The Knowledge Engineering Review, Vol. 15:2, 2000, 171–179
Printed in the United Kingdom

Logical and uncertainty models for information access: current trends

FABIO CRESTANI[1] and MOUNIA LALMAS[2]

[1]*Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, Scotland (email: fabio@dcs.gla.ac.uk)*
[2]*Department of Computer Science, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK (email: mounia@dcs.qmw.ac.uk)*

Abstract

The current trends of research in information access as emerged from the 1999 Workshop on Logical and Uncertainty Models for Information Systems (LUMIS'99) are briefly reviewed in this paper. We believe that some of these issues will be central to future research on theory and applications of logical and uncertainty models for information access.

1 Introduction

The advent of electronic tools for producing and storing information has resulted in an avalanche of computer readable text. The access to all this information has gone through a slow but steady process to adapt to the growth of availability of electronically stored data, and a large number of tools have been developed to enable a user to access and manage these large volumes of information. But what has uncertainty and logic to do with accessing and managing information stored in a computer?

The representation of "information objects" is often uncertain. For example, the extraction of index terms from a document or a query to represent the document or query informative content is a highly uncertain process. The data describing the redness of a red object present in a picture to be stored in a multimedia database is subject to a certain degree of uncertainty too. Uncertainty plays a very important role in the representation, access, and retrieval of information. Probability theory is one way of dealing with uncertainty, but there are many other different approaches, such as, for example, fuzzy logic, Dempster–Shafer's theory of evidence, neural networks, and so on.

Logics, too, may play a very important role in the representation, access and retrieval of information. Logics has proved to be a very powerful modeling and reasoning tool, providing a degree of formality and correctness that can be very useful in the task of dealing with the uncertainty of information objects.

The purpose of the 1999 Workshop on Logical and Uncertainty Models for Information Systems (LUMIS'99), held as part of the Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'99), was to promote discussion and interaction among members of the Information Systems community. In particular, among those members with research interests in logical and uncertainty models for the treatment of semi-structured and unstructured information. This community is made of people coming from different fields (theoretical computer science, databases, information retrieval, hypermedia, digital libraries, to mention just a few) that often find it difficult to talk to each other because of the academic barriers of the different fields of research. We believe we have succeeded in gathering an heterogeneous community that will benefit considerably from exchanging ideas and experiences. We found this in our previous experience in organizing the 1995 and 1996 Workshops on Information Retrieval, Uncertainty and Logics (WIRUL) of which LUMIS can be considered a direct descendent (Lalmas,

1995; Crestani and Lalmas, 1996; Crestani et al., 1998a). We believe that the exchange of ideas and discussions that took place during LUMIS'99, of which the proceedings are a memory[1] have been very stimulating (Crestani and Lalmas, 1999).

In this paper we want to briefly highlight some of the research issues that have emerged from LUMIS'99 since we believe they may be of interest to a wide audience. We believe that some of these issues will be central to future research on the use of logical and uncertainty models for information access.

2 A fuzzy logic for multimedia database querying

Fuzzy set methods have been already applied to the representation of flexible queries in databases systems as well as in information retrieval (see, for example, Kasabov and Kozma (1998) and Crestani and Pasi (2000)). This methodology seems to be even more promising in multimedia databases, an area in which there has been only a few preliminary and specialized papers on the use of fuzzy logic. Multimedia databases have a complex structure, since documents have to be retrieved and selected not only by their contents but also by their appearance, as specified by the user.

The paper by Dubois et al. (1999) presents a preliminary investigation of two potential applications of fuzzy logic in multimedia databases querying. The first one concerns the detection of modifications in semi-structured documents and relies on a graph matching procedure, where subparts of the graph describing the structures have different levels of importance. The second illustrates another aspect of the fuzzy logic methodology allowing for a querying by examples process.

Research on "fuzzy databases" (see Bosc and Kacprzyk (1995), for example) has been carried out for over 20 years. Most work has been mainly concentrated on flexible querying in classical languages, and in object-oriented languages. It has been shown that the flexibility of a query reflects the preferences of the end-user. Using a fuzzy set representation, the extent to which an object described in the database satisfies a request then becomes a matter of degree. Moreover, a query may also allow for some similarity-based tolerance, where close values are often perceived as similar, or interchangeable. In fact, if for instance an attribute value v satisfies an n elementary requirement, a value "close" to v should still somewhat satisfy the requirement. An advantage of fuzzy set-based modeling, is that it is mainly qualitative in nature. Fuzzy set membership functions are convenient tools for modeling user's preference profiles and the large set of fuzzy connectives can capture the different user attitudes concerning the way the different criteria present in his/her query compensate or not (Zadeh, 1987).

In the context of information system applications, one of the main differences between a multimedia database and an ordinary one, from a querying point of view, is that in the first case the request may refer to a document in terms of its appearance, and not only in terms of its information contents. The lack of standardized structure of the documents to be retrieved calls for the use of flexible tools the design of which fuzzy logic is well suited for. This is particularly true for semi-structured documents, like for example HTML or XML documents, that may have some structure, but if it exists, it is enclosed in the instance and is unknown *a priori*. Such structure must be elicited to discover patterns. Dubois et al. (1999) assume that we can elicit a tree-like structure representing the document by recognizing structure items from the content itself, through automatic identification of marks and their rewriting by means of dictionaries. From them, some edges in the tree could be identified which can be considered as more important than others, according to the mark-up process. The level of importance of the edge depends on the recognized mark, its content, the text associated with the pending node, and the depth of the tree representing the structure. It is then possible to evaluate how similar two structures are, and use this results to retrieve and rank documents similar in structure to a given document or a query.

[1]The proceedings of LUMIS'99 are available online at `http://www.dcs.gla.ac.uk/lumis99`.

However, the user is not always able to easily express his/her request, even in a flexible way. Luckily, the same results presented above can be used to carry out a "query by example". Querying based on examples, for eliciting user's preferences, can provide the necessary information for building a query. Thus, the user may say to what extent a few examples of documents are representative of what he/she is looking for by using some finite similarity scale. This process can even be extended to provide counter-examples of what the user is looking for, so that a document will be more relevant to a user need the more similar it is to the examples and the more dissimilar to the counter-examples. Examples and counter-examples are supposed to be provided by the user with their levels of representativity. Finally, complex queries involving combined content specification, examples, and counter-examples can be written, giving the user a very powerful way of expressing his/her information need.

It will be interesting to see some experimental results of the implementation of these ideas. Moreover, the design of a interface enabling the user to specify queries involving content specification, examples and counter-examples is in itself a very interesting research topic.

3 A logic for extended boolean information retrieval systems

Pasi (1999) analyses how logic can be used as a formal basis to formalize the query evaluation process, that is, how to estimate the relevance of a document to a query, for the Boolean and the extended Boolean models. The approach follows a different view from that proposed by van Rijsbergen (1986) or Nie (1988), where the relevance of a document d to a query q is estimated using non-classical uncertain implications $P(d \rightarrow q)$ and $P(q \rightarrow d)$, where P is a probability measure. Pasi starts with the Boolean model, formulates the query evaluation process by means of first order logic, and then formulates the query evaluation process for the extended (or weighted) Boolean model by means of fuzzy logic. The implication is the material one, and relates a term t in a query q with a document $d : q(t) \rightarrow d(t)$. The correspondence between q and d is estimated as the degree of truth of the document under the query interpretation. The direction of the implication derives from how relevance is modeled in the Boolean models: a document is relevant to a query term if its representation contains the term.

One of the main aims of this work is to propose a logical formulation of Boolean models and their extensions, to better understand these models. Particular attention is given to the query weight semantics proposed in the literature. In other words, this work proposes a meta-model for these models. The use of logic as a meta-model to study IR models is an important ally of research (Huibers, 1996). It aims at being able to evaluate IR models not only based on their effectiveness with classical test collections, but also with respect to their qualitative properties. Although no analysis has been done yet (the paper only discusses a possible formalism), the proposed framework can be used to conduct such analysis.

The formulation is based on first-order logic, where predicates are used to express properties between documents and terms (whether the term indexes the document) and between queries and terms (whether the term appears in the query). Although these two predicates are sufficient for the Boolean model and its extension, the formalism seems too powerful for such limited properties. However, it should not be a problem to add predicates for expressing more sophisticated properties (e.g., properties between terms extracted from thesauri or properties between documents derived from clustering them).

The formulation is naturally extended to represent the extended Boolean model. The predicates are fuzzy sets that represent the degree of membership of terms to documents and queries. These degrees model the weights used in the extended Boolean models. Implication is evaluated using the fuzzy implication (Zadeh, 1987).

To conclude, the work uses predicate logic to formalize the Boolean and extended Boolean models, with the view of analyzing these models. Although no analysis has been done yet, and the use of predicate logic seems too expressive to formalize these two models, the proposed approach definitively contribute towards the development of a meta-theory for IR.

4 Probabilistic argumentation systems

One basic assumption of probabilistic models of IR is that terms and documents are independent, although this assumption has been reshaped in recent times (Cooper, 1995). Nevertheless a number of investigations have demonstrated the potential usefulness of incorporating dependencies or relationships between terms (e.g., van Rijsbergen (1977) and Qiu and Frei (1993)) and between documents (for example, Croft and Thompson (1987) and Savoy (1994)). When attempting to incorporate these new features in the matching process, the limits of the traditional term matching approach appear more clearly, and one is often left to the use of ad-hoc schemes. For this reason, and others, it has been argued by a number of researchers that the best way to model the information retrieval process is by the use of an appropriate logic (see Lalmas (1998) for an overview). In fact, the expressiveness of logic makes it a very attractive framework for modeling relationships between terms or documents, although the complexity of its implementation makes it difficult to have large-scale applications. A possible way out of this ''impasse'' could be found in integrating logic in large-scale classic IR systems as a tool for solving specific problems which cannot be formalized within more conventional approaches. The different steps of the retrieval process could be done as usual, and logic could be used as a formal tool for modifying the output of certain components of the retrieval system. In this way the ''logical components'' could be integrated to any retrieval system working on soft term matching, e.g., the vector-space or probabilistic models, even in large-scale applications. The paper by Picard starts from this original premise (Picard, 1999).

Picard's choice of the logic starts from the consideration that IR can be seen both as an inference process under uncertainty involving complex relationships between information items, and as a task of proper assessment of uncertainty. This view is shared by a number of other logical-probabilistic approaches to IR (Crestani et al., 1998b). The originality of Picard is in the choice of the framework used: Probabilistic Argumentation Systems (PAS). PASs are a technique for reasoning under uncertainty which emphasize both the inference process and the assessment of uncertainty, by clearly distinguishing the qualitative and quantitative aspects of uncertainty. PASs represent uncertainty in a clear and easily understandable way: the qualitative part is handled with propositional logic and the quantitative part is treated with probability theory. PASs offer a natural way to model relationships between terms and between documents, and allows complex inferences. Picard presents two applications of PAS:

1. For taking into account existing hypertext links in order to improve an initial ranking of documents.
2. For considering statistical similarities between query terms to improve query weighting.

These two applications can be easily integrated in a retrieval system based on term matching, even for large-scale applications. The interested reader should refer to Picard (1999) for the technical details that cannot be presented here for reasons of space. However, it is worth noticing that even though the symbolic part of uncertain knowledge is naturally modeled with PASs, the numerical assessment of probabilities is often a very difficult problem. It is clear that if the uncertainty is incorrectly assessed, combined and propagated, the inference carried out by the logic will very probably be unable to improve retrieval effectiveness. The transformation of easily estimated statistical similarity information into a measure of uncertainty (e.g., probabilities) is one of the major difficulties found not only by Picard, but also by a number of other researchers (see, for example, Zobel and Moffat (1998)). This is still an open problem. Nevertheless, the work presented in Picard's paper highlights that in IR the numerical and symbolic aspects of uncertainty are profoundly interlaced. It is the author's opinion that a purely symbolic or numerical approach would not bring the same insight in these problems. The theoretical foundations of PASs, which rely on the theory of evidence, make them a reliable technique for approaching problems in which the quantitative and qualitative aspects of uncertainty are of equal importance.

5 A new view of relevance effectiveness

Precision and recall are the two traditional effectiveness measures of IR: precision is the proportion of retrieved documents that are relevant, whereas recall is the proportion of relevant documents that are retrieved. A third less used measure is fallout: the proportion of non-relevant documents that are retrieved. In his paper, Dominich (1999) explains that a surface, which looks similar for all IR systems, can be constructed, such that each point of the surface corresponds to a 3-tuple (precision, recall, fallout), associated to one retrieval process. A sequence of repeatedly applied relevance feedback processes corresponds to a sequence on this surface. Relevance feedback uses information given by a user regarding which of the retrieved documents are relevant or non-relevant, to construct a query that better reflects the user's information need than the original one.

Assuming that the sequence of relevance feedback improves the effectiveness of the IR system, the question addressed by the author is whether the sequence of points tends towards an optimal point, which would corresponds to the optimal retrieval situation (i.e., the optimal query), or in other words, whether relevance feedback can be used to improve retrieval effectiveness until a limit is reached. The paper shows that modeling the sequence with a mathematical structure, referred to as a Diophantine set, yields a point on the effectiveness surface that can be interpreted as the optimal retrieval situation. Dominich shows how the point can be computed. The paper, however, does not say how one can use this result to devise, for instance, the optimal query.

The notion of an optimal retrieval situation in relevance feedback is not new, and has been demonstrated empirically. What this paper adds to this already known fact, is that the result can be obtained formally. In other words, the result concerning the effectiveness of the relevance feedback process can be mathematically proven. This methodology follows the line of work of Huibers (1996) and Lalmas (1998), who suggested the use of a meta-theory for IR. The meta-theory developed by Dominich, based on Diophantine structures, looks promising since it can be used to investigate formally properties of the retrieval process, in terms of how they affect the overall effectiveness of IR systems. The use of a meta-theory for IR is important because it could avoid having to run sometimes cumbersome experiments, and also could help designing better experiments.

6 Query reformulation from belief change

It is well known that users have difficulty in expressing their information need. Poorly specified queries lead to imprecise query results which necessitates query reformulation in order to improve precision. In other words, IR is a process which often involves an initial query being reformulated several times until satisfactory precision is achieved in the result set. Note that the query is not reformulated in an arbitrary way. The user is careful in the way in which he/she modifies the query. For this reason, Amati and Bruza (1999) believe that the process of query reformulation can be considered a process of belief change on the part of the user in the query.

The theory of belief change attempts to explain how a rational agent adjusts and assimilates its beliefs about some real or imagined world (Gärdenfors, 1988). A rational agent is assumed to have a partially ordered set K of beliefs. The ordering captures the intuition that the agent holds some beliefs more strongly than others. Beliefs are assumed to be represented by sentences in propositional logic. Due to the dynamics of the agent's environment, an agent may have to expand, revise or contract its beliefs. Belief expansion refers to the process whereby a new sentence is assimilated into the agent's belief system K, together with the logical consequences of the addition (regardless of whether the new belief system is consistent, or not). Belief revision refers to the process whereby the belief system is adjusted in response to a new sentence that is inconsistent with K. To maintain consistency in the resulting belief system, some sentences in K need to be retracted. Belief contraction is a process in which the agent rejects a belief and in order for the resulting belief system to be closed under logical consequence, some sentences from K may have been retracted. The theory of belief change has been heavily studied within the field of artificial intelligence as an underlying theory for agent technology. In addition, the link between belief

change and nonmonotonic reasoning has been bridged allowing belief change to be characterized from a logical perspective.

Amati and Bruza (1999) assume that beliefs are constructed from a set T of terms. They model query expansion, query contraction (a new terminology that refers to the deletion of some term from the query), and query revision (the replacement of some query term with some other term) using the theory of belief change. The connection between belief change theory and nonmonotonic reasoning is discussed using known results about the monotonicity of classical models of IR. This enables Amati and Bruza to provide the logical foundations of query expansion, contraction, and revision, or what they call a ''query reformulation logic'' that is explained in clear terms in the paper. However, although the connection between belief change theory and query reformulation has intuitive appeal, there are problems when adopting the theory as a formalization of the query reformulation process. For example, negation has a different character in query reformulation than it has in belief change theory. More work is needed to adapt Gärdenfors' (1988) formalization and to check whether its results still hold.

It is interesting to note that this proposal is not simply a theoretical exercise, but it is motivated by a study on user query reformulation for Web searching carried out by one of the authors (Bruza and Dennis, 1997). The examples of query reformulation are therefore quite realistic and are particularly useful to enable any ''not-too-logically-minded'' reader to understand the theoretical points of the paper.

It is also interesting to compare the use of belief revision and conditional logic in IR, since the later too is a current topic in IR research (see, for example, Lalmas (1998) and Crestani and van Rijsbergen (1998)). Conditional logic (Nute, 1980) is a completely different way to realize belief revision. It satisfies the Ramsey test and it has been explored by a number of researchers in IR not just for query expansion, but also for the ranking and retrieval of documents. It has been proved that belief revision and conditional logic yield always different results, unless the information to be added is already in the belief set. Belief revision coincides with a simple expansion when the information to be added is compatible, so that the new information A will become part of the belief system K. Conditional logic applies a more complex revision process. It can happen that new information A is entrenched by K ($K > A$) or not ($\neg(K > A)$). When the information A to be added is incompatible or syntactically inconsistent with K, then conditional logic gives the answer $\neg K > A$. In fact, in all closest possible worlds in which K is true it is not possible that A can be also true, so that $\neg(K > A)$ is true. If the set of closest worlds reduces to one, then we strengthen the conclusion to $K > \neg A$. Therefore, in conditional logic there is no way to accept such an A as a new belief. Belief revision instead operates a contraction, and then an expansion of the set K of beliefs to accommodate for A, making it more suitable for the kind of changes in the user beliefs that occur during the IR process.

7 Directed-line segments for information retrieval evaluation

In Boolean systems, searchers often express their queries using Boolean operators according to their own intuition or experience. However, the ''character'' of the system response to the chosen operators is not foreseeable by the searchers, nor may the searchers be aware of the quantitative difference in effect of choosing a particular Boolean operator. Heine (1999) proposes a framework that can be used to illustrate the effect of Boolean operators in visual terms and to provide a simple quantitative measure of the differences in effect of these operators. The framework uses the notion of directed-line segments (and their associated vectors) to characterize Boolean effects in two contexts: the standard precision and recall evaluation, and, more interestingly, as the author claims, using precision and number of informing documents retrieved.

A directed-line segment is defined as the pair $((n_i, p_i), (n_j, p_j))$, where n_i, n_j are the numbers of informing documents retrieved for the Boolean expression E_i and E_j, respectively, and p_i, p_j are the corresponding precision values. Such pairs are defined for every expression E_j derived from expression E_i by adding to it an AND term or its negation, or adding an OR term. Using the directed-line segments, a number of constructions were investigated (e.g., length of the direct-line

segments, the ratio of the means of the direct-line segments) as a way to measure effectiveness: analysing the effect of adding a conjunct, a disjunct, or the negation of a conjunct to an already existing Boolean expression.

Heine (1999) carried out experiments on the MEDLINE collection using the second evaluation context, since as suggested by the author, it is the most general (i.e., one can retrieve all documents). The results of the experiments are promising although not yet conclusive. Further experiments must be carried out before a real understanding of the effect of using Boolean operators can be obtained.

However, the proposed evaluation framework is very interesting for evaluating Boolean retrieval models. It goes beyond the "controversial" recall paradigm, which can only be used with "fixed" test collections and not with "dynamic" collections such as the web. It also attempts to capture the very important interactive nature of the retrieval process, which is often ignored by standard evaluation measures. This approach could be used, for example, to investigate interactive query expansion with Boolean queries. The development of suitable frameworks to evaluate interactive IR systems is an area of ongoing research (see, for example, MIRA, 1995–1998).

The main limitation of the proposed evaluation framework is that it is not obvious how it can be used for non-Boolean retrieval models, for example, to allow comparisons between different query expansion techniques. Nonetheless, the directed-line segment evaluation framework can give us indications on how to improve Boolean retrieval systems, for example, by allowing users to better understand the effect of their Boolean queries.

8 Supervised learning for text passage classification

Bi et al. (1999) describe a method for text passage classification or extraction by means of supervised machine learning and analytically identifying passages. The underlying characteristic of the method lies in the utilization of the resulting classification, which leads to the classification of the portion of a document in a high dimensional feature space into a low dimensional space which is composed of the features drawn from the document itself. The basic assumption of this approach is that words related to the same topic are semantically close to each other. This assumption is strong, but is at the basis of much of the work on document and term clustering (Rasmussen, 1992).

More formally, given a set of n categories and a document, the task is to split the document into smaller segments and map them into the corresponding categories. The proposed approach consists of a two step process for passage classification. The first step takes the Naïve Bayes learning approach to classify documents via ranking of the posterior probabilities of their features (i.e., terms). The second step refines the classified documents to identify passages which are closer to the known categories than others using the likelihood of the features.

As opposed to standard document classification, this approach employs features which are as informative as possible, regardless of the optimum of feature selection, in order to retain the integrity of the document and the context in which the passage resides. The motivation for this approach comes from the complementary advantages of combining statistically inductive and deductive methods. On the one hand, the inductive method offers the advantage that it requires no explicit knowledge, and learns a function or rule for generalization cases based solely on the training data. This results in a given document being classified into one predefined category. On the other hand, the identified category can be used then as explicit and specific knowledge for uncovering a passage which the local vicinity to the features concentrates more on this category than others, without taking into account the training data. This method leads to an approach to using some results derived from a text classifier for further exploring passage recognition.

An apparent disadvantage of this method is that the deduction can be misled when given an incorrect category, however, minimizing the number of errors of the statistical classification guarantees that the accuracy of classification is highest, and the case of no category assignment is ignored.

Another open issue for this approach is defining the boundaries of a passage. The experimentation reported in the paper aimed more at evaluating the accuracy of such an identification rather the

performance of the document classification. To generate a set of candidate passages, the approach proposed by Bi et al. (1999) take the two aspects of the length of words and paragraphs into account. For the first passage, starting at the beginning of a document, the first k words are considered as a candidate one. For the second passage, starting at an increment in l words of the document, a segment with k words is taken as the second one, and so on. Correspondingly, the feature vector of each passage is also formed following the same procedure, and viewed as a linear concatenation of three segments where each of these consists of l features, each segment approximating one paragraph in the document. The association between an identified category and the passage is ranked without considering the number of paragraphs the passage spans, and the passage with highest-rank will be examined as to how many paragraphs are included in this passage. As a result, the length of this passage may vary. Different values for the above parameters were experimentally investigated.

Concluding, the most interesting aspect of the proposed technique is the combination of inductive learning in a high dimensional feature space with the analytical deduction in a relatively low dimensional feature space. It can guarantee that the objects from the training examples are consistent with the objects to be classified, and techniques developed to work for document classification will work, with suitable modifications, for the task of passage classification.

9 Conclusions

The use of logic to model information access enables one to obtain models that are more general than earlier models. Indeed, some logical models are able to represent within a uniform framework various features of IR systems, such as hypermedia links, multimedia content, users knowledge, performance, and so on. It also provides a common approach to the integration of different information access systems. Moreover, logic makes it possible to reason about a model and its properties. This latter possibility is becoming increasingly important since conventional evaluation methods, although good indicators of the effectiveness of implemented systems, often give results which cannot be predicted, or for that matter, satisfactorily explained.

However, logic by itself cannot fully model information access. It is often necessary to take into account the uncertainty inherent in the information access process. Therefore, new models combining the strength of logical and uncertainty modeling are needed for effective information access. This is an active line of research that we are proud to have followed for long time. We think that it is in this context that new and more effective systems for information access will find their basis.

In this paper, we have briefly surveyed some of the trends in this exciting area of research as emerged from the 1999 Workshop on Logical and Uncertainty Models for Information Systems.

Acknowledgements

We would like to thank the members of the program committee and the reviewers for their invaluable help in setting a standard of quality of papers that we hope we will be able to keep in future LUMIS workshops.

References

Amati, G and Bruza, PD, 1999. "A logical approach to query reformulation motivated from belief change." *Proceedings of the Workshop on Logical and Uncertainty Models for Information Systems* London, UK, 36–45.

Bi, Y, Murtagh, F, McClean, S and Anderson, T, 1999. "Text passage classification using supervised learning" *Proceedings of the Workshop on Logical and Uncertainty Models for Information Systems* London, UK, 22–35.

Bosc, P. and Kacprzyk, J (eds.), 1995. *Fuzzyness in Database Management Systems* Physica-Verlag, Heidelberg.

Bruza, PD and Dennis, S, 1997. "Query-reformulation on the internet: empirical data and the hyperindex

search engine'' *Proceedings of the RIAO Conference: Intelligent Text and Image Handling*, Montreal, Canada, 488–499.

Cooper, WS, 1995. ''Some inconsistencies and misnomers in Probabilistic Information Retrieval'' *ACM Transactions on Information Systems* **13**(1) 100–111.

Crestani, F and Lalmas, M (eds.), 1996. *Proceedings of the Second International Workshop on Information Retrieval, Uncertainty and Logics* Glasgow, Scotland. (Available online at: `http://www.dcs.gla.ac.uk/wirul96/`.)

Crestani, F and Lalmas, M (eds.), 1999. *Proceedings of the Workshop on Logical and Uncertainty Models for Information Systems* London, UK. (Available online at: `http://www.dcs.gla.ac.uk/lumis99/`.)

Crestani, F, Lalmas, M and van Rijsbergen, CJ (eds.), 1998a. *Information Retrieval: Uncertainty and Logics*. Kluwer, Norwell, MA.

Crestani, F, Lalmas, M, van Rijsbergen, CJ and Campbell, I, 1998b. ''Is this document relevant? ... probably. A survey of probabilistic models in Information Retrieval'' *ACM Computing Surveys* **30**(4) 528–552.

Crestani, F and Pasi, G (eds.), 2000. *Soft Computing in Information Retrieval: Techniques and applications*. Physica-Verlag, Heidelberg (in press).

Crestani, F and van Rijsbergen, CJ, 1998. ''A study of probability kinematics in information retrieval'' *ACM Transactions on Information Systems* **16**(3) 225–255.

Croft, WB and Thompson, RH, 1987. ''I^3R: a new approach to the design of Document Retrieval Systems'' *Journal of the American Society for Information Science* **38**(6) 389–404.

Dominich, S, 1999. ''A geometrical view of relevance effectiveness in information retrieval'' *Proceedings of the Workshop on Logical and Uncertainty Models for Information Systems* London, UK, 12–21.

Dubois, D, Prade, H and Sedes, F, 1999. ''Some uses of fuzzy logic in multimedia database querying'' *Proceedings of the Workshop on Logical and Uncertainty Models for Information Systems* London, UK, 46–54.

Gärdenfors, P, 1988. *Knowledge in Flux: Modelling the dynamics of epistemic states* MIT Press, Cambridge, MA.

Heine, MH, 1999. ''Measuring the effects of 'and', 'and not', and 'or' operators in document retrieval systems using directed line segments'' *Proceedings of the Workshop on Logical and Uncertainty Models for Information Systems* London, UK, 55–76.

Huibers, TWC, 1996. ''An Axiomatic Theory for Information Retrieval'' *PhD thesis*, Utrecht University, The Netherlands.

Kasabov, N and Kozma, R (eds.), 1998. *Neuro-fuzzy Techniques for Intelligent Information Systems* Physica Verlag, Heidelberg.

Lalmas, M (ed.), 1995. *Proceedings of the First International Workshop on Information Retrieval, Uncertainty and Logics* Glasgow, Scotland.

Lalmas, M, 1998. ''Logical models in Information Retrieval: introduction and overview'' *Information Processing and Management* **34**(1) 19–33.

MIRA, 1995–1998. ''Evaluation framework for interactive multimedia Information Retrieval applications'' *ESPRIT Working Group Number 20039*.

Nie, JY, 1988. ''An outline of a general model for Information Retrieval'' *Proceedings of ACM SIGIR* Grenoble, France, 495–506.

Nute, D, 1980. *Topics in Conditional Logic* Reidel, Dordrecht.

Pasi, G, 1999. ''A logical formulation of the booolean model and of weighted Boolean model'' *Proceedings of the Workshop on Logical and Uncertainty Models for Information Systems* London, UK, 1–11.

Picard, J, 1999. ''Logic as a tool in a term matching information retrieval system'' *Proceedings of the Workshop on Logical and Uncertainty Models for Information Systems* London, UK, 77–90.

Qiu, Y and Frei, HP, 1993. ''Concept based query expansion'' *Proceedings of ACM SIGIR* Pittsburgh, PA, 160–171.

Rasmussen, E, 1992. ''Clustering algorithms'' in WB Frakes and R Baeza-Yates (eds.), *Information Retrieval: Data structures and algorithms* Prentice Hall, Englewood Cliffs, NJ.

Savoy, J, 1994. ''A learning scheme for Information Retrieval in hypertext'' *Information Processing and Management* **30**(4) 515–533.

van Rijsbergen, CJ, 1977. ''A theoretical basis for the use of co-occurence data in Information Retrieval'' *Journal of Documentation* **33**(2) 106–119.

van Rijsbergen, CJ, 1986. ''A non-classical logic for Information Retrieval'' *The Computer Journal* **29**(6) 481–485.

Zadeh, LA, 1987. *Fuzzy Sets and Applications: Selected Papers* Wiley, New York.

Zobel, J and Moffat, A, 1998. ''Exploring the similarity space'' *ACM SIGIR Forum* **35**(1) 18–34.