Journal of Universal Computer Science, vol. 5, no. 1 (1999), 2-10
submitted: 14/12/98, accepted: 15/1/99, appeared: 28/1/99

Dynamic Background Libraries - New Developments In Distance Education Using HIKS (Hierarchical Interactive Knowledge System)

Thomas Dietinger
(Graz University of Technology, Austria
tdietinger@iicm.edu)

Christian Gütl
(Graz University of Technology, Austria
cguetl@iicm.edu)

Bernhard Knögler
(Graz University of Technology, Austria
bknoeg@iicm.edu)

Dietmar Neussl
(Graz University of Technology, Austria
dneussl@iicm.edu)

Klaus Schmaranz
(Graz University of Technology, Austria
kschmar@iicm.edu)

Abstract: This paper is a short description of HIKS, a working prototype of an interactive knowledge management system. HIKS might be used as a dynamic background library in a Web-based training environment. Research in the field of Web-based learning at the [IICM] has shown that courseware combined only with static background libraries does not satisfy the learners' needs. An additional dynamic background library will always guarantee up-to-date knowledge. Relevant knowledge spaces from the biggest knowledge store, the Internet, will be extracted by HIKS. The core of this system is a sophisticated information gatherer and knowledge area broker system, which will be combined with a Hyperwave-based distance training system. This paper describes the technique of the gatherer and broker and their interaction with the learning process. Furthermore, the realisation of knowledge hierarchy for specific topics and co-operations between organisations is shown.

A short version of this paper was accepted for presentation by the ICCE 98 conference. The research described in this paper was conducted as part of the IICM Knowledge Discovery Project, supported by the Austrian Federal Ministry of Science and Transport.

Key Words: Web based training, dynamic background library, intelligent knowledge broker, information structuring, information relevance assessment, Hyperwave

1 Introduction

During the last two decades, researchers at the [IICM] have carried out a significant amount of work within the field of information and document management, computer based training, digital libraries and electronic publishing [Guetl et al. 98] [Maurer et al. 96b] [Marchionini et al. 95]. The research results and intense co-operation with the industry even led to commercial products such as Hyperwave [Maurer et al. 96a] [Kappe et al. 94]. Based on this experience a group is currently developing GENTLE [Dietinger et al. 98] [Maurer et al. 97], an electronic lecturing system combined with a digital lecture library for teachers and students. GENTLE has already been tested by hundreds of students at the Graz University of Technology. The feedback gained during the lectures is leading to continuous improvements of the system invoking new research concepts.

The introduction of new technologies such as multimedia or hyperlinked objects does not necessarily lead to more efficient learning. Technical environment could support learners' needs for interpersonal communication, the opportunity to ask questions and discuss problems with tutors and co-learners. On the other hand, technologies are useful and necessary for finding proper information and creating courseware. This paper will discuss aspects of accessing desired information to improve the learning process. New possibilities allowing the user to solve complex knowledge-dependent problems are presented.

At the present moment, the majority of self-contained courseware does not meet sufficiently the requirements mentioned above. A background library is likely to increase the usability of the courseware. However, a static background library (a collection of electronic books and journals) alone will not satisfy the users' needs because it cannot deal with the growing stream of information. Therefore, a further component - a dynamic one - has to be provided. The solution is to integrate a world-wide information repository, the information stored in the Internet. About 330,000,000 documents are publicly available at the time of writing. The Internet represents the biggest information source mankind ever had, but unfortunately, it is also the most unstructured one. The lack of structure affects content, quality of information, seriousness, topicality, etc.

One approach taken to handle this chaos is the use of huge index servers like AltaVista or Yahoo just to mention two. Although today's index servers provide at least one way of searching in the information mesh they do not solve the complexity of the problem: Sometimes queries are not narrow enough and results of 50,000 hits and more are quite usual. For this reason, the index servers perform some sort of automatic ranking, mostly just by counting the number of hits in a document. Second, although a query might be narrowed sufficiently, it still returns undesired results because the meaning of a word often depends on the linguistic context. By using context information and an intelligent knowledge broker, this situation can be solved quite elegantly, as we will describe in the following sections of this paper.

The specified problems affect the learning process and the search of relevant information within a Web-based training environment. HIKS provides facilities to combine the dynamics of brand-new information on the web and important additional information, the quality aspects. This concept will always allow up-to-date information in background library with the restriction that only high quality and reliable documents will be provided.

To make the step from an information base towards a knowledge base we are using Hyperwave [Maurer et al. 96a] as the core information system for a Web-based learning system. For the dynamic background system a combination of hierarchical Gatherers together with intelligent knowledge broker are built. In the next chapter the user requirements will be discussed in brief.

2 System Requirements from the Users' Perspective

Generally, parsing and indexing documents is a 100 percent objective process. However, present search attempts often lead to almost endless lists of ridiculous web sites, which contain the searched words but have nothing in common with the desired topics. The question that arises is whether 100 percent objectiveness is really useful. From our point of view it is much more efficient to let subjective human expert knowledge find its way into the system as well. Subjective in this context means classification of documents not only by topics, but also by quality, topicality, relevance for knowledge areas in question, etc. Subjectivity goes even further: users (courseware authors and learners) can define their own profiles including preferred language, areas of interest and skills, etc.

The introduced system HIKS combines document contents, extracted information and server or site information together with additional human expert knowledge and user profiles. This additional information is used to narrow down search queries and is presented as an indicator of relevance together with the query hits. The intelligent knowledge area brokers make decisions about relevance based upon objective and subjective criteria.

User profiles cannot only be relevant for queries and understanding of query results, there is another important point where profiles come into play - the use of background libraries itself. The use of background libraries (static and dynamic ones) is also a result of user feedback from previous projects. Query hits in highly specific knowledge areas very often contain words which the users simply do not know because the users are laymen in the appropriate knowledge area or do not know the language. Our first approach to solve this problem was the integration of searchable digital encyclopaedias and dictionaries into the system. Utilising the power of user profiles is a simple and efficient solution. Depending on the user's knowledge more adequate search results are returned. The current implementation of user profile management stores information given by the users themselves. Some future research will be done to implement a dynamic, self-learning user profile management system that adapts dynamically to the users' behaviour.

Besides profile management we have already discussed expert knowledge integration in this section. From the experts' point of view it has to be as easy as possible to insert categorisation and quality assessments into the system, otherwise knowledge integration is not applicable to huge systems. For this reason, HIKS provides different user interfaces depending on the user's current profile. Consequently, the interface presented to authorised experts vary totally from the one of the average reader. Whenever authorised experts of a knowledge area download a document in their area they automatically also get a set of additional buttons for their quality assessments including style of writing, relevance for the research area, topicality, etc. They also get the possibility to add textual annotations to the document. Readers may also alert other users, if the documents contain errors or present new results. Any information gained from the opinions of the experts then becomes part of the document classification yielding a self-expanding dynamic knowledge base in course of time. This knowledge base may support courseware authors as well as learners.

As a working prototype for a knowledge based information retrieval system, HIKS has to gather all information it can get from different information sources such as the document data itself, the meta data of the documents (e.g. title, author, creation time, language, knowledge area, etc.) and the data of the system (e.g. time of indexing, original document location, etc.). Further, it has to integrate human expert knowledge and make it part of the meta data. Query results and user interfaces have to be adapted dynamically according to the users' preferences and a rating of search hits has to be provided. The following chapters describe the specialised gatherer hierarchy and the intelligent broker system needed to achieve this.

3 Extended Information Gatherer

In order to provide a solid base for the intelligent knowledge area brokers we first have to collect as much information as possible about the indexed documents. The back end for achieving this is the extended information gatherer which is indexing documents from sites which are determined by authorised users of the learning system. Besides simply indexing documents the gatherer also extracts descriptions, context information and other relevant information (quality, language, etc.) if available, for later use with the intelligent knowledge broker. As it will be discussed later, the gatherer is also the collecting point for expert knowledge input. The gatherer itself is divided into several modules: content analyser, keyword extractor, description extractor and expert knowledge integrator.

First of all, gathered information is filtered by the content analyser which extracts meta information and recognises whether Java Applets, ActiveX objects, JavaScript, multimedia objects, etc. are embedded in a document. The content analyser also collects relevant standard data included in the documents (title, headings, etc.). Second step in information gathering is the keyword extractor. It builds a list of keywords describing a document. The keyword extractor also stores the context together with the keywords. This information can then be evaluated later in the intelligent knowledge broker to decide the relevance of the documents.

Keywords and their context are not always enough to describe a document. Therefore the gatherer also consists of a third module which gets even more relevant information out of a document: the description builder. It tries to find authors and abstract in the paper if available. This approach is based on the fact that most scientific papers contain this information either in a section called "Abstract" or have it in the meta-information. If nothing like that can be found the headings of the document are extracted as they provide more information than the usual first few lines of the document. Only if no headings can be found the system falls back on extracting an excerpt of several lines at the top. Further work will be done in the future, integrating AI modules to find proper descriptions.

Another interesting point to consider about site locations comes from document update rates. In course of time the gatherer modules learn how often the content of the sites changes. The description builder can also add information about update frequency to a document to provide the user with the knowledge whether a document could already be outdated and a more recent version could be found on the original site. That is why the gatherers' behaviour is adapted to the update rates: servers with a long mean document lifetime are visited more seldom than the ones which change their contents often.

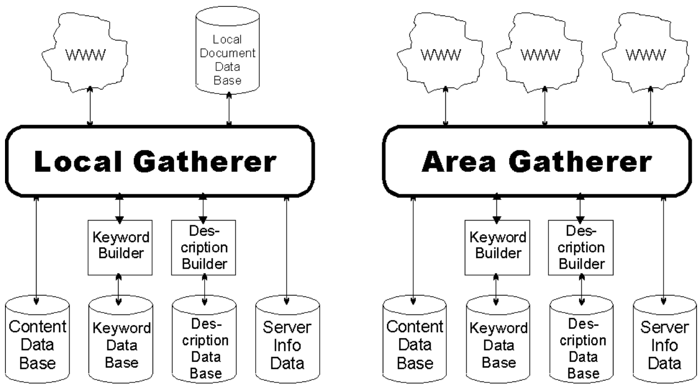

Figure 1: Architecture of the Information Gatherer

After the content of the documents is processed by the Gatherer the whole amount of information is ready to be handled by our intelligent knowledge broker which is described in the following section. Additional research efforts will be taken in the area of indexing not only Web sites but also several different kinds of databases, e.g. Oracle or Informix.

4 Intelligent Knowledge Broker

The intelligent knowledge broker in HIKS is the front end for presenting knowledge to the courseware authors and learners obtained periodically by its gatherers. The HIKS broker is also the channel for bringing expert knowledge back into the system. An adaptive interface provides facilities for creating user profiles, formulating search queries and adding supplementary expert knowledge. Further, there is not only one single broker in the system, but a hierarchical collection of several knowledge area brokers integrated. This concept allows building up proper knowledge spaces for particular courses.

As shown above ordinary search services lack the ability to handle more user information than just the search string or Boolean combinations of strings. This often leads to unwanted, frustratingly large and partly irrelevant results. By giving the system a user description (user profile), a query result can be narrowed down more precisely. For this reason the intelligent knowledge broker presents a form to the users where they can fill out their preferences. These include preferred language, foreign language knowledge, fields of interest, profession, qualifications, education, quality of expected document content, etc. In order not to be forced to fill out the form every time the user profiles are stored together with chosen nicknames for future use.

Present robots do not give information about a Web server, its owner, its purpose, location, etc. This often leads to questionable results. Our intelligent knowledge broker has access to additional information from the extended gatherers. This allows a better document categorisation by quality and relevance on one hand and semantic context on the other. The sites from universities, public libraries and governmental departments are categorised differently from private web sites. This concept allows much more exact search results.

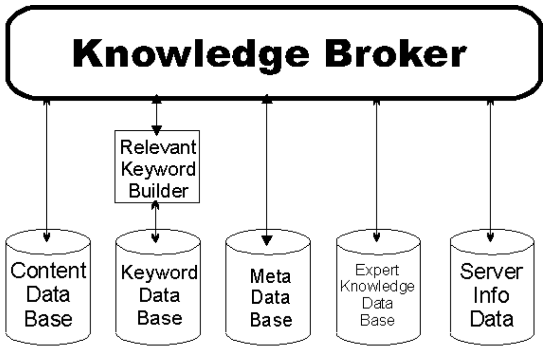

Figure 2: Architecture of the Information Broker System

A single knowledge area broker analyses keywords and contexts extracted by the gatherer to build relevance-based keyword lists. This can be done in quite a simple way as shown in the following example: let us assume a server dedicated to the research area of physics. Most of the documents will have the keyword "physics" somewhere in their body, therefore it cannot be considered a relevant description of the content of a document. For this reason the classification of search hits will not contain the keyword "physics" when found on this server. On the other hand, the keyword "medicine" will not be found in many documents, therefore it will be presented to the user to show that this document has some medical context. The threshold for considering irrelevant keywords is defined by the server administrator. This means that each knowledge broker will provide relevant keywords that depend on the whole set of documents.

It is obviously not enough to provide only keywords to describe documents, even if the keywords are considered relevant. The intelligent knowledge broker takes this into account by evaluating context and user profiles and providing additional features. A medicine student looking for "virus" is rather interested in articles dealing with virology than computer science. We are not trying to implement some a priori behaviour of the broker, because this would be a highly complicated artificial intelligence problem. Instead, HIKS implements a user dialog to make it easy to narrow down the result. In case of the "virus" example the medicine student is able to tell the system that having "virus" in the same context with "computer" is undesirable and the broker will narrow down the search. So the system may learn the users' fields of interest and will be able to make future decisions itself.

Besides keywords and keyword contexts the broker also handles levels of knowledge (beginners, advanced, experts) and desired quality of papers in very much the same way as described above. Users can simply tell the broker which level of quality they expect.

Although there are several approaches to automatic categorisation of documents, until now none of these systems has yielded satisfying results. In order to supply a high quality human expert knowledge it is necessary to classify documents. For this reason, the broker provides an interface allowing a distinct group of authorised experts in specific areas of knowledge to link quality assessments and annotations. Since it is mostly impossible simply to say that a document is good or bad, a more detailed and configurable quality assessment takes place. Authorised experts get "good", "medium" and "bad" buttons for different aspects such as style of writing, depth of knowledge, seriousness etc. as well as the ability to add textual annotations.

Instead of a simple search result list, a more logically ordered representation of information objects is often desirable. The context categorisation of the knowledge system can describe the relations between relevant documents in a hierarchically structured way. Consequently information retrieval can also be a step by step walk through knowledge hierarchies. Furthermore, additional information such as Web site description, keywords as well as quality information will be provided by the system.

5 HIKS as a Cascaded Distributed Network System

In order to provide a flexible, scaleable and efficient system it must be possible to build up a distributed network database instead of building one huge database on a single server. It is also desirable to import information from existing information gatherers and index servers or even treat them as subsystems. For this reason, knowledge brokers can be cascaded. HIKS is conceived for building up sub systems and putting them together to huge knowledge bases. One major advantage is that specialised topics and geographical areas can be built. For example, a campus-wide knowledge system could be installed which consists of several subsystems and itself represents a subsystem of a province-wide knowledge base.

This means that knowledge space related to a specific course could be provided for other organisation units, training organisations or institutes. The advantages of distributed architecture become visible. These are, for example, reduced gathering and indexing process that may lead to higher update rate of the indexes' databases and therefore to a higher grade of consistency between information supply and the brokers' indexes. Further work will include a pricing system for the interchange of such knowledge spaces. This concept will allow for example the co-operation between institutes and universities to provide particular knowledge and consume other knowledge spaces.

Another possibility would be for example to combine all departments of computer science from one country or even several countries to create a big specific knowledge topic. Distributed information systems cause less maintenance for each sub system and specific expert knowledge can be managed easier by local experts.

This concept also leads to further advantages. On the one hand each sub system gathers only a small amount of the whole huge information space and is therefore always able to provide up-to-date data. On the other hand, the introduced system reduces server and network load. The cascaded brokers than deliver all information transparently to the user. Such knowledge topics will also provide a perfect background library for courseware authors as well as learners.

Basically, HIKS is also able to include sets of information from other gatherers as well as it can present gathered information in common formats like SOIF, XML or MCF to already existing services. Consequently, already existing web based training systems could use HIKS by integrating the knowledge base into their own system.

6 HIKS and Web Based Training System Integration

As already mentioned, HIKS could be used as a dynamic background library in a learning environment. At present we are going to make some field experience by using HIKS combined with GENTLE, a Hyperwave-based Web based training system. GENTLE not only provides a set of lessons for learners, but also a discussion forum. GENTLE is also a question and answer system and an annotation system. HIKS as a dynamic background library allows each part of the GENTLE system to reference its background knowledge. HIKS will help courseware authors to produce new lessons or update such. Furthermore, HIKS will provide available knowledge which will help learners to solve problems, to enlarge the content of the lessons and to produce presentations.

7 Conclusions and Future Work

By adding dynamic background libraries we do not only present a profound additional knowledge base for learners but also a huge data collection for courseware authors to create new lessons. Apart from that, the contents of these libraries are automatically organised and updated. Some future research will be done to implement a dynamic, self-learning user profile management system. Improvements in the fields of context extraction and evaluation will be carried out. Further work will be done in the future integrating AI modules to find proper descriptions. At present we are going to work in the area of indexing not only Web sites but also several different kinds of databases, e.g. Oracle or Informix that are widely used to store information.

Acknowledgement

We would like to thank all members of the IICM for their suggestions and help. The research described in this paper was conducted as part of the IICM Knowledge Discovery Project, supported by the Austrian Federal Ministry of Science and Transport.

References

[Dietinger et al. 98] Dietinger, T., Maurer, H.: "GENTLE - General Network Training and Learning Environment"; Proc. of ED-MEDIA/ED-TELECOM98, ISBN: 1-880094-30-4, Freiburg, Germany, (1998), 274-280, http://wbt.iicm.edu/gentle/papers/edmedia98.pdf.

[Guetl et al. 98] Gütl, C.; Andrews, K.; Maurer, H. "Future Information Harvesting and Processing on the Web"; Presentation at European Telematics: advancing the information society, Barcelona, Spain, 4th – 7th February 1998 and http://www2.iicm.edu/cguetl/papers/fihap.

[IICM] Institute for Information Processing and Computer Supported New Media, Technical University Graz, Austria; http://www.iicm.edu.

[Kappe et al. 94] Kappe, F., Andrews, K., Faschingbauer, J., Gaisbauer, M., Maurer, H., Pichler, M., Schipflinger, J.; "Hyper-G: A New Tool for Distributed Multimedia", Proc. of IASTED/ISMM, Honolulu, (1994), 209-214.

[Marchionini et al. 95] Marchionini, G., Maurer, H.: "The Role of Digital Libraries in Teaching and Learning", Communications of the ACM 38, 4 (1995), 67-75.

[Maurer et al. 96a] Maurer, H.: "HyperWave: The Next Generation Web Solution", Addison Wesley Pub.Co..U.K. (1996).

[Maurer et al. 96b] Maurer, H., Scherbakov, N.: "Multimedia Authoring for Presentation and Education: The Official Guide to HM-Card", Addison Wesley Pub.Co., Germany (1996).

[Maurer et al. 97] Maurer H., Dietinger T.: "How Modern WWW Systems Support Teaching and Learning"; Proc. of International Conference on Computers in Education Sarawak Malaysia, (1997), 37-51.