Information Processing & Management Vol. 29, No. 3, pp. 283–285, 1993
Printed in Great Britain.

0306-4573/93 $6.00 + .00

HYPERTEXT AND INFORMATION RETRIEVAL

MARISTELLA AGOSTI
Dipartimento di Elettronica e Informatica, Università di Padova,
Via Gradenigo 6/a, 35131 Padova, Italy

Ideas concerning the concept of hypertexts and information retrieval can be found as early as in the *Atlantic Monthly* of 1945 in the paper entitled "As we may think" by Vannevar Bush, in which the author envisages the possibility of storing and retrieving information by using techniques and capabilities that are proper to hypertext/hypermedia and information retrieval systems, but which are not all as yet effectively and completely available in present-day systems.

Research activities that are currently under way in the field of hypertext and information retrieval focus basically on the design and implementation of systems capable of providing the final user with the properties of the visionary Memex device and system of Vannevar Bush. These properties include:

- storing in the multimedia database of a large collection of textual and multimedia documents, and
- building a network of semantic relationships among the multimedia components of the database,

because the storage and automatic construction of such a database directly from the components of the documents of a network of semantic associations among pieces of information would give the hypothetical final user access to a large depository of knowledge for reading, browsing, and retrieving. Final users of such a device and information base would be given the possibility of satisfying their information needs by using concurrently the different retrieval techniques based upon: *value*, a technique present in a highly specialised way in the majority of existing database management and information retrieval systems; *content*, a feature chiefly available in operational information retrieval systems; and *direct association* for direct presentation of information on different forms of media (text, image, sound, etc.) as in available hypertext systems.

There are important aspects in relation to existing hypertext systems to be used as information retrieval tools, and there are drawbacks concerning use of hypertext systems as efficient information retrieval tools. Aspects to be considered in the design and construction of efficient hypertext information retrieval systems are as follows:

- *Navigation versus direct search: The need for new retrieval models.* The information retrieval modalities carried out by hypertext systems are different from those of traditional information retrieval systems; search is conducted by navigation through the information base, not by direct search by means of a search language. One argument for preferring navigation is the possibility the user is given to dynamically construct an information path by browsing through the pieces of the information base. One argument against this form of retrieval is that this method can be very time-consuming and poorly organised if the information base consists of a large text of multimedia database/hypertext.
- *The possibility to modify the status of a hypertext from passive to active.* This may be accomplished, for example, by attaching a sort of "level of importance" to a link in order to provide a different link relevance depending on the path the user is coming from. In this way the user could be advised to follow one path instead of another, depending on the previous path taken from a node or from an anchor.

This possibility could be implemented by establishing different types of links or using weights on links at construction of the hypertext.

- *The automatic authoring of an information base.* There are no tools available for automatic construction and updating of hypertexts starting from an initial "flat" collection of documents. The hypertext needs to be authored (i.e., built and updated) manually. If the initial collection of documents is of large proportions and consisting mainly of multimedia documents, it could be impossible to author the multimedia database/hypertext; it is therefore important to have tools for automatic generation of links and techniques for segmentation of documents.
- *The possibility to use different techniques for semantic representation of the information being administered in a concurrent manner.* An example is the opportunity to choose between two different techniques for indexing the same information base. This would provide, among other things, for the availability of different semantic interfacing capabilities for those user categories having different knowledge levels in the specific field dealt with by the document collection administered by means of the multimedia database/hypertext system.
- *User modelling and interfaces.* This includes the development of user modelling techniques to build effective interfaces for use of a multimedia database/hypertext.
- *A facility permitting different levels of abstraction for object representation.* This would allow the possibility, provided only on a few hypertext systems, to create nodes and links that would form structures allowing different levels of search depth.
- *The availability of techniques by which to evaluate hypertext information retrieval systems.* The evaluation techniques in information retrieval are not directly usable in the evaluation of characteristics of hypertext systems; it is therefore necessary to develop new procedures and tools that establish a relationship among present evaluation efforts to previous evaluation work in information retrieval, and at the same time be able effectively to evaluate new system capabilities.

Present research work in hypertext and information retrieval is addressing these different aspects, and the papers of this special issue are representative of the work being carried out in order to find solutions to these problems.

The first three papers of this issue focus mainly on the central aspect of developing new retrieval models for a hypertext information retrieval system.

The first paper, "Hypermedia and Free Text Retrieval," by Mark D. Dunlop and Keith van Rijsbergen, presents a hybrid approach combining the capabilities of browsing and querying to retrieve information from a large textual and multimedia collection of documents. Initially, the paper concentrates on the use of contextual information for the retrieval of non-textual documents; the model proposed is presented and its validity is tested by carrying out two experiments that use a text document collection to relate the results to previous findings of the information retrieval area; this part of the paper gives some insights into the development of evaluation techniques for hypermedia systems being used as information-seeking tools. In addition, some more general results from a combination of query-based and browsing-based retrieval are addressed in the paper.

The second paper, "Information Retrieval from Hypertext: An Approach Using Plausible Inference," by Dario Lucarella and Antonella Zanzi, presents a model and architecture of a prototype that combines query-based and browsing-based retrieval methods. The model is based on plausible reasoning, where the hypermedia collection of documents works as an inference network. Within this model the links are labelled by the name of the relationship that exists between the two connected nodes, and a weight is associated to the link to express the strength of the relationship; this is one way of modifying the status of a hypertext from a passive to an active one. Experimental results give some insights into the capabilities of the model.

The third paper, "Retrieval Strategies for Hypertext," by W. Bruce Croft and Howard R. Turtle, deals with the problem of evaluating hypertext information retrieval systems. The paper uncovers the relationship between information contained in hypertext links and improvement of retrieval effectiveness. The results can be useful in the development of a

new hypertext information retrieval model that has to incorporate links and enable automatic construction of links. In fact, the paper concentrates on the comparison of performance of the strategies used in two retrieval models: a probabilistic retrieval model incorporating interdocument links with strategies that ignore the links and the heuristic spreading activation strategy. The presented findings show that a hypertext retrieval model based on inference networks is just as effective as spreading activation, but further work is in progress to support multimedia documents with complex structures.

The next two papers delineate the development of user-based modelling techniques to build effective interfaces that make use of hypertext capabilities.

The first of these two papers, the fourth paper of the issue, "BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval," by Nicholas J. Belkin, Pier Giorgio Marchetti, and Colleen Cool, presents the design of an interface supporting BRowsing And QUEry formulation (BRAQUE) to a large bibliographic information retrieval system. The interface scheme is based upon a progressive development of the capabilities that the final interface is going to have; the framework for the interface is articulated on the basis of an information-seeking strategy model (ISS), a cognitive task analysis (CTA), and a two-level hypertext model for information systems. The paper makes reference to research work currently under way, since the design is going to be translated in a prototype of an operational system that will be evaluated, but the interface design is valid on its own, owing to the fact that it represents the relationships that will be established between hypertext research and previous research in information retrieval to produce new tools for the user.

The second of these papers, that is, the fifth of the issue, "A hypertext-based thesaurus as a subject browsing aid for bibliographic databases," by Richard Pollard, reports on a work that provides an online thesaurus as an interface that helps the final user in his or her information search; the thesaurus is presented to the user as a browsing interface implemented through a hypertext. Through this interface the user will use the information stored in a bibliographic database. The paper presents the design and implementation of the interface established by using a commercially available hypertext system.

The sixth paper of the issue, "Retrieval Hierarchies in Hypertext," by Roy Rada, Weigang Wang, and Alex Birchall, presents the MUCH (Many Using and Creating Hypertext) system developed to permit a group of people to collaborate and reuse information. The system supports the traditional functions of all operational information retrieval systems. In addition to these functions, it supports an information search method that has been called "heuristic" by the authors. A heuristic search for information is based upon a facility that permits different levels of abstraction for the representation of documents and terms, and consequently allows different levels of search depth.

The seventh paper, "Concept-based Retrieval of Hypermedia Information: From Term Indexing to Semantic Hyperindexing," by Hans C. Arents and Walter F.L. Bogaerts, reports another hypertext system, the Active Library on Corrosion (ALC) system. This system is a CD-ROM based hypermedia corrosion information system designed as a standalone hypermedia corrosion handbook for a corrosion engineer; therefore the document collection is highly specialised and well determined. Both approaches to concept-based retrieval implemented in the system rely on the use of thesauri to support browsing search operations; some considerations are given in case the collection to be used should be larger and less specialised.

Acknowledgements—This Special Issue would not have been possible without the effort of the authors in preparing different versions of their papers. Many of them have kindly and patiently waited for the issue to be assembled. Many reviewers have carefully read and re-read the papers, making useful comments and constructive criticism. I would like to thank all of them for having agreed to take part in this effort.