Complex Operations in a Distributed, Network-Like Information Space. Resource Selection, Searching, Structuring and Visualization in an Electronic Market Place

Rainer Kuhlen, Marc Rittberger & Bernard Bekavac
Information Science Department
University of Konstanz
P.O. Box 5560
D-78434 Konstanz
{Rainer.Kuhlen | Marc.Rittberger | Bernard.Bekavac}@uni-konstanz.de

1 Introduction

In searching for information hypertext only offers tools for browsing or navigating through an information space. Obviously, one needs additional tools when interacting with large and growing hypertexts to select, identify and criticize information relevant to a user's needs. Traditional information retrieval techniques [Belkin & Croft 87] can be very helpful in hypertext environments to support explorative interaction limits. If we accept the need for browsing and navigating or chance serendipitist exploration, both hypertext exploration and information retrieval techniques together are necessary to fulfill the demand for finding information in the context of hypertext systems. Agosti [Agosti 88] introduces the combination as a new model for information retrieval. More clearly, one should talk about supplement or integration of information retrieval and hypertext, because tools for searching or navigating should not only be alternatives when entering a hypertext environment, but co-occur in and integrate the information gathering process [Frisse & Cousins 89, Fuhr 90, Croft & Turtle 93, Salton & Allan 93, Kuhlen 92, Dunlop & van Rijsbergen 93, Agosti 93, Agosti & Smeaton 96].

Large distributed data collections used in networked information spaces lead to new problems in retrieving data. These problems heavily depend on the distributed character of the data collections [Fox 93, Lynch 95, Janes & Rosenfeld 96] (c.f. Workshop NIR [Fuhr & Schweppe 96] at Schloss Dagstuhl). The most important subareas of networked information retrieval are resource discovery, searching and data fusion.

2 Project Environment

In the following we will present a project idea which depends on all three disciplines previously mentioned: standard information retrieval, hypertext and networked information retrieval. Also knowledge about value-added services and electronic market places is in the center of the project.

The aim of this project is to improve methods for searching, navigating and presenting information in regional, electronic market places. Therefore, we are not focusing on the World Wide Web itself or national/international information services such as Compuserve, America Online (AOL), Microsoft Network (MSN), T-Online, but instead on more specialized or defined market places restricted by geographical or content criteria. We will start with the 'Electronic Mall Bodensee', a market place located at the Lake Constance region containing information and services from all around the Lake Constance, e.g. from Austria, Germany and Switzerland. The 'Electronic Mall Bodensee' is one of the largest and best known market

places in the German speaking countries. The 'Electronic Mall Bodensee''s services and information servers are distributed around the Lake Constance region, on different WWW-servers, databases, etc.. One can find in each of its sources thousands of information units offered by hundreds of different suppliers.

3 Searching the Web

Identifying relevant information in an electronic market place or on the Internet seems to be becomimg more and more confusing. Therefore the question of how information in these distributed and unstructured information spaces should be found is gaining more and more interest. Just navigating through a network along the links or following some hierarchical organization or using search engines with their traditional retrieval techniques (see [Bekavac 96]) does not solve the problems identified:

- Search engines retrieve their information in so-called roboter-based searches. They recursively run through all web sites, index the information units and build inverted files. The search engines on the web are centralized services and cause a lot of network traffic.
- Search techniques on the web do not take into account the particularities of heterogeneous information sources and distributed information spaces.
- Search techniques on the web, which are based on inverted files do not take into account the network-like hypertext structure. Information about this structure is lost.
- Search engines on the web produce results with high redundancy. Each relevant unit of information is presented. For specialized questions, one can receive hundreds or even thousands of answers.
- Search engines finish their job by presenting the results of a query. There is no information about structural context available for a particular unit. The user receives no additional structural information about where to navigate or where to find similar answers.

During the proposed project, we will build a search service for a regional, distributed, and electronic market place, which will not present isolated information units as results of a search engine's answer set, but will offer a presentation in which users can find the information units embedded in their structural context. The user will see the hypertext structure e.g. of a specific application or company, which means that the hypertext structure is

- identified,
- visualized, and
- embedded into its context.

The following features have been identified as useful for the planned information service:

4 Resource Selection

Content analysis of distributed information spaces by describing the heterogeneous resources in an electronic market place helps to prescan the resources relevant to a user's query (see figure 1)

5 Search

Relevant information units will be searched in the preselected resources. Content analysis of the information will be performed by each information resource at the electronic market place. The content analysis will not be run by a centralized service (see figure 2).

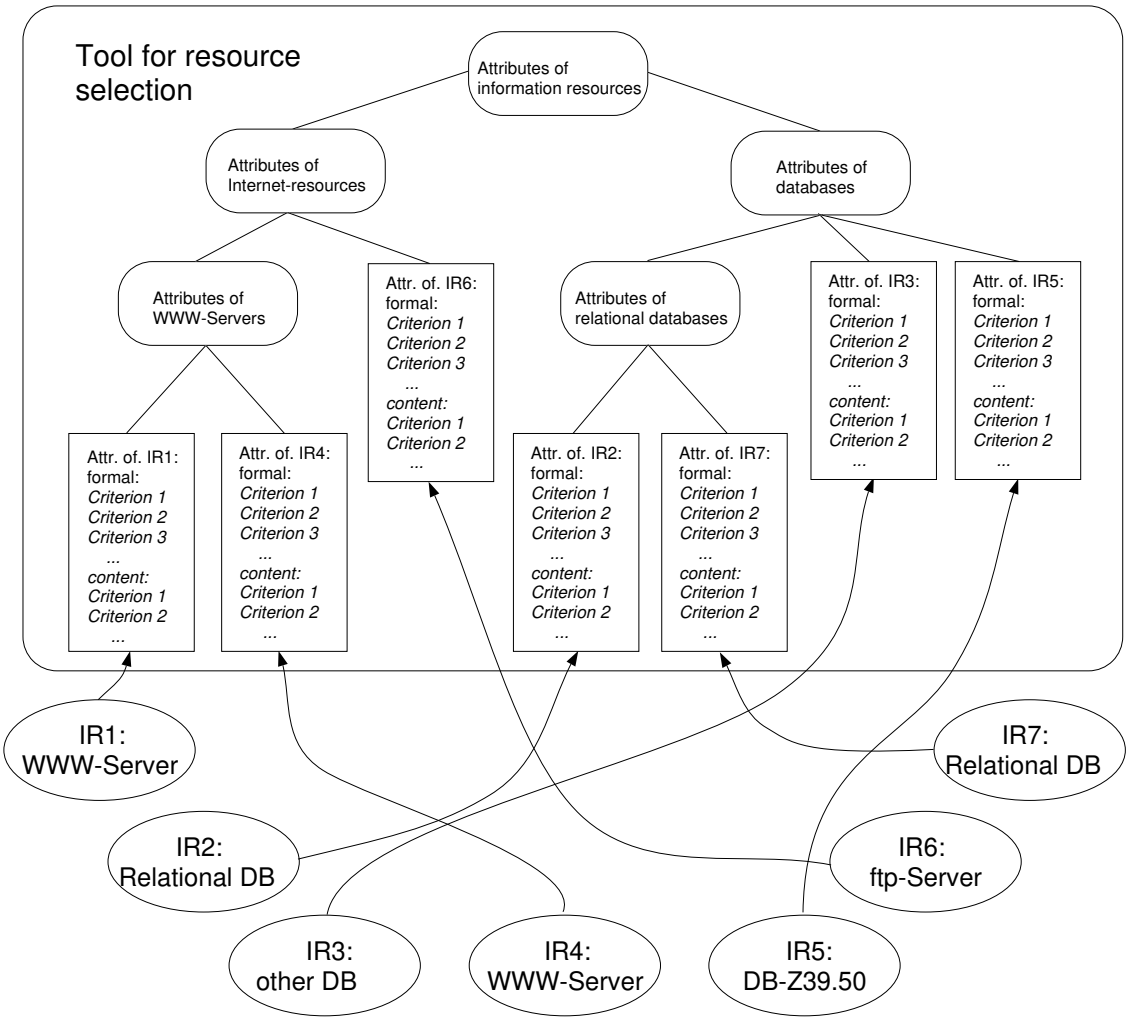

Figure 1: The description of the information resources will be taken from the resources and organized in the polyhierarchical Konstanz Hypertext System [Hammwöhner & Kuhlen 94, Rittberger et al. 94, Hammwöhner & Rittberger 96].

6 Structure

Provided with information about the structure, the system is able to build and visualize a network-like hypertext structure from the retrieved information unit relevant to the query. The information units presented at the visualization tool will be embedded into the context relevant to the query and connected to each other by links. The presentation of context-units can be based on hierarchical or associative information about the current application. If there are more sophisticated content analysis methods to identify the content of a unit or a hypertext they will be used for displaying the context information of the relevant units.

Furthermore, other topics relevant to the query, which have a similar structure will be identified. The results of the retrieval process are not hundreds of isolated information units, but hypertext structures through which a user can navigate to acquire needed information.

The identified hypertext structures will be dynamically visualized. A small application (e.g. a presentation of a hypertext document of a tiny institute) can be presented (see figure 3 for a Fahrrad (=bicycle) query), but in dealing with a large application with perhaps several dozens or hundreds of units, one has to use dynamic zooming, e.g. fish-eye techniques, etc..

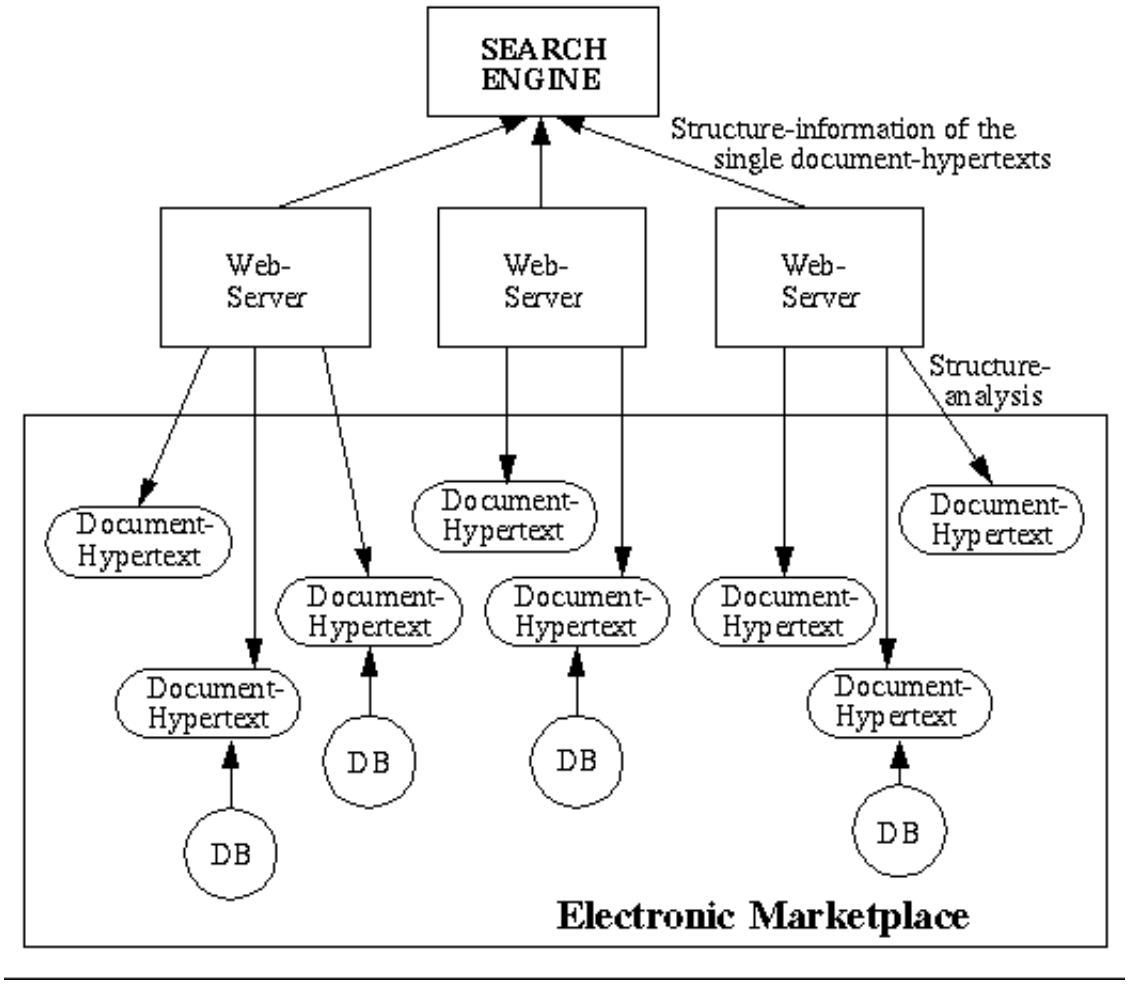

Figure 2: Extracting the structure

7 Context

Using the content analysis of each page, the information about the distributed resources, the information about the structure of the hypertexts, and the orientation facilities of the electronic market, the system should be able to offer other applications which might be relevant to the users particular information need. In our cycling example several applications in the context of cycling would be presented to the user with the help of sophisticated retrieval methods. If the user were to have a more general interest, additional information units and applications would be available and could be presented, e.g. sports wear, cycling tours, hotels, etc. The user could then make up his mind by participating in an interactive dialog, which we call interactive relevance feedback.

8 Conclusion

In the project, progress will be made in the field of networked information retrieval and hypertext, but we will also learn about behaviour and user interaction in an regional electronic market place. Because the Information Science Department at the University of Constance is deeply involved in developing the 'Electronic Mall Bodensee', it is an ideal place for creating such services, evaluating their use and observing the behaviour and benefits of the users.

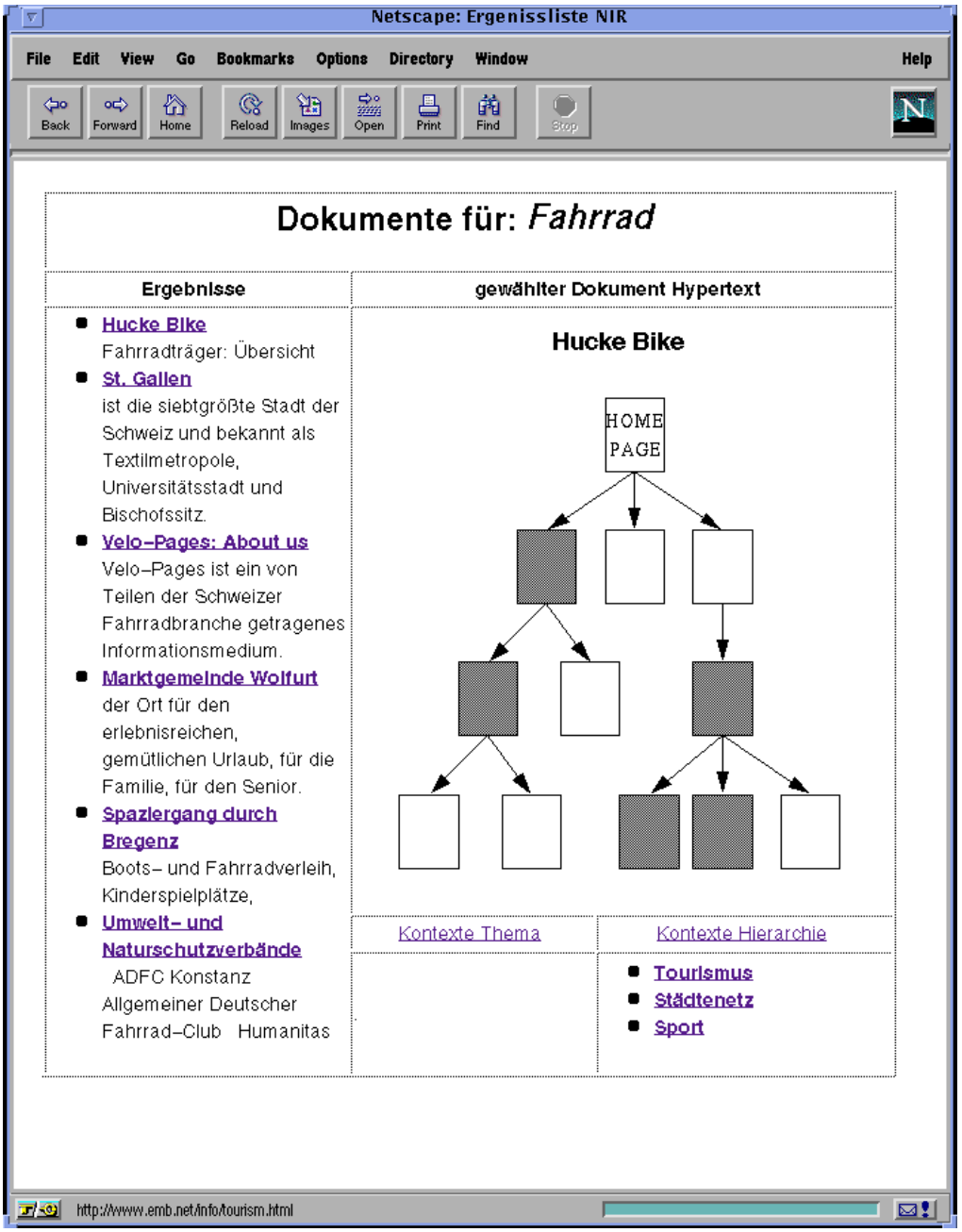

Figure 3: Visualization of the result of a query for the term 'Fahrrad' (=bicycle)

References

[Agosti & Smeaton 96]. M. Agosti and A.F. Smeaton. *Information Retrieval and Hypertext*. Kluwer academic publishers: Boston et.al., 1996.

[Agosti 88]. M. Agosti. Is hypertext a new model of information retrieval? In *Online information 88. Proceedings, 12. International Online Information Meeting*, vol. 1, pp. 57–62, 1988.

[Agosti 93]. M. Agosti (ed.). *Information Processing and Management, special issue: Hypertext and Information Retrieval*, 1993.

[Bekavac 96]. B. Bekavac. Suchverfahren und Suchdienste des World Wide Web. *Nachrichten für Dokumentation*, Vol. 47, No. 4 , 1996.

[Belkin & Croft 87]. N.J. Belkin and W.B. Croft. Retrieval techniques. *Annual Review of Information Science and Technology (ARIST)*, Vol. 22, pp. 110–145, 1987.

[Croft & Turtle 93]. W.B. Croft and H.R. Turtle. Retrieval strategies for hypertext. *Information Processing & Management*, Vol. 29, No. 3 , pp. 313–324, 1993.

[Dunlop & van Rijsbergen 93]. M.D. Dunlop and C.J. van Rijsbergen. Hypermedia and free text retrieval. *Information Processing & Management*, Vol. 29, No. 3 , pp. 287–298, 1993.

[Fox 93]. E. Fox. From information retrieval to networked multimedia information access. In G. Knorz, J. Krause and C. Womser-Hacker (eds.), *Information Retrieval'93. Von der Anwendung zur Modellierung. Proceedings*, pp. 116–124. Universitätsverlag Konstanz, 1993.

[Frisse & Cousins 89]. M.E. Frisse and S.B. Cousins. Information retrieval from hypertext: Update on the dynamic medical handbook project. In *Proceedings of the Hypertext '89 conference. (Pittsburgh, PA, 5.-8. Nov.)*, pp. 199–212. ACM: New York, 1989.

[Fuhr & Schweppe 96]. N. Fuhr and H. Schweppe (eds.). *Dagstuhl Workshop on Networked Information Systems Discovery, Retrieval, Dissemination*, 1996. http://www.inf.fu-berlin.de:80/inst/ag-db/dagstuhl/index.html.

[Fuhr 90]. N. Fuhr. Hypertext und Information Retrieval. In P.A. Gloor and N.A. Streitz (eds.), *Hypertext und Hypermedia. Von theoretischen Konzepten zur praktischen Anwendung*, Informatik Fachberichte 249, pp. 101–111. Springer-Verlag: Berlin, 1990.

[Hammwöhner & Kuhlen 94]. R. Hammwöhner and R. Kuhlen. Semantic control of open hypertext systems by typed objects. *Journal of Information Science*, Vol. 20, No. 3 , pp. 175–184, 1994.

[Hammwöhner & Rittberger 96]. R. Hammwöhner and M. Rittberger. Building application dependent hypertexts. *Information Processing & Management*, Vol. 32, No. 6 , 1996. to be published.

[Janes & Rosenfeld 96]. J.W. Janes and L.B. Rosenfeld. Networked information retrieval and organization: issues and questions. *Journal of the American Society for Information Science*, Vol. 49, No. 9 , pp. 711–715, 1996.

[Kuhlen 92]. R. Kuhlen. Hypertext und Information Retrieval - mehr als Browsing und Suche. In R. Kuhlen (ed.), *Experimentelles und praktisches Information Retrieval. Festschrift für Gerhard Lustig*, pp. 309–324. Universitätsverlag Konstanz: Konstanz, 1992.

[Lynch 95]. C.A. Lynch. Networked information resource discovery: an overview of current issues. *IEEE Journal on Selected Areas in Communications*, Vol. 13, No. 8 , pp. 1505–1522, 1995.

[Rittberger et al. 94]. M. Rittberger, R. Hammwöhner, R. Aßfalg and R. Kuhlen. A homogenous interaction platform for navigation and search in and from open hypertext systems. In *RIAO 94 Conference Proceedings. Intelligent multimedia information retrieval systems and management*, pp. 649–663, New York, 1994. Rockefeller University.

[Salton & Allan 93]. G. Salton and J. Allan. Selective text utilization and text traversal. In *Fifth ACM Conference on Hypertext. Hypertext'93. Seattle, USA*, pp. 131–144, 1993.