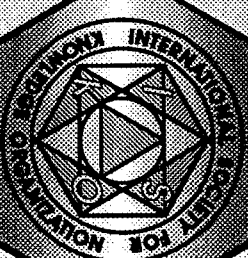


Advances In Knowledge Organization (AKO)

3

International Society for  
Knowledge Organization



**Documentary  
Languages  
and Databases**

**Advances in Knowledge Organization**

Vol. 3 (AKO-3)

published by the

**International Society for Knowledge Organization**

**Editorial Board**

Daniel Benediktsson, Reykjavik, Iceland

Robert Fugmann, Idstein, Germany

Roland Hierppe, Linköping, Sweden

Barbara Kelm, Frankfurt, Germany

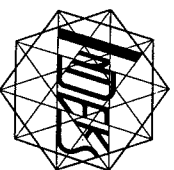
Norbert Meder, Köln, Germany

Roy Rada, Liverpool, England

Yulius Shrejder, Moscow, SSSR

Elaine Svenonius, Los Angeles, USA

Rudolf Ungvary, Budapest, Hungary



ISSN 0938-5495

Advances in Knowledge Organization, Vol.3(1991)

**Documentary Languages  
and**

**Databases**



**Papers  
from the Rome Conference  
Dec.3-4, 1990**

**organized by the**

**Istituto di Studi sulla Ricerca e Documentazione Scientifica, CNR**

**edited**

**by**

**Gliola NEGRINI, Tamara FARNESI**

**Daniel BENEDIKTSSON**

**Frankfurt/Main**

**INDEKS VERLAG 1991**

Maristella AGOSTI

Department of Electronics and Informatics, University of Padua

**USAGE OF CURRENT HYPERTEXT SYSTEMS AS POWERFUL INFORMATION RETRIEVAL TOOLS**

**Abstract:** The fundamental functions of a hypertext system are those regarding handling of non-linear or non-sequential documents. In fact, within a "hypertext" approach one or more documents are stored and managed within a network. If one wishes to view and therefore enquire two nodes of this network in sequence, these are normally connected by means of a link; each sequence of nodes connected together make up what is called a path, which enables exploration of the hypertext.

The usage of hypertext systems in the handling of a collection of documents is one which is normally used; this paper tackles and illustrates the various problems connected to the usage of current hypertext systems as powerful information retrieval tools. Thus the problems relating to the difficulties of using these systems are brought forward, not only from the point of view of people who have to administer a collection of documents, but also in the quality of administrators of networks of concepts and of the interconnections existing between a hypertext and a network of concepts.

**0. Introduction**

The main function of a hypertext system is that which carries out handling operations on non-linear or non-sequential documents. In fact, in a hypertext approach one or more documents are stored and handled within a network; each node of the network contains a part of the document or the entire document. The parts of documents or entire documents which are stored within the nodes can even be text type documents, graphs, or even audio-visual or other forms of data.

If one wishes two specific nodes to be viewed and therefore examined in sequence during use of the same hypertext by the user, the two nodes must be connected by a link; each sequence of nodes connected together by a link make up a path for a prospective exploration or "navigation" along the different nodes. If from a node there are various connections departing, each connection becomes associated to a small part of the node: an anchor. The user who activates an anchor and follows along the associated link actually navigates through the network links existing between the various documents or parts of documents.

The global collection of nodes and links among them make up what we call the hypertext.

The system used for hypertext handling operations is usually and effectively used for handling a collection of documents; this paper illustrates and tackles the problems connected with the use of the hypertext systems currently

available in their quality of powerful information retrieval tools.

### 1. Basic functions of hypertext systems

A hypertext system is a software instrument which is able to handle a hypertext in the same way in which a database system keeps the database as an integrated set of data. The hypertext system is therefore equipped with all the necessary functions for building, handling (updating, modifying and deleting) and consulting a hypertext.

In order to permit creation and administration of a hypertext, the hypertext system must be equipped with the following functions (8):

- a hypertext system must be able to allow and make it easy for the user to navigate along through the various information nodes and therefore it must have provision for a certain set of navigation instruments and mechanisms;
- owing to the fact that navigation operations, even within a vast hypertext database, are so easy for the hypertext user to handle, the user is likely to be persuaded to make intensive use of the navigation function and will soon lose track of his current position; consequently, hypertext systems normally have provision for functions which:
  - a) yield a graphical synthesis of the path the user has taken across the hypertext nodes, by means of appropriate synthesis diagrams; normally synthesis diagrams are known as "overview diagrams";
  - b) permit backtracking along the path followed through the various hypertext nodes that have been examined; in this manner the user who has lost his orientation may be able to return back to a previously examined point of the hypertext. This function is normally known as a "backtrack facility";
  - c) provide the user with a list of all the nodes he has examined during interaction with the system; the nodes are shown in reverse order in the list, i.e. from last to first, due to the fact that even if a hypertext is of non-sequential nature, the user tends to place in a time sequence all the information received even during use of the hypertext; this facility, upon request, allows the user to describe the history of his interaction with the hypertext system, and is generally called the "history list".

### 2. Application areas of hypertext systems

We realised that there are numerous areas of application in which use of a hypertext system could be experimented(9); we decided to present here the fields of application that have been taken into account and put into actual practice, for use of hypertext systems as an information retrieval tool.

Application areas dealt with by current systems:

- creation and management of dictionaries: the linking capabilities of a hypertext system are used mainly for establishing lexicographical links; this specific kind of application has been examined precisely for the intrinsic nature of dictionaries which is based upon a considerable number of cross-references existing between terms;
- electronic encyclopaedias: a normal encyclopaedia is used by consulting an entry connected to a specific starting entry. These entries are, in most traditional encyclopaedias, listed at the end of the starting entry. Using an encyclopaedia by a hypertext system means to bring about the advantage of establishing a direct link between any entry and all the others connected to it, by activation of the links existing between these entries;
- medical texts: a doctor, while formulating a diagnosis, bases his analysis not only on his specific competence but also on the consultation of a considerable amount of prior events or interpretations of other doctors. Handling of this category of information by means of a hypertext tool would allow quicker and more accurate consultation operations if the hypertext is sufficiently extensive, continuously updated and correctly planned to carry out all the links most probably relating to doctors' information needs;
- product catalogues: it is very easy to imagine the way in which a catalogue of production items could lend itself to an automatic handling by means of a hypertext system. In fact, consultation of a catalogue is normally carried out in a hypertextual manner and not in sequential order. Ordinary catalogues are normally organised according to a product type concept, but which usually involve a subsequent search, after having traced down a specific product contained in the product category, in a typically non-sequential fashion, in order to find other auxiliary products connected to the first main product. For example, when consulting a catalogue of personal computer products, we can imagine that after having identified a computer of the desired characteristics, the ease with which the characteristics and costs of peripheral devices, such as a mouse, an appropriate video display, the suitable disks, etc., can be examined. In this case, the needs of the person consulting the catalogue is that of finding information directly relating to products which are normally listed following a strictly sequential organisation;
- help facilities: any person who has used a computer occasionally has at some time used an on-line guidance software facility. The most recent systems are organised according to a hypertext-type philosophy; this means, help is given in tight relationship to the specific environment the user was working on at that moment. Depending upon what just happened and in relation to the kind of help requested by the user, a link displaying different kinds of information is therefore activated;
- technical documentation: as for an encyclopaedia, even for technical documents illustrating sophisticated products, a hypertext version of the written documentation can be made

available in order to permit easy access to specific reading paths;

- teaching: the experiments that have been conducted concerning use of hypertext teaching text books are very promising, due to the fact that they seem to place a certain emphasis on learning by association of ideas;

- software engineering: the aspects of planning and creation of software tools of a certain complexity always involves preparation of a set of specific technical documents which must be kept constantly up-to-date as the planning and creation operations proceed. Hypertext technology seems to be most useful in that it allows adequate development and updating to be carried out on the necessary technical documentation;

- fairs: visits to a fair can be accompanied by consulting an appropriately planned hypertext. In the same way, a visit to a museum can be presented by showing the different possible routes which can be taken depending upon the visitor's interests and prior competences.

Following this brief outlining concerning the possible and specific applications to information retrieval operations with the current hypertext systems, it is now possible to point out the most important aspects displayed by hypertext systems and by information retrieval hypertexts.

### 3. Important aspects concerning information retrieval hypertext systems

Those aspects of current hypertext systems which are most important in information retrieval operations are:

- 1) the different information searching and retrieval modalities which are carried out by the hypertext system;
- 2) the possibility, provided in only a few hypertext systems, to create nodes and links which would form structures allowing different levels of search depth; in general, a facility permitting different levels of abstraction for object representation;
- 3) the possibility to modify the status of a hypertext from passive to active, for example by attaching levels of importance to the links in order to provide different relevance to the different links from a node or from a specific anchor.

### 4. The different roles which a hypertext system can perform

Hypertexts can be used for information retrieval with different purposes and roles:

- A) a hypertext system can be used for storing and handling a collection of non-linear documents;
- B) a hypertext system can also be used for concept/term

semantic administration purposes, in which the informative contents of the collection of documents are represented. As a consequence, the hypertext system can be used not only for storing and handling a collection of documents, but also to administer a semantic structure of concepts to be used in the representation of the domain pertaining to the collection of documents. The possibility provided for handling a network of links existing between concepts can be used, for example, to administer the semantic structure of concepts of a classification scheme or of a thesaurus. A similar dynamic-structured concept handling scheme has been called a hyperconcept database (1, 2, 5);

C) handling of the links (the semantic and structural relationships) existing between the structures of the terms represented within the hyperconcept database and the relevant documents administered by the hypertext.

As far as the planning aspects of the hypertext are concerned, normally referred to as "authoring" by the current literature of this sector, there are certain planning principles which make particular back-reference to the way in which the various pieces, or "chunks", of the hypertext must be structured and how they must be represented. At the present moment, however, no uniform method is adopted for this purpose. It must be kept in mind, though, that a prototype planning scheme has recently been presented in (6, 7) as an extension to a previous methodology proposed for the specific field of information retrieval (4).

Another difficulty related to the realisation of a hypertext concerns its production. Nowadays hypertext production is practically done in a manual way, because it is necessary to actually specify the various links between the parts of text to be linked; the difficulties connected to automatic production of a hypertext is still an open problem. Similarly, the difficulties related to the automatic updating of a hypertext are still to be entirely solved. In actual fact the need to update the information handled by a hypertext is inherent in the very opportunity to administer a document collection through use of a hypertext system. To the present moment the links between the new items of information and those already administered by the hypertext system must be established manually, as no automatic handling tool yet exists. It is needless to say that this unfortunately hampers rapid updating of a hypertext.

### B) Storing and handling a term structure

The storage and handling of a term structure by means of a hypertext system is certainly a new concept in this particular field. In fact, paper (1) illustrates the very first proposal to be made, and in (2, 5) it is treated in further depth.

The term structure which is intended to be administered through a hypertext system must be handled in a concurrent manner together with the collection of documents. In actual fact this term structure must illustrate and give a transparent appearance of the semantic content of the various different

documents in the collection to the final user. This means that the term structure represents the document collection domain.

C) Handling the relationships between the document collection and the term structure  
 Another role that a hypertext system ought to be capable of carrying out, in order to be effectively used as an information retrieval tool, is that of establishing the necessary links between the collection of documents and the term structure representing the knowledge concerning the domain directly related to the collection being administered.

This implies that the hypertext system must be capable of administering the links existing between the nodes, which represent the different information types. This role is not usually implemented in the systems currently available.

#### 5. Hypertext systems of new concept: development requirements

From the preceding comments it appears clear that the current hypertext systems tend to present many drawbacks, too many to be used effectively as efficient information retrieval systems.

In particular the negative aspects which at present seem to require most urgent attention before proceeding in the planning and creation of hypertext systems of new concept are as follows:

- planning new types of links and nodes, so that the system might be capable of adequately performing the three different roles described previously. In particular, the need to make provision for node types capable of administering the terms of the semantic structure representing the domain for the automatic application to be administered by means of the hypertext system in an active manner. It is also necessary to have different kinds of nodes in order to distinguish the link between two nodes containing parts of text, or two terms of the semantic structure, or even between one node containing part of a text and another node representing a term which illustrates the semantic content of the node itself in a partial manner;
- provision for these different types of nodes and links should also allow the creation of a term schema capable of evolving dynamically together with the evolution of the automatic application package;
- the possibility for the end user to retrieve information even with more highly sophisticated retrieval techniques (querying versus browsing);
- the possibility to use the different techniques for semantic representation of the information being administered in a concurrent manner. For example, the opportunity to use two different techniques for indexing the same document collection concurrently. This would permit, among other things, the creation of different semantic interfaces for those user categories having different knowledge levels in the specific field of the document collection administered by means of the hypertext.

#### References

- (1) Agosti, M.: Is hypertext a new model of Information Retrieval? Proc. 12th Int. Online Information Meeting, Learned Information, Oxford, 1988, Vol. I, p.57-62.
- (2) Agosti, M., Gradenigo, G., Mattiello, P.: The hypertext as an effective information retrieval tool for the final user. In: Martino, A.A. (Ed): Pre-proceedings of the III International Conference on Logics, Informatics and Law, Vol. I, Firenze, 1989, p.1-19
- (3) Agosti, M., Archi, A., Colotti, R., Di Giorgi, R.M., Gradenigo, G., Inghirami, B., Mattiello, P., Nannucci, R., Ragona, M.: New prospectives in information retrieval techniques: a hypertext prototype in environmental law. Online Information 89, Proc. 13th International Online Information Meeting, London, England, 1989, p.483-494
- (4) Agosti, M., Crestani, F., Gradenigo, G.: Towards Data Modelling in Information Retrieval. Journal of Information Science, 1989, 15 (6), p.307-319
- (5) Agosti, M., Crestani, F., Gradenigo, G., Mattiello, P.: An approach for the conceptual modelling of IR auxiliary data. Ninth Annual IEEE International Phoenix Conference on Computers and Communications, March 21-23, 1990, Scottsdale, Arizona, p.500-505.
- (6) Colotti, R.: Hyperlaw: Prototipo per un sistema ipertesto per il diritto dell'ambiente. Rapporto, Ottobre 1990, 24p.
- (7) Colotti, R.: Hyperlaw - sistema ipertesto per il diritto dell'ambiente. Guida su Hyperlaw. Rapporto, Novembre 1990, 11p.
- (8) Nielsen, J.: The art of navigating through hypertext. Communications of the ACM, 33(1990)No.3, p.296-310.
- (9) Schneiderman, B., Kearsley, G.: Hypertext hands-on! An introduction to a new way of organizing and accessing information. Addison-Wesley, Reading, MA, 1989.