4 Retrieval Techniques

NICHOLAS J. BELKIN
Rutgers University

W. BRUCE CROFT
University of Massachusetts

INTRODUCTION

Retrieval Techniques

This review is concerned with a specific aspect of research and development (R&D) in information retrieval (IR) systems—that is, the means for identifying, retrieving, and/or ranking texts (or text surrogates or portions of texts), in a collection of texts, that might be relevant to a given query (or useful for resolving a particular problem). In particular, retrieval techniques address the issue of comparing a representation of a query with representations of texts for the above purpose. Although we necessarily discuss different means of representation, our focus is on different techniques for comparison and not on the generation of the representations. Figure 1 indicates the situation with which we are concerned and shows that different representations allow different retrieval techniques without necessarily specifying them.

Our limitation of the topic of this review also means that a number of techniques used for retrieval, or discussed in the IR literature as having to do with retrieval, will be discussed only as aspects of simpler, or basic matching techniques. For instance, in the context of Figure 1, it is clear that "feedback" techniques are not different retrieval techniques but rather methods

The authors would like to thank Barbara Kwaśnik for her invaluable efforts, bibliographic, technical and otherwise supportive, without which this document could not have been produced.

Annual Review of Information Science and Technology (*ARIST*), Volume 22, 1987
Martha E. Williams, Editor
Published for the American Society for Information Science (ASIS)
by Elsevier Science Publishers B.V.

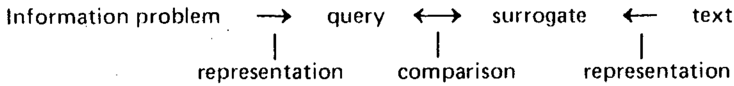

Figure 1 Retrieval technique situation

for enhancing the query or request model, which can then be used with various techniques for eventual or subsequent comparison. Thus, rather than discussing feedback on its own as a retrieval technique, we mention it as a representation mechanism for the various specific retrieval techniques.

Limits on This Review

Although there has been no review in *ARIST* specifically and solely on retrieval techniques as we have defined them, there are two that we consider precursors to this one—namely the chapters by MCGILL & HUITFELDT and by BOOKSTEIN (1985). The former deals with research in information retrieval in general and covers retrieval techniques quite well. It thus provides a starting point for our review. The latter deals with the use of probability and fuzzy set theories in IR and thus also with their related retrieval techniques. Because Bookstein's review is so recent, we do not review in depth all of the post-1980 material he covers. Instead we comment in general on probability and fuzzy set dependent techniques, and provide detailed discussion only of selected work covered by Bookstein, and of material published since his review. Thus, our review covers the R&D literature on retrieval techniques since 1980 except for that dealing with probability theory and fuzzy set theory, which is primarily post-1985.

Our review also omits research concerned primarily with representational issues (whether of query or of text); rather it concentrates on work concerned explicitly with different models for actually comparing query representation and text representation. We also do not treat research on file organization or efficiency of storage and retrieval, concentrating rather on the logic of retrieval. This review covers text retrieval only; it excludes pattern recognition, image matching, numerical representation, etc. as well as chemical structure searching.

In keeping with previous *ARIST* practice, we try to be reasonably comprehensive within these limits in our bibliography, but in the text we discuss at some length selected items as examples of particular approaches. Thus, some items in the bibliography, although read by us, are not cited in the text.

Aims of This Review

This seems to be an especially appropriate time to review R&D in retrieval techniques. The general situation in this field can be characterized as follows. There are a few techniques actually used in large-scale operational IR systems (namely, Boolean or string searching), and there appears to be some general dissatisfaction with them. These techniques have become established more

RP9

through practice than theory. At the same time, there are a number of quite different techniques, usually with a strong theoretical basis, that have been used almost exclusively in experimental settings (e.g., probabilistic retrieval), but these are now well developed in many of their aspects. These "experimental" techniques, when compared with the "operational" techniques in a controlled setting, have almost always performed better on standard measures and often *very* much better. Since these two types of techniques are now well established in their respective settings, why has the experimental experience had so little effect on the operational environment? We attempt to deal with this question on a technical rather than social level.

A related issue is that although there appears to be a general feeling that retrieval techniques can be classed broadly as experimental or operational, the basis for this classification is nothing more than an historical accident of custom or use. To deal with the issues involved in understanding retrieval techniques and the associated problems in IR systems, a more principled and detailed classification seems necessary. One of its aims would be to establish meaningful relationships among the classes of objects involved. We propose such a classification and base the structure of our review on it.

For some time a number of researchers in the field have noted that all retrieval techniques perform better for some queries than for others. In the cases where techniques have been compared with one another on a micro level, it appears that differences in performance on specific queries are masked by evaluation on cumulated results, and that especially, even when specific techniques have been shown to perform more poorly than another technique overall, they may have done much better on at least some individual queries. Thus, are some techniques better for some kinds of queries than others? Although there has not been much research on this issue, there is sufficient experience now to think about an approach to this issue.

Finally, integrated information systems with vastly different problem contexts, databases, and so on within them are now being either constructed or contemplated (e.g., office automation systems, integrated IR and database management systems (DBMS)). Given the experience discussed above, will multiple retrieval techniques be necessary within such systems, and if so, on what bases should they be chosen and used and how can they be integrated within a single system design? With these questions in mind, we hope to provide a framework for work already done and perhaps to provide some guidelines to research on these issues.

Organization of the Review

On the basis of the objectives, the rest of this chapter is structured as follows. First, we present a classification of retrieval techniques and use this classification as the basis for discussing specific R&D in retrieval techniques since about 1980. We then discuss comparative performance studies and the relationship of request representation to retrieval techniques; this leads to the issues of choosing appropriate retrieval techniques and using multiple techniques. The next section deals with architectures for integrated information systems, including so-called "expert" intermediaries, especially from the point of view of problems of retrieval techniques. We conclude with a general discussion of the current status of R&D in retrieval techniques, identification of the issues and problems that appear to be most crucial, and identification of research directions that seem to be the most pressing as well as those that seem most likely to be fruitful in developing retrieval techniques for truly effective information systems.

A CLASSIFICATION OF RETRIEVAL TECHNIQUES

We have defined a retrieval technique as a technique for comparing the query with the document representations. We can further classify retrieval techniques in terms of the characteristics of the retrieved set of documents and the representations that are used. Some techniques do not fall naturally into only a single category in this classification, and others are hybrids of techniques from different categories, but the scheme is useful for discussing the broad distinctions among retrieval techniques. Figure 2 gives a diagrammatic view of the classification. The first distinction that we make among retrieval techniques is whether the set of retrieved documents contains only documents whose representations are an exact match with the query or a partial match with the query. For a partial match, the set of retrieved documents will include also those that are an exact match with the query.

The next level of the classification distinguishes between retrieval techniques that compare the query with individual document representatives and techniques that use a representation of documents that emphasizes connections to other documents in a network. In this category, individual documents are retrieved, but the retrieval is based on connections to other documents and

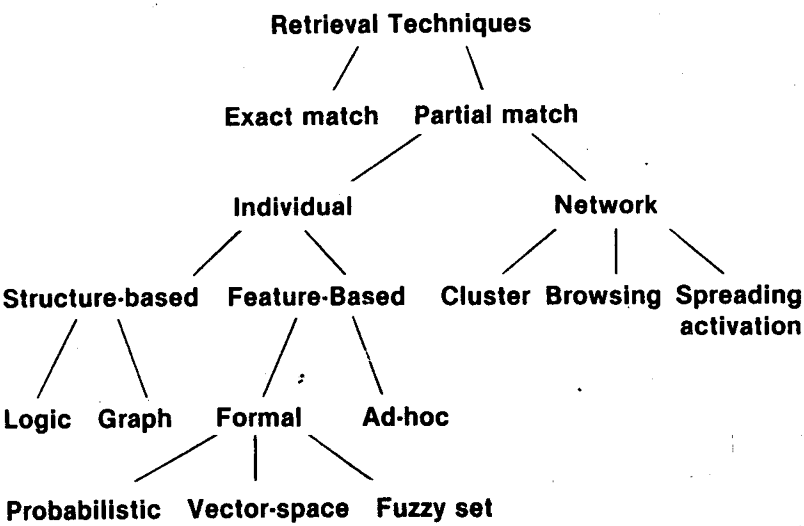

Figure 2 A classification of retrieval techniques

not solely on the contents of an individual document. In the network category, we identify the subcategories of cluster-based searches, searches based on browsing a network of documents, and spreading activation searches.

The individual category breaks down into retrieval techniques that use a feature-based representation of queries and documents and techniques that use a structure-based representation. In a feature-based representation, queries and documents are represented as sets of features, such as index terms. Features can be weighted and can represent more complex entities in the text than single words. The structure-based category is divided into representations based on logic, that is, those in which the meaning of queries and documents are represented using some formal logic, and on representations that are similar to graphs, in which documents and queries are represented by graph-like structures composed of nodes and edges connecting these nodes. Such graphs can be produced by natural language processing (e.g., semantic nets and frames) or statistical techniques.

The feature-based category includes techniques based on formal models (including the vector space model, probabilistic model, fuzzy set model and others) and techniques based on ad-hoc similarity measures. In the following sections, we discuss the techniques that make up these categories in more detail.

EXACT MATCH TECHNIQUES

Exact match retrieval techniques are those that require that the request model be contained, precisely as represented in the query formulation, within the text representation. Implemented as Boolean, full-text, or string searching, this is the retrieval technique in current use in most of the large operational IR systems. The disadvantages of this type of technique are well known and well documented and a variety of aids such as thesauri are required to achieve reasonable performance. In the simple case exact match searching: 1) misses many relevant texts whose representations match the query only partially; 2) does not rank retrieved texts; 3) cannot take into account the relative importance of concepts either within the query or within the text; 4) requires complicated query logic formulation; and 5) depends on the two representations being compared having been drawn from the same vocabulary. BOOKSTEIN (1985) mentions several other undesirable characteristics of Boolean retrieval. Note that we do not consider the provision of "wild card" strings to be true partial matching.

Given its many and obvious objections, why does exact match searching remain the paradigm for operational systems? The traditional answers are that the investment in current systems is so great that changing them is economically unfeasible, that alternative techniques are untested in large-scale environments, and that the results of alternative techniques are not sufficiently better even in experimental environments to justify any changes. A more significant argument, accepted it seems by all parties, is that the structures of Boolean statements represent important aspects of user's queries or problems.

Research in exact match retrieval techniques in the period covered has dealt with all of the problems mentioned above to some extent and with one additional problem: that of efficiency of searching for strings. The major efforts in the logic of exact match retrieval have been in making it less exact, in taking into account relative importance, and in achieving sensible ranking rules. The problem of difficulty of use has been approached mainly via interface design, without reference to the underlying technique, while the vocabulary problem has been considered primarily, although not exclusively, one of request-model elaboration. The efficiency problem has been dealt with by techniques such as file organization (e.g., ARNOW ET AL.), specialized hardware (e.g., CARROLL ET AL.; HOLLAAR, 1984), and text compression (e.g., MOHAN & WILLETT). As far as exact match retrieval techniques themselves are concerned, there has really been no research during the period covered by this review. That is, the only exact match logics that are available are Boolean or simple string matching, and no one has suggested any new exact matching logic. Thus our discussion devolves to brief mention of some attempts to modify exact match searching.

It is possible to relax some constraints on exact match searching by the expedient of specifying parts of the string to be matched which can be ignored. Truncation (that is, ignoring endings of words after some point) and so-called "wild-card" searching are examples of this approach. Since there is really no way to make exact retrieval techniques less exact within their own logic, all attempts to do so have necessarily resulted in hybrids of exact techniques and partial match techniques. Since partial match techniques automatically produce retrieved text rankings, so do these hybrids. Similarly, all attempts to take into account the relative importance of aspects of query and text have relied on partial match techniques of one sort or another. We therefore do not attempt to separate these from one another.

Perhaps the most interesting approach to the question of extending the logic of exact matching has been the recent effort to deal with Boolean techniques as a special case of either vector or probabilistic models. Salton and his colleagues have developed an extended vector model, and CROFT (1986) has proposed a method for making use of the relations established in a Boolean query within a probabilistic search model. For more detail on these approaches, see BOOKSTEIN (1985) and the sections on the "Vector Space Model" and "Probabalistic Model" below.

The vocabulary and efficiency approaches to modifying exact match techniques are more properly seen as attempts to deal with problems common to various retrieval techniques. Thus, for instance, Fox's relational thesaurus for Boolean retrieval (FOX, 1987b) is equally relevant for probabilistic or vector retrieval as are suggestions for a "user thesaurus" (e.g., BATES). Some methods have been developed within operational IR systems to support request-model elaboration. The ZOOM facility on ESA-IRS, for instance, allows frequency-ranked display of terms associated with a retrieved document set (INGWERSEN). All such work deals with the problem of reconciling the query vocabulary with the document or with the index vocabulary in matching and, as such, is relevant to any retrieval technique that faces this issue.

PARTIAL MATCH TECHNIQUES

Individual, Feature-Based Techniques

Techniques in this category are used to compare queries with documents represented as sets of features or index terms. The document representatives are derived from the text of the document either by manual or automated indexing. Similarly, the query terms can either be derived from a query expressed in natural language, or an indexing vocabulary can be used directly for specifying queries. The retrieval techniques used do not depend on the indexing method, and the relative merits of automated and manual indexing are not discussed here. For details on automated indexing techniques and comparisons with manual indexing, see SALTON (1986a), SALTON & MCGILL, and SPARCK JONES.

Features can represent single words, stems, phrases, or concepts and can have weights associated with them. The weights are typically derived from the way the feature is used in an individual document, such as a within document frequency weight, or the way it is used in the document collection, such as the inverse document frequency weight (SALTON & MCGILL, SPARCK JONES).[1] The interpretation of the weights, the way they are used, and the way they are calculated depend on the particular retrieval technique and the underlying retrieval model selected.

In the following discussion of feature-based retrieval techniques, we assume that a document has a representation consisting of a vector of terms $(d_1, d_2, \ldots . d_m)$, where d_i indicates the presence or absence of term i and has the value 1 or 0. Weights associated with these terms are introduced as needed. A query has a similar representation, with q_i referring to the ith query term. Techniques that deviate from the purely feature-based approach but are strongly related to it, such as the partial match Boolean query techniques, are also discussed in this section.

Formal. These retrieval techniques are based on formal models of document retrieval and indexing. In this review, we concentrate on the major modeling approaches that have been used for information retrieval: vector space, probabilistic, and fuzzy set. The reader is referred to BOOKSTEIN (1985), ROBERTSON (1977b), SALTON (1979), and VAN RIJSBERGEN (1979) for further discussions of models of information retrieval. Bookstein's 1985 review, in particular, has described the probabilistic and fuzzy set models in detail, and the presentation in this paper emphasizes the retrieval techniques based on those models, rather than theoretical aspects.

Vector space model. In the vector space model, documents and queries are vectors in an n-dimensional space, where each dimension corresponds to an index term. The model has intuitive appeal and has formed the basis of a

---

[1] The within-document frequency is the number of times an index term (usually a word stem) occurs in the document text (usually the abstract). The inverse document frequency is the inverse of the relative frequency of a term in the collection. The logarithm of this ratio is often used, giving a weight of log N/n, where N is the number of documents in the collection and n is the number of documents that contain the term.

large part of IR research, including the SMART system (SALTON, 1968; SALTON & MCGILL). Although this was one of the first models proposed, modifications to it are still appearing (WONG & RAGHAVAN; WONG ET AL.). From the point of view of retrieval techniques, there have been very specific recommendations about how the model should be applied in operational systems (SALTON, 1981; 1986a). The recommended retrieval process is as follows:

1. Term weights are calculated using a combination of the normalized within-document frequency (*tf*) and the inverse document frequency (*idf*). This *tf.idf* weight can be calculated for document terms either as part of the retrieval process or, less accurately, when the document is indexed.
2. Terms that have poor "discrimination value" (terms that are not useful for distinguishing among documents) are replaced by terms representing thesaurus classes for low-frequency terms and phrases for high-frequency terms. Discrimination values are calculated by observing the document space before and after assignment of a term. If the documents tend to move together after assignment (as measured by the average pairwise similarity), the term is a poor discriminator.
3. Documents are ranked in decreasing order of similarity to the query as measured by the cosine correlation (intuitively retrieving those documents closest to the query in vector space). This is calculated by formula 1:

$$\sum d_i q_i \Big/ \sqrt{\sum d_i^2 q_i^2} \qquad (1)$$

where d_i is the *tf.idf* weight.

There are some important points to make about this retrieval technique that also apply to other techniques. First, the exact formula for calculating term weights, such as *tf.idf*, may vary from one system or set of experiments to another. For example, it is common in calculating the *idf* weight to normalize the document collection frequency with the maximum collection frequency rather than simply the number of documents in the collection; this is done to expand the range of *idf* weights that result. Similarly, in some experiments the *tf* weight is normalized with the maximum within-document frequency, and in others it is not normalized. These details can have a significant impact on the effectiveness of a system, and it is essential to check their validity.

Second, when similarity measures are used, the query is not directly compared with every document in the collection to produce a ranking. Typically, an inverted file is used to exclude documents that have no terms in common with the query. Refinements of this technique have been devised to reduce search time further (BUCKLEY & LEWIT; CROFT & PARENTY; SMEATON & VAN RIJSBERGEN, 1981).

Finally, although the term weighting done in steps 1 and 2 is often regarded as part of the indexing process, it can also be done during retrieval. The weights used follow directly from the retrieval models described in the next section, and because the collection is dynamic, they can be more accurately calculated during retrieval. The identification of important relationships among words and the expansion of terms to include thesaurus classes can also be part of retrieval since it is those relationships and terms that are relevant to a particular query that should be identified (CROFT & THOMPSON, 1987).

An important extension of the retrieval techniques based on the vector space model is extended Boolean retrieval (FOX, 1983; SALTON, 1983, 1985; SALTON & VOORHEES; SALTON ET AL., 1983b; SALTON ET AL., 1985). This technique overlaps the structure-based category because it uses structured (Boolean) queries. That is, the query is formulated with index terms and the Boolean operators AND, OR, and NOT. The prevalence of systems that use Boolean queries has led researchers to consider the problem of producing ranked output from Boolean specifications. In this approach to the problem, a similarity measure is defined that ranks documents, giving precedence to those that match all or part of the Boolean specification. For example, consider a situation in which document 1 contains terms A and B, document 2 contains terms B and C, and the query is A AND (B OR C). Ignoring the effect of term weights, the standard cosine correlation would rank documents 1 and 2 equally because both have two query terms. The extended Boolean similarity measure would rank document 1 higher because it matches the Boolean query specification. The addition of term weights makes the calculation more complex, but the general effect is the same.

Research based on the vector space model has led to other techniques, such as relevance feedback, clustering, and document space modification (SALTON, 1968). These techniques are discussed in other sections.

Probabilistic model. The version of the probabilistic model that is discussed most often in research papers was introduced by ROBERTSON & SPARCK JONES and VAN RIJSBERGEN (1979). Other forms of the model are discussed by BOOKSTEIN (1985), but these have not contributed significantly to the development of new retrieval techniques. The advantages of the probabilistic model are the insights it gives into techniques that have been used in previous research and that it is a powerful framework for developing new techniques.

The retrieval techniques based on the probabilistic model are very similar to those developed from the vector space model. The basic aim is to retrieve documents in order of their probability of relevance to the query (ROBERTSON, 1977a). If we assume that document term weights are either 1 or 0 and that terms are independent of each other, this can be shown to be achieved by ranking documents according to Formula 2:

where q_i is a weight equal to log pr_i (1-pnr_i)/pnr_i (1-pr_i) where pr_i is the probability that term i occurs in the relevant set of documents, and pnr_i is the probability that term i occurs in the nonrelevant set of documents.

The problem in applying this ranking function is to estimate the probabilities in the query term weights. Experimental results have shown that pnr is best estimated from the entire collection (HARPER & VAN RIJSBERGEN). This means that the formula, $pnr_i = n_i/N$ (where n_i is the number of documents that contain term i and N is the number of documents), is a reasonable estimate. If a set of relevant documents is available (for example, after feedback), the best estimate for pr_i is r_i/R, where r_i is the number of relevant documents that contain term i and R is the total number of relevant documents. This is another case in which details are important. The actual estimation formula that provides the best results is $(r_i + 0.5)/(R + 1)$ (SPARCK JONES & WEBSTER). When r_i is 0, this estimate is too high and a value of 0.05 for pr_i is used. If a set of relevant documents is not available, as in an initial search, the retrieval technique suggested by the simple probabilistic model is approximately equivalent to using the *idf* weight in Formula 2 (CROFT & HARPER).

The probabilistic model can be extended to use within-document frequency information (CROFT, 1983b). In this case, the term weight used in Formula 2 is *ts.idf*, where *ts* is a term significance weight that measures the importance of a term to a particular document. The *ts* weight is best estimated with the normalized within-document frequency. This form of the ranking function is virtually identical to that developed from the vector space model (CROFT, 1984). Another approach to incorporating term weights is described by FUHR.

A number of proposals have been made to remove the term independence assumption of the probabilistic model. HARPER & VAN RIJSBERGEN, VAN RIJSBERGEN (1977), and YU ET AL. (1979; 1983) describe retrieval techniques that involve calculating correlations between terms (or term dependencies) in the document collection. These dependencies are then used to expand queries and change estimates of the relevance of documents. It has been shown that if there is sufficient information about the occurrence of terms in relevant documents, these retrieval techniques could significantly improve effectiveness. In practice, however, estimation problems make it difficult to realize any of the potential benefits (VAN RIJSBERGEN ET AL.; YU ET AL., 1983).

Another approach is to identify important dependencies in the query and to use the presence of those dependencies to modify the document scores according to a probabilistic model that assumes term dependence (CROFT, 1986). This avoids the calculation of dependencies that are not used in queries and identifies those dependencies most likely to affect retrieval effectiveness. The retrieval technique based on this model modifies a ranking produced with Formula 2 and *ts.idf* weights by adding a correction factor to a document score for each set of dependent terms that the document contains. The dependencies used by this technique can be identified using groups of query terms joined with the Boolean AND, thereby allowing structured queries to be used with a probabilistic retrieval model.

The maximum entropy approach suggested by COOPER and COOPER & HUIZINGA can be interpreted (as it has been by them) as effectively simulating Boolean relations within a statistical retrieval environment. For details on this work, see BOOKSTEIN (1985).

Fuzzy set. A fuzzy set approach to information retrieval has been discussed in many papers (BOOKSTEIN, 1985). The main contribution of this work in terms of retrieval techniques has been the integration of Boolean queries with ranking techniques. This integration is limited, however, when compared with extended Boolean retrieval based on the vector space model or the use of term dependencies in probabilistic models.

Ad hoc. A number of similarity measures for comparing queries and documents have been proposed in the literature (MCGILL ET AL.; SALTON, 1968). Many of them were developed in the context of numerical taxonomy (SNEATH & SOKAL). Similarity measures typically consist of a measure of the overlap of the query and document sets of terms normalized by the size of the sets involved. For example, Dice's coefficient is $2(Q \cap D)/(|Q| + |D|)$ for queries and documents represented as sets of unweighted index terms. Although these measures are similar to those described in the last section (we do not consider the cosine correlation to be *ad hoc*), they are not based on a particular model of document retrieval. Thus, there is no means for comparing measures apart from exhaustive evaluations such as that done by McGill et al. These types of evaluations are never conclusive, and a more appropriate motivation for using a particular technique for ranking documents is to base the choice on a well-founded retrieval model. Small differences in weights can lead to significant differences in results, and approaching the design of similarity measures in an ad-hoc manner can lead to a confusing collection of results.

Individual, Structure-Based Techniques

In this category of retrieval techniques, either the query or the documents or both are represented by more complicated structures than the sets of terms used in feature-based techniques. We have already encountered retrieval techniques designed to deal with Boolean queries, although these were still primarily feature-based. The types of techniques described here typically rely on a much richer representation of the knowledge in the subject domain covered by the documents and queries. This domain knowledge can be regarded as a more complex form of the thesaurus information found in many systems.

Logic. It is theoretically possible to represent the information conveyed by the text in documents as sentences in a formal logic. For example, the statement, "DEC sells computers," could be represented in first-order predicate calculus as (sells dec computer). Similarly, the statement, "If a company sells computers, it is financially viable," could be represented as (forall (x) (if (sells x computer) (viable x))). More complex sentences require more complex logic representations (CHARNIAK & MCDERMOTT). Given a logic representation of document content, a query in the same logic could be answered by inference using the rules associated with that logic. For example, the query (viable ?) can be answered by forward chaining (a form of *modus ponens*) from the sentences given above. This approach to information retrieval has been studied by WALKER & HOBBS and SIMMONS and is related to the natural language research of people like Schank (SCHANK; SCHANK & ABELSON). Simmons has represented a portion of the *Handbook of Artificial Intelligence* (COHEN & FEIGENBAUM) in logic and retrieves answers to queries in this domain (SIMMONS). The critical problem with this approach is the translation of the text into logic. In current experimental systems, this is done manually.

VAN RIJSBERGEN (1986a) has proposed a framework for information retrieval based on logic. He describes retrieval as a process of determining if a query (expressed in logic) can be inferred from a document's content (expressed in logic). In many cases, this inference cannot be made directly because information is missing in the document; in these cases the inference is uncertain. This framework can be used to describe other models of retrieval and may lead to further insights but currently has not produced new retrieval techniques.

The notion of uncertain inference is also the basis of the RUBRIC system (TONG ET AL.). A part of this system provides standard full-text document retrieval. Queries, however, are represented as rules that describe how pieces of evidence in the document text can be used to infer the relevance of the document. Numbers are attached to the rules to represent the certainty of the statement. For example, a query may be represented by the rules

"information" AND "retrieval" → information-retrieval (0.6)
'information" ADJACENT "retrieval" →
  information-retrieval (0.9)
"probabilit" → probabilistic-model (0.5)
"information-retrieval" AND "probabilistic-model" →
  probabilistic-information-retrieval (0.9)

If a particular document contains the sentence, "Retrieval of information with high probability of relevance is desirable," the rules above will be used to infer that the document is about information retrieval with a certainty of .6, probabilistic models with a certainty of .5, and probabilistic information retrieval with a certainty of .45 (.5 x .9). This type of rule-based representation has also been used for thesaurus information to infer concepts that are related to terms in a query (CROFT & THOMPSON, 1987; SHOVAL).

Graph. A number of structures fall into this general category. The general characteristic of a graph-like representation is a set of nodes and edges (or links) connecting these nodes. Specific examples include semantic nets and frames (CHARNIAK & MCDERMOTT), which are typically produced by natural language processing. Simpler network structures can also be produced by statistical techniques, such as those used in the ASK (anomalous states of knowledge) project (BELKIN ET AL., 1982; BELKIN & KWASNIK). Retrieval techniques in this category must look for similarities in the structures of query and document graphs. This similarity can be used directly to determine if a document should be retrieved or to modify a document ranking.

BELKIN & KWASNIK, for instance, describe the identification of regions of interest in graphs generated by a co-occurrence analysis of narrative "problem statements" (ASK representations). These areas are identified as specific kinds of structures within the graph, such as groups of highly interconnected nodes at high association strengths or two such groups weakly connected with one another. The structural nature of the ASK graph is then used to determine first the terms that will be matched against the database of documents, then the ranking of the retrieved documents. The first retrieval stage takes little account of the relationships of the terms within the document structures (computed in the same way as the ASK structures), using instead the ASK structure to identify terms that either must or may appear in the text structures. In the second stage, the candidate set of retrieved documents is then ranked according to how well each satisfies desirable criteria of term position, importance, and relationship to other terms as established by rules associated with the structures identified in the ASK representation. This retrieval method has not been tested in a formal experiment, but it appears to have some possibilities of providing a way to use graphical representations to choose different retrieval strategies. It may also be of use in specifying term dependencies that can then be used by other retrieval techniques.

The general method of first making a rough pass at the text collection, omitting graph matching of any sort, and then using the graphs to order the retrieved set is typical of graph- and structure-based techniques. The reason for this is that graph matching of any sort is computationally difficult, and searching the entire database is almost always an intractable problem. Another approach is to use very general characteristics of the graphs to define the search space and then progressively to refine the search. This has been used successfully in similarity searching for chemical structures (e.g., WILLETT ET AL.).

Network

Cluster. A cluster is a group of documents whose contents are similar. A particular clustering method gives a more detailed definition of a cluster and provides a technique for generating them. The use of clustering for information retrieval was a major topic in the SMART project (SALTON, 1968). The approach used was to form a cluster hierarchy using an ad-hoc clustering technique. The cluster hierarchy was formed by dividing documents into a few large clusters, dividing these clusters into smaller clusters, and so on. A top-down search of the cluster hierarchy is performed by comparing (using a similarity measure) the query to cluster representatives of the top-level (largest) clusters, choosing the best clusters, comparing the query with representatives of lower-level clusters within these clusters, and so on until a ranked list of lowest-level clusters is produced. The documents in the top-ranked clusters are then ranked individually for presentation to the user. A cluster representative can be generated in various ways, but in general it represents the average properties of documents in the cluster.

Jardine and Van Rijsbergen also used a top-down search of a cluster hierarchy, with the difference that the hierarchy was produced using a formal clustering method (single-link) and clusters were retrieved in their entirety without individual document ranking (JARDINE & VAN RIJSBERGEN). The cluster hypothesis was introduced as a basis for using cluster searches to improve retrieval effectiveness relative to ranking individual documents.

CROFT (1980) described a probabilistic model of cluster searching and introduced the bottom-up retrieval technique. Here the query is compared with representatives of the lowest-level clusters directly, and documents in the top-ranked clusters are retrieved. The emphasis on small, well-defined clusters has led to the development of retrieval techniques based on the generation of the document's nearest neighbors (CROFT & PARENTY; GRIFFITHS ET AL., 1986). A document's nearest neighbors are those most similar to it, and a cluster of nearest neighbors is very similar to the lowest-level single-link clusters (WILLETT, 1984b). Given a network of documents connected to their nearest neighbors, it is possible to generate clusters and their representatives at search time with considerable storage savings (CROFT & PARENTY).

A retrieval technique that has strong similarities to those based on nearest neighbors is Goffman's indirect retrieval method (GOFFMAN). This technique has been used in some recent research (BADRAN). Other recent research on clustering techniques has compared the relative effectiveness and efficiency of different types of clusters (VOORHEES).

Browsing. If the documents, terms, and other bibliographic information are represented in the system as a network of nodes and connections, the user can browse through this network with system assistance. Browsing is an interesting retrieval technique in that it places less emphasis on query formulation than do other techniques and relies heavily on the immediate feedback provided by user browsing decisions. The THOMAS system (ODDY) uses index terms as starting points in a simple network of documents and terms. Through dialog with the user, the system uses the network to build a model of the user's information need that includes relevant documents found during the process. The browsing component of the I^3R system (CROFT & THOMPSON, 1987) contains nodes that represent documents, index terms, domain knowledge, authors, and journals. The links represented include indexing information, thesaurus information, nearest neighbors, citations, and authorship. The system makes browsing recommendations based on the number and types of connections between and among documents but allows users to choose any path.

Other research in browsing concentrates on the use of visual representations of the document database to acquire information from the user interactively (FREI & JAUSLIN).

Spreading activation. Spreading activation is a retrieval technique that has some similarities to browsing. A query is used to "activate" parts of a network that describes the contents of documents and how they are related to each other. In the simplest case, the query would activate index term nodes that are connected to document nodes and other terms. In more knowledge-intensive networks, the links and nodes represent concepts from the subject domain and how they relate to each other as well as the documents that contain those concepts (COHEN & KJELDSEN; RAU). From the "start nodes" provided by the query, other nodes connected to those nodes are in turn "activated" (hence, the term "spreading activation"). Criteria, such as threshold values that decrease as the activation propagates through the network or rules about the reasonableness of the inference implied by using a particular link, are used to control the spread of activation. Activation can converge on particular document nodes from a number of links. These highly activated nodes are retrieved.

In a simple network of documents and terms, the documents that have a high level of activation after the first links from the query nodes are followed will be those documents that have a high number of terms in common with the query. If the activation spreads to other terms connected to those documents and then to other documents, the documents retrieved in this second phase will be similar to those found by a cluster search based on nearest neighbors. When the activation is refined using inference rules and more link types, it is more difficult to relate the retrieved documents to those found with conventional techniques. The retrieval technique in this case is more similar to structure-based techniques.

FEEDBACK METHODS

Relevance feedback techniques are not considered retrieval techniques by our criteria. Rather they are used to refine the request model, which is then used for another search. Feedback techniques are, however, an extremely important part of ensuring that a document retrieval system will be effective.

These techniques were primarily developed in the context of feature-based retrieval (SALTON, 1968) although the principles apply to any retrieval technique. The main part of relevance feedback is the adjustment of weights associated with query terms. This adjustment is done on the basis of term occurrences in the documents identified by the user as relevant. The probabilistic model has a particularly strong motivation behind this weight adjustment in that the identified relevant documents provide a sample to estimate the *pr* values. The query (or request model) can also be changed by the addition of new terms from relevant documents. Some control is needed over the number of new terms added, and it seems that the most reliable method is to have users identify interesting terms in relevant documents.

Other types of modifications based on feedback are possible, such as adding term dependencies identified in relevant documents or modifying the "document space" (document indexing) to make relevant documents more similar to the queries. Attempts have also been made to use adaptive mechanisms to select retrieval techniques appropriate for a particular query (CROFT & THOMPSON, 1984). There have also been some attempts to use this type of information for feedback in operational systems, presenting it to the user, who then adjusts the query manually (e.g., INGWERSEN).

RELATIVE PERFORMANCE OF RETRIEVAL TECHNIQUES

Comparative Performance Studies

The evaluation of IR systems has been a major topic of research for a number of years (VAN RIJSBERGEN, 1979). Although there have been and continue to be problems with the evaluations that have been done, these results, together with theoretical results derived from underlying models, provide valuable information about the relative performance of retrieval techniques. Again, we do not argue the relative merits of automated vs. manual indexing in this paper. In a recent article, SALTON (1986a) summarizes these arguments and other results. In making the comparisons in this section, we shall not discuss the particular effectiveness measures used. When a difference is described as significant, we are following the generally accepted "rule of thumb" guidelines of at least 10 percent increases in recall and precision.

The first important result is that all available evidence points to the superiority of partial match techniques over exact match techniques (in particular, see SALTON ET AL., 1983b). Although there are problems with making direct comparisons between the sets of retrieved documents, it appears that the difference in effectiveness is significant. For feature-based retrieval, the evidence indicates that the best performance is provided by the probabilistic retrieval strategy incorporating term significance weights or its equivalent, the *tf.idf*/cosine correlation combination. This retrieval technique uses a simple similarity measure (the inner product) and the effectiveness is due entirely to the index term weights used. Results that indicate superiority of one technique over another in this context (e.g., SALTON, 1986b) can be interpreted in terms of the estimates used for the weights. The use of good estimates is the major factor in obtaining the best performance from these techniques.

The use of term dependencies to modify document rankings can also improve performance but only if the dependencies are accurately identified by the user or natural language processing techniques (CROFT, 1986). The same problems occur with the extended Boolean retrieval technique (SALTON & VOORHEES). Techniques that rely on identifying dependencies in document collections independently of a particular query do not seem to give significant improvements (VAN RIJSBERGEN ET AL.; YU ET AL., 1983). The automated use of thesaurus information to expand queries appears to be effective but only if the terms expanded and the type of thesaurus information used are tightly controlled. Relevance feedback can give very good results even when few relevant documents are identified (SPARCK JONES & WEBSTER).

Cluster-based searches can achieve levels of performance that are similar to individual feature-based searches (CROFT, 1980; GRIFFITHS ET AL., 1986) but in general they tend to be better for high-precision results. The primary advantage of cluster searching, however, is that it retrieves different relevant documents than, say, a *tf.idf* search, and for some queries it works much better (CROFT & HARPER; GRIFFITHS ET AL., 1986). Cluster-based retrieval, therefore, is a good alternative technique to the individual feature-based method. Systems have been designed to allow both retrieval techniques to be used simultaneously or for cluster searches to be used when other techniques fail (CROFT & THOMPSON, 1987).

Although the techniques described so far can achieve reasonable levels of performance and can be implemented efficiently in operational systems, there is still a lot of room for improvement in terms of absolute performance. To obtain much higher levels of performance, it is apparently necessary to consider knowledge-intensive techniques such as structure-based retrieval or spreading activation. The problem is that because these techniques are knowledge-intensive, they are difficult to implement and have been tested only with very small collections of documents in very specific domains. Techniques that use some form of natural language processing to construct representations of document and query content have been studied in IR for some time (e.g., SPARCK JONES, 1974). Experiments with these techniques have never achieved significant performance benefits, often because the information derived from natural language processing was used in inappropriate ways. Hybrid systems (CROFT & LEWIS; SPARCK JONES & TAIT; TONG ET AL.) that combine knowledge-intensive techniques with efficient full-text retrieval or ranking strategies appear to have significant promise. The RUBRIC approach seems particularly suited to users who are prepared to spend a lot of effort in constructing queries and would not be appropriate in a general environment. Cohen's recent paper (COHEN & KJELDSEN) provides a detailed evaluation of a spreading activation technique that achieved good results. The problem of translating text into the representations used, however, remains unsolved.

Relationship of the Request Model to Retrieval Techniques

It has probably become clear that it is difficult to separate the retrieval technique from the form of representation used in the request model. There is, however, a clear message from the evaluations performed in IR research. Whatever retrieval technique is used, the quality of the results depends almost entirely on the accuracy of the information in the request model. More sophisticated retrieval techniques can use more detailed request models but constructing these models requires more user effort. It is this process of an intermediary's interaction with the user to formulate the query that is the heart of current retrieval systems, and it is also crucial for more advanced retrieval systems. This obvious fact has led to a lot of research in expert intermediary systems (e.g., BELKIN ET AL., 1983; CROFT & THOMPSON, 1987; MARCUS). These systems engage the user in a dialog with a variety of facilities to acquire a detailed request model. The systems also assist the user in evaluating the retrieved documents and, in some cases, in selecting retrieval techniques.

Predicting Appropriate Retrieval Techniques

A number of experimental results have indicated that although different retrieval techniques appear to have similar results, they often retrieve different relevant documents for the same query. The use of alternative content representations, such as citation information, can also result in the retrieval of different documents. Different techniques also vary in their performance for different queries. If the best results from different techniques for individual queries could be selected, very high performance would be possible (CROFT & THOMPSON, 1984; GRIFFITHS ET AL., 1986). The problem is then to identify which technique (and representation) is appropriate for a particular query. Unfortunately, growing evidence shows that this is extremely difficult if not impossible (e.g., MCCALL & WILLETT). One solution is to design systems to use alternative strategies (such as probabilistic, cluster, and browsing) and alternative representations (such as index terms, citations, and semantic nets) (BELKIN ET AL., 1982; CROFT & THOMPSON, 1987). The selection of a particular strategy and representation can be guided by rules in consultation with the user. Given such a system, retrieval can be viewed as a form of plausible inference (VAN RIJSBERGEN, 1986a) or gathering of evidence about the relevance of documents from various sources.

ARCHITECTURES AND TECHNIQUES FOR INTEGRATED SYSTEMS

It seems, therefore, that there are strong arguments either for using specific retrieval techniques for specific kinds of queries—on the assumption that some techniques are more appropriate for some queries than for others—or for using many retrieval techniques on a single query in the hope that the combination will result in a satisfactory response when no single technique is adequate. The latter position does not really require any specific kind of system architecture or design to be implemented other than having representations that allow the various techniques to be used. On the other hand, the task of choosing a specific technique for specific circumstances presents a more significant problem.

The SIRE system (KOLL ET AL.) is an example of a system that provides several retrieval techniques that can be used singly or in combination for any specific query. In this system the user chooses the technique. Systems that automatically choose the technique or that have special techniques or data structures associated with specific kinds of queries or problems present are more problematic. This approach has arisen in two basic contexts. One is the context associated with so-called "expert" information systems in which it is assumed that different techniques will be required for different kinds of queries; the other is in office automation and/or integration of DBM and IR systems. Here different data types or structures seem to imply the need for different techniques, and there are many different user populations with presumably different data needs. We discuss some approaches to these problems below.

"Expert" Information Intermediaries and Systems

Several "expert" or "intelligent" information systems have been proposed that have assumed that multiple retrieval strategies will be necessary to respond to different query types. We have already discussed one such proposal, the ASK approach (BELKIN ET AL., 1982). It is based on the hypothesis that there will be different categories of anomalous states of knowledge that will require different retrieval strategies. The architecture of this proposal was never fully specified, although it was clear that an elaborate request model would be necessary to do the correct request classification. BELKIN & KWASNIK suggest some strategies and note how they can be chosen according to structural representations of the user's ASK. For details of the kinds of techniques, and the types of characteristics of ASKs, see the section on *Graphs* above. Here, it is sufficient to say that the system design requires a representation of the texts in structural terms, as well as of the queries, and that the alternative retrieval techniques are chosen in a rule-based stepwise manner, operating directly on the text representations.

There have been several suggestions for intelligent information systems based on a "distributed expert" model, in which a number of functions (e.g., building up a request model, building a user model, and choosing a retrieval technique) are isolated as separate processes that communicate with one another. BELKIN ET AL. (1983) suggested the logical design for such a system, and CROFT & THOMPSON (1987) and FOX (1987a) have built prototypes of similar systems. Although the details of these systems differ, they share some significant characteristics in terms of what functions are specified and how they are related. From our point of view, they all assume that different retrieval strategies will be necessary for different user situations, and therefore a major function of their retrieval strategy or technique expert is to choose one of several techniques available, based on information provided by the other experts about the user. Croft and Thompson have available in their system both probabilistic and cluster searching as well as a separate browsing component that the user can choose to instigate, and Fox has implemented separate *p*-norm search (a form of extended Boolean retrieval—see FOX, 1983) and browsing experts. As yet there is no strong reasoning capability in any of these systems for choosing one or another retrieval technique, but the general architecture of separation of functions seems to be useful in this regard. This work in general has progressed on the assumption that the more elaborate the request model, the more likely that: 1) any retrieval technique will work well, and 2) that a most appropriate technique can be selected. These assumptions have yet to be rigorously tested.

Other intelligent interface or system designs have paid less attention to incorporating multiple retrieval techniques, relying instead on elaboration of the request model through interaction (e.g., MARCUS) or through natural language access to the system (e.g., GUIDA & TASSO) in order to use a single technique already in place. KRAWCZAK ET AL. use a hierarchical representation of the knowledge associated with a particular domain (in this case environmental pollution), in order to guide this interaction.

Multimedia and "Integrated" Information Systems

For some time there has been active research in attempting to integrate DBM and IR systems into a single system or model (e.g., SCHEK). The impetus for such work has been twofold. First, the two models offer different capabilities for, and have different problems in, data management and query formulation; these might complement one another in various ways if they were incorporated into one system. Second, it has been suggested that the two systems might provide different retrieval techniques and interfaces appropriate to different classes of users or to the same users with different problems. CROFT (1982) provides an overview of these positions and suggests that knowledge-based or "expert" systems be included in such an amalgamation.

Proposals for integration have generally taken the form of putting an IR interface on top of a DBM system (DBMS), thus effectively translating the IR-based query into a DBMS "exact match" retrieval technique. This technique may be useful for managing very large databases efficiently, but as a multiple architecture it leaves something to be desired. A different approach is exemplified by the HAM-ANS (HAMburg- Application-Oriented Natural language System) project (HOEPPNER ET AL.), in which a single interface is used to access three different kinds of databases: factual, document, and "knowledge." Although the project did not actually manage to integrate this access entirely, the system was meant to be able to choose an appropriate database (and therefore its appropriate retrieval technique) according to characteristics of the user's query.

Research in office automation has also led to proposals for new architectures for information systems (VAN RIJSBERGEN, 1986c). Here the problem is twofold. First, the documents in the office environment are of many types, and even within a single type, they are often multimedia (i.e., made up of mixtures of text, images, data, tables, voice, and so on). Standard DBMS retrieval techniques are inadequate in this environment. Second, the range of users and uses of such systems is very broad, meaning that it might be necessary to identify particular uses with specific retrieval techniques or at least types of documents. The typical response in the first case has been to define strictly the general office document or object and its parts. One can then presumably identify from the query what kind of document, or what part of the document, is desired and search accordingly (e.g., CHRISTODOULAKIS; HARPER ET AL.). This approach has led to proposals for architectures that include DBMS and conventional IR techniques and sometimes others, such as browsing (for a review of such systems, see SMEATON & VAN RIJSBERGEN, 1986). In the case of the second issue, very little has been proposed since up to now there has been little systematic investigation of the users on which to base an architecture.

Overall, little seems to have been done to attack the problem of how, in principle, to choose the techniques that should be incorporated in such systems, although the work on document types is promising. Even less seems to have been done on how to design a system that will choose an appropriate technique automatically- i.e., without direction from the user.

CURRENT STATUS AND FUTURE RESEARCH IN IR TECHNIQUES

The Situation Now

One can conclude from this review that there is a disquieting disparity between the results of research on IR techniques, which demonstrate fairly conclusively, on both theoretical and empirical grounds, the inadequacy of the exact match paradigm for effective information retrieval and the status of operational IR systems, which use almost exclusively just one exact match technique--Boolean logic. This, of course, is not exactly news. What is perhaps new is the directions in which IR technique research is going. In the past, most research in retrieval techniques was rather far removed from operational environments and to some extent even from operational constraints. The past several years seem to show movement in this research in three directions; two respond to some extent to what may have been problems in acceptance of research results; the third seems to offer promise in greatly improving retrieval performance.

First there has been a good deal of work on relating partial match techniques to exact matching, as in extended Boolean searching and the use of Boolean-derived dependencies in probabilistic searching. This can be seen as a response to the demands of the operational environment. Similarly, there are at least a few studies being carried out of partial-match techniques in operational environments (e.g., ROBERTSON ET AL.), which are explicitly designed to attend to the criticism that such techniques have never been demonstrated to be worthwhile in large systems. There are also several microcomputer-based systems that use partial match techniques in operational, albeit smaller, environments—e.g., the SIRE system (KOLL ET AL.).

The second response seems to be the realization that no one technique will be adequate for all purposes and that either a mix of techniques or a principled choice of techniques is required to improve IR system performance. Thus, there is perhaps less in-fighting among the various camps and more willingness to accept the usefulness of particular techniques in specific circumstances. This tendency may have been encouraged by the special problems of the office automation environment and by the many results indicating that although different techniques perform similarly, they retrieve different relevant documents.

The third new direction we note is increasingly complex representations of the request or user's problem. As noted, although retrieval techniques are not the same as representations, the techniques one can use are determined by the representation. The more complex the representation, one might think, the more kinds of retrieval techniques are possible. Much, although not all, of the work of this type has had its roots in the knowledge representation schemes associated with artificial intelligence research. In general, this work seems to have arisen because of an increased understanding of the importance of the request model (or understanding of the user) to all of IR.

Open Problems

Some issues that have been raised by current research in IR techniques and some that have been around for a while have become particularly important. These open problems will need to be resolved before there will be great improvements in IR system performance, but at least some IR technique research is now attacking them.

First, and extremely important, is our understanding of the gaps between perfect, optimal, and current performance of IR systems. We know more or less what current performance is (although we may not be happy with the evaluation measures). We do not know, however, what perfect performance is now that it is generally accepted that the ideal of all and only the relevant documents is not the goal for all users at all times. Nor do we have any understanding of what optimal performance might be other than a widespread feeling that none of the current techniques seems able to achieve it alone and that current performance therefore is suboptimal. This, of course, is an assumption, which we also hope is valid. Thus, the issues that need to be seriously addressed are: 1) what is perfect performance and why is it perfect; 2) what is optimal performance and why is it optimal; and 3) how can optimal performance be achieved.

There is a growing feeling that it is inappropriate to treat all queries with just one technique and that multiple techniques could be used on single queries. This might be a way to approach the problem of optimal performance, but first some questions must be answered. For example, different retrieval techniques seem to operate differently on queries. We need to discover what aspects of the request model the various techniques capture or reflect. Then we can begin to learn how to choose techniques appropriate to different user situations and decide whether and how to apply several techniques to the same request.

Also we need some ideas about how to develop new retrieval techniques. If the previous questions have been answered, at least to some extent, it may be possible to suggest techniques that directly approach the issue of optimal performance or that respond to aspects of the user's problem that other techniques do not but that theory indicates are important. This, of course, will depend on the development of representations that will allow such techniques to be useful.

There is now some reason for optimism because, as this review has shown, not only is there dissatisfaction with current retrieval techniques, but significant work is at least beginning to attack all of the questions we have raised. Thus, we expect to see much more work in the near future on the relationship between technique and request model, on the testing of experimental techniques in operational environments, on the specification and implementation of new architectures for multiple retrieval techniques, and especially on methods and effects of request model elaboration and their use in the development of new retrieval techniques.

BIBLIOGRAPHY

APPELRATH, HANS-JURGEN. 1985. Die Erweiterung von DB- und IR-Systemen zu Wissensbasierten Systemen [The Extension of Data Base and Information Retrieval Systems to Knowledge Based Systems]. Nachrichten für Dokumentation (West Germany). 1985; 36(1): 13-21. ISSN: 0027-7436.

ARNOW, DAVID M.; TENENBAUM, AARON M.; WU, CONNIE. 1985. P-Trees: Storage Efficient Multiway Trees. See reference: ASSOCIATION FOR COMPUTING MACHINERY. 111-121.

ASSOCIATION FOR COMPUTING MACHINERY (ACM). SPECIAL INTEREST GROUP ON INFORMATION RETRIEVAL (SIGIR). 1985. Research and Development in Information Retrieval: [Proceedings of the] 8th Annual International ACM SIGIR Conference; 1985 June 5-7; Montreal, Canada. New York, NY: ACM, Inc.; 1985. 288p. Available from: ACM Order Department, P.O. Box 64145, Baltimore, MD 21264 (Order no. 606850). ISBN: 0-89791-159-8.

BADRAN, ODETTE MAROUN. 1985. An Alternative Search Strategy to Improve Information Retrieval. In: Parkhurst, Carol A., ed. ASIS '85: Proceedings of the American Society for Information Science (ASIS) 48th Annual Meeting; 1985 October 20-24; Las Vegas, NV. White Plains, NY: Knowledge Industry Publications, Inc. for ASIS; 1985. 137-140. ISSN: 0044-7870; ISBN: 0-86729-176-1; CODEN: PAISDQ.

BARBI, E.; CALVO, F.; PERALE, C.; SIROVICH, F.; TURINI, F. 1984. A Conceptual Approach to Document Retrieval. In: Ellis, Clarence, ed. [Proceedings of the] 2nd Association for Computing Machinery, Special Interest Group on Office Automation (ACM-SIGOA) Conference on Office Information Systems; 1984 June 25-27; Toronto, Canada. New York, NY: ACM, Inc.; 1984. 219-226. (Published as Vol. 5(1-2) of SIGOA Newsletter). Available from: ACM Order Department; P.O. Box 64145, Baltimore, MD 21264 (Order no. 611840). ISBN: 0-89791-140-7.

BARR, AVRON; FEIGENBAUM, EDWARD A., eds. 1981. The Handbook of Artificial Intelligence: Volume 1. Los Altos, CA: William Kaufmann, Inc.; 1981. 409p. ISBN: 0-86576-004-7.

BATES, MARCIA J. 1986. Subject Access in Online Catalogs: A Design Model. Journal of the American Society for Information Science. 1986 November; 37(6): 357-376. ISSN: 0002-8231; CODEN: AISJB6.

BELKIN, N. J.; KWASNIK, B. H. 1986. Using Structural Representations of Anomalous States of Knowledge for Choosing Document Retrieval Strategies. See reference: RABITTI, FAUSTO, ed. 11-22.

BELKIN, N. J.; ODDY, R. N.; BROOKS, H. M. 1982. ASK for Information Retrieval: Part I. Background and Theory; Part II: Results of a Design Study. Journal of Documentation. 1982 June; September; 38(2-3): 61-71; 145-164. ISSN: 0022-0418.

BELKIN, N. J.; SEEGER, T.; WERSIG, G. 1983. Distributed Expert Problem Treatment as a Model for Information System Analysis and Design. Journal of Information Science. 1983; 5: 153-167. ISSN: 0165-5515.

BELL, D. A. 1985. An Architecture for Integrating Data, Knowledge, and Information Bases. In: Informatics 8: Advances in Intelligent Retrieval: Proceedings of a Conference Jointly Sponsored by Aslib, the Aslib Informatics Group, and the Information Retrieval Specialist Group of the British Computer Society; 1985 April 16-17; Wadham College, Oxford, England. London, England: Aslib, The Association for Information Management, Information House; 1985. 240–257. Available from: Aslib, 26–27 Boswell Street, London WC1N3JZ, England. ISBN: 0-85142-195-4.

BILLER, HORST. 1983. On the Architecture of a System Integrating Data Base Management and Information Retrieval. In: Salton, Gerard; Schneider, Hans-Jochen, eds. Research and Development in Information Retrieval: Proceedings [of the 5th International Conference of the Association for Computing Machinery, Special Interest Group on Information Retrieval (ACM-SIGIR)]; 1982 May 18–20; Berlin. Berlin, West Germany: Springer-Verlag; 1983. 80–97. (Lecture Notes in Computer Science: 146). ISBN: 3-540-11978-7 (Berlin); ISBN: 0-387-11978-7 (New York).

BISWAS, GAUTAM; SUBRAMANIAN, VISWANATH; BEZDEK, JAMES C. 1985. A Knowledge Based System Approach to Document Retrieval. In: The Engineering of Knowledge-Based Systems: [Proceedings of the] 2nd Conference on Artificial Intelligence Applications; 1985 December 11–13; Miami Beach, FL. Washington, DC: Institute of Electrical and Electronics Engineers (IEEE) Computer Society Press; 1985. 455–460. Available from: IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854 (Catalog no. 85CH2215-2). ISBN: 0-8186-0688-6; LC: 85-62321.

BISWAS, GAUTAM; SUBRAMANIAN, VISWANATH; MARQUES, M. M.; BEZDEK, JAMES C. 1985. A Document Retrieval System Using a Fuzzy Expert Systems Approach. In: Proceedings of the International Conference on Cybernetics and Society; 1985 November 12–15; Tucson, AZ. New York, NY: Institute of Electrical and Electronics Engineers, Inc.; 1985. 126–130. (IEEE Catalog no. 85CH2253-3). ISSN: 0360-8913.

BIVINS, KATHLEEN T.; ERIKSSON, LENNART. 1982. Reflink: A Microcomputer Information Retrieval and Evaluation System. Information Processing & Management. 1982; 18(3): 111–116. ISSN: 0306-4573; CODEN: IPMADK.

BLAIR, DAVID C. 1984. The Data-Document Distinction in Information Retrieval. Communications of the ACM. 1984 April; 27(4): 369–374. ISSN: 0001-0782.

BLAIR, DAVID C. 1986a. Full Text Retrieval: Evaluation and Implications. International Classification (West Germany). 1986; 13(1): 18–23. ISSN: 0340-0050.

BLAIR, DAVID C. 1986b. Indeterminacy in the Subject Access to Documents. Information Processing & Management. 1986; 22(2): 229–241. ISSN: 0306-4573; CODEN: IPMADK.

BLAIR, DAVID C.; MARON M. E. 1985. An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System. Communications of the ACM. 1985 March; 28(3): 289–299. ISSN: 0001-0782.

BOOKSTEIN, ABRAHAM. 1980. Fuzzy Requests: An Approach to Weighted Boolean Searches. Journal of the American Society for Information Science. 1980 July; 31(4): 240–247. ISSN: 0002-8231; CODEN: AISJB6.

BOOKSTEIN, ABRAHAM. 1982. Recent Developments in the Theory of Information Retrieval. 1982 December. 28p. (NTIS Report TRITA-LIB-6019; Based on a seminar presented at the Royal Institute of Technology Library [Sweden], 1982 November 10). Also published as ISSN: 0346-9042.

BOOKSTEIN, ABRAHAM. 1983. Explanation and Generalization of Vector Models in Information Retrieval. In: Salton, Gerard; Schneider, Hans-Jochen, eds. Research and Development in Information Retrieval: Proceedings [of the 5th International Conference of the Association for Computing Machinery, Special Interest Group on Information Retrieval (ACM-SIGIR)]; 1982 May 18-20; Berlin. Berlin, West Germany: Springer-Verlag; 1983. 118-132. (Lecture Notes in Computer Science: 146). ISBN: 3-540-11978-7 (Berlin); ISBN: 0-387-11978-7 (New York).

BOOKSTEIN, ABRAHAM. 1985. Probability and Fuzzy-Set Applications to Information Retrieval. In: Williams, Martha E., ed. Annual Review of Information Science and Technology: Volume 20. White Plains, NY: Knowledge Industry Publications, Inc. for the American Society for Information Science; 1985. 117-151. ISSN: 0066-4200; ISBN: 0-86729-175-3; LC: 66-25096; CODEN: ARISBC.

BOSE, PRASANTA K.; RAJINIKANTH, M. 1985. KARMA: Knowledge-Based Assistant to a Database System. In: The Engineering of Knowledge-Based Systems: [Proceedings of the] 2nd Conference on Artificial Intelligence Applications; 1985 December 11-13; Miami Beach, FL. Washington, DC: Institute of Electrical and Electronics Engineers (IEEE) Computer Society Press; 1985. 467-472. Available from: IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854 (Catalog no. 85CH2215-2). ISBN: 0-8186-0688-6; LC: 85-62321.

BROOKS, H. M. 1983. Information Retrieval and Expert Systems- Approaches and Methods of Development. In: Informatics 7: Intelligent Information Retrieval. London, England: Aslib; 1983. 65-75. ISBN: 0-85142-187-5.

BUCKLEY, CHRIS. 1985. Implementation of the SMART Information Retrieval System. Ithaca, NY: Cornell University. Department of Computer Science; 1985 May. 37p. (Technical Report TR 85-686). Available from: Chris Buckley, Department of Computer Science, Cornell University, Ithaca, NY 14853.

BUCKLEY, CHRIS; LEWIT, ALAN F. 1985. Optimization of Inverted Vector Searches. See reference: ASSOCIATION FOR COMPUTING MACHINERY. 97-110.

BUELL, DUNCAN A. 1985. A Problem in Information Retrieval with Fuzzy Sets. Journal of the American Society for Information Science. 1985; 36(6): 398-401. ISSN: 0002-8231; CODEN: AISJB6.

CARROLL, DAVID M.; POGUE, CHRISTINE A.; WILLETT, PETER. 1987. Bibliographic Pattern Matching Using the ICL Distributed Array Processor. Journal of the American Society for Information Science. 1987. (In press). ISSN: 0002-8231; CODEN: AISJB6.

CHARNIAK, EUGENE; MCDERMOTT, DREW. 1985. Introduction to Artificial Intelligence. Reading, MA: Addison-Wesley; 1985. 701p. ISBN: 0-201-11945-5.

CHRISTODOULAKIS, S. 1984. Framework for the Development of an Experimental Mixed-Mode Message System. See reference: VAN RIJSBERGEN, C. J., ed. 1984. 1-20.

CLEVELAND, DONALD B.; CLEVELAND, ANA D.; WISE, OLGA B. 1984. Less than Full-Text Indexing Using a Non-Boolean Searching Model. Journal of the American Society for Information Science. 1984; 35(1): 19-28. ISSN: 0002-8231; CODEN: AISJB6.

COHEN, PAUL R.; FEIGENBAUM, EDWARD A. eds. 1982. The Handbook of Artificial Intelligence: Volume 3. Los Altos, CA: William Kaufmann, Inc.; 1982. 639p. ISBN: 0-86576-007-1.

COHEN, PAUL R.; KJELDSEN, RICK. 1987. Information Retrieval by Constrained Spreading Activation in Semantic Networks. Information Processing & Management. 1987. (In press). ISSN: 0306-4573; CODEN: IPMADK.

COOPER, WILLIAM S. 1983. Exploiting the Maximum Entropy Principle to Increase Retrieval Effectiveness. Journal of the American Society for Information Science. 1983; 34(1): 31-39. ISSN: 0002-8231; CODEN: AISJB6.

COOPER, WILLIAM S.; HUIZINGA, P. 1982. The Maximum Entropy Principle and Its Application to the Design of Probabilistic Retrieval Systems. Information Technology: Research and Development. 1982; 1: 99-112. ISSN: 0144-817X.

CROFT, W. BRUCE. 1980. A Model of Cluster Searching Based on Classification. Information Systems. 1980; 5(3): 189-195. ISSN: 0306-4379; CODEN: INSYD6.

CROFT, W. BRUCE. 1982. An Overview of Information Systems. Information Technology: Research and Development. 1982 January; 1(1): 73-96. ISSN: 0144-817X.

CROFT, W. BRUCE. 1983a. Applications for Information Retrieval Techniques in the Office. In: Kuehn, Jennifer J., ed. Proceedings of the 6th Annual International Association for Computing Machinery Special Interest Group on Information Retrieval (ACM-SIGIR) Conference on Research and Development in Information Retrieval; 1983 June 6-8; Bethesda, MD. New York, NY: ACM, Inc.; 1983. 18-23. (A publication of SIGIR Forum; 1983 Summer; 17(4)). Available from: ACM Order Department, P.O. Box 64145, Baltimore, MD 21264. ISBN: 0-89791-107-5.

CROFT, W. BRUCE. 1983b. Experiments with Representation in a Document Retrieval System. Information Technology: Research and Development. 1983 January; 2(1): 1-21. ISSN: 0144-817X.

CROFT, W. BRUCE. 1984. A Comparison of Cosine Correlation. Information Technology: Research and Development. 1984; 3: 113-114. ISSN: 0144-817X.

CROFT, W. BRUCE. 1986. Boolean Queries and Term Dependencies in Probabilistic Retrieval Models. Journal of the American Society for Information Science. 1986 March; 37(2): 71-77. ISSN: 0002-8231; CODEN: AISJB6.

CROFT, W. BRUCE; HARPER, D. J. 1979. Using Probabilistic Models of Document Retrieval Without Relevance Information. Journal of Documentation. 1979 December; 35(4): 285-295. ISSN: 0022-0418.

CROFT, W. BRUCE; LEWIS, D. 1987. An Approach to Natural Language Processing for Document Retrieval. In: [Proceedings of the] Association for Computing Machinery Special Interest Group on Information Retrieval (ACM-SIGIR) 10th International Conference on Research and Development in Information Retrieval; 1987 June 3-5; New Orleans, LA. 26-32. ISBN: 0-89791-232-2. New York, NY: ACM, Inc., 1987.

CROFT, W. BRUCE; PARENTY, THOMAS J. 1985. A Comparison of a Network Structure and a Database System Used for Document Retrieval. Information Systems. 1985; 10(4): 377-390. ISSN: 0306-4379; CODEN: INSYD6.

CROFT, W. BRUCE; THOMPSON, ROGER H. 1984. The Use of Adaptive Mechanisms for Selection of Search Strategies in Document Retrieval Systems. See reference: VAN RIJSBERGEN, C. J., ed. 1984. 95–110.

CROFT, W. BRUCE; THOMPSON, ROGER H. 1985. An Expert Assistant for Document Retrieval. Amherst, MA: University of Massachusetts, Department of Computer and Information Science; 1985. 52p. (COINS Technical Report 85-05). Available from: W. Bruce Croft, Dept. of Computer and Information Science, University of Massachusetts, Amherst, MA 01003.

CROFT, W. BRUCE; THOMPSON, ROGER H. 1987. I^3R: A New Approach to the Design of Document Retrieval Systems. Journal of the American Society for Information Science. 1987. (In press). ISSN: 0002-8231; CODEN: AISJB6.

DEERWESTER, SCOTT; DUMAIS, SUSAN T.; FURNAS, GEORGE, W.; LANDAUER, THOMAS K.; HARSHMAN, RICHARD. 1987. Indexing by Latent Structure Analysis. 1987. To appear. 33p. Available from: Susan T. Dumais, Bell Communications Research, 435 South St., MRE 2L-371, Morristown, NJ 07960.

EL-HAMDOUCHI, A.; WILLETT, PETER. 1986. Hierarchic Document Clustering Using Ward's Method. See reference: RABITTI, FAUSTO, ed. 149–156.

ELKERTON, JAY; WILLIGES, ROBERT C. 1984. Information Retrieval Strategies in a File-Search Environment. Human Factors. 1984; 26(2): 171–184. ISSN: 0018-7208.

ENSER, P. G. B. 1985a. Automatic Classification of Book Material Represented by Back-of-the-Book Index. Journal of Documentation. 1985 September; 41(3): 135–155. ISSN: 0022-0418.

ENSER, P. G. B. 1985b. Experimenting with the Automatic Classification of Books. In: Informatics 8: Advances in Intelligent Retrieval: Proceedings of a Conference Jointly Sponsored by Aslib, the Aslib Informatics Group, and the Information Retrieval Specialist Group of the British Computer Society; 1985 April 16–17; Wadham College, Oxford, England. London, England: Aslib, The Association for Information Management, Information House; 1985. 68–83. Available from: Aslib, 26–27 Boswell Street, London WC1N3JZ, England. ISBN: 0-85142-195-4.

FINDLER, NICHOLAS V. 1980. Vers l'Optimisation Interactive de l'Organisation de la Base des Donnees [Toward Interactive Optimization of Database Organization]. In: Diday, E.; Lebart, L.; Pages, J. P.; Tomassone, R., eds. Data Analysis and Informatics: Proceedings of the 2nd International Symposium on Data Analysis and Informatics; 1979 October 17–19; Versailles, France. Amsterdam, The Netherlands: North-Holland Publishing Co.; 1980. 411–421. (In English). ISBN: 0-444-86005-3.

FOX, EDWARD A. 1983. Extending the Boolean and Vector Space Models of Information Retrieval with P-Norm Queries and Multiple Concept Types. Ithaca, NY: Cornell University; 1983. (Ph.D. dissertation). 386p. Available from: University Microfilms Int., Ann Arbor, MI.

FOX, EDWARD A. 1987a. Development of the CODER System: A Testbed for Artificial Intelligence Methods in Information Retrieval. Information Processing and Management. 1987; 23(4). (In press). ISSN: 0306-4573; CODEN: IPMADK.

FOX, EDWARD A. 1987b. Improved Retrieval Using a Relational Thesaurus Expansion of Boolean Logic Queries. In: Evens, Martha, ed. Proceedings of a Workshop on Relational Models of the Lexicon; 1984 July; Stanford, CA. (To appear). Available from: Edward A. Fox, Department of Computer Science, Virginia Tech, Blacksburg, VA 24061.

FREI, H. P.; JAUSLIN, J.-F. 1983. Graphical Presentation of Information and Services: A User-Oriented Interface. Information Technology: Research and Development. 1983 January; 2(1): 23–42. ISSN: 0144-817X.

FUHR, NORBERT. 1986. Two Models of Retrieval with Probabilistic Indexing. See reference: RABITTI, FAUSTO, ed. 249–257.

FUHR, NORBERT; KNORZ, G. E. 1984. Retrieval Text Evaluation of a Rule Based Automatic Indexing (AIR/PHYS). See reference: VAN RIJSBERGEN, C. J., ed. 1984. 391–408.

GOFFMAN, W. 1968. An Indirect Method of Information Retrieval. Information Storage and Retrieval. 1968; 4: 361–373. ISSN: 0020-0271.

GRIFFITHS, ALAN; LUCKHURST, H. CLAIRE; WILLETT, PETER. 1986. Using Interdocument Similarity Information in Document Retrieval Systems. Journal of the American Society for Information Science. 1986 January; 37(1): 3–11. ISSN: 0002-8231; CODEN: AISJB6.

GRIFFITHS, ALAN; ROBINSON, LESLEY A.; WILLETT, PETER. 1984. Hierarchic Agglomerative Clustering Methods for Automatic Document Classification. Journal of Documentation. 1984 September; 40(3): 175–205. ISSN: 0022-0418.

GUIDA, GIOVANNI; TASSO, CARLO. 1983. An Expert Intermediary System for Interactive Document Retrieval. Automatica (Great Britain). 1983; 19(6): 759–766. ISSN: 0005-1098.

HAHN, UDO. 1985. Expertensysteme als intelligente Informationsysteme: Konzepte für die funktionale Erweiterung des Information Retrieval [Expert Systems as Intelligent Information Systems: Perspectives for Functional Extension of Information Retrieval]. Nachrichten für Dokumentation (West Germany). 1985; 36(1): 2–12. ISSN: 0027-7436.

HARPER, D. J.; DUNNION, J.; SHERWOOD-SMITH, M.; VAN RIJSBERGEN, C. J. 1986. Minstrel-ODM: A Basic Office Data Model. Information Processing & Management. 1986; 22(2): 83–107. (Special Issue). ISSN: 0306-4573; CODEN: IPMADK.

HARPER, D. J.; VAN RIJSBERGEN, C. J. 1978. An Evaluation of Feedback in Document Retrieval Using Co-Occurrence Data. Journal of Documentation. 1978; 34(3): 189–216. ISSN: 0022-0418.

HOBBS, JERRY R.; WALKER, DONALD E.; AMSLER, ROBERT A. 1982. Natural Language Access to Structured Text. In: Horecky, Jan, ed. COLING 82: Proceedings of the 9th International Conference on Computational Linguistics; 1982 July 5–10; Prague, Czechoslovakia. Amsterdam, The Netherlands: North Holland Publishing Co.; 1982. 127–132. ISBN: 0-444-86393-1 (U.S.); LC: 82-7960.

HOEPPNER, WOLFGANG; MORIK, KATHARINA; MARBURGER, HEINZ. 1986. Talking It Over: The Natural Language Dialog System HAM-ANS. In: Bolc, Leonard; Jarke, Matthias, eds. Cooperative Interfaces to Information Systems. Berlin, West Germany: Springer-Verlag; 1986. 189–258. ISBN: 0-387-16599-1 (U.S.); LC: 86-13965.

HOLLAAR, LEE A. 1979. Unconventional Computer Architectures for Information Retrieval. In: Williams, Martha E., ed. Annual Review of

Information Science and Technology: Volume 14. White Plains, NY: Knowledge Industry Publications, Inc. for the American Society for Information Science; 1979. 129-151. ISSN: 0066-4200; ISBN: 0-914236-44-X.

HOLLAAR, LEE A. 1984. The Utah Text Retrieval Project: A Status Report. See reference: VAN RIJSBERGEN, C. J., ed. 1984. 123-132.

HOLLAAR, LEE A. 1985a. A Testbed for Information Retrieval Research: The Utah Retrieval System Architecture. See reference: ASSOCIATION FOR COMPUTING MACHINERY. 1985. 227-232.

HOLLAAR, LEE A. 1985b. The Utah Search Engine. In: Parkhurst, Carol A., ed. ASIS '85: Proceedings of the American Society for Information Science (ASIS) 48th Annual Meeting: Volume 22; 1985 October 20-24; Las Vegas, NV. White Plains, NY: Knowledge Industry Publications, Inc. for ASIS; 1985. 369. ISSN: 0044-7870; ISBN: 0-86729-176-1; CODEN: PAISDQ.

INGWERSEN, PETER. 1984. A Cognitive View of Three Selected Online Search Facilities. Online Review. 1984; 8(5): 465-492. ISSN: 0309-314X.

JARDINE, N.; VAN RIJSBERGEN, C. J. 1971. The Use of Hierarchic Clustering in Information Retrieval. Information Storage and Retrieval. 1971; 7: 217-240. ISSN: 0020-0271.

KANTOR, PAUL B. 1983. Minimal Constraint Implementation of the Maximum Entropy Principle in the Design of Term-Weighting Systems. In: Vondran, Raymond, F.; Caputo, Anne; Wasserman, Carol; Diener, Richard A.V., eds. Productivity in the Information Age: Proceedings of the American Society for Information Science (ASIS) 46th Annual Meeting: Volume 20; 1983 October 2-6; Washington, DC. White Plains, NY: Knowledge Industry Publications, Inc. for ASIS; 1983. 28-31. ISSN: 0044-7870; ISBN: 0-86729-072-2; CODEN: PAISDQ.

KOLL, MATTHEW B.; NOREAULT, TERRY; MCGILL, MICHAEL J. 1984. Enhanced Retrieval Techniques on a Microcomputer. In: Williams, Martha E.; Hogan, Thomas H., comps. Proceedings of the National Online Meeting; 1984 April 10-12; New York, NY. Medford, NJ: Learned Information, Inc.; 1984. 165-170. ISBN: 0-938734-07-5.

KOLODNER, JANET L. 1983. Indexing and Retrieval Strategies in Natural Language Fact Retrieval. ACM Transactions on Database Systems. 1983 September; 8(3): 434-464. ISSN: 0362-5915.

KRAWCZAK, DEBORAH A.; SMITH, PHILIP J.; SHUTE, STEVEN J.; CHIGNELL, MARK. 1985. EP-X: A Knowledge-Based System to Aid in Searches of the Environmental Pollution Literature. In: The Engineering of Knowledge-Based Systems: [Proceedings of the] 2nd Conference on Artificial Intelligence Applications; 1985 December 11-13; Miami Beach, FL. Washington, DC: Institute of Electrical and Electronics Engineers (IEEE) Computer Society Press; 1985. 455-460. Available from: IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08904 (Catalog no. 85CH2215-2). ISBN: 0-8186-0688-6; LC: 85-62321.

KROPP, D.; WALCH, G. 1981. A Graph Structured Text Field Based on Word Fragments. Information Processing & Management. 1981; 17(6): 363-376. ISSN: 0306-4573; CODEN: IPMADK.

KWOK, K. L. 1985a. A Probabilistic Theory of Indexing and Similarity Measure Based on Cited and Citing Documents. Journal of the American Society for Information Science. 1985 September; 36(5): 342-351. ISSN: 0002-8231; CODEN: AISJB6.

KWOK, K. L. 1985b. A Probabilistic Theory of Indexing Using Author-Provided Relevance Information. In: Parkhurst, Carol A., ed. ASIS'85: Proceedings of the American Society for Information Science (ASIS) 48th Annual Meeting: Volume 22; 1985 October 20-24; Las Vegas, NV. White Plains, NY: Knowledge Industry Publications, Inc. for ASIS; 1985. 59-63. ISSN: 0044-7870; ISBN: 0-86729-176-1; CODEN: PAISDQ.

LEVINSON, ROBERT ARLEN. 1985. A Self-organizing Retrieval System for Graphs. Austin, TX: University of Texas at Austin; 1985. (Ph.D. Dissertation). 101p. Available from: University Microfilms Int., Ann Arbor, MI; No. 8527601.

MARCUS, RICHARD S. 1983. An Experimental Comparison of the Effectiveness of Computers and Humans as Search Intermediaries. Journal of the American Society for Information Science. 1983 November; 34(6): 381-404. ISSN: 0002-8231; CODEN: AISJB6.

MARON, M. E. 1982. Associative Search Techniques versus Probabilistic Retrieval Models. Journal of the American Society for Information Science. 1982 September; 33(5): 308-310. ISSN: 0002-8231; CODEN: AISJB6.

MARON, M. E.; CURRY, SEAN; THOMPSON, PAUL. 1986. An Inductive Search System: Theory, Design, and Implementation. Institute of Electrical and Electronics Engineers (IEEE) Transactions on Systems, Man, and Cybernetics. 1986 January/February; SMC-16(1): 21-28. ISSN: 0018-9472.

MCCALL, FIONA M.; WILLETT, PETER. 1986. Criteria for the Selection of Search Strategies in Best Match Document Retrieval Systems. International Journal of Man-Machine Studies. 1986 September; 25(3): 317-326. ISSN: 0020-7373.

MCCUNE, BRIAN P.; TONG, RICHARD M.; DEAN, JEFFREY S.; SHAPIRO, DANIEL G. 1983. RUBRIC: A System for Rule-Based Information Retrieval. In: COMPSAC 83: Proceedings of the Institute of Electrical and Electronics Engineers (IEEE) Computer Society's 7th International Computer Software and Applications Conference; 1983 November 7-11; Chicago, IL. Silver Spring, MD: IEEE Computer Society Press; 1983. 166-172. Available from: IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854 (Catalog no. 83CH1940-6). ISSN: 0730-3157; ISBN: 0-8186-0509-X; LC: 83-640060.

MCCUNE, BRIAN P.; TONG, RICHARD M.; DEAN, JEFFREY S.; SHAPIRO, DANIEL G. 1985. RUBRIC: A System for Rule-Based Information Retrieval. Institute for Electrical and Electronics Engineers (IEEE) Transactions on Software Engineering. 1985 September; SE-11(9): 939-944. ISSN: 0098-5589.

MCGILL, MICHAEL J.; HUITFELDT, JENNIFER. 1979. Experimental Techniques of Information Retrieval. In: Williams, Martha E., ed. Annual Review of Information Science and Technology: Volume 14. White Plains, NY: Knowledge Industry Publications, Inc. for the American Society for Information Science; 1979. 93-127. ISSN: 0066-4200; ISBN: 0-914236-44-X; LC: 66-25096; CODEN: ARISBC.

MCGILL, MICHAEL J.; KOLL, MATTHEW B.; NOREAULT, TERRY. 1979. An Evaluation of Factors Affecting Document Ranking by Information Retrieval Systems. Syracuse, NY: Syracuse University, School of Information Studies; 1979. (Technical Report). Available from: Syracuse University, School of Information Studies, Huntington Hall, Syracuse, NY 13244.

METZLER, DOUGLAS P.; NOREAULT, TERRY; HEIDORN BRYAN. 1983. Syntactic Parsing for Information Retrieval. In: Vondran, Raymond F.; Caputo, Anne; Wasserman, Carol; Diener, Richard A. V., eds. Proceedings of the American Society for Information Science (ASIS) 46th Annual Meeting; 1983 October 2-6; Washington, DC. White Plains, NY: Knowledge Industry Publications, Inc.; 1983. 269-273. ISSN: 0044-7870; ISBN: 0-86729-072-2; CODEN: PAISDQ.

MIYAMOTO, SADAAKI; NAKAYAMA, K. 1986. Fuzzy Information Retrieval Based on a Fuzzy Pseudothesaurus. Institute of Electrical and Electronics Engineers (IEEE) Transactions on Systems, Man, and Cybernetics. 1986 March/April; SMC-16(2): 278-282. ISSN: 0018-9472.

MOHAN, KONDRAHALLI C.; WILLETT, PETER. 1985. Nearest Neighbor Searching in Serial Files Using Text Signatures. Journal of Information Science. 1985; 11(1): 31-39. ISSN: 0165-5515.

MUHLHAUSER, G. 1985. Dawn of Next Generation Information Retrieval. In: 9th International Online Information Meeting; 1985 December 3-5; London, England. Oxford, England: Learned Information, Ltd.; Medford, NJ: Learned Information, Inc.; 1985. 365-371. ISBN: 0-904933-50-4.

MUKHOPADHYAY, UTTAM; STEPHENS, LARRY M.; HUHNS, MICHAEL N.; BONNELL, RONALD D. 1986. An Intelligent System for Document Retrieval in Distributed Office Environments. Journal of the American Society for Information Science. 1986 May; 37(3): 123-135. ISSN: 0002-8231; CODEN: AISJB6.

MURTAGH, F. 1985. Clustering and Nearest Neighbor Searching. In: Informatics 8: Advances in Intelligent Retrieval: Proceedings of a Conference Jointly Sponsored by Aslib, the Aslib Informatics Group, and the Information Retrieval Specialist Group of the British Computer Society; 1985 April 16-17; Wadham College, Oxford, England. London, England: Aslib, The Association for Information Management, Information House; 1985. 54-65. Available from: Aslib, 26-27 Boswell Street, London WC1N3JZ, England. ISBN: 0-85142-195-4.

NOREAULT, TERRY; CHATHAM, R. 1982. A Procedure for the Estimation of Term Similarity Coefficients. Information Technology: Research and Development. 1982; 1(3): 189-196. ISSN: 0144-817X.

NOREAULT, TERRY; MCGILL, MICHAEL J.; KOLL, MATTHEW B. 1981. A Performance Evaluation of Similarity Measures, Document Term Weighting Schemes and Representations in a Boolean Environment. In: Oddy, R. N.; Robertson, S. E.; Van Rijsbergen, C. J.; Williams, P. W., eds. Information Retrieval Research: Papers given at the 1st Joint British Computer Society (BCS) and Association for Computing Machinery (ACM) Symposium: Research and Development in Information Retrieval; 1980 June; St. John's College, Cambridge, England. London: Butterworths; 1981. 57-76. ISBN: 0-408-10775-8.

ODDY, R. N. 1977. Information Retrieval through Man-Machine Dialogue. Journal of Documentation. 1977; 33: 1-14. ISSN: 0022-0418.

PANYR, JIRI. 1986a. Probabilistische Modelle in Information-Retrieval-Systemen [Probabilistic Models in Information-Retrieval-Systems]. Nachrichten für Dokumentation (West Germany). 1986 April; 37(2): 60-66. ISSN: 0027-7436.

PANYR, JIRI. 1986b. Die Theorie der Fuzzy Mengen und Information-Retrieval-Systeme [Fuzzy Set Theory and Information Retrieval Systems]. Nachrichten für Dokumentation (West Germany). 1986; 37(3): 163-167. ISSN: 0027-7436.

POGUE, CHRISTINE A.; WILLETT, PETER. 1984. An Evaluation of Document Retrieval from Serial Files Using the ICL Distributed Array Processor. Online Review. 1984; 8(6): 569–584. ISSN: 0309-314X.

POGUE, CHRISTINE A.; WILLETT, PETER. 1987. Use of Text Signatures for Document Retrieval in a Highly Parallel Environment. Parallel Computing. 1987. (In press). ISSN: 0167-8191.

POLLITT, A. S. 1983. End User Searching for Cancer Therapy Literature A Rule Based Approach. In: Kuehn, Jennifer J., ed. Proceedings of the 6th Annual International Association for Computing Machinery Special Interest Group on Information Retrieval (ACM SIGIR) Conference on Research and Development in Information Retrieval; 1983 June 6–8; Bethesda, MD. New York, NY: ACM, Inc.; 1983. 136–145. (A publication of SIGIR Forum; 1983 Summer; 17(4)). Available from: ACM Order Department, P.O. Box 64145, Baltimore, MD 21264. ISBN: 0-89791-107-5.

POLLITT, A. S. 1985. End User Boolean Searching on Viewdata Using Numeric Keypads. In: 9th International Online Information Meeting; 1985 December 3–5; London, England. Oxford, England: Learned Information, Ltd.; Medford, NJ: Learned Information, Inc.; 1985. 373–379. ISBN: 0-904933-50-4.

POPOV, I. I.; KRAVCHENKO, A. E.; PAVLOV, A. N. 1985. Implementing Associative Search Strategies in a Documentary-Lexical Information Base. Automatic Documentation and Mathematical Linguistics. 1985; 19(1): 13–22. (Translated from: Nauchno-Tekhnicheskaya Informatsiya, Ser. 2. 1985; 19(1): 9–15). ISSN: 0005-1055.

RABITTI, FAUSTO, ed. 1986. [Proceedings of the] Association for Computing Machinery (ACM) Conference on Research and Development in Information Retrieval; 1986 September 8–10; Pisa, Italy. 283p. Available from: ACM Order Department, P.O. Box 64145, Baltimore, MD 21264 (Order no. 606860); or ACM-CNR CONFERENCE, IEI- Via S. Maria, 46, 56100 Pisa, Italy. ISBN: 0-89791-187-3.

RADECKI, TADEUSZ. 1982. Reducing the Perils of Merging Boolean and Weighted Retrieval Systems. Journal of Documentation. 1982 September; 38(3): 207–211. ISSN: 0022-0418.

RADECKI, TADEUSZ. 1983. Incorporation of Relevance Feedback into Boolean Retrieval Systems. In: Salton, Gerard; Schneider, Hans-Jochen, eds. Research and Development in Information Retrieval: Proceedings [of the 5th International Conference of the Association for Computing Machinery, Special Interest Group on Information Retrieval (ACM-SIGIR)]; 1982 May 18–20; Berlin, West Germany. Berlin: Springer-Verlag; 1983. 133–150. (Lecture Notes in Computer Science: 146). ISBN: 3-540-11978-7 (Berlin); ISBN: 0-387-11978-7 (New York).

RADECKI, TADEUSZ. 1985. A Theoretical Framework for Defining Similarity Measures for Boolean Search Request Formulations, Including Some Experimental Results. Information Processing & Management. 1985; 21(6): 501–524. ISSN: 0306-4573; CODEN: IPMADK.

RAU, LISA F. 1987. Knowledge Organization and Access in a Conceptual Information System. Information Processing & Management. 1987. (In press). ISSN: 0306-4573; CODEN: IPMADK.

ROBERTSON, S. E. 1977a. The Probability Ranking Principle in IR. Journal of Documentation. 1977 December; 33(4): 294–304. ISSN: 0022-0418.

ROBERTSON, S. E. 1977b. Theories and Models in Information Retrieval. Journal of Documenation. 1977 June; 33(2): 126-148. ISSN: 0022-0418.

ROBERTSON, S. E.; SPARCK JONES, KAREN. 1976. Relevance Weighting of Search Terms. Journal of the American Society for Information Science. 1976 May/June; 27(3): 129-146. ISSN: 0002-8231; CODEN: AISJB6.

ROBERTSON, S. E.; THOMPSON, C. L.; MACASKILL, M. J.; BOVEY, J. D. 1986. Weighting, Ranking and Relevance Feedback in a Front-End System. Journal of Information Science. 1986; 12(1-2): 71-75. ISSN: 0165-5515.

SACCO, GIOVANNI MARIA. 1984. OTTER: An Information Retrieval System for Office Automation. In: Ellis, Clarence A., ed. [Proceedings of the] 2nd Association for Computing Machinery, Special Interest Group on Office Automation (ACM-SIGOA) Conference on Office Information Systems; 1984 June 25-27; Toronto, Canada. New York, NY: ACM, Inc.; 1984. 104-112. (Published as: SIGOA Newsletter. 1984; 5(1-2)). Available from: ACM Order Department, P.O. Box 64145, Baltimore, MD 21264 (Order no. 611840). ISBN: 0-89791-140-7.

SALTON, GERARD. 1968. Automatic Information Organization and Retrieval. New York, NY: McGraw-Hill; 1968. 514p. LC: 68-25664.

SALTON, GERARD. 1979. Mathematics and Information Retrieval. Journal of Documentation. 1979; 35(1): 1-29. ISSN: 0022-0418.

SALTON, GERARD. 1981. A Blueprint for Automatic Indexing. SIGIR Forum. 1981; 16: 22-38. (A publication of the Association for Computing Machinery Special Interest Group on Information Retrieval). Available from: ACM SIGIR, 11 W. 42nd St., New York, NY 10036.

SALTON, GERARD. 1983. On the Representation of Query Term Relations by Soft Boolean Operators. In: Proceedings of the Association for Computational Linguistics (ACL) European Chapter 2nd Conference; 1985 March 27-29; Geneva, Switzerland. [n.p.]: ACL; 1985. 116-122. Available from: Donald E. Walker (ACL), Bell Communications Research, 445 South Street, MRE 2A379, Morristown, NJ 07960.

SALTON, GERARD. 1985. Some Characteristics of Future Information Systems. SIGIR Forum. 1985 Fall; 18(2-4): 28-39. (A publication of the Association for Computing Machinery Special Interest Group on Information Retrieval). Available from: ACM SIGIR, 11 W. 42nd St., New York, NY 10036.

SALTON, GERARD. 1986a. Another Look at Automatic Text-Retrieval Systems. Communications of the ACM. 1986 July; 29(7): 648-656. ISSN: 0001-0782.

SALTON, GERARD. 1986b. Recent Trends in Automatic Information Retrieval. See reference: RABITTI, FAUSTO, ed. 1-10.

SALTON, GERARD; BUCKLEY, CHRIS; YU, C. T. 1983a. In: Salton, Gerard; Schneider, Hans-Jochen, eds. Research and Development in Information Retrieval: Proceedings [of the 5th International Conference of the Association for Computing Machinery, Special Interest Group on Information Retrieval (ACM-SIGIR)]; 1982 May 18-20; Berlin, West Germany. Berlin: Springer-Verlag; 1983. 151-165. (Lecture Notes in Computer Science: 146). ISBN: 3-540-11978-7 (Berlin); ISBN: 0-387-11978-7 (New York).

SALTON, GERARD; FOX, EDWARD A.; VOORHEES, ELLEN M. 1985. Advanced Feedback Methods in Information Retrieval. Journal of the American Society for Information Science. 1985 May; 36(3): 200-210. ISSN: 0002-8231; CODEN: AISJB6.

SALTON, GERARD; FOX, EDWARD A.; WU, HARRY. 1983b. Extended Boolean Information Retrieval. Communications of the ACM. 1983 November; 26(11): 1022-1036. ISSN: 0001-0782.

SALTON, GERARD; MCGILL, MICHAEL J. 1983. Introduction to Modern Information Retrieval. New York, NY: McGraw-Hill; 1983. 400p. ISBN: 0-07-054484-0.

SALTON, GERARD; VOORHEES, ELLEN M. 1985. Automatic Assignment of Soft Boolean Operators. See reference: ASSOCIATION FOR COMPUTING MACHINERY. 54-69.

SARACEVIC, TEFKO. 1983. On a Method for Studying the Structure and Nature of Requests in Information Retrieval. In: Vondran, Raymond F.; Caputo, Anne; Wasserman, Carol; Diener, Richard A. V., eds. Productivity in the Information Age: Proceedings of the American Society for Information Science (ASIS) 46th Annual Meeting: Volume 20; 1983 October 2-6; Washington, DC. White Plains, NY: Knowledge Industry Publications, Inc.; 1983. 22-25. ISSN: 0044-7870; ISBN: 0-86729-072-2; CODEN: PAISDQ.

SCHANK, ROGER C. 1975. Conceptual Information Processing. Amsterdam, The Netherlands: North-Holland; New York, NY: American Elsevier; 1975. 374p. ISBN: 0-444-10773-8 (American Elsevier).

SCHANK, ROGER C.; ABELSON, ROBERT P. 1977. Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures. Hillsdale, NJ: L. Erlbaum Associates; 1977. ISBN: 0-470-99033-3; LC: 76-051963.

SCHEK, HANS-JOERG. 1984. Nested Transactions in a Combined IRS-DBMS Architecture. See reference: VAN RIJSBERGEN, C. J., ed. 1984. 55-70.

SHAW, W. M., JR. 1986. An Investigation of Document Partitions. Information Processing & Management. 1986; 22(1): 19-28. ISSN: 0306-4573; CODEN: IPMADK.

SHOVAL, PERETZ. 1985. Principles, Procedures and Rules in an Expert System for Information Retrieval. Information Processing & Management. 1985; 21(6): 475-487. ISSN: 0306-4573; CODEN: IPMADK.

SIMMONS, ROBERT F. 1987. A Text Knowledge Base from the AI Handbook. Information Processing & Management. 1987. (In press). ISSN: 0306-4573; CODEN: IPMADK.

SMEATON, A. F.; VAN RIJSBERGEN, C. J. 1981. The Nearest Neighbor Problem in Information Retrieval: An Algorithm Using Upper Bounds. Proceedings of the Association for Computing Machinery Special Interest Group on Information Retrieval (ACM SIGIR) 4th International Conference on Information Storage and Retrieval. 1981 May 31-June 2; Oakland, CA. New York, NY: ACM, Inc., 1981. 83-87. ISBN: 0-89791052-4.

SMEATON, A. F.; VAN RIJSBERGEN, C. J. 1986. Information Retrieval in an Office Filing Facility and Future Work in Project Minstrel. Information Processing & Management. 1986; 22(2): 135-149. (Special issue). ISSN: 0306-4573; CODEN: IPMADK.

SNEATH, PETER H. A.; SOKAL, ROBERT R. 1973. Numerical Taxonomy: The Principles and Practice of Numerical Classification. San Francisco, CA: W. H. Freeman; 1973. 573p. ISBN: 0-7167-0697-0; LC: 72-001552.

SPARCK JONES, KAREN. 1974. Automatic Indexing. Journal of Documentation. 1974; 30(4): 393-432. ISSN: 0022-0418.

SPARCK JONES, KAREN; TAIT, J. I. 1984. Automatic Search Term Variant Generation. Journal of Documentation. 1984 March; 40(1): 50-66. ISSN: 0022-0418.

SPARCK JONES, KAREN; WEBSTER, C. A. 1980. Research on Relevance Weighting. Cambridge, England: University of Cambridge, Computer Laboratory; 1980. (British Library R&D Report 5553). Available from: Computer Laboratory, University of Cambridge, Corn Exchange Street, Cambridge CB2 3QG, England.

SUBRAMANIAN, VISWANATH; BISWAS, GAUTAM; BEZDEK, JAMES C. 1986. Document Retrieval Using a Fuzzy Knowledge-Based System. Optical Engineering. 1986 March; 25(3): 445-455. ISSN: 0091-3286.

TONG, RICHARD M.; ASKMAN, VICTOR N.; CUNNINGHAM, JAMES F.; TOLLANDER, CARL J. 1985. See reference: ASSOCIATION FOR COMPUTING MACHINERY. 243-251.

TONG, RICHARD M.; SHAPIRO, DANIEL G. 1985. Experimental Investigations of Uncertainty in a Rule-Based System for Information Retrieval. International Journal of Man-Machine Studies. 1985; 22(3): 265-282. ISSN: 0020-7373.

VAN RIJSBERGEN, C. J. 1977. A Theoretical Basis for the Use of Co-Occurrence Data in Information Retrieval. Journal of Documentation. 1977 June; 33(2): 106-119. ISSN: 0022-0418.

VAN RIJSBERGEN, C. J. 1979. Information Retrieval. 2nd ed. London, England: Butterworths; 1979. 208p. ISBN: 0-408-70929-4.

VAN RIJSBERGEN, C. J., ed. 1984. Research and Development in Information Retrieval: Proceedings of the 3rd Joint British Computer Society (BCS) and Association for Computing Machinery (ACM) Symposium; 1984 July 2-6; Cambridge, England. Cambridge, England: Cambridge University Press; 1984. 433p. (The British Computer Society Workshop Series). ISBN: 0-521-26865-6; LC: 84-45234.

VAN RIJSBERGEN, C. J. 1986a. A New Theoretical Framework for Information Retrieval. See reference: RABITTI, FAUSTO, ed. 194-200.

VAN RIJSBERGEN, C. J. 1986b. A Non-classical Logic for Information Retrieval. Computer Journal. 1986 December; 29(6): 481-485. ISSN: 0010-4620.

VAN RIJSBERGEN, C. J., ed. 1986c. Office Automation. Information Processing & Management. 1986; 22(2). (Special Issue). ISSN: 0306-4573; ISBN: 0-08-034217-5; CODEN: IPMADK.

VAN RIJSBERGEN, C. J.; ROBERTSON, S. E.; PORTER, M. F. 1980. New Models in Probabilistic Information Retrieval. Cambridge, England: University of Cambridge, Computer Laboratory; 1980. 123p. (British Library R&D Report 5587). Available from: Computer Laboratory, University of Cambridge, Corn Exchange Street, Cambridge CB2 3QG, England.

VLADUTZ, G.; COOK, J. 1984. Bibliographic Coupling and Subject Relatedness. In: Flood, Barbara; Witiak, Joanne; Hogan, Thomas H., comps. 1984: Challenges to an Information Society: Proceedings of the American Society for Information Science (ASIS) 47th Annual Meeting: Volume 21; 1984 October 21-25; Philadelphia, PA. White Plains, NY: Knowledge Industry Publications, Inc.; 1984. 204-207. ISSN: 0044-7870; ISBN: 0-86729-115-X; CODEN: PAISDQ.

VOORHEES, ELLEN M. 1986. The Efficiency of Inverted Index and Cluster Searches. See reference: RABITTI, FAUSTO, ed. 164–174.

WALKER, DONALD E.; HOBBS, JERRY R. 1981. Natural Language Access to Medical Text. Menlo Park, CA: SRI International, Artificial Intelligence Center, Computer Science and Technology Division; 1981 March. 20p. (Technical Note 240; Project 1944). Available from: SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025.

WATTERS, C. R.; SHEPHERD, M. A.; GRUNDKE, E. W.; BODORIK, P. 1985. Integration of Menu Retrieval and Boolean Retrieval from a Full-Text Database. Online Review. 1985;9(5): 391–402. ISSN: 0309-314X.

WILLETT, PETER. 1984a. A Nearest Neighbour Search Algorithm for Bibliographic Retrieval from Multilist Files. Information Technology: Research and Development. 1984 April;3(2): 78–83. ISSN: 0144-817X.

WILLETT, PETER. 1984b. A Note on the Use of Nearest Neighbors for Implementing Single Linkage Document Classifications. Journal of the American Society for Information Science. 1984 May; 35(3): 149–152. ISSN: 0002-8231; CODEN: AISJB6.

WILLETT, PETER. 1985a. An Algorithm for the Calculation of Exact Term Discrimination Values. Information Processing & Management. 1985; 21(3): 225–232. ISSN: 0306-4753; CODEN: IPMADK.

WILLETT, PETER. 1985b. Query-Specific Automatic Document Classification. International Forum on Information and Documentation. 1985 April; 10(2): 28–32. ISSN: 0304-9701.

WILLETT, PETER. 1985c. Ranked Output Searching in Textual and Structural Data Bases. In: 9th International Online Information Meeting; 1985 December 3–5; London, England. Oxford, England: Learned Information, Ltd.; Medford, NJ: Learned Information, Inc.; 1985. 343–353. ISBN: 0-904933-50-4.

WILLETT, PETER; WINTERMAN, VIVIENNE; BAWDEN, DAVID. 1986. Implementation of Nearest-Neighbor Searching in an Online Chemical Structure Search System. Journal of Chemical Information and Computer Sciences. 1986; 26(1): 36–41. ISSN: 0095-2338; CODEN: JCISD8.

WINETT, SHEILA G.; FOX, EDWARD A. 1985. Using Information Retrieval Techniques in an Expert System. In: The Engineering of Knowledge-Based Systems: [Proceedings of the] 2nd Conference on Artificial Intelligence Applications; 1985 December 11–13; Miami Beach, FL. Washington, DC: Institute of Electrical and Electronics Engineers (IEEE) Computer Society Press; 1985. 230–235. Available from: IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854 (Catalog no. 85CH2215-2). ISBN: 0-8186-0688-6; LC: 85-62321.

WONG, S. K. M.; RAGHAVAN, VIJAY V. 1984. Vector Space Model of Information Retrieval– A Reevaluation. See reference: VAN RIJSBERGEN, C. J., ed. 1984. 167–185.

WONG, S. K. M.; ZIARKO, WOJCIECH. 1985. On Generalized Vector Space Model in Information Retrieval. Annales Societatis Mathematicae Polonnae: Series IV: Fundamenta Informaticae. 1985; 8(2): 253–267. Available from: Fundamenta Informaticae, P. O. Box 22, Warsaw 00-901, Poland.

WONG, S. K. M.; ZIARKO, WOJCIECH; WONG, P. C. N. 1985. See reference: ASSOCIATION FOR COMPUTING MACHINERY. 18–25.

YU, C. T.; BUCKLEY, CHRIS; LAM, K.; SALTON, GERARD. 1983. A Generalized Term Dependence Model in Information Retrieval. Information Technology: Research and Development. 1983 October; 2(4): 129-154. ISSN: 0144-817X.

YU, C. T.; LEE, T. C. 1986. Non-binary Independence Model. See reference: RABITTI, FAUSTO, ed. 265-268.

YU, C. T.; LUK, W. S.; SIU, M. K. 1979. On Models of Information Retrieval Processes. Information Systems. 1979; 4(3): 205-218. ISSN: 0306-4379; CODEN: INSYD6.

ZARRI, GIAN PIERO. 1983. RESEDA, an Information Retrieval System Using Artificial Intelligence and Knowledge Representation Techniques. In: Kuehn, Jennifer J., ed. Proceedings of the 6th Annual International Association for Computing Machinery, Special Interest Group on Information Retrieval (ACM-SIGIR) Conference on Research and Development in Information Retrieval; 1983 June 6-8; Bethesda, MD. New York, NY: ACM, Inc.; 1983. 189-195. (A publication of SIGIR Forum, 1983 Summer; 17(4)). Available from: ACM Order Department, P. O. Box 64145, Baltimore, MD 21264. ISBN: 0-89791-107-5.

ZARRI, GIAN PIERO. 1984. Some Remarks about the Inference Techniques of RESEDA, an "Intelligent" Information Retrieval System. See reference: VAN RIJSBERGEN, C. J., ed. 1984. 281-300.

ZARRI, GIAN PIERO. 1985. Interactive Information Retrieval: An Artificial Intelligence Approach to Deal with Biographical Data. In: Informatics 8: Advances in Intelligent Retrieval: Proceedings of a Conference Jointly Sponsored by Aslib, the Aslib Informatics Group, and the Information Retrieval Specialist Group of the British Computer Society; 1985 April 16-17; Wadham College, Oxford, England. London, England: Aslib, The Association for Information Management, Information House; 1985. 101-119. Available from: Aslib, 26-27 Boswell Street, London WC1N3JZ, England. ISBN: 0-85142-195-4.

ZENNER, REMBRAND B. R. C.; DE CALUWE, RITA M. M.; KERRE, ETIENNE E. 1985. A New Approach to Information Retrieval Systems Using Fuzzy Expressions. Fuzzy Sets and Systems (The Netherlands). 1985; 17(1): 9-22. ISSN: 0165-0114.