, an intelligent search for objects can be achieved. This so-called intelligence offers alternatives in case the actual search fails (zero results).

As mentioned , the sorts in the Q-Vocabulary describe a multidimensional space (Figure 4). The dimensions of this space consist of the properties of scales. These properties are, if possible, sorted according to their similarity i.e. they are sorted in an ascending or descending order (e.g. publication year = <1993, 1993, 1994, >1994). We can take advantage of the ordering, because similar objects lie next to each other in a dimension on the border of the search space. We have to scan these borders during the systematic search for alternatives.

If a search does not return a result, it can be for one or more reasons. A user may be looking for a document in English about knowledge management written in 1994. It may be that there are no documents with all those attributes, only papers about knowledge management written after 1994 in English or in German, for example. In that case, the search request of the user returns an empty set. One of the criterions for the failure of the search was the specification of the publication year. However, we get a set of alternative results if we take the adjacent value on the publication year scale, >1994.

If several properties are the cause of the failure, the most irrelevant property will be determined according to user preference. We change this property while trying to keep the other properties the same. This preference can either be set by the system or by the user. The user can specify in which order the different attributes are relevant. In the example of the document search, the most important attribute for the tourist may be the publication year. The user prefers to have a different language or a different subject.

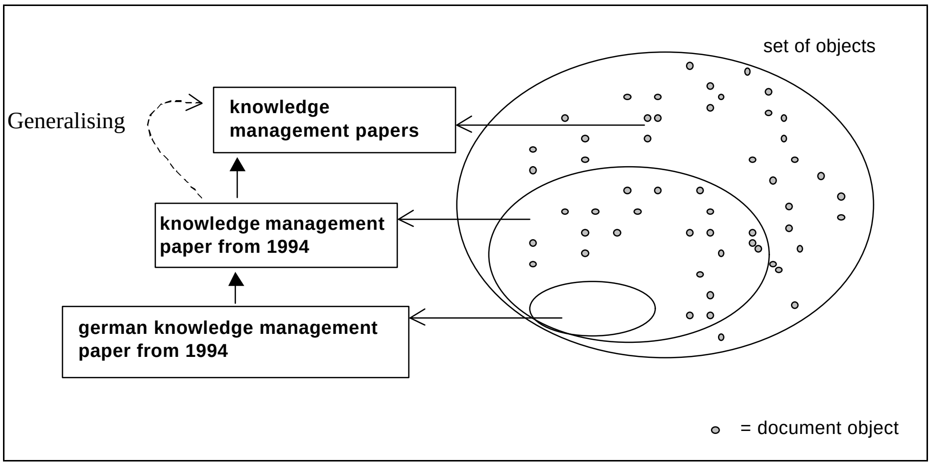

Figure 5: Intelligent search

In certain cases, the designation of the similarity is not possible, since the scale cannot always be precisely sorted or its ordering does not correspond to the wishes of the user. Language is such a scale (Language = English, German, French). The order of the languages may not reflect the preferences of the user. For a German and French speaking user English is not an alternative although it is closer to the German language than French. In such cases, a user's preference for the scale value is required to receive qualitative alternative results.

If no object is available in the complete search area of a term, we can use the hierarchy of the terms to search in the upper hierarchy (Figure 5). This method, applied to the example of the document search, would lead to a search under knowledge management paper.

5 Current State of Implementation and Application

5.1 Current State of Implementation

We have currently implemented a Q-Server and client in Java. We use it for the retrieval and classification of a variety of objects. These objects range from product descriptions to documents. Figure 6 gives an overview of this architecture. The descriptions of the data objects are stored in a database. The actual objects can be stored on a variety of mediums. The query for an object originates in a WWW browser where a user specifies a query selecting one or more attributes from a list (see Figure 7). The query is sent to the Q-Server, which resolves the attributes and matches them to a query on a database. The results are returned and the browser displays a list of documents in the form of a table. The user can then click on one of the results to view or retrieve the document.

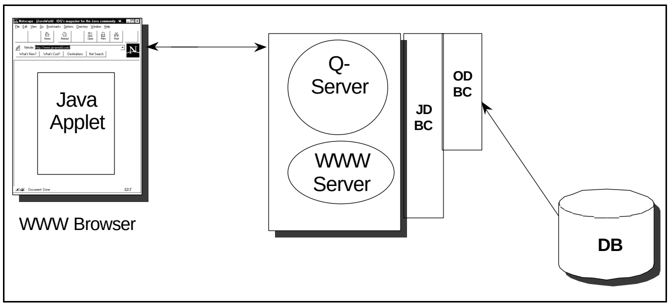

Figure 6: Q-Technology Architecture

We are still manually classifying documents. The process has still not been smoothly integrated within the knowledge creation cycle. Nonetheless, solutions have been developed and a Java GUI will be available shortly.

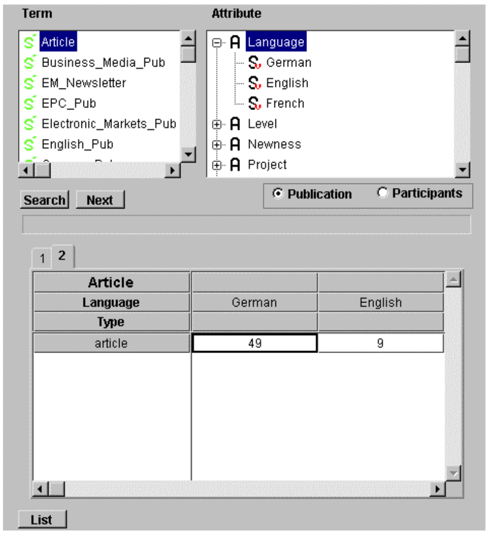

Figure 7: Q-Technology query interface

One specific instantiation of the Q-Technology is the knowledge medium NetAcademy. Here we are applying the Q-Technology for the classification and retrieval of documents. It will be briefly introduced in the following section.

5.2 Practical Application of the Q-Technology in the NetAcademy

The current state of publishing on the Internet is not acceptable, especially for scientific publications, while the normal process of publishing is often slow and cumbersome. The quality of information retrieval, as well as the linkage and reference of information, are essential technologies needed to provide a possibility for electronic publishing via the medium Internet. Scientific communities need media and channels to accumulate and disseminate knowledge with the characteristics and advantages of both traditional and new media [6]. The NetAcademy (http://www.netacademy.org) is designed to provide such a medium.

The NetAcademy is of interest because it is a generic knowledge medium [9]. The approach is generic and is applicable to a wide area of environments. In the case of the NetAcademy the general template for a knowledge medium is simply instantiated with the NetAcademy. The same can be done for similar environments, including the organisation. The NetAcademy and all knowledge mediums in general, provide tools for the internalisation and externalisation of knowledge.

6 Outlook and Conclusion

The evaluation of different information retrieval and classification systems in this paper has shown that most available systems are not efficient at all aspects of the knowledge creation cycle. Some systems are more efficient at the internalisation of knowledge, while others are better suited for the externalisation. The Q-System tries to overcome some of the shortcomings of current information retrieval and classification systems.

The Q-System described in this paper is very efficient at the internalisation process of knowledge in the knowledge creation cycle. It is now important to make the externalisation as efficient as the internalisation. Mechanisms and interfaces need to be implemented, which allow for an intuitive classification of documents during the entire externalisation process. The process has to be smoothly integrated into the overall working process. We would like to implement this functionality within the NetAcademy project.

The current Q-System only comprises one homogeneous vocabulary. Hence, different worlds cannot communicate with each other. The classification of documents used in one environment will always be different from the one used in another environment, due to the different context in different worlds. But generally, different worlds are related and often have things in common. But, to retrieve documents from different systems, a query has to be initiated on each system. The goal now is to develop a system of heterogeneous vocabularies. This means, two or more logically related worlds with different, but similar, vocabulary can communicate and interact. Ideally, the integration of different worlds will have to be made as smooth as possible, requiring the least possible effort.

References

1. N. Belkin and W. Croft. *Information Filtering and Information Retrieval: Two Sides of the Same Coin?* Communications Of The ACM, Vol. 35, No. 12, December 1992
2. D. Harris. *Creating a Knowledge Centric Information Technology Environment*. 1996, URL: http://www.htcs.com/ckc.html
3. J. Krause. *Intelligentes Information Retrieval*. in: Kuhlen, R. Experimentelles und praktisches Information Retrieval. Universitätsverlag Konstanz 1992
4. R. Kuhlen. *Hypertext - Ein nicht-lineares Medium zwischen Buch und Wissensbank.* Springer Verlag 1991
5. R. Kuhlen. *Hypertext und Information Retrieval,* in: Kuhlen, R. Experimentelles und praktisches Information Retrieval. Universitätsverlag Konstanz, 1992
6. U. Lechner and K. Stanoevska. *The NetAcademy – A New Concept for Online Publishing and Knowledge Management*. ACOS 98 Workshop, Lisboa, Portugal 1998
7. I. Nonaka. *The Knowledge Creating Company*. Harvard Business Review, November-December 1991. Pp. 96-104

8. G. Saliton. *Information Retrieval – Grundlegendes für Informationswissemschaftler*. McGrawHill, 1987

9. B. Schmid. *Wissensmedien*. Gabler Verlag, 1998 (To appear)

10. B. Schmid and G. Geyer and W. Wolff and R. Schmid and K. Stanoevska-Slabeva. *Representation and automatic evaluation of empirical, especially quantitative knowledge*, 1996

11. T.R. Smith. *Call for Papers: International Journal of Digital Libraries Special Issue on Metadata and Digital Libraries.* 1996, URL: http://www.llnl.gov/liv_comp/metadata/mail_dir/0027.html

12. V. Thompson. *Corporate Memories*. Byte, September 1997

13. S. Wess. *Intelligente Systeme für den Customer-Support – Fallbasiertesschliessen in Help-Desk und Call-Center-Anwendungen*. Wirtschafts Informatik, vol. 1 February 1996

14. ISO 8613. Information Processing – text and office systems – office document architecture (ODA) and interchange format, 1998. International Organisation for Standardisation