A new fuzzy logic based information retrieval model *

K. Nowacka
Doctoral Studies
Systems Research Institute
Polish Academy of Sciences
01-447 Warsaw, Poland
ul. Newelska 6
kasia.nowacka@gmail.com

S. Zadrożny
Systems Research Institute
Polish Academy of Sciences
01-447 Warsaw, Poland
ul. Newelska 6
zadrozny@ibspan.waw.pl

J. Kacprzyk
Systems Research Institute
Polish Academy of Sciences
01-447 Warsaw, Poland
ul. Newelska 6
kacprzyk@ibspan.waw.pl

Abstract

We propose a comprehensive model of information retrieval (IR) based on Zadeh's linguistic statements. Its characteristic feature is a capability to take into account both the imprecision and uncertainty pervading the textual information representation. It extends earlier IR models based on broadly meant fuzzy logic. Moreover, some techniques for obtaining quantitative representations of documents and queries are proposed.

Keywords: information retrieval model, fuzzy logic, imprecision, uncertainty.

1 Introduction

Generally, *(textual) information retrieval* deals with broadly meant storage and processing of textual information. A basic task is here the retrieval of those documents from a collection which are *relevant*, i.e., match *information needs* of a user expressed as a *query*. The relevance may be meant as binary, i.e., a document is then regarded as either relevant or irrelevant. Thus the answer to a query is a set of documents considered to be relevant. More generally, a *matching degree* is computed for each document meant as an assessment of its relevance. Then an answer to a query is a list of documents non-increasingly ordered against their matching degree.

The relevance is evaluated by an information retrieval systems (IRS) using some representations of a document and a query. The most popular representation is by using some combinations of *keywords*. The keywords are used both to represent the content of documents in a process of *indexing* and to create a query by the user. The former often proceeds automatically while the latter is done either automatically or "manually". The representations of documents, queries and methods to match them vary leading to different *models of information retrieval* exemplified by the traditional Boolean, vector space and probabilistic ones. All of them have some pros and cons, which implies extensions (cf., e.g., [1]).

We propose a new fuzzy logic based model somehow inspired by all three traditional models. We intend to obtain a comprehensive treatment of imprecision and uncertainty pervading the information retrieval process (being to some extent also the main postulate of the probabilistic model) in the general framework of the Boolean model (with the most powerful querying language) and referring to some methods from the vector space model.

In this paper imprecision and uncertainty are related to how particular keywords are *important* to express the meaning of documents and queries. These information deficiencies require some modelling tools, and should be accounted for while computing the matching degree. We propose a comprehensive model

*This work was partially supported by the Ministry of Science and Higher Education under Grant 3 T11C 052 27

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 1749–1756
Torremolinos (Málaga), June 22–27, 2008

of information retrieval based on the core concepts of fuzzy logic – Zadeh's linguistic statements.

2 Imprecision and Uncertainty and Their Modelling in Information Retrieval Systems

Particular keywords represent the content of a document to a different degree. Usually, the notion of *importance* is used to describe the role of a keyword in this respect but it is not clear how to measure it. It may be conveniently assumed that importance is expressed by a real number from $[0, 1]$.

Realistically we may abandon an artificial quest for *precision* in expressing how *important* a given keyword is for the representation of document meaning or information needs of a user. Moreover there is *uncertainty* related to the assessment of the importance of particular keywords due to, e.g., a user's hesitation while a query is composed or a partial reliability of an expert or algorithm in the indexing process.

Many aspects of imperfect information related to the information retrieval have been studied. The probabilistic model (cf., e.g. [1]) deals primarily with the uncertainty as to the relevance of a document against a query. The classical Boolean model does not provide any means to deal either with imprecision or uncertainty in the representation or matching. In this model a query takes the form of a logical formula and the document is treated as an interpretation, i.e., a valuation of the propositional variables composing the query formula. The importance of keywords in the representation of documents and queries, and the relevance of a document for a query are binary concepts. Among many extensions of the original Boolean model to alleviate this deficiency, one can cite the best known p-norm model by Salton et al. [18], Losada and Barreiro's [13] model in which a distance is defined in the space of interpretations (in the classical propositional calculus) and used to determine the matching between a document and a query, Losada and Barreiro's [14] model taking into account degrees of importance (weights) of keywords while computing the matching degree, etc. More similar to our approach is a possibilistic approach by Liau and Yao [12] who start with a fuzzy similarity relation in the space of documents (i.e., interpretations) – cf. also Losada and Barreiro [13]. Then, each document generates a possibility distribution in the space of interpretations so that a possibility degree is equated with the fuzzy similarity relation membership function value. Then the matching degree between a document and a query is via the pair of possibility and necessity measures of the set of models of the query representing formula.

The first step towards the application of fuzzy logic in IR is to employ multivalued logic instead of the binary one. First, a document is treated as a *fuzzy set* of keywords and the membership of a keyword reflects its importance in representing the meaning of the document; cf., e.g., [6, 10, 15, 11, 9, 2, 4, 16].

The next step is to allow for weights to be associated with keywords in a query. This goes beyond the syntax of the classical, even multivalued, logic and calls for the use of an extended formalism in the spirit of Pavelka's logic. In [22] we proposed a similar extension in the relational database querying framework and later [16] in the context of the IR. An even more important problem is a proper interpretation of weights; cf., e.g. [2, 4]. Basically, 3 most popular interpretations (semantics) of weights in queries are:

- a *relative importance*: if the weight of a keyword in a query is high, then its presence in a document (i.e., high weight there) is required for this document to match (to a high degree) the query,
- an *ideal weight*: the keyword is expected to have in a document a similar weight to that in the query,
- a *threshold*: the keyword is expected to have in a document a weight at least as high as that in the query.

(1)

In [16] we show how both syntactic and semantic aspects of the query weights may be taken into account within the same logical formalism.

In order to provide the extended Boolean model with a capability to handle imprecision, Bordogna and Pasi [3, 5, 9] proposed to treat the importance of the keywords in a *query* as Zadeh's [19] *linguistic variable* postulating the use of *linguistic terms* such as *important*, *very important* etc.

Here we further develop the extended fuzzy Boolean model by:

- assuming linguistic terms as importance weights of keywords also in *documents*,
- considering the linguistic terms in the representation of documents and queries as well as their matching in Zadeh's fuzzy logic,
- discussing the pragmatic aspects of the proposed model.

3 Fuzzy logic in the sense of Zadeh

We assume a standard notation in which a fuzzy set F in the universe U is defined by its *membership function* $\mu_F : U \to [0,1]$, $\mu_F(x) \in [0,1]$, $\forall x \in U$, and $\mu_F(x)$ is meant as the *membership degree* of x to the set F. The family of all fuzzy sets defined in U will be denoted $\mathcal{F}(U)$. The membership degree has its counterpart in the *truth value* in multivalued logic. For example, the statement "John is young", depending on the actual age of John, denoted x, may be treated as *true to a certain degree*, with this truth degree (value) equated with the membership degree $\mu_F(x)$, where F is a fuzzy set modelling the linguistic term *young*.

In the fuzzy logic in the sense of Zadeh a more abstract form of such statements as above is considered. The atomic formula in this logic is:

$$X \text{ IS } A \tag{2}$$

where X denotes a (linguistic) variable and $A \in \mathcal{F}(U)$ a fuzzy set; it is a fuzzy predicate. The truth value of (2) considered under a valuation of the base variable [19] of X depends on the membership function of A.

Then, the truth of (2) is considered when a valuation (knowledge) of the value of X is expressed by another expression of type (2), say X IS B, $B \in \mathcal{F}(U)$. X IS B is assumed to generate a *possibility distribution* [20] $\pi_B : U \to [0,1]$ in the space of possible values of X's base variable, and

$$\pi_B(x) = \mu_B(x) \tag{3}$$

Now the truth of X IS A under the valuation (knowledge) X IS B is expressed with a *fuzzy truth value* τ, i.e., a fuzzy set in [0,1] given by:

$$\mu_\tau(t) = \sup_x \{\mu_A(x) \mid \mu_B(x) = t\} \tag{4}$$

assuming $\sup \emptyset = 0$.

The fuzzy truth values are often replaced in applications by a pair of values of *possibility* and *necessity*, Π_B and N_B, related to the possibility distribution π_B, i.e.,

$$\Pi_B(A) = \sup_{x \in U} \min(\pi_B(x), \mu_A(x)) \tag{5}$$

$$\mathrm{N}_B(A) = \inf_{x \in U} \max(1 - \pi_B(x), \mu_A(x)) \tag{6}$$

Linguistic statements (2) may be combined using logical connectives, for example the conjunction of X_i IS B_i, $i = 1, \ldots, n$, $B_i \in \mathcal{F}(U_i)$, each generating π_{B_i} by (3), generates a joint possibility distribution π on $U = U_1 \times \ldots \times U_n$ such that:

$$\pi(x) = \min(\pi_{B_1}(x_1), \ldots, \pi_{B_n}(x_n)), \quad x = (x_1, \ldots, x_n) \in U \tag{7}$$

assuming the *non-interactiveness* [20] of the particular variables X_i. Then the truth of the conjunction of n linguistic statements X_i IS A_i, $i = 1, \ldots, n$, $A_i \in \mathcal{F}(U_i)$, may be assessed by the pair of measures:

$$\Pi_{B_1 \times \ldots \times B_n}(A_1 \times \ldots \times A_n) = \min(\Pi_{B_1}(A_1), \ldots, \Pi_{B_n}(A_n)) \tag{8}$$

$$\mathrm{N}_{B_1 \times \ldots \times B_n}(A_1 \times \ldots \times A_n) = \min(\mathrm{N}_{B_1}(A_1), \ldots, \mathrm{N}_{B_n}(A_n)) \tag{9}$$

Zadeh introduced also the extended forms of statements (2) [20, 8], i.e., *qualified statements* which will be useful for our purposes.

In particular, *certainty qualified statements* contain a qualifier determining the minimal degree of certainty $\alpha \in [0,1]$ in the truth of statement "X IS A":

$$X \text{ IS } A \text{ is at least } \alpha\text{-certain} \qquad (10)$$

The statement (10) may be identified with a simple, qualifier-free statement "X IS A'" with μ'_A defined as:

$$X \text{ IS } A,\ \alpha \quad \longmapsto \quad X \text{ IS } A' \qquad (11)$$

$$\mu_{A'} = f(\alpha, \mu_A) \qquad (12)$$

where function f may take different forms [8], e.g.:

$$\mu_{A'}(x) = \max(\mu_A(x), 1-\alpha) \qquad (13)$$

4 The model

Our starting point is the classical Boolean model (cf., [1]) and its fuzzy logic based extension, notably due to Bordogna and Pasi [4].

We employ the statements of the type (2) to represent documents and queries. Thus we treat the importance of keywords as linguistic variables and the statement "X_i IS A" is meant as a generic form of the expressions exemplified by: "Keyword t_i is *fairly important* for the representation of the content of the document (query)" so that we can model imprecision concerning the actual importance of the keywords. To also grasp uncertainty, the certainty qualified statements (10) are employed.

Document representation Each document is represented as a compound linguistic statement built of:

$$X_i \text{ IS } B_j \text{ is at least } \alpha\text{-certain} \qquad (14)$$

where X_i is a linguistic variable corresponding to the importance in the document of the keyword t_i, and B_j is a linguistic term such as "very important", "important to a degree around 0.6", "fairly important" etc., while $\alpha \in [0,1]$ is a certainty degree as to the importance of the keyword. Particular linguistic terms are modelled by fuzzy sets defined in the interval $[0,1]$ which is used as the range of importance degrees. During computations the statements (14) are transformed to a qualifier-free form using (11), in particular (13). Thus each such a statement generates a possibility distribution π_{B_j} on the space of importance degrees of given keyword, according to (3).

The conjunction of statements (14) will be a typical form of the representation of a document. Then, using (7) a joint possibility distribution on the Cartesian product $[0,1]^n$ is determined, where n is the number of keywords under consideration. The document may be then treated as a multidimensional possibility distribution:

$$\pi_D(x_1,\ldots,x_n) = \min(\mu_{B_1}(x_1),\ldots,\mu_{B_n}(x_n)) \qquad (15)$$

assuming the non-interactiveness of the variables corresponding to the importances of keywords.

Query representation To preserve the syntactical homogeneity of the representation of documents and queries, like in the classical Boolean model, a query is also represented as a compound linguistic statement built of the statements (14). Thus a linguistic assessment of selected keywords importance is given, each accompanied by a cetainty degree. The statements (14) are then transformed to a qualifier-free form via (13). The entire query is treated as a fuzzy set Q in a multidimensional space $[0,1]^n$ such that:

$$\mu_Q(x_1,\ldots,x_n) = \min(\mu_{A_1}(x_1),\ldots,\mu_{A_n}(x_n)) \qquad (16)$$

Evaluation of the relevance - the computation of the matching degree To evaluate the relevance of a document against a query, we compute – as in the classic Boolean model – the truth degree of the statement q representing the query under the assumption that the statement d representing the document is true. We use the pair of necessity and possibility values (5)–(6), and obtain:

$$(\mathrm{N}_D(Q), \Pi_D(Q)) \qquad (17)$$

The answer to a query is a list of documents lexicographically ordered on these pairs.

Thus, in our model the matching degree expresses the possibility and necessity of matching between a document and a query.

5 Pragmatic aspects of the model

The question of the actual form of (14) used to represent documents and queries goes beyond the proposed model. However we have experimented with some possible ways of determining them on two levels of abstraction. Namely we have made some tests with: the general shapes (templates) of the membership functions of the linguistic terms appearing in (14) and the ways of automatic determining concrete forms of these membership functions during the indexing process. Here we will only discuss the former one. Moreover, since "min" in (7) does not lead to satisfactory results, we have also tested some alternatives.

Basically, to effectively and efficiently implement the approach proposed, a proper user interface is needed. For lack of space it will not be discussed here.

As to some numerical experience, we were testing the following pairs of the shapes for the membership functions of the linguistic terms in documents and queries.

Variant A

document

$$\mu_B(x) = 1 - k \mid x - d_0 \mid \qquad (18)$$

query

$$\mu_A(x) = \begin{cases} 1 - q_0 & x \leq 1 - q_0 \\ x & x > 1 - q_0 \end{cases} \qquad (19)$$

Variant B

document

$$\mu_B(x) = 1 - k \mid x - d_0 \mid \qquad (20)$$

query

$$\mu_A(x) = 1 - k \mid x - q_0 \mid \qquad (21)$$

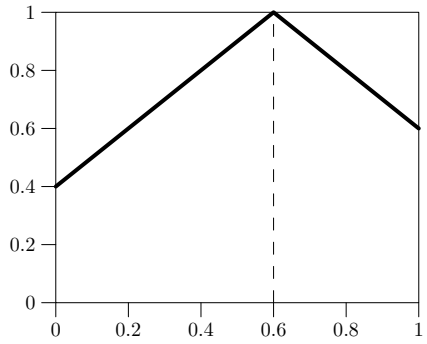

Figure 1: A membership function representing the linguistic term: "*Important to a degree around* d_0 "; d_0=0.6; cf. eq. (18), $k = 1$

Variant C

document

$$\mu_B(x) = \begin{cases} 1 - d_0 & x = 0 \\ 1 & x = 1 \end{cases} \qquad (22)$$

query

$$\mu_A(x) = \begin{cases} 1 - q_0 & x = 0 \\ 1 & x = 1 \end{cases} \qquad (23)$$

with (18), (19) and (22) shown in Figs. 1, 3 and 4, respectively. For all, their characteristic points are denoted by d_0 and q_0 in case of documents and queries, respectively. These characteristic points were computed for a given keyword in a document/query using standard keyword weighting schemes of the vector space IR model [17].

Each of these membership functions should be treated as a *template* which is instantiated by the user using the IRS interface or automatically during the indexing. Namely, the template given by (18) represents the linguistic term "important to a degree around d_0". Besides d_0 there is an additional parameter k which determines how "narrow" or "wide" the membership function is around d_0. In Fig. 1 this template is illustrated with $d_0 = 0.6$ and $k = 1$; this may be described in the user interface as "somewhat important". Figure 2 illustrates this template when used in a certainty qualified linguistic statement (cf. (10)) with $\alpha = 0.8$; $d_0 = 0.6$ and $k = 8$.

The template given by (19) represents "important with certainty at least q_0". Its under-

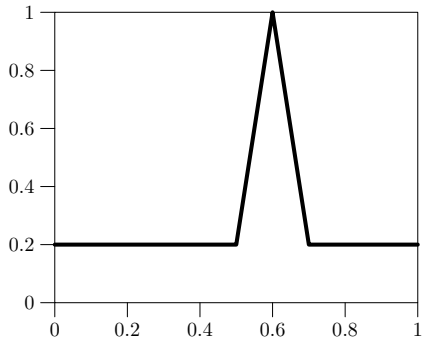

Figure 2: A membership function representing: "*Important to a degree around* d_0*"* with certainty at least α"; d_0=0.6, α=0.8

lying membership function $\mu_A(x) = x$ corresponds to a general meaning of "important": the higher the degree of importance (i.e., x) the more compatible it is with this term. The template (19) combines the interpretation of importance with uncertainty as to the actual importance of the keyword so that it also fits the scheme of the certainty qualified linguistic statements (10) and in Fig. 3 is exemplified for $\alpha = q_0 = 0.6$.

Finally, the template given by (22) is a simplified version of the previous one with the importance of a keyword treated as a binary concept but again the uncertainty as to whether the keyword is actually important is quantified with a number from $[0, 1]$; cf. Fig. 4 for $\alpha = d_0 = 0.8$.

Now, for variants A–C defined by (18)–(23) built of the above mentioned templates, A is best suited for a typical retrieval scenario when documents are indexed automatically and queries are composed manually by the user. Then the representation of documents is determined by, e.g., a weighting schemes using the frequency of terms, inverted document frequency, etc. [17]. The user is assumed to only select keywords that are important to him or her, but possibly hesitating as to their sure importance to a degree q_0 (weight).

Let us consider a query being a conjunction of n linguistic statements X_i IS A_i, where A_i's are represented by (19). Then, as q_0 tends to 0, the membership function μ_{A_i} tends to 1 for any $x \in U_i$. Thus, the possibility and necessity measures (5)–(6) tend to 1 too (A in this formula is approaching the crisp set comprising the whole interval), and further the keyword t_i gets a lower and lower influence on the matching of the query and a document, cf. (8)–(9). Thus the interpretation of the linguistic terms in a query given by (19) is here in the spirit of the relative importance weights semantics, cf. (1). Also when "min" in (8)–(9) is replaced with another aggregation operator (as suggested by our computational experiments), then still this semantics is to some extent preserved. For example, when the average is used instead of the minimum, then a given keyword in a query contributes to the matching degree of all documents to more or less the same extent (the same for $q_0 = 0$) because the matching of *all* documents with respect to this keyword (both in terms of possibility and necessity) is very high (even 1). As we are primarily concerned with the ordering of documents according to their matching degrees, the influence of such a keyword is very limited (or none).

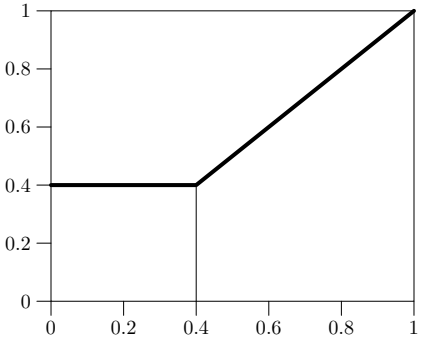

Figure 3: A membership function representing: "*Important with certainty at least* q_0*"*; q_0=0.6

Variant B employs the same templates (20)–(21) to represent both the documents and queries. This is best suited for automatic indexing of both the documents and queries. Many standard test document collections contain queries as short texts, which may be indexed like documents to obtain some representation. The possibility measure of matching, which is used here alone, is the higher the more similar are the shapes of instances of the templates (20) and (21), i.e., the closer are the

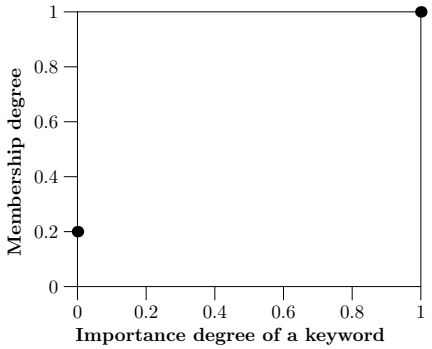

Figure 4: A membership function representing: "*Completely important with certainty at least* α"; α=0.8

values of d_0 and q_0 therein. Thus in this variant the interpretation of the linguistic terms in a query is in the spirit of the ideal weights semantics, cf. (1).

Variant C is a generalization of the classical Boolean model. It preserves the original interpretation of binary importance but allows to express some reservation as to the decision on the importance of a given keyword or its lack. In this variant the formulas for the matching degree are a little bit simpler (i.e., specific forms of (5) – (6) obtained using (22) – (23)) than in case of other variants, slightly reducing the computational costs. Moreover this variant may be directly cast within possibilistic logic [7] which makes a further analysis and enhancements easier (cf. [21]).

As mentioned earlier, in general the queries in our model, like in the classical Boolean model, are composed of linguistic statements of type (10) with the use of logical connectives. The conjunction and disjunction are processed according to the general rules of the possibility theory – cf. comments under (5)–(6). A separate discussion is needed in case of negation. In the classical Boolean model the situation is fairly obvious: if a keyword is negated in a query then it should not appear in the matching documents. In the calculus briefly presented in section 3 the negation of a linguistic statement should be understood in the following way:

$$\neg X \text{ IS } A \longmapsto X \text{ IS } \overline{A} \qquad (24)$$

where $\overline{A}$ denotes a complement of a fuzzy set A. The templates (19) and (23) may be easily adopted for this interpretation. First of all the linguistic term "unimportant" is represented in case of the template (19) by the the membership function:

$$\mu_A(x) = -x + 1$$

and by:

$$\mu_A(x) = \begin{cases} 1 & \text{for } x = 0 \\ 0 & \text{for } x = 1 \end{cases}$$

in case of the template (23). Then using (13) one obtains the formulas for the negated counterparts of the templates (19) and (23), i.e. for $\neg X$ IS A, α

$$\mu_A(x) = \begin{cases} 1 - \alpha & x > \alpha \\ -x + 1 & x \leq \alpha \end{cases}$$

and

$$\mu_A(x) = \begin{cases} 1 & x = 0 \\ 1 - \alpha & x = 1 \end{cases}$$

In the above formulas we denote the certainty level with α rather than with q_0 as these queries are rather meant to be constructed manually while q_0 refers primarily to the result of an automatic indexing of a short document playing the role of a query.

In case of template (19) it is not clear what its negated version should mean. Neither a complement nor an antonym of a fuzzy set seems to provide a reasonable semantics. We leave it as an open question if the negated version of this template makes sense and should be included somehow in the model. In fact this template is primarily meant as to be used in the automatic indexing of documents and queries. From this point of view its negated version is of a lesser importance.

6 Concluding remarks

We presented a new fuzzy logic based information retrieval model to directly represent imprecision and uncertainty of the IR processes within the formal framework of fuzzy logic. Pragmatic aspects of the proposed model are discussed. Three templates for the representation of keyword importance are proposed. The

results of the computational experiments on standard test collections with various weighting schemes and aggregation operators will be presented during the conference. A detailed description and discussion of the model will be given in a forthcoming journal paper.

References

[1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern information retrieval*. ACM Press and Addison Wesley, 1999.

[2] G. Bordogna, P. Carrara, and G. Pasi. Query term weights as constraints in fuzzy information retrieval. *Information Processing & Management*, 27:15–26, 1991.

[3] G. Bordogna, P. Carrara, and G. Pasi. Extending Boolean information retrieval: a fuzzy model based on linguistic variables. In *First IEEE Int. Conf. on Fuzzy Systems*, pages 769–776, San Diego, CA, USA, 1992.

[4] G. Bordogna, P. Carrara, and G. Pasi. Fuzzy approaches to extend Boolean information retrieval. In P. Bosc and J. Kacprzyk, editors, *Fuzziness in Database Management Systems*, pages 231–274. Physica Verlag, Heidelberg, 1995.

[5] G. Bordogna and G. Pasi. A fuzzy linguistic approach generalizing Boolean information retrieval: A model and its evaluation. *Journal of the American Society for Information Science*, 44(2):70–82, 1993.

[6] D. Buell D. and D.H. Kraft. Threshold values and Boolean retrieval systems. *Information Processing & Management*, 17:127–136, 1981.

[7] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In Gabbay D.M. et al., editor, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3, pages 439–513. Oxford University Press, Oxford, UK, 1994.

[8] D. Dubois and H. Prade. Fuzzy sets in approximate reasoning, part 1: Inference with possibility distributions. *Fuzzy Sets and Systems*, 40:143–202, 1991.

[9] D.H. Kraft, G. Bordogna, and G. Pasi. An extended fuzzy linguistic approach to generalize Boolean information retrieval. *Journal of Information Sciences*, 2(3):119–134, 1994.

[10] D.H. Kraft and D.A. Buell. Fuzzy sets and generalized Boolean retrieval systems. *International Journal on Man-Machine Studies*, 19:45–56, 1983.

[11] D.H. Kraft and D.A. Buell. Fuzzy sets and generalized Boolean retrieval systems. In D. Dubois D., H. Prade, and R.R. Yager, editors, *Readings in Fuzzy Sets for Intelligent Systems*. Morgan Kaufmann Publishers, San Mateo, 1992.

[12] Ch.-J. Liau and Y.Y. Yao. Information retrieval by possibilistic reasoning. In *LNCS 2113*, LNCS, pages 52–61. Springer, 2001.

[13] D. E. Losada and A. Barreiro. A logical model for information retrieval based on propositional logic and belief revision. *The Computer Journal*, 44(5):410–424, 2001.

[14] D. E. Losada and A. Barreiro. Embedding term similarity and inverse document frequency into a logical model of information retrieval. *JASIST*, 54(4):285–301, 2003.

[15] T. Radecki. Fuzzy set theoretical approach to document retrieval. *Information Processing and Management*, 15(5):247–260, 1979.

[16] S. Zadrożny S. and J. Kacprzyk. An extended fuzzy boolean model of information retrieval revisited. In *Proc. of FUZZ-IEEE 2005*, pages 1020–1025, Reno, NV, USA, May 22-25, 2005. IEEE.

[17] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24:513–523, 1988.

[18] G. Salton, E.A. Fox, and H. Wu. Extended Boolean information retrieval. *Communications of ACM*, 26(11):1022–1036, 1983.

[19] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. Part I-III. *Information Sciences*, 8,8,9:199–249,301–357,43–80, 1975.

[20] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

[21] S. Zadrożny, K. Nowacka, and J. Kacprzyk J. A concept of a possibilistic logic based information retrieval model. In *11th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 992–999, Paris, France, 2006.

[22] S. Zadrożny and J. Kacprzyk. Fuzzy querying of relational databases: a fuzzy logic view. In *EUROFUSE Workshop on Information Systems*, pages 153–158, Villa Monastero, Varenna, Italy, 2002.