

# Concept-based indexing and retrieval of hypermedia information

Hans C. Arents and Walter F.L. Bogaerts

Materials Information Processing Systems (MIPS) group  
Department of Metallurgy and Materials Engineering (Dept. MTM)  
Katholieke Universiteit Leuven, W. de Croylaan 2, B-3001 Leuven, Belgium  
E-mail: [Hans.Arents@mtm.kuleuven.ac.be](mailto:Hans.Arents@mtm.kuleuven.ac.be), [Walter.Bogaerts@mtm.kuleuven.ac.be](mailto:Walter.Bogaerts@mtm.kuleuven.ac.be)

## Abstract

The key to unlocking the information retrieval potential of hypermedia systems lies in the design of effective indexing structures for the multimedia documents stored in these systems, and in the development of appropriate retrieval mechanisms which use these index structures to overcome the shortcomings of the basic hypertext navigation mechanism. In this report, we review in detail a number of different approaches to the concept-based indexing and retrieval of hypermedia information, focusing in particular on the organization of the underlying index structure and on the visualization of the query and retrieval process. We give a number of examples for each indexing and retrieval approach, and we conclude by pointing out a number of open research problems.

## I. Introduction and background

Hypermedia systems are capable of storing and presenting vast amounts of multimedia documents, that are densely interconnected by a rich variety of hypertextual links. The great challenge however is to make this wealth of information and this richness of interconnections more effectively accessible to the user of the system. Historically, most hypermedia research efforts have concentrated on addressing implementation issues and technical problems, such as building robust and reliable linking mechanisms, or satisfying the timing and synchronization demands of digital video data. Only recently have hypermedia researchers begun to address the far thornier issues of how to better support the user during the navigation and retrieval of hypermedia information. The present reliance on “leap and look” navigation (or “browse till you get bored”) as the basic hypertext access and retrieval mechanism leads to the well-documented usability problems of “cognitive overhead” and “navigational disorientation” (1), and to the “embedded digression problem” and the “art museum syndrome” (2). For an in-depth discussion of these usability problems and their solutions we refer to the overview of Gygi (3).

These serious navigation and retrieval problems are usually addressed through the design of more intuitive user interfaces (4), or through the development of better navigation tools (5). Such hypermedia system improvements include offering “bookmarks” and “breadcrumbs” (6), incorporating tools such as active “agents” or “guides” that assist the user during the navigation process (7, 8), and developing ever more refined graphical “browsers” or “maps” which present to the user a clearer visual overview of the nodes in the hypernetwork (9, 10). Usability research has shown however that such system improvements, although clearly useful because of their beneficial influence on the overall usability of the hypermedia system, do not offer a truly satisfactory solution (11, 12).

In this report we do not discuss these usability problems, but we focus instead on the research efforts devoted to extending the basic hypertext navigation mechanism by introducing powerful concept-based indexing and retrieval mechanisms for hypermedia information. The structure of this report is as follows: in section 2 of this report we present a system architecture that will allow us to better grasp the issues involved in the indexing and retrieval of hypermedia information. In section 3 we discuss the various concept-based index structures that have been proposed, as well as the various mechanisms that have been developed for the acquisition of these indices. In section 4 we review three fundamentally different mechanisms for the concept-based retrieval of hypermedia information, and we discuss the various efforts that have been made to try to visualize the now most widely used mechanism: query by navigation. In the last section of this report we draw some conclusions about the results that have been achieved so far, and we point out a number of open research problems.

Note that in this report we assume that the reader is already familiar with the basic concepts, tools and techniques relating to hypertext and hypermedia. The readers for whom this is not the case are referred to the bibliography at the end of this report which contains references to a number of excellent books and review articles.

## **II. Indexing and retrieval of hypermedia information**

The field of information retrieval has focused historically on the development and evaluation of retrieval models for text documents, such as those found in bibliographic or full-text databases. These retrieval models specify index-based retrieval mechanisms for comparing documents with a given query, typically resulting in ranked output. For a detailed discussion of these retrieval models, we refer to the recent overview of Turtle and Croft (13).

Hypermedia systems however consist of very flexible organisations of multimedia documents, connected through a variety of user-specified or system-generated links. As a result, they emphasize the use of retrieval mechanisms that are based on *navigation* in this hypernetwork of nodes and links, whereby retrieving information consists of scanning, browsing, searching, exploring or simply wandering through the hypernetwork (14). In recent years it has become increasingly clear that improving the retrieval effectiveness of hypermedia systems will require integrating conventional index-based retrieval mechanisms with these navigation-based retrieval mechanisms. The problem then becomes how to build in support for these index-based retrieval mechanisms into the basic hypermedia system architecture. We first briefly review this basic architecture, and we then discuss how this architecture can be extended to provide support for index-based retrieval of hypermedia information.

### A. Hypermedia without index-based retrieval

A hypermedia system is essentially a union of a data management system (containing multimedia documents and the hypertext links between these documents), and a user interface (that is used to navigate through the network of links and to consult the documents). The corresponding layered system architecture is shown in Figure 1. In the bottom layer we can make a distinction between the *database* (the collection of multimedia documents) and the *linkbase* (the collection of hypertext links), which together constitute the *hyperbase*. In the top layer we can make a distinction between the *navigation manager* (responsible for controlling navigation) and the *presentation manager* (responsible for presenting information), which together constitute the *interface* of the system. Using this layered architecture, the hypermedia system can basically only provide support for navigation through the links that have been made between the multimedia documents, and for the presentation of the multimedia documents themselves.

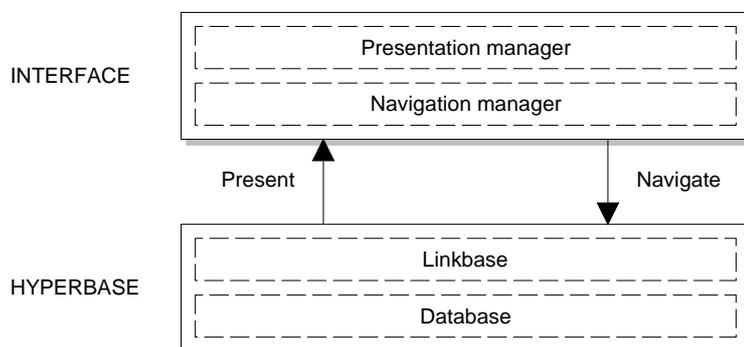


Figure 1. The basic system architecture of a hypermedia system.

## B. Hypermedia with index-based retrieval

It was not obvious to the designers of the first generation of hypermedia systems that a need existed for a separate storage of index information, which would enable a better query and retrieval of the hypermedia data itself. The first to introduce a distinction between the *data* (the contents of the nodes in the hypernetwork) and the *paradata* (the data that is used to index the nodes' contents) in hypermedia systems was Agosti (15, 16). She proposed a system architecture (17) which associated documents to concepts expressed by index terms. A similar model was later advocated by Bruza and van der Weide (18), who proposed to separate the hypermedia data into a bottom *hyperbase* and a top *hyperindex*. In another, similar model proposed by Lucarella (19), the hypernetwork was also organized as a layered structure, here consisting of a *document network* and a *concept network*.

Fundamentally, all these different proposals can be captured in an extended system architecture, which includes an additional layer responsible for the indexing of the multimedia documents (Figure 2). In this additional layer, we can make a distinction between the *index elements* (the items used to index the multimedia documents) and the *index structure* (the organization of these index elements), which together constitute the *index space* on top of the hyperbase. Using this index space, it now becomes possible to provide support for conventional index-based retrieval mechanisms. The challenge however is to design new indexing methods, and new retrieval mechanisms using those indexing methods, that integrate better with the basic navigation mechanism of hypertext.

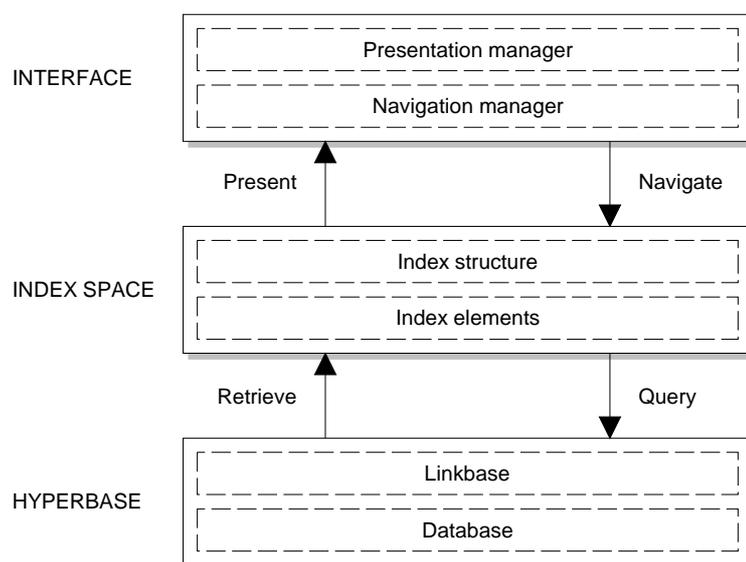


Figure 2. The extended system architecture of a hypermedia system.

### III. Concept-based indexing of hypermedia information

From the above discussion it should have become clear that developing new indexing methods for hypermedia information revolves around two important design decisions: what will be used as index elements, and what will be the index structure defined between those index elements. In hypertext and hypermedia information retrieval, each node in the hypernetwork is generally assumed to contain a single unit of information. In the majority of the present generation of hypermedia systems, indexing of the contents of the nodes is done by using simple index terms, i.e. *keywords* representing (part of) the meaning of the node's contents. Information retrieval using these index terms consists of formulating a query, which specifies the desired node(s) using a subset of the index terms and Boolean operators. This specification is then matched against the index terms that were attributed to each node. The corresponding nodes are retrieved and then offered to the user as the starting points for further hypertext navigation. This keyword-based retrieval is conceptually a very simple retrieval model, and one that is familiar from information retrieval in classic, non-hypermedia information systems. However, many researchers have contended that this simple indexing technique is too limited to capture the full richness of hypermedia, and have argued in favour of using more domain knowledge in indexing the contents of the multimedia documents and indexing the meaning of the hypertext links (20, 21, 22). Recent hypermedia systems have seen a move towards the use of *concepts* instead of keywords as index elements. Concepts differ from simple index terms by the fact that they are not independent, unrelated keywords, but are part of a larger index structure that is used to capture and represent knowledge about the contents of the hypermedia information and its possible uses. We review a number of index structures for concept-based indexing that have been proposed recently, and we also discuss the various approaches that have been developed for the computer-assisted acquisition of such indices.

#### A. Index structures

In order to be able to do something useful with the concepts used to index the hypermedia information, some type of formal *index structure* has to be imposed on these concepts. Such an index structure has to try to capture and represent what the hypermedia information is about and what it can be used for, but it is also to be used as a search and retrieval structure (see section IV). An index structure encompasses both the internal structure defined for the concepts (their attributes, allowable attribute values, etc.) and the external structure defined between these concepts (the relations that are defined between the concepts, how these relations are represented, etc.). The design requirements that such a concept-based index structure has to fulfil are very demanding (23, 24):

- the index has to find a correct balance between representational power and practical usability: it should capture the useful semantics of the hypermedia information, while remaining understandable by the users
- the index has to represent a model which is similar to the user's model of the hypermedia information: it should correspond with the user's view on the contents, reducing the effort needed to understand and use it
- the index has to be more intelligent than the contents of the hypermedia information itself: it should be able to capture and represent every possible navigation path the user might want to take through the contents

Several concept-based index structures have been explored these last couple of years, each trying to address these difficult demands in different ways. Some index structures are familiar from classic information retrieval (thesaurus, faceted thesaurus, concept lattice), some have been developed specifically to index hypermedia information (hyperindices, semantic hyperindices) and some have been derived from A.I. knowledge representation formalisms (inference network, semantic network). We discuss each of these index structures in turn.

## 1. Thesaurus

A thesaurus is the most widely used index structure in conventional information retrieval systems (25). A thesaurus consists of a set of concepts and a limited set of relationships between these concepts. Only three types of inter-concept relationships are represented: *equivalence* (preferred/non-preferred equivalent concepts), *hierarchical* (broader/narrower concepts) and *associative* (related concepts). As a result, the thesaurus consists of a standardised, controlled vocabulary of concepts that are hierarchically structured into a single inheritance tree. The major advantages of a thesaurus-based index structure are its flexibility and its intelligibility. There often exists a natural hierarchy in the concepts that are used to index the contents of a hypermedia document, and a thesaurus can easily capture this hierarchy. A major drawback is the effort involved in thesaurus construction and validation (26). Although tools and techniques have been developed for the computer-assisted creation of thesauri (27, 28), most existing thesauri have been carefully hand-crafted (e.g. the well-known MeSH thesaurus).

*Examples:* In the TACHIR system (29), a prototype tool for the automatic construction of hypertexts for information retrieval, an existing collection of concepts (e.g. a commercially available electronic thesaurus) can be used to extract keywords from a collection of documents and automatically build a corresponding hypertext. The system associates to each concept a number of keywords, which are related to each other based on a statistical analysis of keyword occurrence in the documents. The user can browse through the concept space by first using

the thesaurus of concepts to select a concept, and by then descending to the list of keywords associated with this concept he can finally locate the multimedia documents he is interested in. In the TraverseNet system (30), a graphical thesaurus-based retrieval interface for document databases, the user can select from a number of different thesauri the one that he would like to use. TraverseNet then displays a hierarchy window, which shows a concept at its centre, surrounded by its children concepts, in their turn surrounded by their children concepts, etc. Using this hierarchy window, the user can navigate the thesaurus and select concepts to formulate a query.

## 2. Faceted thesaurus

It is difficult for any complex collection of documents to accommodate into one single hierarchical index structure all the concepts that are used to index those documents. *Facet analysis* is an indexing technique where concepts are classified into separate hierarchical structures, where each hierarchy captures a different viewpoint on the documents (Figure 3). A faceted thesaurus therefore consists of a number of different thesauri, and each of these thesauri is used to index the documents with respect to some different knowledge domain (31). The advantage of using a faceted thesaurus is that this index structure allows for greater exhaustivity and precision in the hypermedia indexing process, since the documents can now be indexed with respect to all aspects that are judged relevant (32). The disadvantages of a faceted thesaurus index structure are the same as those of a conventional thesaurus: creation of the thesauri and validation of their hierarchical structure is very time-consuming.

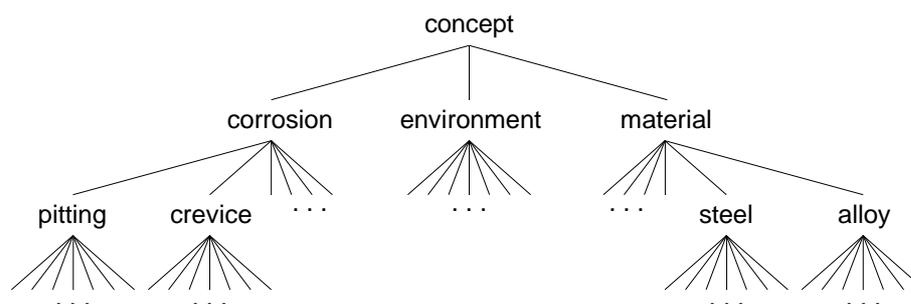


Figure 3. An example of a faceted thesaurus.

*Examples:* In the PRESS system (33, 34), a hypertext system designed to support software reuse, software components such as subroutines or modules are abstracted into software concepts. These software concepts are characterized with respect to application-oriented, implementation-oriented and historical *attributes*. Each of these attributes represents a facet of the corresponding concept, and the values that are allowed for such a facet are themselves organised in a thesaurus of index terms. Using these different thesauri, the user can navigate

through the different allowable values for each of the facets of a software concept, and in this way formulate a query for a specific piece of software. In the Talaria system (35), a hypermedia training and reference tool for healthcare providers managing patients with cancer pain, information is divided up into information units that are assigned a location in a context space. The co-ordinates of an information unit in this context space are derived by expressing the strength of association between the information unit and the usage *traits* which characterize this unit. The values that are allowed for such a trait can themselves be organised in a thesaurus of index terms. Information units that have similar ratings on a large number of traits are close together in the context space and as a result are linked. This approach guarantees that when the user is navigating through the information using these links, information that is related from a usage point of view will be located closeby.

### 3. Concept lattice

A concept lattice is a powerful extension of the thesaurus index structure. Mathematically speaking, a concept lattice is a partially ordered set of concepts in which every pair of concepts has a both a greatest lower bound (a unique narrower concept) and a least upper bound (a unique broader concept). The resulting index structure is similar to a thesaurus, but it extends this structure by the fact that a concept can have broader/narrower concepts which are not necessarily one level higher/lower in the concept hierarchy, and by the fact that there exists a single lowest concept which is narrower than any other concept (for a detailed introduction to concept lattices and formal concept analysis, we refer to 36). The major advantage of a concept lattice is that it can represent more flexible hierarchical structures than an ordinary or faceted thesaurus. There also exists a complete set of mathematical techniques that can be used to create concept lattices and check their internal consistency (37).

*Examples:* In the WorldViews system (38), a system designed to process electronic news articles and abstracts of technical reports, documents are automatically indexed and classified with respect to a lattice of concepts derived from the IEEE Inspec thesaurus. The WorldViews retrieval engine interprets a user's query relative to this lattice of concepts, and then restricts the lattice to the sublattice relevant to the query. Using this sublattice, it can find the narrower concepts that can be used to extend the scope of the original query. The WorldViews query interface also uses the lattice to facilitate iterative user navigation through neighbouring concepts of the query concepts. In the BRAQUE system (39, 40), a system for information retrieval from on-line bibliographic databases, a special type of concept lattices is used, called *relationship lattices*, to maintain an extensible personal thesaurus that the user can use during the information retrieval process. In this personal thesaurus, the

user can define personal concepts and personal relationships between these concepts, that are automatically related to the actual keywords and actual relations used in the bibliographic databases. In this way, the user can impose his own ideas about the most appropriate index structure for the bibliographic documents in the database, but can still use the original keyword-based retrieval mechanisms of the database itself.

#### 4. Hyperindices

Hyperindexing is an indexing technique that was specifically developed for hypermedia information (41). In the hyperindexing method (42), the contents of a document is characterized by constructing a so-called index expression (a set of index terms and connectors between these index terms) from the title of the document. From such an index expression one can derive the so-called *power index expression*, which forms a lattice-like structure of index expressions, that can then be used as a hypertext of indices (Figure 4). Each vertex in this lattice can be considered as a pre-defined query to the document space that can be enlarged (made less specific) or refined (made more specific) by moving respectively to the descendant or ancestor vertices in the lattice of index expressions. Bosman *et al.* (43) have shown experimentally that information retrieval using hyperindices is at least as effective as information retrieval using a faceted thesaurus. They believe however that hyperindices are superior with regards to both *collocation* (the degree to which relevant index terms are near to each other) and *exhaustivity* (the degree to which the contents of the documents are reflected in the index terms).

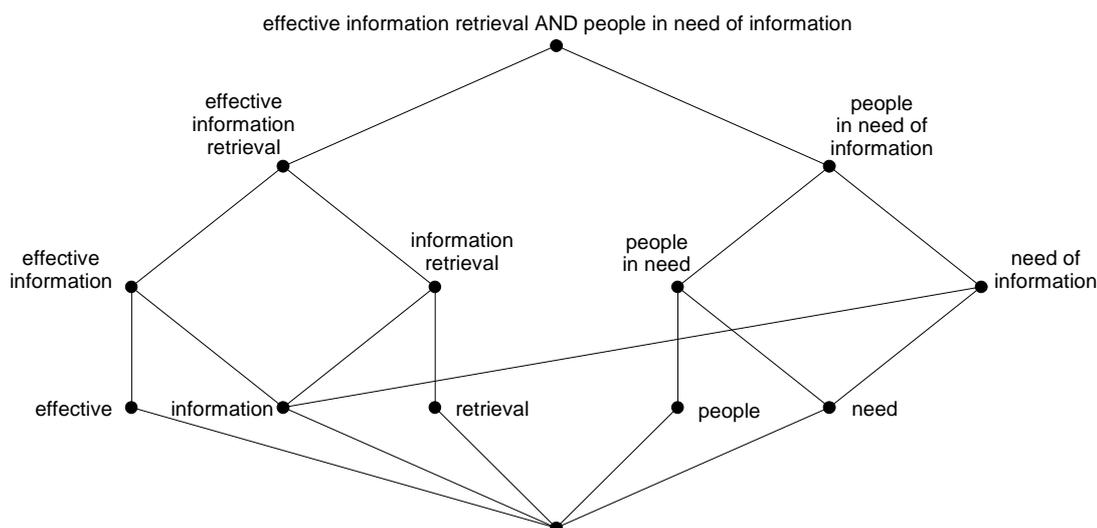


Figure 4. An example power index expression (adapted from 41).

*Example:* In his doctoral thesis (44), Bruza describes the development of a hypertext-based information retrieval tool for an art slides library, which uses the IconClass faceted thesaurus (developed for the characterization of Western art) to create a hyperindex layer on top of the slides collection. Using the concepts of IconClass, the titles of the slides are parsed to extract index terms. These index terms are then used to derive a lattice of hyperindices which capture all the possible combinations of index terms found in all the slide titles. The user can navigate through this lattice of possible slide titles using a graphical interface, enlarging or refining the slide title he is considering by removing or adding related index terms. Whenever a slide title matches the title of an existing slide, the user can “beam down” from the hyperindex to the slides database to retrieve the corresponding slide.

## 5. Semantic hyperindices

The strength of the hyperindexing technique lies in the fact that the lattice of hyperindices can be generated automatically from the concepts characterizing the node contents. However, when building these hyperindices the technique does not take into account how these concepts may possibly relate to each other semantically. To overcome this limitation, we have ourselves developed a more semantics-aware version of these hyperindices, so-called *semantic hyperindices* (45). The semantic hyperindexing technique introduces the use of associations, or relationships between concepts belonging to different knowledge domains. These associations try to express which combinations of concepts are inherently valid, or are potentially interesting from a usage point of view. They circumscribe the subsets of concepts that can be meaningfully taken together at the same time. Domain-specific associations express which combinations of concepts are inherently valid with respect to the knowledge domains to which these concepts belong. E.g. certain combinations of concepts are excluded since they are not possible in theory or not pertinent in practice. Usage-specific associations express which combinations of concepts should be considered together for specific kinds of readers and for specific kinds of tasks. This use of associations allows us to fine-tune the lattice of hyperindices, by excluding certain combinations of concepts that were generated by the hyperindexing technique, and by including other combinations that would never have been generated by the hyperindexing technique. Using these semantic indices it is also possible to develop useful numerical metrics to characterize the degree of information overlap of the nodes' contents (46).

*Example:* In the IKON system (47), a knowledge-based hypermedia system for use by materials engineers, semantic hyperindices are used to index the contents of documents with corrosion and materials engineering information. The associations between concepts are used primarily to capture those relations between corrosion,

material and environment concepts, which are appropriate for addressing typical materials engineering problems. This allowed us to keep the number of associations limited in number but still comprehensive in overall scope. These associations are used for generating on the fly presentation views for the documents and traversal trails between these presentation views, resulting in a flexible mechanism for task-driven and user-directed navigation. The way these traversal trails are defined means that in IKON there are no longer any links between the nodes in the conventional sense of the word. There are only associations between concepts characterizing the nodes, resulting in navigation movements between nodes that are performed either at the user's discretion, or as part of a predefined traversal trail. The resulting browsing process is no longer rigid and deterministic, which unfortunately also means that some of the tractability and analysability of the browsing process is lost.

## 6. Inference network

In an inference network, nodes represent concepts and links represent dependence relations between these concepts. An inference network consists of two component networks: a *document network*, which represents the document collection, and a *query network*, which represents the user's information need (Figure 5). The two component networks are joined by links between document concepts and query concepts, and during query processing the query concepts are matched with the document concepts through probabilistic inference (48). The use of inference networks as a hypermedia index structure was first proposed by Croft and Turtle (49), who have shown experimentally that this index structure is indeed very effective for hypertext information retrieval (50).

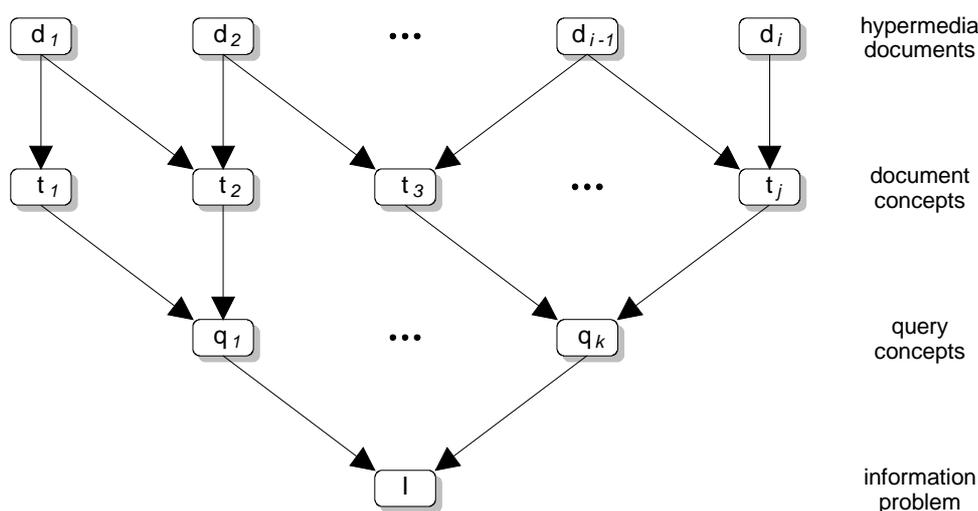


Figure 5. The organisation of an inference network (adapted from 50).

*Examples:* In the FIRST system (19), a prototype system for knowledge-based information retrieval, concept nodes link single concepts to document nodes in which this concept is referenced. The concept nodes are linked into a concept network using weights which express the strength of the semantic association between pairs of concepts. The system acts as an question-answering system that, given a request, returns the best matching documents, by reasoning on the concept network as a knowledge base through a process of spreading activation. In the Dynamic Medical Handbook (51), a system that is used as a testbed for the design of effective information retrieval methods for large-scale biomedical hypertexts, documents are indexed using a hierarchical index space where concepts are joined by probabilistic dependencies. Using these dependencies, user feedback about the appropriateness of a given concept as a representation of document contents is propagated to all other related concepts. In this way, the structure of the index space gradually adapts itself to the user's retrieval preferences.

## 7. Semantic network

In a semantic network, nodes represent concepts and links represent semantic relations among these concepts (Figure 6). Compared to the restricted number of relations in a thesaurus, semantic networks have a rich internal organization of relations which can support different reasoning mechanisms (52). More importantly, the node-link-node structure of a semantic network is conceptually very close to the structure of the hypertext network itself and therefore supports browsing in a very natural way (53). Also, Rada *et al.* (54) have shown that using a semantic network, it is possible to develop more robust and efficient retrieval mechanisms, provided the relations between the concepts are chosen with the user's typical retrieval tasks in mind. However, identifying the important concepts in a knowledge domain and the relations between these concepts is a challenging task.

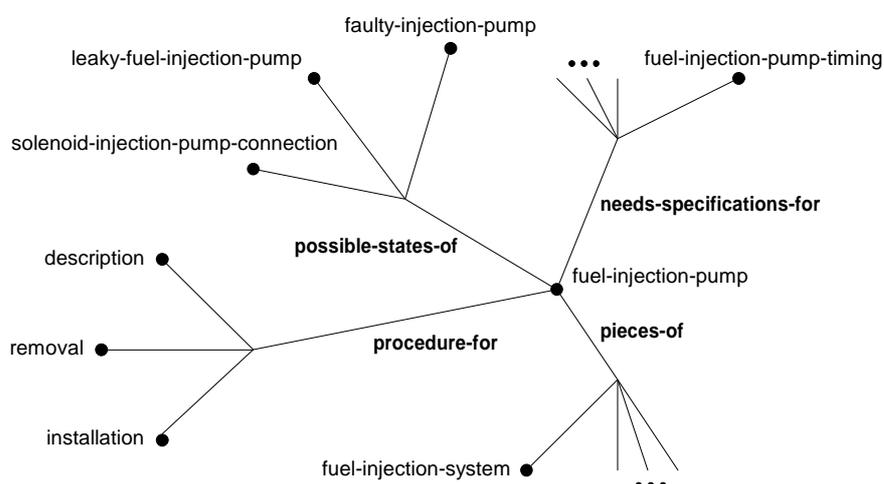


Figure 6. An example semantic network for a car repair manual (adapted from 55).

*Examples:* In the VISAR system (56), that was designed to act as an intelligent literature survey assistant for researchers, technical citations of journals and conference articles are indexed by deriving concepts from the citation titles. These concepts are organized into a semantic network, and the VISAR system allows a structured exploration of the resulting concept space, by first matching a *personal information representation* against the concepts and relationships between the concepts, and then retrieving the corresponding citations. In the MacWeb system (57), a knowledge-based hypertext system for document production applications, every document is divided up into nodes that are indexed by manually giving them distinct types. Relationships between these node types are expressed using typed links. Together, these *node types* and *typed links* form a semantic network, and by attaching scripts to these types, the system can support contextual or task-driven access to documents (58).

## B. Index acquisition

Once the decision has been made to use a certain index structure for the indexing of hypermedia information, one is still faced with the problem of how to acquire the concepts which will be part of this index structure, and how to acquire the meaningful relations that have to be represented between these concepts. This is often the most difficult and time-consuming phase in the development of a hypermedia system (59). Fortunately, a number of promising approaches for the computer-assisted acquisition of concept-based index structures for hypermedia documents have been developed recently. We discuss each of these approaches in turn.

### 1. Principles

The process of creating concept-based index structures for hypermedia documents involves three major steps:

1. extraction of index terms from the hypermedia documents
2. refinement of these index terms into a controlled vocabulary of concepts
3. creation of an index structure that represents relationships between these concepts

This of course assumes that some form of text is available in the multimedia documents (either in the document itself, or in a short description associated with each document). Step 1 basically involves the removal of stop words, lemmatizing of the remaining words into keywords, etc. using well-known techniques from information retrieval (25). In step 2, potential concepts are collected using existing paper-based lists of keywords, or are defined by the future users (60), or are generated automatically by applying statistics-based cluster analysis techniques to the extracted keywords (61). The crucial step during index acquisition is of course step 3. Some

attempts have been made to derive concepts and relationships through semantic text analysis (e.g. 62), but most hypermedia researchers have focused their attention on the development of techniques that require some form of user feedback to interactively create new collections of concepts, or extend and refine existing index structures of concepts. Since hypermedia systems are characterized by their strong reliance on active user interaction, most, if not all, concept acquisition approaches described below have tried to adhere to that same interaction paradigm.

## 2. Approaches

**Indexing in context:** in the CID system (63), a document management system that enables the integration of various technical documents in a hypermedia framework, the system is capable of providing better navigational support by learning from user feedback. The hypermedia documents are indexed using concepts that are specifically designed to provide meaningful entries for a search in the documentation. The correspondence relations between these concepts and the documents they refer to are modified by using interactive user feedback to either reinforce or correct the system's knowledge in case of success or failure. In this way, the concepts are modified incrementally, so that the system will later remember what the user found useful in a particular context.

**Question-based indexing:** in the DEDAL system (64, 65), a hypermedia system that facilitates the indexing and retrieval of design documents in technical engineering, the system can acquire conceptual indices of text, graphics and videotaped documents on the basis of the user's questions. A user formulates a query to the system, and if there is no corresponding set of indexing concepts, DEDAL uses the underlying domain model and a set of retrieval heuristics to approximate the query concepts, asking for confirmation from the user. If the user finds the retrieved information relevant, DEDAL acquires a new set of indexing concepts based on the query.

**Conversational indexing:** in the Trans-ASK system (66), a large hypermedia system in the domain of military transportation planning, the user is guided through hypermedia documents on the basis of a conversational model of hypertext navigation. To support this navigation mechanism, the hypermedia documents are indexed using concepts which express the conversational topics of the documents. These concepts are derived by segmenting the documents into self-sufficient units and having human indexers enumerate questions for which the information unit is likely to provide a good answer. These questions and the concepts answering these questions are then categorized, and used to manually link together units which raise and answer specific questions.

**Agglomerative hierarchic clustering:** in the SMART system (67), an experimental information retrieval system that provides tools for textual analysis and concept clustering, a hypertextual interface was built which uses concept cluster hierarchies to improve the navigational search process. Concept classes, that are chosen from a concept thesaurus, are used to represent documents as concept vectors in a vector space. These documents are then repeatedly merged into clusters on the basis of the similarity between the concept vectors describing them. In this way, hierarchic clusters of concepts are created, that can be used for interactive browsing searches.

**Interactive taxonomic classification:** in the HyperSet system (68, 69), a set-based hypermedia system designed to support taxonomic reasoning, nodes are organized in sets on the basis of their similarity with respect to one or more attributes. The user can sort nodes into sets based on a particular number of attributes, examine the different sets that a node is a member of, and generate new sets from old ones. In this way, he can determine interactively which attributes have the greatest discriminatory power, and define concepts which uniquely represent combinations of particular values for these attributes. These concepts can then be used to index the nodes.

## IV. Concept-based retrieval of hypermedia information

We have examined a number of different concept-based indexing structures for hypermedia information, and we have seen how different approaches have been developed for the computer-assisted acquisition of the required concepts and the relations between these concepts. Finding and constructing an appropriate concept-based indexing structure for the hypermedia information solves only half of the hypermedia information retrieval problem though. We also have to find a retrieval mechanism that uses this concept-based indexing effectively, and which integrates in a natural way with navigation (which is, after all, the defining interaction mechanism of hypertext).

### A. Retrieval mechanisms

The two principal retrieval mechanisms in hypermedia systems are *retrieval by query* (as in conventional, non-hypermedia information systems) and *retrieval by navigation* or browsing. Marchionini and Shneiderman (70) have argued that users subjectively prefer browsing search strategies, because most users are either unable or unwilling to cogently formulate their search objectives, and because browsing places less severe cognitive demands on the user. As a result, most hypermedia systems developers have focused their attention on this search by navigation process, since this fits best within the whole hypertext interaction model. Only recently have efforts

been made to integrate both retrieval mechanisms into one single mechanism: *query by navigation*. We discuss retrieval by query and retrieval by navigation in detail below, highlighting the shortcomings of both mechanisms, and we then focus our attention on the now widely used query by navigation mechanism.

## 1. Retrieval by query

The retrieval by query mechanism for hypermedia information is basically an extension of the classic document retrieval mechanism towards the retrieval of multimedia data (Figure 7). The hypermedia documents are indexed (e.g. using one of the concept-based approaches described above), and the user is faced with the task of translating his information problem into a query that can be understood by the retrieval engine. The query terms are compared with the index terms, and the corresponding documents are retrieved. These documents are then offered to the user as the starting points for further navigation in the hypernetwork of nodes and links (e.g. as in 71).

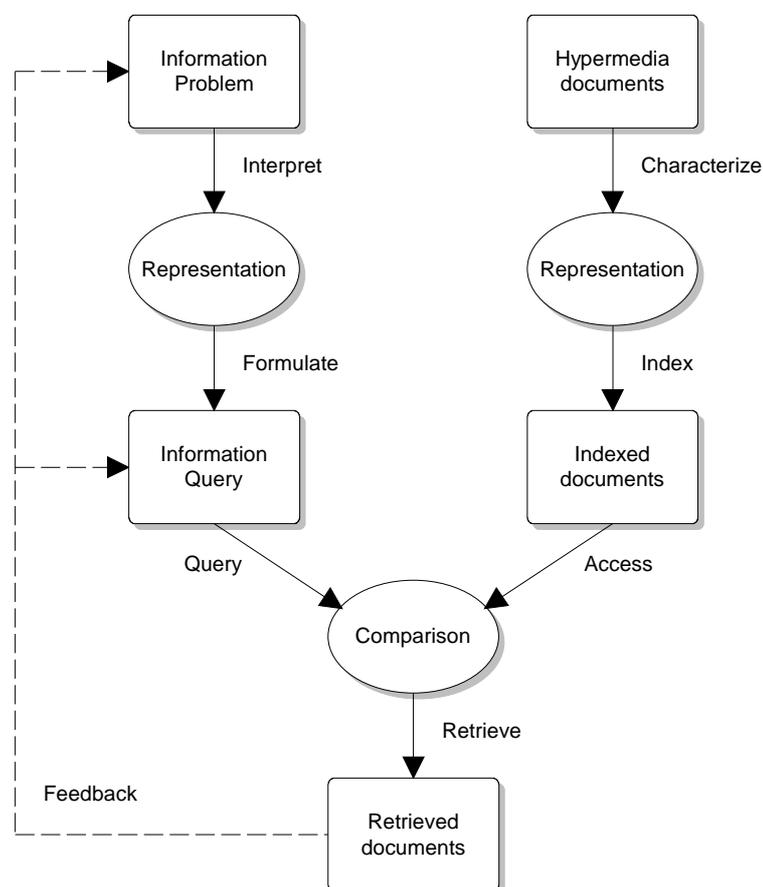


Figure 7. Hypermedia information retrieval by query.

From the user's point of view, there are two big problems with the retrieval by query mechanism. First, he has to interpret his information problem in terms of the model captured in the indexing representation. This model almost certainly does not correspond with his own ideas about the structure of the data. Secondly, he has to use this own, imperfect interpretation of the information problem to formulate a query in the query language understood by the retrieval engine. He most likely does not have sufficient experience with the requisite query terms or the query language to do so. The user will get feedback through the retrieved hypermedia documents about the accurateness of his interpretation and the correctness of his query formulation, but that does not necessarily help him in gaining a better understanding of the indexing representation, or in improving his querying dexterity

## 2. Retrieval by navigation

The basic retrieval mechanism of hypermedia is of course navigation, the process of following links between multimedia documents until the information one is searching for has been found (Figure 8). The user tries to solve his information problem by directly navigating through the hypernetwork, and by changing his area of search in response to the documents he finds (e.g. as in 72). The main advantage of the retrieval by navigation mechanism is that users no longer have to worry about a correct problem interpretation or query formulation, since retrieval is realised by simply browsing around. This works fine for small collections of hypermedia documents, where the user can build his own mental map of the hypernetwork, or for hypernetworks which have a well-designed link structure, where the user can predict where links will take him. However, when the document collection is just too big, or the hypernetwork just too complex, this mechanism rapidly breaks down (73, 74).

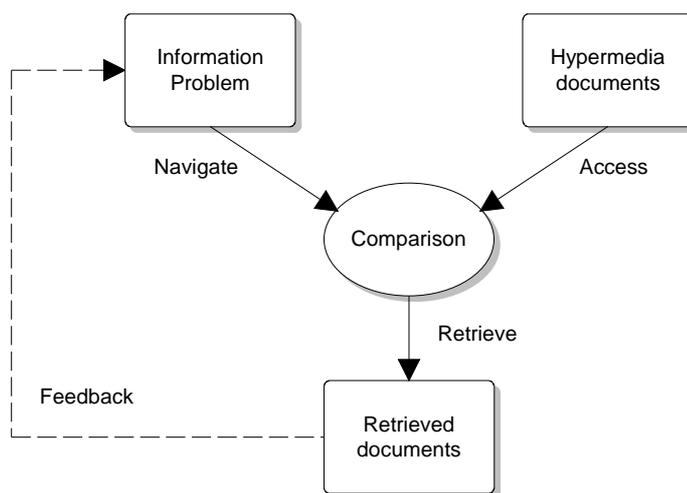


Figure 8. Hypermedia information retrieval by navigation.

### 3. Query by navigation

Clearly, both retrieval by query and retrieval by navigation have serious shortcomings. However, by merging these mechanisms, it should be possible to develop a retrieval mechanism which combines the expressive power of retrieval by query with the ease of use of retrieval by navigation. This was first proposed by Bruza (41) and was coined “query by navigation” (Figure 9). In query by navigation, the user still performs simple navigation actions, but now not only in the hypernetwork, but also in the index space itself (e.g. as in 75). The user directly expresses his information problem by navigating in the index space, and now it becomes the responsibility of the hypermedia engine to translate these navigation movements into a query that can be understood by the retrieval engine. The query itself remains hidden from the user. This eliminates the need for the user to really understand the model captured in the indexing representation, and also relieves him from the burden of learning a difficult query language. However, all the usual problems associated with navigation, such as disorientation and cognitive overhead (1), now also pop up at the level of the index space instead of only in the hypernetwork, so we now face the challenge of developing an effective user interface for this query by navigation mechanism.

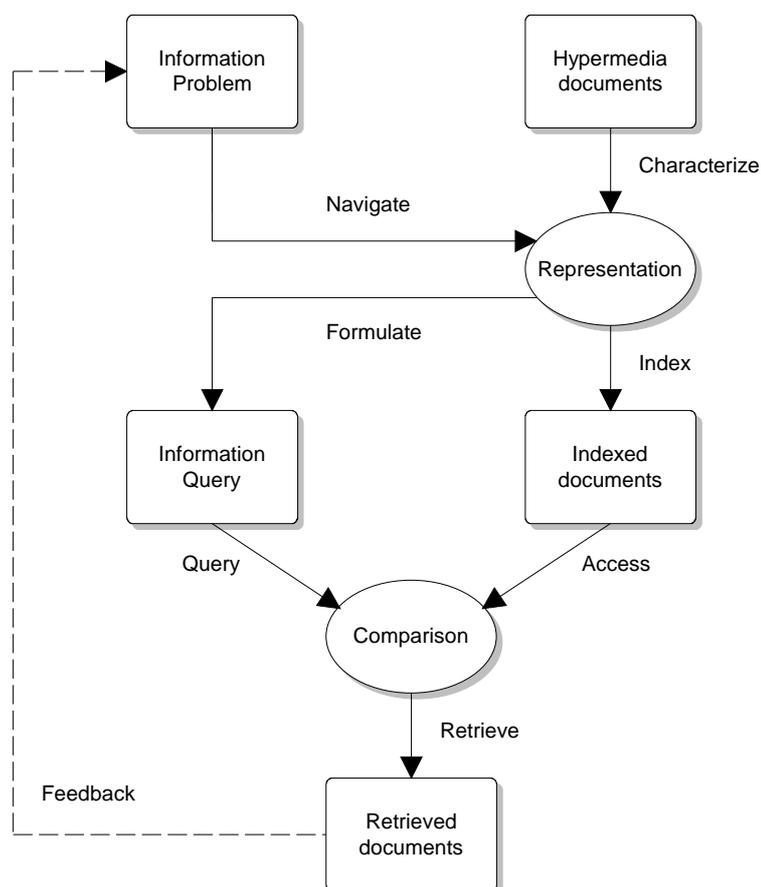


Figure 9. Hypermedia information query by navigation

## B. Retrieval visualization

Agosti (24) has pointed out that the difficulty of presenting to the user in a transparent way the index elements together with their index structure is one of the major deficiencies of hypermedia systems when it comes to effectively supporting query by navigation operations. In most conventional, non-hypermedia information retrieval systems which use an index structure, the user can just see a single index term or a list of index terms during the query formulation. Only a few systems (e.g. 76, 77) provide the user with a direct way of visualizing the index structure itself, as a means of facilitating the information retrieval process. Shum (78) has argued convincingly that especially in hypermedia systems the use of *spatial visualization*, i.e. the use of a visual structure to reflect conceptual structure, has a number of important benefits to offer to users in terms of the retrieval of information. Amongst these benefits are the possibility to use our well-developed senses of distance and direction, the ability to locate known information and allocate meaningful positions to unknown information in relation to the whole of the index space, and the intuitive intelligibility of a well-chosen spatial representation. Another important cognitive advantage is also that visualization can serve as an unobtrusive means of instructing the user about the concepts which exist in the index space, and their relationships.

### 1. Principles

Ideally, the steps involved in browsing and retrieving hypermedia documents should be hidden behind an interface that lets users search at a conceptual, descriptive level instead of at a concrete, procedural level (79). Although some researchers (80) have questioned the need for graphical overviews of the structure of hypermedia data, browsers or overview diagrams are still considered to be one of the best tools for orientation and navigation in an index structure (81). The basic idea behind the graphic display of index structures is to allow users to grasp these structures more readily by making use of a spatial metaphor. By presenting a map of the index structure, they allow the users to see where they are, what other index elements are available, and how to combine these index elements to formulate a query and access the underlying documents. In visualizing such index structures, a number of problems (82) have to be addressed, such as what kinds of concepts and relations between these concepts are to be visualized, how nodes and relations are to be represented and positioned on the display, etc. In general, no optimal solution can be found for these problems, and each of the approaches described below tries to address a different set of user and task requirements. What all these approaches do have in common is that they use layout to express both the index structure itself and what is allowable as query combinations of indices.

## 2. Approaches

**Tree-structured visualization:** in the SYRIUS system (83), a prototype hypertext information retrieval system, indexing links connect thesaurus nodes to document nodes, but the system also uses classification links which aggregate documents into classes according to different criteria. These classification criteria are visualized in a tree-like structure of classes (Figure 10), that can be used to select additional criteria and narrow the scope of the retrieval to a subset of documents. The resulting set of documents and their structural and referral links can then be further examined by navigating through the corresponding hypernetwork of documents and links.

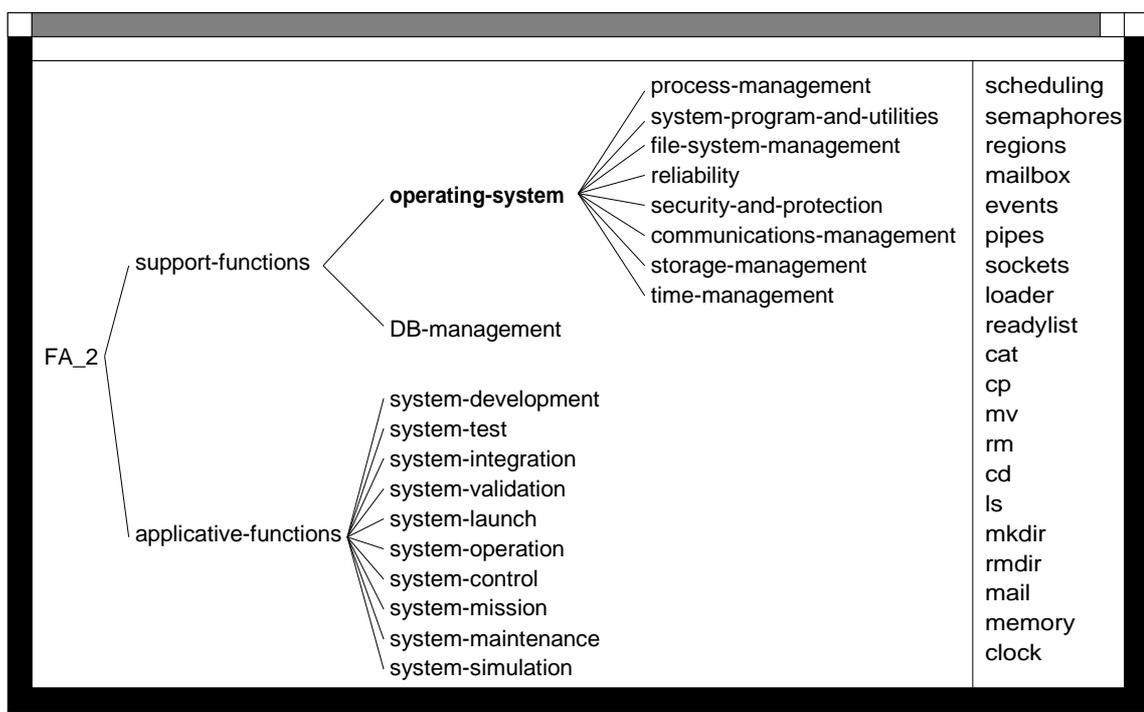


Figure 10. Performing a multi-criteria search in the SYRIUS system (adapted from 83).

The user can choose one or several classification criteria that seem best suited to his consultation. The browser shows a classification criterion as a tree-like structure of classes (left subwindow). Choosing a class in this classification hierarchy triggers the visualization of general information about the set of associated documents (number of documents and their descriptors). The titles of the documents attached to the chosen class (in bold) are displayed in the right subwindow.

**Interactive dynamic maps:** in the SHADOCS system (84), a document retrieval system for sharing documents between different users, user can access and navigate the information using topic interactive dynamic maps which represent the semantic contents of sets of documents. Topic interactive dynamic maps provide an overview of the

topics present in a collection of documents, their importance, and the similarities/correlations among them (Figure 11). Queries can be issued by selecting topics directly on these maps, that are translated into a real query. This finally results in a subset of documents, for which a document interactive dynamic map is displayed.

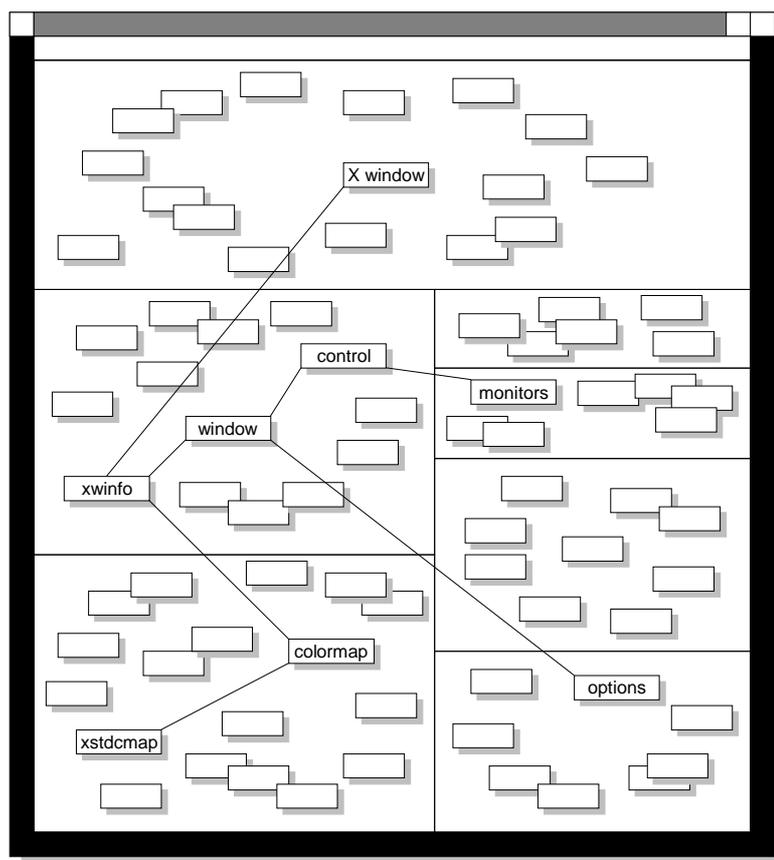


Figure 11. A topic interactive dynamic map in the SHADOCS system (adapted from 84).

A topic interactive dynamic map provides an overview of a large number of documents by extracting semantic information from them rather than displaying the documents themselves. The areas of a topic interactive dynamic map are the classes of a thesaurus. Each class contains a set of topics represented by “cities” on the map, which are depicted by icons. “Roads” between cities represent relationships between topics. Users can issue queries by selecting regions, cities and roads.

**Graph representations:** in the MORE system (85), a visual environment for multimedia information retrieval, a graph representation is used to visualize both the conceptual schema (i.e. the semantic network of concepts and relations that is used to represent the multimedia document contents) and the user queries that can be formulated on that conceptual schema. The user manipulates on-screen concept graphs and visual representations of objects

to formulate queries in terms of the concepts used (Figure 12). This visual query mechanism combines browsing and querying under a uniform interface, maintaining one and the same interaction style throughout.

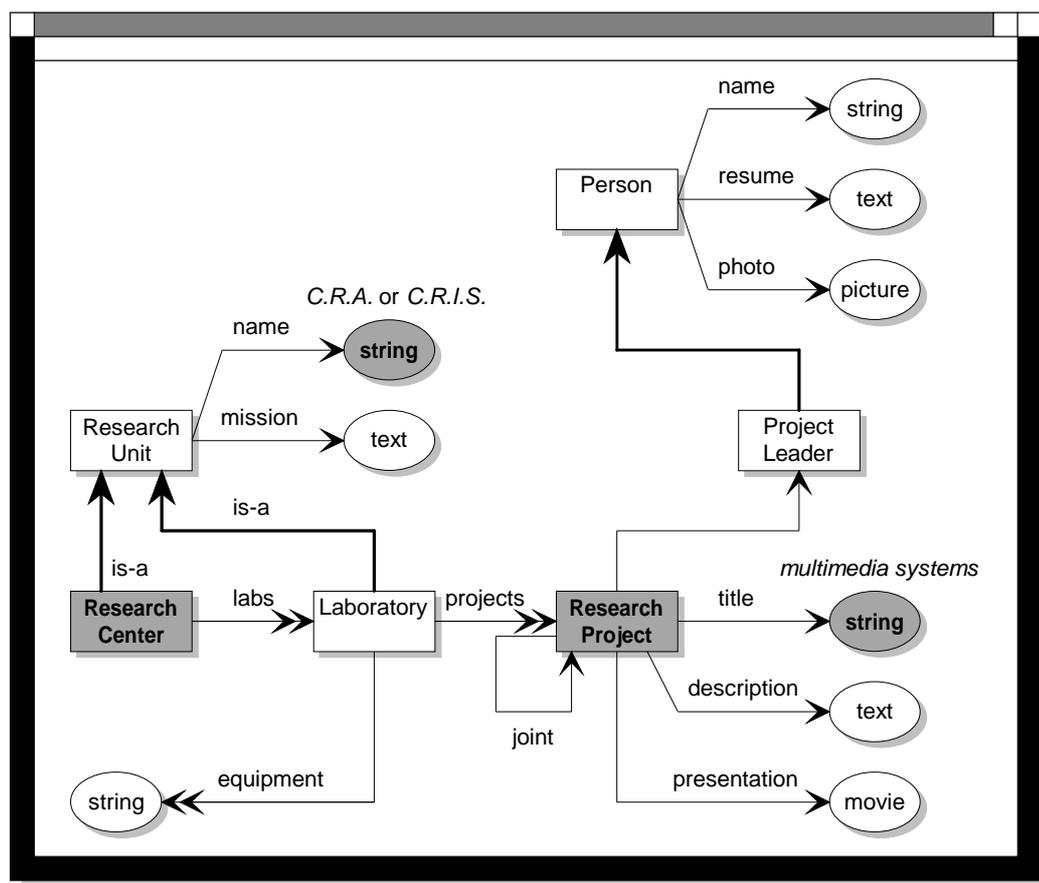


Figure 12. A graph representation of a visual query in the MORE system (adapted from 85).

A (part of) a conceptual schema graph. Rectangular nodes represent a class of complex objects, oval nodes represent simple objects. Labelled arrows depict the properties of a class (multi-valued properties are depicted with double-headed arrows). The bold lines express the inheritance *is-a* relationship from a subclass to its superclass. In this particular example, the following query has been visually specified: “I want to know if, in the research centres named *C.R.A.* or *C.R.I.S.*, there are research projects in the field of *multimedia systems*, and, if this is the case, I want to see the laboratories where these research activities are carried out and who are the project leaders involved.”

**Interactive clustering overviews:** in the Navigational View Builder (86), a tool for the construction of overview diagrams of hypermedia collections, interactive clustering techniques are used to generate overviews for large sets of documents. Each node is given a set of attributes, whose values are used to index the nodes. The user can interactively specify which attributes have to be used to generate clusters of nodes, based on the similarity of the

corresponding attribute values. In this way, the user can repeatedly cluster together sets of nodes into different abstraction layers, that are then visualized in a 3-dimensional browser (Figure 13). In this browser, the user can repeatedly shift his eye-point in real-time to bring other parts of the abstraction hierarchy into focus.

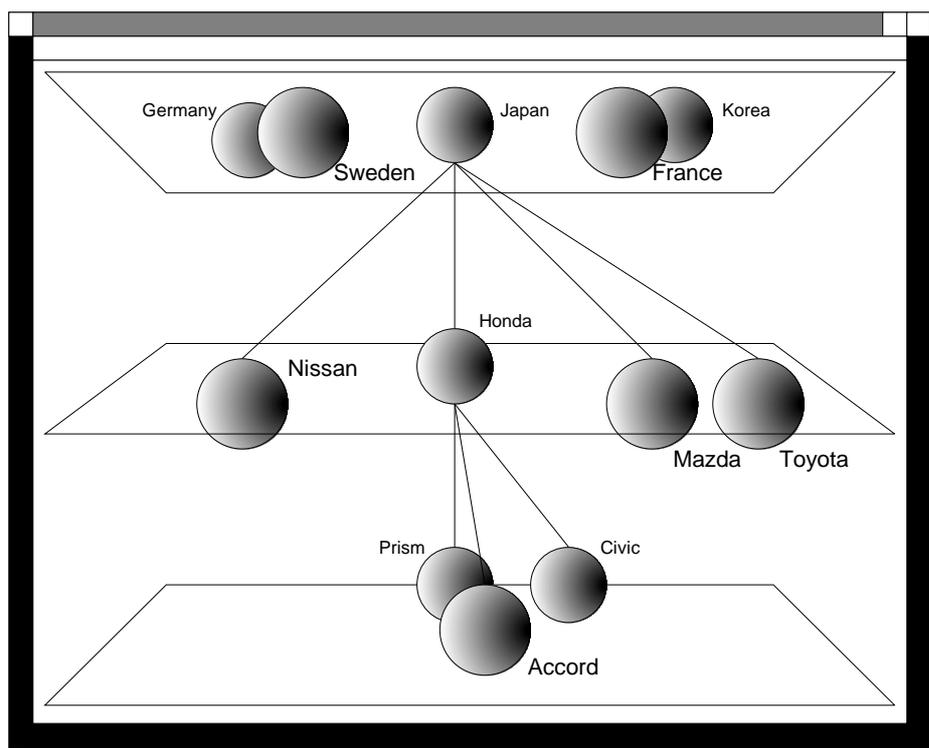


Figure 13. A view of abstraction layers in the Navigational View Builder (adapted from 86).

A top view of a hierarchy of different abstraction layers defined on a hyperbase about automobiles. The user first wanted to see details on Japanese cars, and then on Honda. The browser shows links between the parents and the children in the abstraction hierarchy. It can also show links between those clusters (e.g. “Germany” and “Sweden”) whose children have again links in-between them. The nodes “Japan” and “Honda” are expanded into their lower-lying child nodes.

**Dynamically bounding overviews:** in the GALOIS Bound&Browse system (87), a document retrieval system which uses concepts organised into a Galois concept lattice to index documents, the user can visually formulate a query by bounding or restricting the displayed concepts to a specific subset of all available concepts (Figure 14). By introducing constraints on the displayed concepts, the user can prune away concepts that he considers to be irrelevant, and dynamically bound the overview of the concept lattice to the concepts he is interested in. In this way, the user can use the browser to gradually focus on those concepts which may be useful for his query.

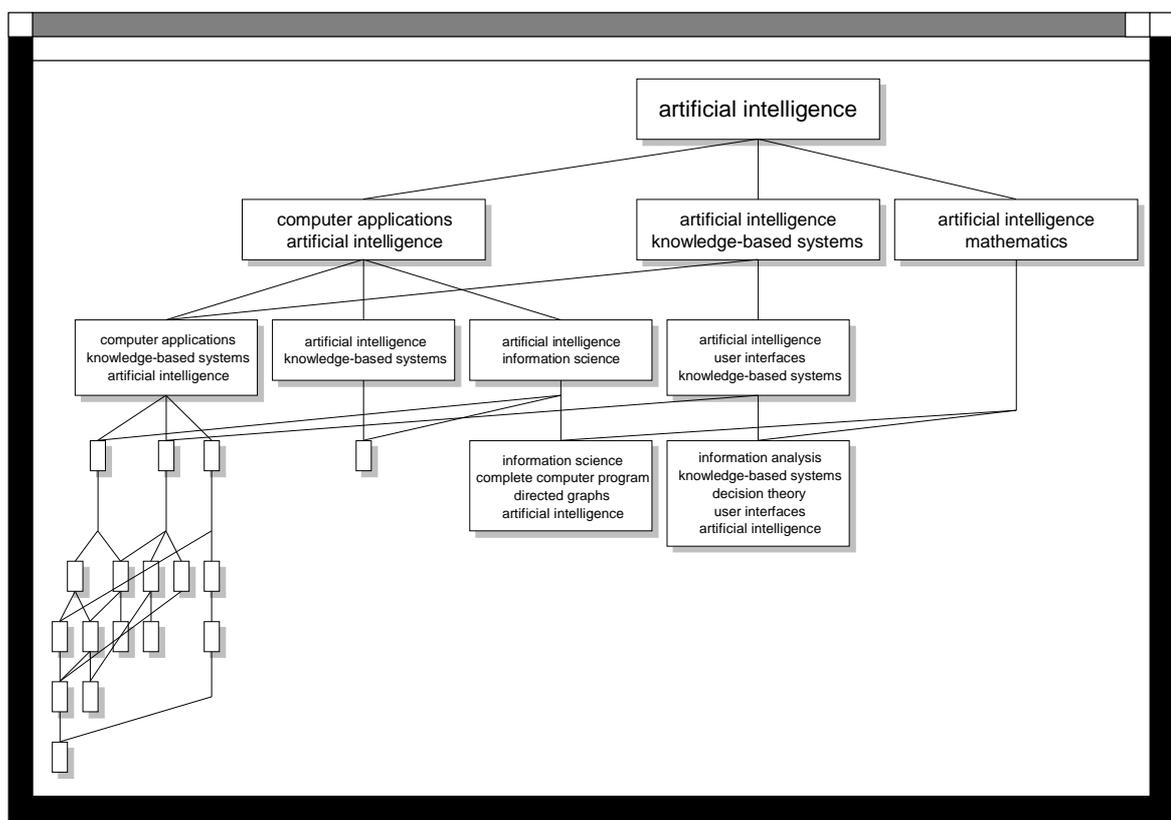


Figure 14. A view of a bounded lattice in the GALOIS Bound&Browse system (adapted from 87).

A fisheye view on the index space, where the concept which is the current focus is shown at the top of the window, and the other, related concepts are shown in varying levels of detail depending on the distance from the focus. The user may bound the concept lattice that is being shown to a smaller concept sublattice by introducing restraints which enlarge or refine one or more concepts.

## V. Conclusions and future research

Classic information systems have focused traditionally on effective *access* to knowledge resources, hypermedia systems now focus more on effective *interaction* with knowledge resources. We have reviewed how the use of query by navigation as a retrieval mechanism tries to give us the best of both worlds, and how a number of different concept-based indexing techniques have been developed that support such a retrieval mechanism. Concept-based indexing and retrieval of hypermedia information clearly show great promise in enhancing the information retrieval capabilities of hypermedia systems. A lot of important research questions remain open however. As far as the concept-based indexing of hypermedia information is concerned, the concept-based indexing techniques developed so far have focused on indexing the meaning of the *contents* of the hypermedia nodes, and

have not attempted to index the meaning of the *structure* of the links between those nodes. However, the links between the nodes are an integral part of the hypermedia information structure, and Halasz has argued very convincingly that we should not only be capable of performing *contents search*, but also *structure search* (88). The user of a hypermedia system should be capable of asking such queries as “Show me all the nodes that refute what is in this node.” or “What is the sequence of nodes which leads to the conclusions in this node?” Answering such queries will require the development of techniques for indexing the meaning of different link types, and the meaning of different link configurations. Another problem with the concept-based indexing techniques developed so far is that they do not really take into account what *kind of user* is looking for information, and what *kind of task* he is trying to perform using that information. Ideally, the index space should dynamically adapt to the user and the task, e.g. by changing what concepts are made available during query by navigation or even by altering the whole organization of the index structure itself. This will require the formulation of reading models and utilization models that are specifically targeted towards coping with linked, multimedia information.

As far as the concept-based retrieval of hypermedia information is concerned, Halasz has argued (89) that we should also consider another perspective on the query by navigation mechanism. He wondered if we could develop retrieval mechanisms that would support *navigation by query*, where links would be generated on the fly on the basis of query specifications. This would enable us to do away altogether with frozen, static links and replace them by fluid, dynamic links, which would result in more robust and more open hypermedia systems. A major problem is the tremendous shortage of good, reliable usability studies on hypermedia system functionality in general, and hypermedia information retrieval functionality in particular. The main difficulty here is that no one has yet come up with the hypermedia equivalent of the *precision* and *recall* measures of classic information retrieval. As a result, most studies into hypermedia information retrieval report on qualitative rather than on quantitative evaluations of retrieval efficiency. This makes it very hard to compare different retrieval interfaces with respect to the efficacy of their user interaction model, or to come to a conclusion as to how well the structure and capabilities of the underlying concept-based index space get translated into an effective visualization.

Hypermedia systems can learn a lot from the research results in classic information systems, but the reverse is also true: the challenges that global, on-line, multi-user hypermedia systems like the World-Wide Web (90) pose for efficient indexing and effective retrieval are formidable. Indexing and retrieval have always been at the core of classic information research, they will be at the core of hypermedia research for the foreseeable future as well.

## Acknowledgements

The first author wishes to acknowledge the financial support provided by the Flemish Institute for Scientific and Technological Research in Industry (I.W.T.), who made this research into intelligent concept-based indexing and retrieval of hypermedia information possible as part of the IKEM project.

## Bibliography

- Begoray, J.A. (1990). An introduction to hypermedia issues, systems and application areas. *International Journal of Man-Machine Studies*, **33** (2), 121-147.
- Berk, E. and Devlin, J. (1991). The hypertext/hypermedia handbook. McGraw-Hill Software Engineering Series, Intertext Publications, McGraw-Hill Publishing Company, Inc., New York, 571 pp.
- Bornman, H. and von Solms, S.H. (1993). Hypermedia, multimedia and hypertext: Definitions and overview. *The Electronic Library*, **11** (4/5), 259-268.
- Nielsen, J. (1990). *Hypertext and Hypermedia*. Academic Press Inc., Boston, San Diego, New York, London, Sydney, Tokyo, Toronto, 263 pp.
- Parsaye, K., Chignell, M., Khoshafian, S. and Wong, H. (1989). *Intelligent databases: Object-oriented, deductive hypermedia technologies*. John Wiley & Sons Inc., New York, Chichester, Brisbane, Toronto, Singapore, 479 pp.
- Parsaye, K. and Chignell, M. (1993). *Intelligent database tools & applications: Hyperinformation access, data quality, visualization, automatic discovery*. John Wiley & Sons Inc., New York, Chichester, Brisbane, Toronto, Singapore, 541 pp.
- Rada, R. (1991). *Hypertext: From text to expertext*. McGraw-Hill Book Company, London, New York, 237 pp.

## References

1. Conklin, J. (1987). Hypertext: An introduction and survey. *IEEE Computer*, **20** (9), 17-40.
2. Foss, C.L. (1989). Tools for reading and browsing hypertext. *Information Processing & Management*, **25** (4), 407-418.
3. Gygi, K. (1990). Recognizing the symptoms of hypertext ... and what to do about it. In *The art of human computer design* (Ed. Laurel, B.). Addison-Wesley Publishing Company Inc., Reading (Massachusetts), Menlo Park (California), New York, pp. 279-287.
4. Wright, P. (1989). Interface alternatives for hypertext. *Hypermedia*, **1** (2), 146-166.
5. Simpson, A. (1990). Navigation in hypertext: Design issues. In *Proceedings of the 13th International Online Information Meeting* (London, December 12-14) (Ed. Learned Information Ltd.). Learned Information Ltd., Oxford, New Jersey, pp. 241-255.
6. Bernstein, M. (1988). The bookmark and the compass: Orientation tools for hypertext users. *ACM SIGOIS Bulletin*, **9** (4), 34-45.
7. Frisse, M.E. and Cousins, S.B. (1990). Guides for hypertext: An overview. *Artificial Intelligence in Medici-*

- ne, 2 (4), 303-314.
8. Oren, T., Salomon, G., Kreitman, K. and Don, A. (1990). Guides: Characterizing the interface. In *The art of human computer design* (Ed. Laurel, B.). Addison-Wesley Publishing Company Inc., Reading (Massachusetts), Menlo Park (California), New York, pp. 367-381.
  9. Feiner, S. (1988). Seeing the forest for the trees: Hierarchical display of hypertext structure. In *Proceedings of the ACM Conference on Office Information Systems* (Palo Alto, California, March 23-25) (Ed. Allen R.B.). ACM Press, New York, pp. 205-212.
  10. Pintado, X. and Tsichritzis, D. (1990). Sa<sup>T</sup>ellite: A visualization and navigation tool for hypermedia. In *Proceedings of the Conference on Office Information Systems* (Cambridge, Massachusetts, April 25-27) (Eds Lochovsky, F.H. and Allen, R.B.). Special issue of *ACM SIGOIS Bulletin*, 11 (2-3), 271-280.
  11. McKnight, C., Dillon, A. and Richardson, J. (1989). Problems in hyperland? A human factors perspective. *Hypermedia*, 1 (2), 167-178.
  12. Nielsen, J. (1990). The art of navigating through hypertext. *Communications of the ACM*, 33 (3), 296-310.
  13. Turtle, H.R. and Croft, W.B. (1992). A comparison of text retrieval models. *The Computer Journal*, 35 (3), 279-290.
  14. Canter, D., Rivers, R. and Storrs, G. (1985). Characterizing user navigation through complex data structures. *Behaviour and Information Technology*, 4 (2), 93-102.
  15. Agosti, M. (1988). Is hypertext a new model of information retrieval? In *Proceedings of the 12th International Online Information Meeting* (London, December 6-8) (Ed. Learned Information Ltd.), Vol. I. Learned Information Ltd., Oxford, New Jersey, pp. 57-62.
  16. Agosti, M., Colotti, R., Gradenigo, G., Matiello, P., Archi, A., Di Giorgi, R.M., Inghirami, B., Nannucci, R. and Ragona, M. (1989). New perspectives in information retrieval techniques: A hypertext prototype in environmental law. In *Proceedings of the 13th International Online Information Meeting* (London, December 12-14) (Ed. Learned Information Ltd.). Learned Information Ltd., Oxford, New Jersey, pp. 483-494.
  17. Agosti, M., Gradenigo, G. and Marchetti, P.G. (1991). Architecture and functions for a conceptual interface to very large online bibliographic collections. In *Proceedings of RIAO '91: Intelligent text and image handling* (Barcelona, April 2-5) (Ed. Lichnerowicz, A.), Vol. 1. Elsevier, Amsterdam, pp. 2-24.
  18. Bruza, P.D. and van der Weide, Th.P. (1990). Two level hypermedia - an improved architecture for hypertext. In *Proceedings of the Data Base and Expert System Applications Conference DEXA '90* (Eds Tjoa, A M. and Wagner, R.). Springer-Verlag, Berlin, Heidelberg, New York, pp. 76-83.
  19. Lucarella, D. (1990). A model for hypertext-based information retrieval. In *Hypertext: Concepts, systems and applications* (Eds Rizk, A., Streitz, N. and André, J.), The Cambridge Series on Electronic Publishing. Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney, pp. 81-94.
  20. Sølvsberg, I., Nordbø, I. and Aamodt, A. (1991). Knowledge-based information retrieval. *Future Generations Computer Systems*, 7 (4), 379-390.
  21. Arents, H.C. and Bogaerts, W.F.L. (1992). Information structuring for intelligent hypermedia: A knowledge engineering approach. In *Proceedings of the 3rd International Conference on Database and Expert Systems Applications* (Valencia, September 2-4) (Eds Tjoa, A M. and Ramos, I.). Springer-Verlag, Wien, New York, pp. 369-372.
  22. Soergel, D. (1992). *Information structure management: A unified framework for indexing and searching in database, expert, information-retrieval, and hypermedia systems*. College of Library and Information

Services Technical Report. University of Maryland, Maryland, USA, 60 pp.

23. Forrester, M.A. (1993). Hypermedia and indexing: Identifying appropriate models from user studies. In *Proceedings of the 17th International Online Information Meeting (London, December 7-9)* (Eds Raitt, D.I. and Jeapes, B.). Learned Information Ltd., Oxford, New Jersey, pp. 313-323.
24. Agosti, M. (1991). New potentiality of hypertext systems in information retrieval operations. In *Human aspects in computing: Design and use of interactive systems and work with terminals* (Ed. Bullinger, H.-J.), *Advances in Human Factors/Ergonomics*, Vol. 18A. Elsevier Science Publishers B.V., Amsterdam, London, New York, Tokyo, pp. 317-321.
25. Salton, G. and McGill, M.J. (1987). *An introduction to modern information retrieval* (3rd print). McGraw-Hill Book Company, New York, 448 pp.
26. Savoy, J. (1993). Searching information in hypertext systems using multiple sources of evidence. *International Journal of Man-Machine Studies*, **38** (6), 1017-1030.
27. Crouch, C.J. (1990). An approach to the automatic construction of global thesauri. *Information Processing & Management*, **26** (5), 629-640.
28. Crouch, C.J. and Yang, B. (1992). Experiments in automatic statistical thesaurus construction. In *Proceedings of SIGIR '92: Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 77-88.
29. Agosti, M., Melucci, M. and Crestani, F. (1994). TACHIR: A Tool for the Automatic Construction of Hypertexts for Information Retrieval. In *Proceedings of RIAO '94: Intelligent multimedia information retrieval systems and management* (New York, October 11-13) (Eds CASIS Inc. and CID), pp. 338-357.
30. McMath, C.F., Tamaru, R.S. and Rada, R. (1989). A graphical thesaurus-based information retrieval system. *International Journal of Man-Machine Studies*, **31** (2), 121-147.
31. Rockmore, M. (1992). Structuring a flexible faceted thesaurus record for corporate information retrieval. In *Classification research for knowledge representation and organization* (Eds Williamson, N.J. and Hudon, M.). Elsevier Science Publishers B.V., Amsterdam, London, New York, Tokyo, pp. 319-328.
32. Duncan, E.B. (1989). A faceted approach to hypertext? In *Hypertext: Theory into practice* (Ed. McAleese, R.). Blackwell Scientific Publications Ltd., Oxford, pp. 157-163.
33. Albrechtsen, H. (1991). Subject representation of software concepts: A semi-automatic indexing approach. In *Proceedings of the World Congress on Expert Systems* (Orlando, Florida, December 16-19) (Ed. Liebowitz, J.), Vol. 4. Pergamon Press, New York, Oxford, Seoul, Tokyo, pp. 2776-2784.
34. Albrechtsen, H. (1992). PRESS: A thesaurus-based information system for software reuse. In *Classification research for knowledge representation and organization* (Eds Williamson, N.J. and Hudon, M.). Elsevier Science Publishers B.V., Amsterdam, London, New York, Tokyo, pp. 137-144.
35. Madigan, D., Chapman, C.R., Gavrin, J., Villumsen, O. and Boose, J. (1994). Repertory hypergrids: An application to clinical practice guidelines. In *ECHT '94 Proceedings* (Edinburgh, United Kingdom, September 18 - 23) (Eds Chambel, T. and Moreno, C.). ACM Press, New York, pp. 117-125.
36. Wille, R. (1992). Concept lattices and conceptual knowledge systems. *Computers & Mathematics with Applications*, **23** (6-9), 493-515.
37. Scheich, P., Skorsky, M., Vogt, F., Wachter, C. and Wille, R. (1992). Conceptual data systems. In *Information and classification: Concepts, methods and applications* (Eds Opitz, O., Lausen, B. and Klar, R.). Springer-Verlag, Berlin, Heidelberg, New York, pp. 72-84.
38. Ginsberg, A. (1993). A unified approach to automatic indexing and information retrieval. *IEEE Expert*, **8**

- (5), 46-56.
39. Pedersen, G.S. (1993). A browser for bibliographic information retrieval, based on an application of lattice theory. In *Proceedings of SIGIR '93: Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Pittsburgh, Pennsylvania, June 27 - July 1) (Eds Korfhage, R., Rasmussen, E. and Willett, P.). Special issue of *ACM SIGIR Forum*, **27**, 270-279.
  40. Belkin, N.J., Marchetti, P.G. and Cool, C. (1993). BRAQUE: Design of an interface to support user interaction in information retrieval. *Information Processing & Management*, **29** (3), 325-344.
  41. Bruza, P.D. (1990). Hyperindices: A novel aid for searching in hypermedia. In *Hypertext: Concepts, systems and applications* (Eds Rizk, A., Streitz, N. and André, J.), The Cambridge Series on Electronic Publishing. Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney, pp. 109-122.
  42. Bruza, P.D. and van der Weide, Th.P. (1991). The modelling and retrieval of documents using index expressions. *ACM SIGIR Forum*, **25** (2), 91-103.
  43. Bosman, F.J.M., Bouwman, R.W.T., and Bruza, P.D. (1991). The effectiveness of navigable information disclosure systems. In *Proceedings of the Informatiewetenschap 1991 Conference* (Ed. G.A.M. Kempen), pp. 55-69.
  44. Bruza, P.D. (1993). *Stratified information disclosure: A synthesis between hypermedia and information retrieval* (Ph.D. dissertation). Thesis Publishers, Amsterdam, 159 pp.
  45. Arents, H.C. and Bogaerts, W.F.L. (1993). Concept-based retrieval of hypermedia information: From term indexing to semantic hyperindexing. *Information Processing & Management*, **29** (3), 373-386.
  46. Arents, H.C. and Bogaerts, W.F.L. (1994). Knowledge-based indexing of hypermedia information for task-related navigation. In *Moving toward expert systems globally in the 21st century: Proceedings of the Second World Congress on Expert Systems* (Estoril, January 10-14) (Ed. Liebowitz, J.). Scholium International Inc., Port Washington, New York, pp. 850-859.
  47. Arents, H.C. and Bogaerts, W.F.L. (1993). Navigation without links and nodes without contents: Intensional navigation in a third-order hypermedia system. *Hypermedia*, **5** (3), 187-204.
  48. Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufman Publishers, San Mateo, California.
  49. Croft, W.B. and Turtle, H. (1989). A retrieval model for incorporating hypertext links. In *Hypertext '89 Proceedings* (Pittsburgh, Pennsylvania, November 5-8) (Ed. Meyrowitz, N.). ACM Press, New York, pp. 213-224.
  50. Croft, W.B. and Turtle, H.R. (1993). Retrieval strategies for hypertext. *Information Processing & Management*, **29** (3), 313-324.
  51. Frisse, M.E. and Cousins, S.B. (1989). Information retrieval from hypertext: Update on the Dynamic Medical Handbook project. In *Hypertext '89 Proceedings* (Pittsburgh, Pennsylvania, November 5-8) (Ed. Meyrowitz, N.). ACM Press, New York, pp. 199-212.
  52. Woods, W.A. (1975). What's in a link: Foundations for semantic networks. In *Representation and understanding: Studies in cognitive science* (Eds Bobrow, D.G. and Collins, A.). Academic Press Inc., New York, San Francisco, London, pp. 35-82.
  53. Rada, R., Mhashi, M. and Barlow, J. (1990). Hierarchical semantic nets support retrieving and generating hypertext. *Information and Decision Technologies*, **16** (2), 117-136.
  54. Rada, R., Barlow, J., Potharst, J., Zanstra, P. and Bijstra, D. (1991). Document ranking using an enriched

- thesaurus. *Journal of Documentation*, **47** (3), 240-253.
55. Collier, G.H. (1987). Thoth-II: Hypertext with explicit semantics. In *Hypertext '87 Proceedings* (Chapel Hill, North Carolina, November 13-15) (Eds Smith, J.B. and Halasz, F.). ACM Press, New York, pp. 269-289.
56. Clitherow, P., Riecken, D. and Muller, M. (1989). VISAR: A system for inference and navigation in hypertext. In *Hypertext '89 Proceedings* (Pittsburgh, Pennsylvania, November 5-8) (Ed. Meyrowitz, N.). ACM Press, New York, pp. 293-304.
57. Nanard, J. and Nanard, M. (1991). Using structured types to incorporate knowledge in hypertext. In *Hypertext '91 Proceedings* (San Antonio, Texas, December 15-18) (Eds Stotts, P.D. and Furuta, R.K.). ACM Press, New York, pp. 329-343.
58. Nanard, J. and Nanard, M. (1993). Should anchors be typed too? An experiment with MacWeb. In *Hypertext '93 Proceedings* (Seattle, Washington, November 14-18) (Eds Kacmar, C.J. and Schnase, J.L.). ACM Press, New York, pp. 51-62.
59. Arents, H.C., Bogaerts, W.F.L. and Agema, K.S. (1990). Authoring a CD-ROM hypermedia system for corrosion engineers. In *Proceedings of the 14th International Online Information Meeting* (London, December 11-13) (Ed. Learned Information Ltd.). Learned Information Ltd., Oxford, New Jersey, pp. 13-24.
60. Albrechtsen, H. (1991). Subject representation of software concepts: A semi-automatic indexing approach. In *Proceedings of the World Congress on Expert Systems* (Orlando, Florida, December 16-19) (Ed. Liebowitz, J.), Vol. 4. Pergamon Press, New York, Oxford, Seoul, Tokyo, pp. 2776-2784.
61. Chen, H., Lynch, K.J., Basu, K. and Dorbin Ng, T. (1993) Generating, integrating, and activating thesauri for concept-based document retrieval. *IEEE Expert*, **8** (2), 25-34.
62. Di Nubila, B., Gagliardi, I., Macchi, D., Milanese, L., Padula, M. and Pagani, R. (1994). Concept-based indexing and retrieval of multimedia documents. *Journal of Information Science*, **20** (3), 185-196.
63. Boy, G.A. (1991). Indexing hypertext documents in context. In *Hypertext '91 Proceedings* (San Antonio, Texas, December 15-18) (Eds Stotts, P.D. and Furuta, R.K.). ACM Press, New York, pp. 51-61.
64. Baudin, C., Kedar, S., Underwood, J.G. and Baya, V. (1993). Question-based acquisition of conceptual indices for multimedia design documentation. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*. AAAI Press, pp. 452-458.
65. Baudin, C., Pell, B. and Kedar, S. (1994). Increasing levels of assistance in refinement of knowledge-based retrieval systems. *Knowledge Acquisition*, **6** (2), 179-196.
66. Bareiss, R. and Osgood, R. (1993). Applying AI models to the design of exploratory hypermedia systems. In *Hypertext '93 Proceedings* (Seattle, Washington, November 14-18) (Eds Kacmar, C.J. and Schnase, J.L.). ACM Press, New York, pp. 94-105.
67. Crouch, D.B., Crouch, C.J. and Andreas, G. (1989). The use of cluster hierarchies in hypertext information retrieval. In *Hypertext '89 Proceedings* (Pittsburgh, Pennsylvania, November 5-8) (Ed. Meyrowitz, N.). ACM Press, New York, pp. 225-237.
68. Parunak, H.V.D. (1991). Don't link me in: Set based hypermedia for taxonomic reasoning. In *Hypertext '91 Proceedings* (San Antonio, Texas, December 15-18) (Eds Stotts, P.D. and Furuta, R.K.). ACM Press, New York, pp. 233-242.
69. Parunak, H.V.D. (1993). Hypercubes grow on trees (and other observations from the land of hypersets). In *Hypertext '93 Proceedings* (Seattle, Washington, November 14-18) (Eds Kacmar, C.J. and Schnase, J.L.). ACM Press, New York, pp. 73-81.

70. Marchionini, G. and Shneiderman, B. (1988). Finding facts vs. browsing knowledge in hypertext systems. *IEEE Computer*, **21** (1), 70-80.
71. Herczeg, J., Hohl, H. and Ressel, M. (1991). HyperQuery - Ein Anfragesystem mit Graphischer Benutzeroberfläche. In *Information Retrieval: Proceedings of the GI/GMD-Workshop* (Darmstadt, 23-24 June), Informatik-Fachberichte, Vol. 289. Springer-Verlag, Berlin, Heidelberg, New York, pp. 152-162.
72. Kupka, I. and Fiege, G. (1992). Navigational retrieval for ceramic materials information. *Swiss Materials*, **4** (1), 5-11.
73. Dillon, A., McKnight, C. and Richardson, J. (1990). Navigation in hypertext: A critical review of the concept. In *Proceedings of the IFIP TC 13 Third International Conference on Human-Computer Interaction INTERACT '90* (Cambridge, August 27-31) (Eds Diaper, D., Gilmore, D., Cockton, G. and Shackel, B.). Elsevier Science Publishers B.V. (North-Holland), Amsterdam, New York, Oxford, Tokyo, pp. 587-592.
74. Bernstein, M., Brown, P.J., Frisse, M., Glushko, R., Landow, G. and Zellweger, P. (1991). Structure, navigation, and hypertext: The status of the navigation problem. In *Hypertext '91 Proceedings* (San Antonio, Texas, December 15-18) (Eds Stotts, P.D. and Furuta, R.K.). ACM Press, New York, pp. 363-366.
75. Duval, E. and Olivie, H. (1993). Towards the integration of a query mechanism and navigation for retrieval of data on multimedia documents. *ACM SIGIR Forum*, **26** (2), 8-25.
76. Pollard, R. (1993). A hypertext-based thesaurus as a subject browsing aid for bibliographic databases. *Information Processing & Management*, **29** (3), 345-357.
77. Thompson, R.H. and Croft, W.B. (1989). Support for browsing in an intelligent text retrieval system. *International Journal of Man-Machine Studies*, **30** (6), 639-668.
78. Shum, S. (1990). Real and virtual spaces: Mapping from spatial cognition to hypertext. *Hypermedia*, **2** (2), 133-158.
79. Fox, E.A., Chen Q.-F. and France, R.K. (1991). Integrating search and retrieval with hypertext. In *Hypertext/Hypermedia handbook* (Eds Berk, E. and Devlin, J.). McGraw-Hill Software Engineering Series, Intertext Publications, McGraw-Hill Publishing Company, Inc., New York, pp. 329-355.
80. Brown, P.J. (1989). Do we need maps to navigate round hyperdocuments? *Electronic publishing: Origination, dissemination and design*, **2** (2), 91-100.
81. Utting, K. and Yankelovich, N. (1989). Context and orientation in hypermedia networks. *ACM Transactions on Office Information Systems*, **7** (1), 58-84.
82. Craven, T.C. (1992). Concept relation structures and their graphic display. In *Classification research for knowledge representation and organization* (Eds Williamson, N.J. and Hudon, M.). Elsevier Science Publishers B.V., Amsterdam, London, New York, Tokyo, pp. 49-59.
83. Aboud, M., Chrisment, C., Razouk, R., Sedes, F., and Soule-Dupuy, C. (1993). Querying a hypertext information retrieval system by the use of classification. *Information Processing & Management*, **26** (3), 387-396.
84. Zizi, M. and Beaudouin-Lafon, M. (1994). Accessing hyperdocuments through interactive dynamic maps. In *ECHT '94 Proceedings* (Edinburgh, United Kingdom, September 18 - 23) (Eds Chambel, T. and Moreno, C.). ACM Press, New York, pp. 126-135.
85. Lucarella, D., Parisotto, S. and Zanzi, A. (1993). MORE: Multimedia Object Retrieval Environment. In *Hypertext '93 Proceedings* (Seattle, Washington, November 14-18) (Eds Kacmar, C.J. and Schnase, J.L.). ACM Press, New York, pp. 39-50.
86. Mukherjea, S., Foley, J.D., Hudson, S.E. (1994). Interactive clustering for navigating in hypermedia systems.

In *ECHT '94 Proceedings* (Edinburgh, United Kingdom, September 18 - 23) (Eds Chambel, T. and Moreno, C.). ACM Press, New York, pp. 136-145.

87. Carpineto, C. and Romano, G. (1994). Dynamically bounding browsable retrieval spaces: An application to Galois lattices. In *Proceedings of RIAO '94: Intelligent multimedia information retrieval systems and management* (New York, October 11-13) (Eds CASIS Inc. and CID), pp. 533-547.
88. Halasz, F.G. (1988). Reflections on NoteCards: Seven issues for the next generation of hypermedia systems. *Communications of the ACM*, **31** (7), 836-852.
89. Halasz, F.G. (1991). Seven issues: Revisited. Slides of closing lecture at *Hypertext '91* (San Antonio, Texas, December 15-18).
90. Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H.F. and Secret, A. (1994). The World-Wide Web. *Communications of the ACM*, **37** (8), 76-82.