Information Retrieval as Interactive Cognitive Processes

Peter Ingwersen

Department of Information Retrieval Theory, Royal School of Librarianship,
Birketinget 6, DK 2300 Copenhagen S, Denmark. e-mail: pi@db.dk

Abstract. This paper provides an over-view of the cognitive perspectives for the design and implementation of information retrieval systems. It argues for an holistic view of information retrieval that is capable of encompassing the algorithmic as well as user-centered research approaches to information retrieval interaction.

1. Introduction

The subject of information retrieval, or IR, involves the development of computer systems for the storage and retrieval of (predominantly) textual information. IR techniques were initially developed for the retrieval of references to documents from bibliographic databases. and the discussion that follows assumes this form of textual information. However, the techniques that have been developed for searching bibliographic databases are equally applicable to any sort of textual information, such as reports of meetings, legal contracts, newswire stories, film scripts, technical manuals and, increasingly over the last few years, multimedia information systems.

Interactive IR from bibliographic databases has now been available for some two decades, either *via* in-house systems or *via* dial-up to online hosts. While the number and the size of the databases have increased hugely over this period, the great majority of them have continued to employ the familiar Boolean retrieval model, in which the query terms are linked by the logical operators (AND, OR and NOT) and in which there is a range of supplementary pattern-matching facilities for truncation and proximity searching. Similar comments apply to many of the CD-ROM-based retrieval systems that have been introduced in the last few years.

The Boolean model is well understood, but has several inherent limitations that lessen its attractiveness for text searching[1-3]. There has thus been substantial interest in the development of alternative methods for text searching that are more appropriate for end-users: IR systems based on such methods are normally referred to as *best-match*, *nearest-neighbour*, *ranked-output*, *vector-processing* or *probabilistic* retrieval systems[3-5]. The research that has been undertaken in this area focuses principally on the algorithms and data structures that are needed to maximise *retrieval effectiveness*, *i.e.*, the ability of the system to retrieve documents from a database that are relevant to a user's query, whilst maintaining a reasonable level of *retrieval efficiency*, *i.e.*, the ability of the system to carry out its functions with the minimal use of machine resources.

An algorithmic focus, whether Boolean or best-match, is not inappropriate if one considers the design of IR systems for trained professionals who can be expected to make themselves fully conversant with the particular systems that they need to use in their day-to-day business. Examples of such professionals are librarians, lawyers, online intermediaries and an increasing number of academic researchers. However, such a focus neglects many of the social and cognitive processes that are involved in IR, and these processes are likely to be of great significance if one is to design effective retrieval systems for inexperienced users, for whom database searching is of only minor importance. Specifically, the algorithmic approach has two principal limitations, as detailed below.

The first limitation is that no account is taken of the large body of studies that have been carried out on users' *information seeking behaviour* (*i.e.*, on the formation, nature and properties of a user's *information need*[6-8]); and the second limitation is that there is an almost-total lack of real-life investigations of the impact of the algorithmic techniques on users in socio-organisational contexts. These limitations have provided the driving force for a range of communicative and psycho-sociological studies of IR systems. The studies have been motivated by the belief that an understanding of user behaviour and user-system communication will permit the construction of *knowledge-based intermediary systems* that can support an individual's search for information in various ways, *e.g.*, by identifying a suitable combination of retrieval techniques[9]. Thus far, these studies have considered only large-scale Boolean systems but they have sufficed to show that the user's background knowledge of the information that is being sought can play a vital role in the retrieval process, as do the reasons for the information request and the subject domain. As a result, several models of *intermediary functionality* have been formulated and partially tested over the last few years.[10,11]

Research on user-centred approaches to IR led to the observation that individual information needs may be stable, but that they may also change during the course of an interaction with an IR system; moreover, these needs may be ill-defined owing to a lack of appropriate background knowledge. The research that has been carried out has also shown that it is necessary to *contextualise* the information need by means of supplementary information on intent, purpose and goals. Information seeking and the formation of the information need are hence assumed to be a *process of cognition* by the individual searcher, in which the retrieval system and the intermediary functionalities are the crucial components of the contextualisation process. An immediate consequence of this approach to information retrieval is that the wide range of representational and searching techniques now available are seen as complementary information structures of different nature and cognitive origin. This, in turn, leads to the notion of a *cognitive theory of information retrieval*, which signifies an attempt to globalise information retrieval by regarding all of its components as representing cognitive structures of varying degrees of complexity that co-operate in an interactive communication process.[12]

The next section of this paper outlines the cognitive perspective that has been introduced above, with the relationships between the various components of the two approaches being summarised by the diagram shown in Figure 1. The paper concludes by noting some of the current research areas that may help further to define these two very different, but complementary, approaches to the design of IR systems.

It is not possible, in a brief survey paper such as this, to provide detailed accounts of the algorithmic and cognitive approaches; however, the listed references should provide an entry point to the very large body of research that has been undertaken to date. More detailed accounts from the algorithmic viewpoint are provided by Belkin and Croft[1], Frakes and Baeza-Yates[4] and by Salton[5], while Ellis[13] and Ingwersen[14] provide comparable accounts from the cognitive viewpoint. Very recently, Ingwersen and Willett have outlined the most pre-dominant features of both these approaches[15]. Current research in both of these areas is reported in the proceedings of the *International Conference on Research and Development in Information Retrieval*, which is held annually under the auspices of the Specialist Group in Information Retrieval of the Association for Computing Machinery.

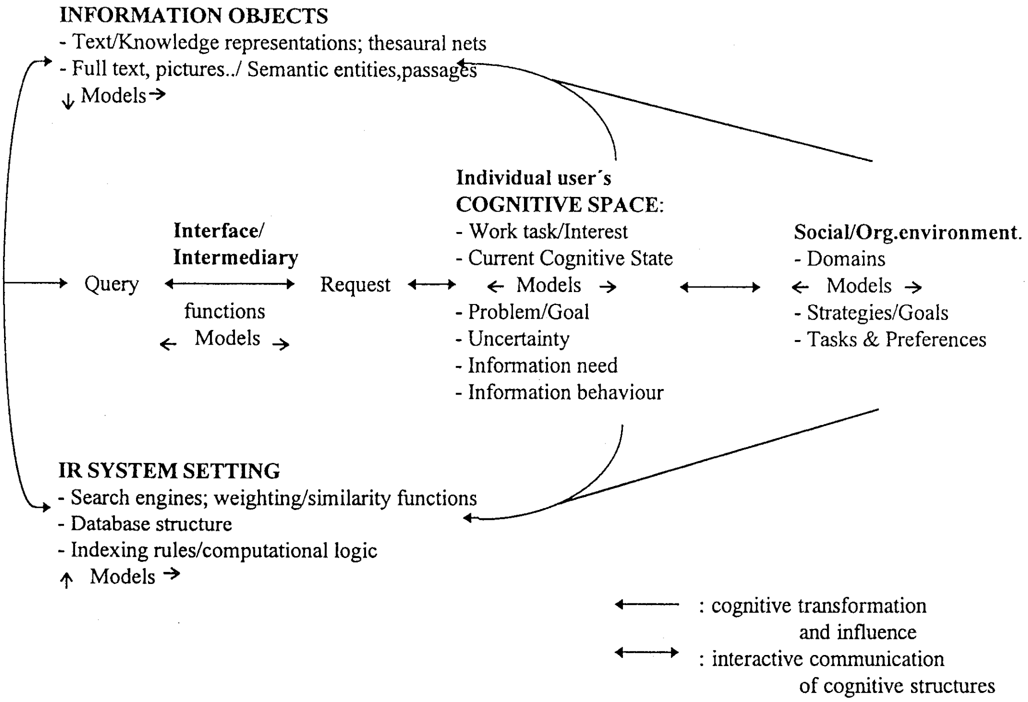

Fig. 1. Cognitive model of IR interaction. Extension of [14, p. 23].

2. User-oriented and cognitive approaches

The user-centred approach to IR is principally based on cognitive psychology and social science methods. The approach has provided substantial insights into users' mental behaviour and in their information seeking characteristics, both on an individual basis and in social and/or organisational contexts (as shown in the centre and the right-hand side of Figure 1). It has also supplied a fair amount of information about inter-human information interactions, such as the interaction between a librarian or information specialist and a user. Finally, the role of the (human) intermediary has been defined in relation to User and Request Model building by means of search interviewing and feedback from IR systems. However, just as the traditional algorithmic approach disregards the dynamic role of the user, so the user-oriented tradition does not encompass the full range of IR system factors.

Until the mid-eighties, no investigations had taken place that involved non-Boolean retrieval and different methods of representation as well as intermediaries and users[16]. This 'monolithic' situation seems understandable, since without established models of searcher (users and intermediaries) behaviour, such advanced experiments could not yield results that were valid for design and test purposes or for the development of IR theories. Such models are now becoming available.

2.1 Information retrieval interaction - the cognitive turn

The complex nature of the information need[12,14] makes it obvious that research on IR techniques alone cannot provide a complete understanding of the entire process of retrieving information. This process must be seen in its totality by incorporating the *system characteristics*, including the representational and retrieval techniques that characterise algorithmic approaches, with *the user's situational characteristics* and the necessary *intermediary functionalities* (see Figure 1). In IR interaction, the intermediary (or user-interface) is the principal mechanism linking the system and the user. The Monstrat Model[10], which has been mentioned previously, was an attempt to functionalise the human intermediary behaviour, but was mainly directed towards the user. The later, Mediator Model[14] suggests that equal importance should be given to both the user and the system.

In this framework one may observe two 'schools': *intelligent IR* tries to *simulate* the human behaviour of mediation by means of extensive user model building, Request Model Building (RQMB) and computational inference techniques; and the *supportive approach to IR* tries to *stimulate* a user's mental processes during IR by means of tailored conceptual feedback from the system driven by the underlying domain model and RQMB. Both approaches provide a common platform for researchers in IR and AI. Since IR is too broad an environment for expert-system-like AI solutions, current efforts are concentrating on finding an appropriate balance between model building, inference, and user support of conceptual nature.

In addition to the RQMB functionality Mediator stresses the minimal application of user model building. This latter type of model building is assumed to encompass only two dimensions of expertise by extraction from the current user: *conceptual expertise* relating to the actual topic or domain; and the user's *current retrieval competence*. The construction of more elaborate user models (based, *e.g.*, on general knowledge, education, age, *etc.*) seems to be of use only in very narrow and consistent domains. Depending on the answers that are received to the questions about the user's current expertise, the level of support and mode of man-machine dialogue may be determined and adopted by the system; an example of this approach is provided by the I^3R system described by Croft and Thomson[16]. Depending on what is known of seeking characteristics from the mandatory domain analysis, information on expertise may also be used to infer how the available IR techniques should be applied, *e.g.*, it might be decided to use all of the available techniques in the system for immediate and specific retrieval in the case of a user who has extensive domain knowledge. This concurrent application of several algorithmic techniques leads to various kinds of data (dif)fusion, of retrieval overlaps and of types of relevance feedback during IR.

The RQMB stage involves elaborating the cognitive characteristics of the request, *i.e.*, the underlying problem situation, work task, and domain or interest - see Fig. 1, centre and right-hand side. Search preferences form part of this model-building functionality. RQMB is meant to provide the system with additional structured contexts, and is not simply concerned with the request formulation itself. The assumption is made that *several simultaneous representations* of the same personal cognitive space may yield improved retrieval results and feedback for further modifications of request, problem or task. The advantage is that the intermediary mechanism is free to perform a kind of *cognitive fusion* of the representations, or to make separate use of each individual representative structure.

2.2 Cognitive IR Theory - information in context

The user-oriented research that has been carried out thus far gives rise to a significant question: how deeply do we need to understand what the user really means. Logically, this is only theoretically possible if the intermediary has found that the current user's information need belongs to the verificative or conscious topical types. If the need is vague or intrinsically ill-defined such an understanding is impossible, regardless of the questions that are posed to the user. This gives rise to the paradoxical situation in which if the user actually possesses a great deal of knowledge about his need for information, he is more capable of assessing the usefulness of the retrieval outcome intelligently than the system is capable of estimating the meaning of the actual need (since the system's background knowledge is severely limited and insufficient). Thus, the conclusion is that IR systems may, at most, provide some *support* at a structural linguistic level by stimulating the user's associative and intuitive thinking processes, which are at a cognitive and pragmatic linguistic level.

The development of a cognitive theory for IR is thus an attempt to understand these uncertainty situations and paradoxes in an *holistic* manner, and to propose a framework for workable solutions. A conceivable way to achieve such a framework would be to make simultaneous use of the *variety* of information structures which are to be found associated with the Information Objects, the System Setting, and the Cognitive Space of users (as depicted in Fig. 1). The basic assumption is that, by supplying structures of suitable contextual nature to all three retrieval components during interaction, uncertainty can be reduced and improved support can be provided for heuristic searching.

Cognitive IR models suggest that we should view IR interactions as the interactions of various types of cognitive structures, as demonstrated in Figure 1. Cognitive structures are generally understood as manifestations of human cognition, reflection or ideas [12]. In IR they take the form of transformations generated by a variety of human actors, *i.e.*, belonging to a variety of different *intentionalities* and *cognitive origins*. These include systems designers and producers, IR technique developers, indexing rule constructors, indexers, authors of texts and images, intermediary mechanism designers, and users in a domain-related societal or organisational context. In the System Setting an IR system designers' cognitive structures may be represented by specific database architectures and one or several matching algorithms or logics. Human indexers' cognitive structures are represented by the index terms added to the original Information Objects. These terms are essentially the result of an intellectual interpretation of an author's text or images, and their assignment is often guided by pre-defined rules and a thesaurus containing semantic relations and knowledge representations that have been developed by other people. Similar problems arise in automatic indexing, where any different weighting function or similarity meassure can also be regarded as a form of transformed cognitive structure. Authors' texts, which include titles, captions, headings, or cited works, are representations of cognitive structures that are intended to be communicated. Later citations pointing to that particular text imply different kinds of interpretations, each carrying its own cognitive background and intentionality. Fig. 2 below demonstrates the *variety of cognitively different access* points to information space. Specific portions of the texts, *e.g.*, titles, abstracts, figures, the introduction, or the full-text sections demonstrate *different functional styles*. Each type of document exhibits an analogous set of differences, as does each domain, and should thus be treated differently.

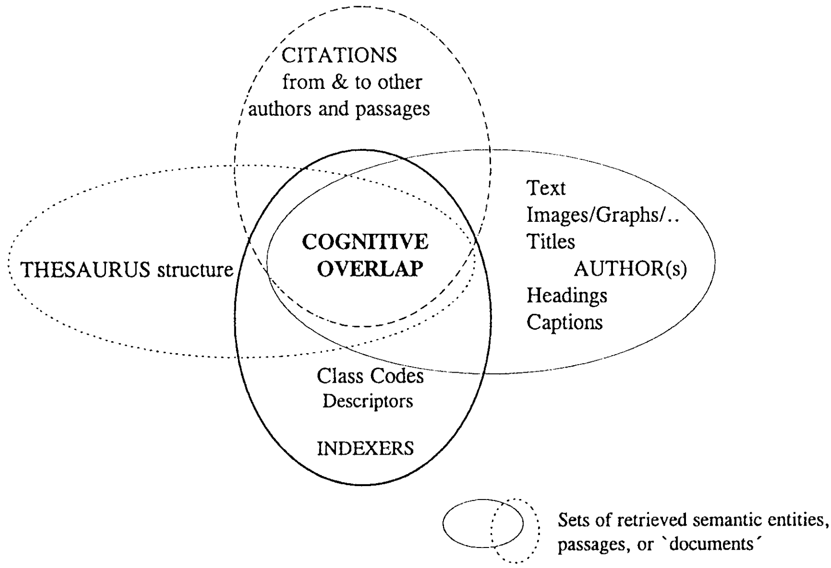

Fig. 2. Overlaps of cognitively different access points provided by one retrieval engine in information space, and associated with *one searcher statement*.

Further cognitive structures are involved in the manipulation of user requests into query formulations during RQMB and retrieval by an intermediary (whether human or computerised), as shown in the centre of Figure 1 where the cognitive structures might, for example, be those of the Monstrat or Mediator models. The right-hand side of the figure summarises the major different cognitive structures of individual users. It is these structures that are identified by an intermediary mechanism and (re)presented to an IR system, *i.e.*, these are the actual work tasks or interests that lead to the current cognitive state and that may be included in the final problem or uncertainty state for the actual user. These mental activities take place in the context of epistemological, social or organisational domains that not only influence the current searcher in a 'historical' socio-semantic sense but also maintain a continuous influence on the authors of texts and on attitudes to system design (the arrows pointing from the right to the left, Fig. 1). The simplest type of a domain is an academic subject field, which is essentially a *social construct* represented by the collective cognitive structures of the individuals forming that field. Other types of domain include industrial sectors, individual firms and organisations, or professional groupings, such as journalists.

One consequence of the cognitive modelling of IR interaction is the demonstration of its fundamentally *polyrepresentative* nature, in particular in relation to full-text IR[12]; another is the recognition of the futility of performance competition between the different algorithmic and logical approaches to retrieval. This can be overcome by replacing it with investigations of their exact characteristics when interacting with the cognitive space of users *and* the types of Information Objects. Accordingly, we must consider how best to fit together such representations and structures during IR.

One recent step forward has been the introduction of *passage retrieval* in full-text systems[17]. Another step has been to allow for *manual query modification* during experimentation, *e.g.*, in the ongoing large-scale TREC[18] and OKAPI[19] experiments. Manual query modification is necessary for two reasons: Firstly, the feedback from the system provides the basis for relevance and utility judgements of text portions, *e.g.*, by means of marking-up the relevant portions of an information object like passages of text. Secondly, it also provides the basis for *improved cognition* by the user of his actual need for information, and, possibly, of his underlying problem or goal, by forcing him to interpret the search outcome. This outcome does not have to be monolithic, *i.e.*, one simple ranked list, but may also contain pointers to *several* conceivable routes into information space, *e.g.*, by hypertext links, condensed or structured lists of concepts, and analogous means of conceptual feedback. Any modification of the request, or problem or work task statements by selections and/or deletions of concepts then mirrors the altered intrinsic formations and conceptions of the need, problem or task. In this prospective framework the well-known issue of *inter-indexer inconsistency*, for instance, *then becomes an asset*, rather than a problem. Similar inconsistencies have been traced in relation to Boolean and algorithmic retrieval techniques[1,12].

In fact, a cognitive theory would favour *all* kinds of inconsistencies and, in particular, the *retrieval overlaps* between the variety of different cognitive structures involved in IR, see Fig. 2-3. The assumption is that the more remote in cognitive origin, logic, functionality, and in time, the smaller the overlap, Fig. 2; and the better and probably more relevant the retrieval outcome[12]. The conceptions of cognitive retrieval overlaps as well as of data and request *fusion* and *diffusion*, Fig. 3, are thus essential elements of a theory framed by the cognitive perspective.

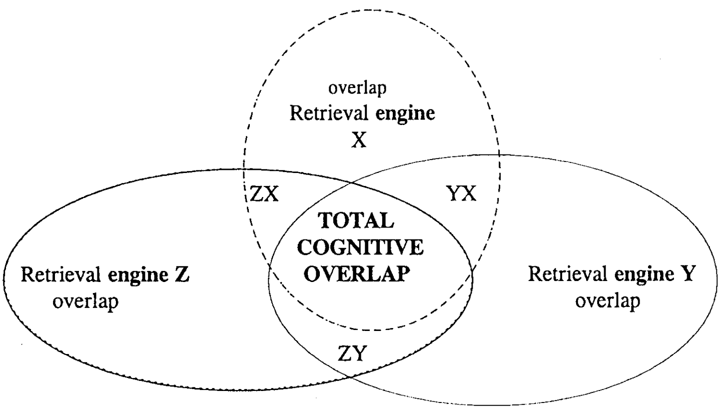

Fig. 3. Cognitive overlaps of passages in information space provided by the intersection of the polyrepresentative retrieval results from different retrieval engines, and associated with *one searcher statement*, and generating one set of total cognitive overlap. From[12].

3. *Conclusions*

It is clear that we shall continue to see developments in the algorithmic approach to IR, but it is most unlikely that these developments will, in themselves, suffice to enable effective searching of the increasing amounts of text that are becoming available in machine-readable form. Efforts are thus being made to embed these new best-match techniques in an holistic model that takes account of all of the factors that are of importance in the retrieval process. This model, which is illustrated in Figure 1, has been discussed from a relatively user-centred approach as well as from the prospective global cognitive view. Human intermediary characteristics are well-understood, principally in relation to user and request model building, and other user-oriented primary functions. The various types of knowledge which are necessary for implementation in (or teaching to) intermediaries has been successfully specified and modelled. Knowledge of human search strategies and tactics, of seeking behaviour, and of conceptual domains and types of information needs, is also fairly well established. For request model building, it seems sufficient to separate the user's underlying work task and problem from the resulting information need, thus allowing or driving the user to elaborate on these cognitive representations during the IR interaction. From a global cognitive approach the mutual advantage for the intermediary (mechanism) is that it is free to perform a *cognitive fusion* of these representative structures, or to make separate use of each in relation to the system. The holistic view suggests the *simultaneous use* of different algorithmic techniques and modes of indexing in information space: this may lead to various kinds of data (dif)fusion, retrieval overlaps, and relevance-feedback possibilities which may support interpretation and elaboration activities in the mental space of the user.

New questions and problems emerge in all rapidly developing sciences. Thus far, the requests in *all* laboratory experiments have been pre-defined sets of *simulated* well-defined and static information needs, in order to define the exact recall (or topical relevance) ratio for each request. The re-introduction of users into the non-Boolean experimental settings brings several profound methodological issues into play:

1. The users, of course, interpret the simulated requests differently during both initial query formulations and later query modifications[18]; that is, they become an uncontrollable variable in an otherwise invariable environment. Problem: how to deal experimentally with the simple conception of topical relevance assessment measures? Any fixed pre-established measure is in principle unreliable in a statistical sense if not carried out by a panel. We may expect inter-panel inconsistencies analogous to those observed previously in studies of inter-indexer consistency.
2. Issues of relevance and evaluation methodology across different systems become still more controversial in experiments in which users pose real-life requests, which are likely to be more variable and more ill-defined in nature than the simulated ones that have been used previously. Problem: only post assessments can be performed, and the statistical population of users, requests and panel participants has to be large.
3. It is likely that the concept of relevance will need to encompass 'relative' as well as 'partial', *i.e.*, non-binary, assessments, differentiated into situational 'usefulness' or 'utility', and 'topicality'. Problem: how can we manage the range of variables introduced by real-life experimentation without introducing a whole range of sociological methodologies?

Other substantive problems that must be faced include issues relating to the definition of appropriate combinations of retrieval logics or algorithms for handling full-text and multimedia information objects and the range of intrinsic information needs associated with such objects? Finally: what is the role of data fusion[20] in this landscape, which fusion techniques should be used, and how can these encompass cognitive retrieval overlaps?

We believe that the algorithmic and the user-centred approaches are complementary in nature, in that algorithmic techniques are necessary if one wishes to search a computerised database and that cognitive techniques are necessary if one wishes to take full account of the contexts in which the searchers operate and the retrieved information is to be used. Accordingly, we hope that this review will help to bridge the divide that still exists between these two approaches to the continuing problem of retrieving information from textual databases.

References

1. Belkin, N.J. and Croft, W.B.: "Retrieval Techniques". *Annual Review of Information Science and Technology*, 22 (1987): 109-145.
2. Cooper, W.S.: "Getting Beyond Boole". *Information Processing and Management*, 24 (1988): 243-248.
3. Willett, P. (editor): *Document Retrieval Systems* (Taylor Graham, London, 1988).
4. Frakes, W.B. and Baeza-Yates, R. (editors): *Information Retrieval: Data Structures and Algorithms* (Prentice-Hall, Englewood Cliffs NJ, 1992).
5. Salton, G.: *Automatic Text Processing. The Transformation, Analysis and Retrieval of Information by Computer* (Addison-Wesley, Reading, Mass., 1989).
6. Su, L.: "The Relevance of Recall and Precision in User Evaluation". *Journal of the American Society for Information Science*, 45 (1994): 207-217.
7. Kuhlthau, C. *Seeking Meaning.* (Ablex Publ., New York, 1993).
8. Ingwersen, P. "Search Procedures in the Library Analysed from the Cognitive Point of View". *Journal of Documentation*, 38 (1982): 165-191.
9. Ingwersen, P.: "Polyrepresentation of Information Needs and Semantic Entities: Elements of a Cognitive Theory for Information Retrieval Interaction". In Croft, W.B. and van Rijsbergen, C.J. (editors): *SIGIR '94. Proceedings of the Seventeenth Annual International Conference on Research and Development in Information Retrieval* (Springer Verlag, Berlin, 1994). 101-110.
10. Belkin, N.J. Brooks, H. and Daniels, P.J.: "Knowledge Elicitation Using Discourse Analysis". *International Journal of Man-Machine Studies*, 27 (1987): 127-144.
11. Fox, E.A., Hix, D., Nowell, L., Bruneni, D., Wake, W., Heath, L. and Rao, D.: "Users, User Interfaces, and Objects: Envision, a Digital Library". *Journal of the American Society for Information Science*, 44 (1993): 480-491.
12. Ingwersen, P.: "Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory". *Journal of Documentation*, 51(1) (1996): 3-50.
13. Ellis, D.: "A Behavioural Approach to Information Retrieval Systems Design". *Journal of Documentation*, 45 (1989): 171-212.
14. Ingwersen, P.: *Information Retrieval Interaction* (Toppan Publ, Tokyo, 1995)(4-8101-8917-1).
15. Ingwersen, P. and Willett, P. An introduction to algorithmic and cognitive approaches for information retrieval. *Libri,* 45 (1995): 160-177.
16. Croft, W.B. and Thomson, R. "I^3R: A New Approach to the Design of Document Retrieval Systems". *Journal of the American Society for Information Science*, 38 (1987): 389-404.
17. Salton, G., Allan, J. and Buckley, C.: "Automatic Structuring and Retrieval of Large Text Files". *Communications of the ACM*, 37(2) (1994): 97-108.
18. Callan, J.P. and Croft, W.B. "An evaluation of query processing strategies using the TIPSTER collection". In Korfhage, R., Rasmussen, E.M. and Willett, P. (editors): *Proceedings of the Sixteenth Annual International Conference on Research and Development in Information Retrieval* (ACM, New York, 1993), 347-356.
19. Hancock-Beaulieu, M.: "Query Expansion: Advances in Research in Online Catalogues". *Journal of Information Science*, 18 (1992): 99-103.
20. Belkin, N.J., Kantor, P., Fox, E.A. and Shaw, J.A.: "Combining the Evidence of Multiple Query Representation for Information Retrieval". *Information Processing and Management*, 31 (1995): 431-448.