Information processing & Management Vol. 15, pp. 247–259
Pergamon Press Ltd., 1979. Printed in Great Britain

FUZZY SET THEORETICAL APPROACH TO DOCUMENT RETRIEVAL†

TADEUSZ RADECKI
Wrocław Technical University, Main Library and Scientific Information Centre, Wybrzeże St. Wyspiańskiego 27, 50–370 Wrocław, Poland

(Received 7 May 1979)

Abstract—The aim of a document retrieval system is to issue documents which contain the information needed by a given user of an information system. The process of retrieving documents in response to a given query is carried out by means of the search patterns of these documents and the query. It is thus clear that the quality of this process, i.e. the pertinence of the information system response to the information need of a given user depends on the degree of accuracy in which document and query contents are represented by their search patterns. It seems obvious that the weighting of descriptors entering document search patterns improves the quality of the document retrieval process.

A mathematical apparatus which takes into consideration, in a natural manner, the fact that the grades of importance of the descriptors in document search patterns are of the continuum type, that is an apparatus adequate to the description of a retrieval system of documents indexed by weighted descriptors is—among known mathematical methods—the theory of fuzzy sets, formulated by L. A. Zadeh.

It is the aim of this paper to present a new method of document retrieval based on the fundamental operations of the fuzzy set theory. We start by introducing basic notions, then the syntax and semantics of the proposed language for document retrieval will be given and an algorithm allocating documents to particular queries will be described and its properties discussed.

The basic advantage of the use of the fuzzy set theory for document retrieval system description is that it takes into consideration, in a simple way, the differentiation of the importance of descriptors in document search patterns and the differentiation of the formal relevance grades of particular documents of an information system to a given query. Documents of the highest grades (in the given information system) of formal relevance to the given query may be retrieved by means of the application of simple operations of the fuzzy set theory.

1. INTRODUCTION

The aim of a document retrieval system is the issue of documents which contain information required by a given user of the system. The process of document retrieval is such that in response to a given query those documents are issued whose search patterns correspond best to the query search pattern, according to a given criterion. Most usually document search patterns are sets of index-terms representing the ideas contained in the subject-matter of the documents, or sets of index-terms with numerical weights assigned to the terms according to their importance. Sometimes the search patterns of queries are created in a similar way but more usually they are constructed by using the Boolean operators AND, OR and NOT to connect the index-terms of queries. Often while creating the search patterns of documents and queries one uses a thesaurus which is a set of terms on which specific kinds of relations are defined, e.g. the relation of synonymity, the relation of hierarchy, the relation of affinity. It is intuitively obvious that the weights assigned to index-terms and the use of a thesaurus during the creation of search patterns of documents and queries improve this process as well as the effectiveness of the document retrieval process. This has also been confirmed experimentally (see e.g.[1]).

† A preliminary version of this paper was presented at the Sixth Cranfield International Conference on Mechanised Information Storage and Retrieval Systems (26–29 July 1977).

In this connection a matter of vital importance is the use, for developing document retrieval strategies, of a mathematical method which will allow for the differentiation of descriptor importance in document and query search patterns. The desired mathematical method should also take into account that the information system documents should ideally have varying estimations of the pertinence, relevance and formal relevence grades and also the fact that semantic relations defined on a set of index-terms representing ideas in a given field of knowledge do not, as a rule, have sharp boundaries of membership. Generally speaking, a mathematical method for describing the document retrieval systems should be able to take into account the differentiation of the importance of a given feature in particular elements in the universe of the objects considered.

However, in literature these postulates are not normally considered and most often set theory[2–11], probability mathematics and theory of statistics (see, e.g. [8]) and other mathematical theories (see e.g.[8]) are used to describe document retrieval systems. Of the widely known retrieval methods of documents indexed by weighted descriptors, the most adequate seem to be the methods of associative retrieval (see e.g.[1, 7, 8, 10]).

Of the known mathematical methods, the method best fulfilling the postulates formulated above, and therefore adequate for an analysis of document retrieval systems is the theory of fuzzy sets, whose bases L. A. Zadeh has given in[12–15]. The idea of the theory of fuzzy sets is that the grades of membership of particular elements of the universe in a given fuzzy set are determined by the so-called membership function which is a generalization of the characteristic function. The transition from membership to non-membership of the universe elements in the fuzzy set, in contrast to the ordinary set theory, is continuous.

Many papers have already been written on investigations into the possibility of creating a uniform document retrieval system theory based on the theory of fuzzy sets. Besides the present author[16–20] many other specialists have also dealt with this question[21–27]. In paper[21] C. V. Negoita used the theorem on the separation of fuzzy sets[12] to divide a set of document search patterns into clusters where each cluster is made up of those document search patterns whose grades of membership in that particular cluster are not smaller than the established threshold value. The idea of using the theory of fuzzy sets, or, to be precise, the concept of the similarity relation to formulate an algorithm for the division of a set of document search patterns into clusters has also been used in[18, 22, 25]. In[22, 25]—making direct use of the definition of the max–min composition[13] of fuzzy relations— a way is suggested of dividing the set of document search patterns into clusters where each cluster is made up of those document search patterns whose grades of similarity are not smaller than the established threshold value. One disadvantage of the way of organizing the document file, suggested in these papers, is that in the case of a large set of documents the process of dividing the set of document search patterns into clusters is very time-consuming and also expensive. This inconvenience can be significantly attenuated by using the method of organizing the document file proposed in paper[18] based on the notion of the maximum spanning tree. In paper[23] as in paper[24] C. V. Negoita defines the response of an information retrieval system as a fuzzy set and describes the ralationships between various responses of the system in terms of the theory of fuzzy sets. Retrieval methods of documents indexed by weighted descriptors, which are a natural generalization of the set theory methods, have been described in papers[16, 17, 19, 20] by the author. In paper[26] W. M. Sachs draws attention to the possibility of defining associative retrieval in terms of the fuzzy set theory, but does not provide any new solutions however. On the other hand, paper[27] by V. Tahani, based on an idea similar to that expressed by the author in paper[16], contains a description of the organization of document file and a strategy for the retrieval of documents using basic notions and operations of the theory of fuzzy sets.

The aim of the present paper is to describe a generalized method (in comparison to the strategies presented in papers[16, 27]) of document retrieval. In the writing of this paper ideas contained in previous papers by the author[16, 17, 19, 20, 28] were utilized. Before entering a detailed description of the proposed method of document retrieval, we will present the basic notions used in the rest of the paper. We will then describe the proposed document retrieval language and present an algorithm for the allocation of documents to particular queries and describe the properties of the language and the algorithm. The proposed document retrieval

strategy will also be illustrated by an example. Finally the results of the present paper will be summarized and modifications to the document retrieval method presented will be discussed.

2. BASIC NOTIONS

Let us begin by defining a document retrieval system. By this, we understand the quadruple:

$$I = \langle D, Q, T, \psi \rangle,$$

where D is a set of documents of an information system of the cardinality $|D| = n$, Q is a set of queries directed towards the information system of the cardinality $|Q| = m$, T is a set of descriptors of the cardinality $|T| = k$, whereas ψ is a maping in the form:

$$\psi: Q \rightarrow 2^D.$$

We will call the mapping ψ the algorithm for the assignment of documents to particular queries. For a given query $q \in Q$ we shall call the set $\psi(q) \subset D$ the response of the information system. On the other hand, the response documents issued to a given query $q \in Q$ in the order of decreasing suitability of that query's search pattern to the search patterns of documents $d \in D$ will be symbolized by Ord $\psi(q)$.

Let us assume that we know relation F of the document description which is a binary fuzzy relation in the form:

$$F = \{\langle d, t, \mu_F(d, t)\rangle | d \in D, t \in T\},$$

where μ_F: $D \times T \rightarrow [0, 1]$ is a function specifying, for each ordered pair $\langle d, t\rangle$; $d \in D$, $t \in T$, the importance of the descriptor $t \in T$ in the description of the document $d \in D$. On the basis of the binary fuzzy relation F in the Cartesian product $D \times T$, written as $F \overset{f}{\subset} D \times T$, we can define the search pattern of document $d \in D$, i.e. its description or representative.

The search pattern of a given document $d \in D$ is the fuzzy set F_d in the set T of the descriptors, written as $F_d \overset{f}{\subset} T$, is the form:

$$F_d = \{\langle t, \mu_{F_d}(t) = \mu_F(d, t)\rangle | t \in T\}.$$

In the case of a large set D of documents in an information system, a matter of vital importance is to formulate such an algorithm for allocating documents to particular queries that the time taken for the information system to respond to these queries is acceptable. In our case optimalization of the retrieval process of documents can be carried out by the optimal choice of the value of λ^* with regard to the time and the quality of the retrieval process, and then by operating on λ^*-level document search patterns.

By the λ^*-level search pattern of a given document we will understand the λ^*-level fuzzy set $F_{d(\lambda^*)}$ defined as follows:

$$F_{d(\lambda^*)} = \{\langle t, \mu_{F_{d(\lambda^*)}}(t) = \mu_{F_d}(t)\rangle | t \in F_d(\lambda^*)\},$$

where $F_d(\lambda^*)$ is the λ^*-level set:

$$F_d(\lambda^*) = \{t | \mu_{F_d}(t) \geq \lambda^*, t \in T\}.$$

Of course, for $\lambda^* = 0$, $F_{d(\lambda^*)} = F_d$.

3. DOCUMENT RETRIEVAL LANGUAGE

The set T^* of correct expressions of a document retrieval language is defined as the smallest set containing the set T of descriptors and such that if $t', t'' \in T^*$ then $t' \wedge t''$, $t' \vee t''$, $]t'$ and $]t''$ also belong to the set T^* of correct expressions, where the symbols $\wedge$, $\vee$ and $]$ represent the

Boolean operators of conjunction, disjunction and negation, respectively. The elements of the set T^* will be called complex descriptors of the document retrieval language. In our case, however, the meanings of complex descriptors are fuzzy sets which are defined on the basis of the relation F of the description of the information system's documents $d \in D$. Bearing in mind the time taken to retrieve documents in response to particular queries, we will define the meanings of the complex descriptors $t \in T^*$ on the basis of the λ^*-level relation F_{λ^*} of the description of the documents $d \in D$:

$$F_{\lambda^*} = \{\langle d, t, \mu_{F_{\lambda^*}}(d, t) = \mu_F(d, t)\rangle | \langle d, t\rangle \in F(\lambda^*)\},$$

where $F(\lambda^*)$ is the λ^*-level set:

$$F(\lambda^*) = \{\langle d, t\rangle | \mu_F(d, t) \geqslant \lambda^*\}.$$

Of course, the λ^*-level relation F_{λ^*} of the description of the documents $d \in D$ is the union of the λ^*-level search patterns of the documents $d \in D$, i.e. $F_{\lambda^*} = \bigcup_{d \in D} F_{d(\lambda^*)}$.

Below we define the λ^*-level meanings $M_{t(\lambda^*)}$ of the correct expressions $t \in T^*$ of our document retrieval language.

Namely, the λ^*-level meaning $M_{t(\lambda^*)}$ of descriptor $t \in T \subset T^*$ in the document retrieval system I is defined as follows:

$$M_{t(\lambda^*)} = \{\langle d, \mu_{M_{t(\lambda^*)}}(d) = \mu_{F_{t(\lambda^*)}}(d)\rangle | d \in F_t(\lambda^*)\}.$$

Therefore, $M_{t(\lambda^*)}$ is a λ^*-level fuzzy set which equals $F_{t(\lambda^*)}$, i.e. $M_{t(\lambda^*)} = F_{t(\lambda^*)}$.

If $t' \in T^*$, then the λ^*-level meaning $M_{t(\lambda^*)}$ of a complex descriptor $t = \rceil t'$ in the document retrieval system I is defined as follows:

$$M_{t(\lambda^*)} = \{\langle d, \mu_{M_{t(\lambda^*)}}(d) = 1 - \mu_{M_{t'(\lambda^*)}}(d)\rangle | 1 - \mu_{M_{t'(\lambda^*)}}(d) \geqslant \lambda^*, d \in D\}.$$

In accordance with fuzzy set theory terminology the λ^*-level fuzzy set $M_{t(\lambda^*)}$ is the complement of the λ^*-level fuzzy set $M_{t'(\lambda^*)}$, i.e. $M_{t(\lambda^*)} = \overset{f}{\rceil} M_{t'(\lambda^*)}$.

If $t', t'' \in T^*$, then the λ^*-level meaning $M_{t(\lambda^*)}$ of a complex descriptor $t = t' \wedge t''$ in the document retrieval system I is defined as follows:

$$M_{t(\lambda^*)} = \{\langle d, \mu_{M_{t(\lambda^*)}}(d) = \min[\mu_{M_{t'(\lambda^*)}}(d), \mu_{M_{t''(\lambda^*)}}(d)]\rangle | d \in M_{t'}(\lambda^*) \cap M_{t''}(\lambda^*)\},$$

where $M_t(\lambda^*)$ is a λ^*-level set, i.e. $M_t(\lambda^*) = \{d | \mu_{M_t}(d) \geqslant \lambda^*, d \in D\}$.

And so, the λ^*-level fuzzy set $M_{t(\lambda^*)}$ being the λ^*-level meaning of the complex descriptor $t = t' \wedge t''$, where $t', t'' \in T^*$, in the document retrieval system I, is the intersection of the λ^*-level fuzzy sets $M_{t'(\lambda^*)}$ and $M_{t''(\lambda^*)}$, i.e. $M_{t(\lambda^*)} = M_{t'(\lambda^*)} \overset{f}{\cap} M_{t''(\lambda^*)}$.

If $t', t'' \in T^*$, then the λ^*-level meaning $M_{t(\lambda^*)}$ of a complex descriptor $t = t' \vee t''$ in the document retrieval system I is defined as a λ^*-level fuzzy set in the form:

$$M_{t(\lambda^*)} = \{\langle d, \mu_{M_{t(\lambda^*)}}(d) = \max[\mu_{M_{t'(\lambda^*)}}(d), \mu_{M_{t''(\lambda^*)}}(d)]\rangle | d \in M_{t'}(\lambda^*) \cup M_{t''}(\lambda^*)\}.$$

Consequently, the λ^*-level fuzzy set $M_{t(\lambda^*)}$ being the λ^*-level meaning of the complex descriptor $t = t' \vee t''$, where $t', t'' \in T^*$, in the document retrieval system I, is the union of the λ^*-level fuzzy sets $M_{t'(\lambda^*)}$ and $M_{t''(\lambda^*)}$, i.e. $M_{t(\lambda^*)} = M_{t'(\lambda^*)} \overset{f}{\cup} M_{t''(\lambda^*)}$.

The λ^*-level meaning $M_{t(\lambda^*)}$ $(\lambda^* > 0)$ of a complex descriptor $t \in T^*$ in the document retrieval system I is empty, written as $M_{t(\lambda^*)} = \overset{f}{\square}_{\lambda^*}$, iff $M_t(\lambda^*) = \phi$.

$$(\forall t \in T^*)(M_{t(\lambda^*)} = \overset{f}{\square}_{\lambda^*}) \quad \Leftrightarrow \quad (\forall d \in D)(\mu_{M_t}(d) \leq \lambda^*),$$

where $M_t = M_{t(0)}$.

We will say that the λ^*-level meaning $M_{t(\lambda^*)}$ $(\lambda^* > 0)$ of a complex descriptor $t \in T^*$ in the document retrieval system I is full, written as $M_{t(\lambda^*)} = \overset{f}{\square}_{\lambda^*}$, iff

$$M_{t(\lambda^*)} = \{(d, \mu_{M_{t(\lambda^*)}}(d) = 1) | d \in M_{t(\lambda^*)}\}.$$

In the set T^* of complex descriptors, we can define the following relations. The relation $E^I_{\lambda^*} \subset T^* \times T^*$ of a λ^*-level indistinguishability of complex descriptors $t \in T^*$ in the document retrieval system I is defined as follows:

$$(\forall t', t'' \in T^*)(t' E^I_{\lambda^*} t'' \Leftrightarrow M_{t'(\lambda^*)} = M_{t''(\lambda^*)}).$$

So complex descriptors $t', t'' \in T^*$ are indistinguishable in the domument retrieval system I, written as $t' E^I_{\lambda^*} t''$, iff their λ^*-level meanings in I are equal to each other. From the definition of the relation $E^I_{\lambda^*}$ it follows that it is an equivalence relation.

On the other hand, the relation $G^I_{\lambda^*} \subset T^* \times T^*$ of a λ^*-level generalization of complex descriptors $t \in T^*$ in the document retrieval system I is defined as follows:

$$(\forall t', t'' \in T^*)(t' G^I_{\lambda^*} t'' \Leftrightarrow M_{t''(\lambda^*)} \overset{f}{\subset} M_{t'(\lambda^*)}),$$

where $M_{t''(\lambda^*)} \overset{f}{\subset} M_{t'(\lambda^*)}$ means that the λ^*-level fuzzy set $M_{t''(\lambda^*)}$ being the λ^*-level meaning of the complex descriptor $t'' \in T^*$ is the subset of the λ^*-level fuzzy set $M_{t'(\lambda^*)}$, being the λ^*-level meaning of the complex descriptor $t' \in T^*$, i.e.

$$(M_{t''(\lambda^*)} \overset{f}{\subset} M_{t'(\lambda^*)}) \quad \Leftrightarrow \quad (\forall d \in M_{t''}(\lambda^*))(\mu_{M_{t''(\lambda^*)}}(d) \leq \mu_{M_{t'(\lambda^*)}}(d)).$$

It directly follows from the definition of the relation $G^I_{\lambda^*}$ that it is a relation of partial order.

Proposition 1. In the document retrieval system I the following relationships hold:

(1) $(\forall t', t'' \in T^*)(\forall \lambda \in [\lambda^*, 1])(t' G^I_{\lambda^*} t'' \Rightarrow M_{t''(\lambda^*)}(\lambda) \subset M_{t'(\lambda^*)}(\lambda))$,

where $M_{t(\lambda^*)}(\lambda)$ is the λ-level set, i.e.

$$M_{t(\lambda^*)}(\lambda) = \{d | \mu_{M_{t(\lambda^*)}}(d) \geq \lambda, d \in M_t(\lambda^*)\}, \lambda \in [\lambda^*, 1].$$

(2) $(\forall t', t'' \in T^*)(\forall \lambda \in [\lambda^*, 1])(t' G^I_{\lambda^*} t'' \Rightarrow M_{(t' \wedge t'')(\lambda^*)}(\lambda) = M_{t'(\lambda^*)}(\lambda))$.

(3) $(\forall t', t'' \in T^*)(\forall \lambda \in [\lambda^*, 1])(t' G^I_{\lambda^*} t'' \Rightarrow M_{(t' \vee t'')(\lambda^*)}(\lambda) = M_{t''(\lambda^*)}(\lambda))$.

(4) $(\forall t', t'' \in T^*)(\forall \lambda \in [\lambda^*, 1])(t' G^I_{\lambda^*} t'' \Rightarrow M_{(\neg t' \wedge t'')(\lambda^*)}(\lambda) = M_{\neg t'(\lambda^*)}(\lambda))$.

(5) $(\forall t', t'' \in T^*)(\forall \lambda \in [\lambda^*, 1])(t' G^I_{\lambda^*} t'' \Rightarrow M_{(\neg t' \vee t'')(\lambda^*)}(\lambda) = M_{\neg t'(\lambda^*)}(\lambda))$.

Proof. The proof results from the definition of the relation $G^I_{\lambda^*}$ and from the definition of the λ-level set $M_{t(\lambda^*)}(\lambda)$.

Proposition 2. The operations $\overset{f}{\cap}$, $\overset{f}{\cup}$ and $\overset{f}{\rceil}$ applied to the λ^*-level meanings $M_{t(\lambda^*)}$ of the complex descriptors $t \in T^*$ in the document retrieval system I satisfy the following laws:

(1) Idempotency, i.e.

$$M_{t(\lambda^*)} \overset{f}{\cap} M_{t(\lambda^*)} = M_{t(\lambda^*)},$$

$$M_{t(\lambda^*)} \overset{f}{\cup} M_{t(\lambda^*)} = M_{t(\lambda^*)}.$$

(2) Commutativity, i.e.

$$M_{t'(\lambda^*)} \overset{f}{\cap} M_{t''(\lambda^*)} = M_{t''(\lambda^*)} \overset{f}{\cap} M_{t'(\lambda^*)},$$

$$M_{t'(\lambda^*)} \overset{f}{\cup} M_{t''(\lambda^*)} = M_{t''(\lambda^*)} \overset{f}{\cup} M_{t'(\lambda^*)}.$$

(3) Associativity, i.e.

$$M_{t'(\lambda^*)} \overset{f}{\cap} (M_{t''(\lambda^*)} \overset{f}{\cap} M_{t'''(\lambda^*)}) = (M_{t'(\lambda^*)} \overset{f}{\cap} M_{t''(\lambda^*)}) \overset{f}{\cap} M_{t'''(\lambda^*)},$$

$$M_{t'(\lambda^*)} \overset{f}{\cup} (M_{t''(\lambda^*)} \overset{f}{\cup} M_{t'''(\lambda^*)}) = (M_{t'(\lambda^*)} \overset{f}{\cup} M_{t''(\lambda^*)}) \overset{f}{\cup} M_{t'''(\lambda^*)}.$$

(4) Absorption, i.e.

$$M_{t'(\lambda^*)} \overset{f}{\cap} (M_{t'(\lambda^*)} \overset{f}{\cup} M_{t''(\lambda^*)}) = M_{t'(\lambda^*)},$$

$$M_{t'(\lambda^*)} \overset{f}{\cup} (M_{t'(\lambda^*)} \overset{f}{\cap} M_{t''(\lambda^*)}) = M_{t'(\lambda^*)}.$$

(5) Distributivity, i.e.

$$M_{t'(\lambda^*)} \overset{f}{\cap} (M_{t''(\lambda^*)} \overset{f}{\cup} M_{t'''(\lambda^*)}) = (M_{t'(\lambda^*)} \overset{f}{\cap} M_{t''(\lambda^*)}) \overset{f}{\cup} (M_{t'(\lambda^*)} \overset{f}{\cap} M_{t'''(\lambda^*)}),$$

$$M_{t'(\lambda^*)} \overset{f}{\cup} (M_{t''(\lambda^*)} \overset{f}{\cap} M_{t'''(\lambda^*)}) = (M_{t'(\lambda^*)} \overset{f}{\cup} M_{t''(\lambda^*)}) \overset{f}{\cap} (M_{t'(\lambda^*)} \overset{f}{\cup} M_{t'''(\lambda^*)}).$$

(6) De Morgan's laws, i.e.

$$\overset{f}{\rceil}(M_{t'(\lambda^*)} \overset{f}{\cap} M_{t''(\lambda^*)}) = \overset{f}{\rceil} M_{t'(\lambda^*)} \overset{f}{\cup} \overset{f}{\rceil} M_{t''(\lambda^*)},$$

$$\overset{f}{\rceil}(M_{t'(\lambda^*)} \overset{f}{\cup} M_{t''(\lambda^*)}) = \overset{f}{\rceil} M_{t'(\lambda^*)} \overset{f}{\cap} \overset{f}{\rceil} M_{t''(\lambda^*)}.$$

Proof. The laws above directly follow from the definition of operations on the λ^*-level meanings $M_{t(\lambda^*)}$ of the complex descriptors $t \in T^*$ in the document retrieval system I.

Proposition 3. The operation $\overset{f}{\rceil}$ applied to the λ^*-level meanings $M_{t(\lambda^*)}$ of the complex descriptors $t \in T^*$ in the document retrieval system I satisfy the law of involution, i.e.

$$\overset{f}{\rceil}(\overset{f}{\rceil} M_{t(\lambda^*)}) = M_{t(\lambda^*)}$$

$$\text{iff } (\forall t \in T)(\forall d \in M_t(\lambda^*))(1 - \mu_{M_{t(\lambda^*)}}(d) \geq \lambda^*).$$

Proof. Proof of this proposition results directly from the definition of the complement of a level fuzzy set and from the definitions of union and intersection of level fuzzy sets.

Proposition 4. The λ^*-level meanings $M_{t(\lambda^*)}$ of the complex descriptors $t \in T^*$ in the document retrieval system I with operations $\overset{f}{\rceil}$, $\overset{f}{\cup}$ and $\overset{f}{\cap}$ do not constitute Boolean algebra.

Proof. The proof follows from the fact that even if two such complex descriptors $t', t'' \in T^*$ exist, for which $M_{t'(\lambda^*)} = \overset{f}{\square}_{\lambda^*}$ and $M_{t''(\lambda^*)} = \overset{f}{\square}_{\lambda^*}$, then the operations $\overset{f}{\rceil}$, $\overset{f}{\cap}$ and $\overset{f}{\cup}$ applied to the λ^*-level meanings $M_{t(\lambda^*)}$ of the complex descriptors $t \in T^*$ in the document retrieval system I do not fulfil the laws of complementarity, i.e.

$$M_{t(\lambda^*)} \overset{f}{\cap} \overset{f}{\rceil} M_{t(\lambda^*)} \neq M_{t'(\lambda^*)}$$

and

$$M_{t(\lambda^*)} \overset{f}{\cup} \overset{f}{\rceil} M_{t(\lambda^*)} \neq M_{t''(\lambda^*)}.$$

4. DESCRIPTION OF THE DOCUMENT RETRIEVAL PROCESS

We will here present an algorithm for assigning documents to particular queries. In the system proposed document retrieval in response to given queries takes place by means of λ^*-level document search patterns and the search pattern of the query in question.

If we assume that the λ^*-level meanings $M_{t(\lambda^*)}$ of descriptors $t \in T$ in the system I are previously defined, the process of document retrieval in response to a given query $q_j \in Q$, $j = 1, 2, \ldots, m$, leads to the following stages:

(1) The creation (according to the syntactic rules of the document retrieval language) of a complex descriptor $t_j \in T^*$ which represents the search pattern of that query.

(2) The specification of the λ^*-level meaning of the complex descriptor t_j in accordance with the semantic rules of the document retrieval language, where the properties of that language are taken into account. The λ^*-level meaning of the complex descriptor t_j which is the search pattern of the query q_j, is a λ^*-level fuzzy set $M_{t(\lambda^*)}$ in the form:

$$M_{t_j(\lambda^*)} = \{\langle d, \mu_{M_{t_j(\lambda^*)}}(d)\rangle | d \in M_{t_j}(\lambda^*)\}.$$

(3) The issue of documents in decreasing order according to the grade of their membership in $M_{t_j(\lambda^*)}$. If we do not take this order into account, the documents issued in response to a given query $q_j \in Q$, whose search pattern is defined by the complex descriptor $t_j \in T^*$, constitute a λ^*-level set in the form:

$$M_{t_j}(\lambda^*) = \{d | \mu_{M_{t_j}}(d) \geq \lambda^*, d \in D\}.$$

And thus

$$\psi(q_j) = M_{t_j}(\lambda^*).$$

Considering the expense and the perceptual limitations of the user of an information system, it is not always feasible to issue the entire set $M_{t_j}(\lambda^*)$ of documents in response to a given query $q_j \in Q$, but assuming the possibility that not all documents of a sufficient grade of pertinence are issued, it is advisable to limit the information system response $\psi(q_j)$ to those documents with the highest grade of formal relevance to the query. This can be done in two ways: either by establishing the threshold value λ_j of formal relevance or by fixing the maximum permissible number N_j of documents which will constitute the response $\psi(q_j)$ to the query $q_j \in Q$. In the former case

$$\psi(q_j) = M_{t_j(\lambda^*)}(\lambda_j), \lambda_j \in [\lambda^*, 1], j = 1, 2, \ldots, m$$

and in the latter

$$\psi(q_j) = M_{t_j(\lambda^*)}(\lambda_{N_j}) = \max_{\substack{\lambda \in [\lambda^*, 1] \\ |M_{t_j(\lambda^*)}(\lambda)| \leq N_j}} M_{t_j(\lambda^*)}(\lambda), j = 1, 2, \ldots m.$$

It seems more natural to limit the size of the information system response $\psi(q_j)$ to a given query $q_j \in Q$ by establishing the maximum permissible number N_j of the documents of which this response is composed. The following cases can then be distinguished:

Case 1. $(\forall \lambda \in [\lambda^*, 1])(|M_{t_j(\lambda^*)}(\lambda)| = 0), j = 1, 2, \ldots, m.$

Case 2. $(\forall \lambda \in [\lambda^*, 1])(0 < |M_{t_j(\lambda^*)}(\lambda)| < N_j), j = 1, 2, \ldots, m.$

Case 3. $(\exists \lambda \in [\lambda^*, 1])(|M_{t_j(\lambda^*)}(\lambda)| = N_j), j = 1, 2, \ldots, m.$

Case 4. $|M_{t_j(\lambda^*)}(\lambda = \max\{\alpha | \alpha \in \sharp M_{t_j(\lambda^*)}\})| > N_j, j = 1, 2, \ldots, m,$

where $\sharp M_{t_j(\lambda^*)}$ stands for the range of the λ^*-level fuzzy set $M_{t_j(\lambda^*)}$ defined as follows:

$$\sharp M_{t_j(\lambda^*)} = \{\mu_{M_{t_j(\lambda^*)}}(d) | \langle d, \mu_{M_{t_j(\lambda^*)}}(d)\rangle \in M_{t_j(\lambda^*)}\}.$$

When Case 1 occurs, there are no documents in the set D of documents which have sufficient grades of formal relevance to the query $q_j \in Q$. In Case 2 the information system provides the user with all those documents whose grades of formal relevance are not smaller than the established permissible value λ^*, where the number of documents issued is smaller than the number required. In Case 3 the information system completely fulfils the requirements of the user, both from the point of view of the number of documents issued and of the grade of formal relevance of these documents. When Case 4 occurs, then according to the previously described algorithm for assigning documents to queries we have $\psi(q_j) = \phi$. However, in this case it would be more proper to assume that $\psi(q_j) = M_{t_j(\lambda^*)}(\lambda = \max\{\alpha | \alpha \in \sharp M_{t_j(\lambda^*)}\})$ or that the information system response $\psi(q_j)$ to the query q_j represented by the complex descriptor t_j constitutes a certain N_j element subset of the set $M_{t_j(\lambda^*)}(\lambda = \max\{\alpha | \alpha \in \sharp M_{t_j(\lambda^*)}\})$.

From now on we shall assume that the accepted method of limiting the size of a response $\psi(q_j)$ is based on the determination of the maximum permissible number N_j of documents. Since each threshold value λ_j of formal relevance corresponds to a certain number N_j of documents, therefore in the case of limiting the size of the response $\psi(q_j)$ by the establishement of a threshold value λ_j, the discussion concerning the document retrieval process would be similar to that for the case of limiting the size of the information system response by fixing the maximum number N_j of documents.

The relation $G^t_{\lambda^*} \subset T^* \times T^*$ of the λ^*-level generalization of complex descriptors $t \in T^*$ in the document retrieval system I implies the relation $G^q_{\lambda^*} \subset Q \times Q$ which partially orders the set Q of queries.

The relation $G^q_{\lambda^*} \subset Q \times Q$ of a λ^*-level pattern generalization of queries $q \in Q$ in the document retrieval system I is defined as follows:

$$(\forall q_i, q_j \in Q)(q_i G^q_{\lambda^*} q_j \Leftrightarrow t_j G^t_{\lambda^*} t_i),$$

where the complex descriptor $t_i \in T^*$ is the search pattern of the query $q_i \in Q$ and the complex descriptor $t_j \in T^*$—the search pattern of the query $q_j \in Q$.

Proposition 5. If a set Q of queries is partially ordered by the relation $G^q_{\lambda^*} \subset Q \times Q$ of the λ^*-level pattern generalization then for $|M_{t_i(\lambda^*)}(\lambda_{N_i})| \leq |M_{t_j(\lambda^*)}(\lambda_{N_j})|$, where $i, j = 1, 2, \ldots, m$, the document retrieval system I is inclusive, i.e.

$$(\forall q_i, q_j \in Q)(q_i G^q_{\lambda^*} q_j \Rightarrow \psi(q_i) \subset \psi(q_j)).$$

Proof. By assumption, the set Q of queries is partially ordered by the relation $G^q_{\lambda^*} \subset Q \times Q$ of the λ^*-level pattern generalization and the maximum permissible numbers N_i and N_j ($i, j = 1, 2, \ldots, m$) of documents making up the responses $\psi(q_i)$ and $\psi(q_j)$ to queries q_i and q_j are such that $|M_{t_i(\lambda^*)}(\lambda_{N_i})| \leq |M_{t_j(\lambda^*)}(\lambda_{N_j})|$. So, for each pair of queries $q_i, q_j \in Q$ such that $q_i G^q_{\lambda^*} q_j$ if there exists a document $d \in D$ such that $d \in M_{t_i(\lambda^*)}(\lambda_{N_i})$, i.e. $d \in \psi(q_i)$ then, by virtue of the definition of the relation $G^q_{\lambda^*} \subset Q \times Q$ and the definition of the information system response, we have $d \in M_{t_j(\lambda^*)}(\lambda_{N_j})$, i.e. $d \in \psi(q_j)$. Thus it results that $\psi(q_i) \subset \psi(q_j)$.

Conclusion 1. If $q_1 G^q_\lambda \cdot q_2$, $q_2 G^q_\lambda \cdot q_3, \ldots, q_{i-1} G^q_\lambda \cdot q_i$ and $|M_{t_1(\lambda^*)}(\lambda_{N_1})| \leq |M_{t_2(\lambda^*)}(\lambda_{N_2})| \leq \ldots \leq |M_{t_i(\lambda^*)}(\lambda_{N_i})|$ then $\psi(q_1) \subset \psi(q_2) \subset \ldots \subset \psi(q_i)$. So, for a given query $q_k \in Q$ the following relationship holds:

$$\psi(q_k) \subset \{d | d \in \psi(q_j), q_k G^q_\lambda \cdot q_j, q_j \in Q, q_k \neq q_j, |M_{t_k(\lambda^*)}(\lambda_{N_k})| \leq |M_{t_j(\lambda^*)}(\lambda_{N_j})|\}.$$

Proposition 6. If $\lambda_{N_i} \geq \lambda^*$ and $\lambda^* \leq \min\{\lambda | \lambda \in \complement F \cup \complement \overset{f}{\neg} F, \lambda > 0\}$, where $\complement F$ stands for the range of the relation F of the description of information system documents $d \in D$, i.e.

$$\complement F = \{\mu_F(d, t) | (d, t, \mu_F(d, t)) \in F\}$$

whereas $\overset{f}{\neg} F$ stands for the complement of the binary fuzzy relation F, i.e.,

$$\overset{f}{\neg} F = \{(d, t, \mu_{\neg F}(d, t) = 1 - \mu_F(d, t)) | d \in D, t \in T\}$$

then the ordered responses Ord $\psi(q_j)$, $j = 1, 2, \ldots, m$, to queries $q_j \in Q$ found on the basis of the λ^*-level meanings of descriptors $t \in T$ in the document retrieval system I are equal to the ordered responses retrieved on the basis of the meanings $(\lambda^* = 0)$ of those descriptors in the system I.

Proof. If the condition $\lambda^* \leq \min\{\lambda | \lambda \in \complement F \cup \complement \overset{f}{\neg} F, \lambda > 0\}$ is fulfilled, it then follows that operation $\overset{f}{\neg}$ applied to the λ^*-level meanings $M_{t(\lambda^*)}$ of the complex descriptors $t \in T^*$ in the document retrieval system I satisfies the law of involution (Proposition 3) and

$$(\forall t \in T)(\forall \lambda \in [\lambda^*, 1])(\forall d \in M_{t(\lambda^*)}(\lambda^*))(\mu_{M_{t(\lambda^*)}}(d) = \mu_{M_t}(d)).$$

From the above, and from the assumption $\lambda_{N_j} \geq \lambda^*$, the stated hypothesis is true.

We will say that an ordered response

$$\text{Ord } \psi(q_i) = \langle K^i_1, K^i_2, \ldots, K^i_n \rangle,$$

where K_r $(r = 1, 2, \ldots, n)$ denotes the set (being the subset of the set $M_{t_i}(\lambda_{N_i})$) of documents with the same grades of membership in the λ^*-meaning $M_{t_i(\lambda^*)}$ of the complex descriptor t_i representing the query q_i, is contained in an ordered response

$$\text{Ord } \psi(q_j) = \langle K^j_1, K^j_2, \ldots, K^j_n, K^j_{n+1}, \ldots, K^j_s \rangle,$$

written as Ord $\psi(q_i) \preccurlyeq$ Ord $\psi(q_j)$, iff $K^i_r = K^j_r$ for $r = 1, 2, \ldots, n$.

Proposition 7. Let Ord $\psi(q_j)_{\lambda^*}$ stand for an ordered response to a query $q_j \in Q$ in the case where the λ^*-level search patterns of documents are utilized. If operation $\overset{f}{\neg}$ applied to the λ^*-level meanings $M_{t(\lambda^*)}$ of those descriptors $t \in T$ which constitute the search pattern of the query $q_j \in Q$ fulfils the law of involution then if $\lambda_{N_j} \geq \lambda^*_1$ and $\lambda_{N_j} \geq \lambda^*_2$ the following equation holds:

$$\text{Ord } \psi(q_j)_{\lambda^*_1} = \text{Ord } \psi(q_j)_{\lambda^*_2}, \qquad j = 2, \ldots, m.$$

Proof. This equation results directly from the method of assigning the information system response to a given query.

Conclusion 2. If operation $\bar{\ }$ applied to the λ^*-level meanings $M_{t(\lambda^*)}$ of those descriptors $t \in T$, which constitute the search pattern of a given query $q_j \in Q$ does not fulfil the law of involution, then Proposition 7 remains valid, if the search pattern of that query is a complex descriptor $t_j \in T^*$ created from descriptors $t \in T$ and the Boolean operators $\wedge$ and $\vee$.

Conclusion 3. If the assumptions of Proposition 7 are fulfilled, and if $\lambda_{N_j'} \geq \lambda_{N_j''}$ and $\lambda_{N_j'} \geq \max\{\lambda_1^*, \lambda_2^*\}$ and $\lambda_{N_j''} \geq \max\{\lambda_1^*, \lambda_2^*\}$, then Ord $\psi(q_{N_j'}) \leq$ Ord $\psi(q_{N_j''})$, where Ord $\psi(q_{N_j})$ stands for the information system's ordered response to a query $q_j \in Q$ in case the maximum permissible number of documents constituting that response has been established as N_j.

Proposition 8. In the document retrieval system I the following relationships hold:

(1) $N_j' > N_j'' \Rightarrow M_{t_j(\lambda^*)}(\lambda_{N_j'}) \subset M_{t_j(\lambda^*)}(\lambda_{N_j''}) \Rightarrow \lambda_{N_j'} \leq \lambda_{N_j''}$, $j = 1, 2, \ldots, m$.

(2) $(\forall \lambda \in [\lambda^*, 1])(\exists N_j)(M_{t_j(\lambda^*)}(\lambda) = M_{t_j(\lambda^*)}(\lambda_{N_j}))$, $j = 1, 2, \ldots, m$.

(3) $(\exists N_j)(M_{t_j(\lambda^*)}(\lambda_{N_j}) = M_{t_j}(\lambda^*))$, $j = 1, 2, \ldots, m$.

(4) The number of various possible sets of documents which are potential information system responses to a given query $q_j \in Q$ is equal to the number of various values $\mu_{M_{t_j(\lambda^*)}}(d)$ in the range of the λ^*-level fuzzy set $M_{t_j(\lambda^*)}$.

Proof. These relationships result directly from the definition of a fuzzy set and the definition of a level set.

Proposition 9. The λ^*-level meaning $M_{t_j(\lambda^*)}$ of a complex descriptor $t_j \in T^*$ in the document retrieval system I, which is the search pattern of a given query $q_j \in Q$ can be presented in the form of the following resolution:

$$M_{t_j(\lambda^*)} = \overset{f}{\bigcup}{}_{i=1}^{M} [M_{t_j(\lambda^*)}(\lambda_{N_j^i})]_{\lambda_{N_j^i}}, \quad i = 1, 2, \ldots, m,$$

where $[M_{t_j(\lambda^*)}(\lambda_{N_j^i})]_{\lambda_{N_j^i}}$ is a fuzzy set for which the grades of membership of all the documents belonging to the $\lambda_{N_j^i}$-level set $M_{t_j(\lambda^*)}(\lambda_{N_j^i})$ are equal to one another, i.e. $\lambda_{N_j^i}$. Thus

$$[M_{t_j(\lambda^*)}(\lambda_{N_j^i})]_{\lambda_{N_j^i}} = \{\langle d, \lambda_{N_j^i}\rangle | d \in M_{t_j(\lambda^*)}(\lambda_{N_j^i})\}.$$

On the other hand

$$\overset{f}{\bigcup}{}_{i=1}^{M} [M_{t_j(\lambda^*)}(\lambda_{N_j^i})]_{\lambda_{N_j^i}} = \{\langle d, \max_i \lambda_{N_j^i}\rangle | d \in M_{t_j}(\lambda^*)\},$$

where $\lambda_{N_j^i}$, $i = 2, \ldots, M$, are different values $\mu_{M_{t_j(\lambda^*)}}(d)$ in $\not\subset M_{t_j(\lambda^*)}$.

Proof. The proposition is true since

$$\{\langle d, \max_i \lambda_{N_j^i}\rangle | d \in M_{t_j(\lambda^*)}\} = \{\langle d, \mu_{M_{t_j(\lambda^*)}}(d)\rangle | d \in M_{t_j}(\lambda^*)\} = M_{t_j(\lambda^*)}.$$

Example 1. Let us take a given set $D = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8\}$ of information system documents and the set $T = \{t_1, t_2, t_3, t_4, t_5\}$ of descriptors. The relation F of the document description is presented in Table 1.

Table 1.

		Documents							
		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
Descriptors	t_1	0.1	0.7	0.6	0.4	0.8	0.6	0.3	0.6
	t_2	0.3	0.8	0.9	0.8	0.6	0.5	0.7	0.2
	t_3	0.8	1	0.2	0.6	0.8	0.2	0.8	0.5
	t_4	0.4	0.6	0.7	0.5	0.7	0.3	0.1	0.9
	t_5	0.7	0.2	0.8	0.4	0.6	0.2	0.8	0.4

Table 2.

		Documents.							
		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
Descriptors	t_1		0.7	0.6	0.4	0.8	0.6	0.3	0.6
	t_2	0.3	0.8	0.9	0.8	0.6	0.5	0.7	
	t_3	0.8	1		0.6	0.8		0.8	0.5
	t_4	0.4	0.6	0.7	0.5	0.7	0.3		0.9
	t_5	0.7		0.8	0.4	0.6		0.8	0.4

Taking $\lambda^* = 0.3$, we get the 0.3-level binary fuzzy relation $F_{\lambda^*=0.3}$ in the form of Table 2. The search pattern of the query q has been specified by the complex descriptor as follows:

$$t = (t_1 \wedge \rceil t_2) \vee (t_2 \wedge \rceil t_3 \wedge t_4).$$

So the 0.3-level meaning of that complex descriptor is as follows:

$$M_{t(0.3)} = \{\langle d_3, 0.7\rangle, \langle d_4, 0.4\rangle, \langle d_5, 0.4\rangle, \langle d_6, 0.5\rangle, \langle d_7, 0.3\rangle, \langle d_8, 0.6\rangle\}.$$

Let the maximum permissible number of documents constituting the response of the given information system to the query q represented by the complex descriptor t be $N = 4$. Thus

$$\text{Ord}\ \psi(q) = \langle\{d_3\}, \{d_8\}, \{d_6\}\rangle.$$

Noting that the number of the different values $\mu_{M_{t(0.3)}}(d)$ in $\not\subset M_{t(0.3)}$ is $M = 5$, the number of various possible sets of documents which are potential responses of the information system to the query equals 5.

The 0.3-level meaning $M_{t(0.3)}$ of the complex descriptor t can be presented in the form of the following resolution:

$$M_{t(0.3)} = \overset{f}{\bigcup}_{i=1}^{5} [M_{t(0.3)}(\lambda_{N^i})]_{\lambda_{N^i}}$$

$$= [M_{t(0.3)}(0.3)]_{0.3} \overset{f}{\cup} [M_{t(0.3)}(0.4)]_{0.4} \overset{f}{\cup} [M_{t(0.3)}(0.5)]_{0.5}$$

$$\overset{f}{\cup} [M_{t(0.3)}(0.6)]_{0.6} \overset{f}{\cup} [M_{t(0.3)}(0.7)]_{0.7}$$

$$= [\{d_3, d_4, d_5, d_6, d_7, d_8\}]_{0.3} \overset{f}{\cup} [\{d_3, d_4, d_5, d_6, d_8\}]_{0.4}$$

$$\overset{f}{\cup} [\{d_3, d_6, d_8\}]_{0.5} \overset{f}{\cup} [\{d_3, d_8\}]_{0.6} \overset{f}{\cup} [\{d_3\}]_{0.7}.$$

Of course

$$M_{t(0.3)}(0.7) \subset M_{t(0.3)}(0.6) \subset M_{t(0.3)}(0.5) \subset M_{t(0.3)}(0.4) \subset M_{t(0.3)}(0.3).$$

5. FINAL REMARKS

The basic advantage of the document retrieval method discussed in the present paper is that it takes into account, in a simple way, the differentiation of descriptor importance in particular document search patterns, together with the differentiation in the grades of the formal relevance of particular documents in a document retrieval system to a given query. This advantage is a direct result of the application of the fuzzy set theory to the description of document retrieval strategy when document search patterns are sets of weighted descriptors.

By making use of the simple operations of the fuzzy set theory the retrieval of documents with the highest grades of formal relevance (in a given information system) in relation to a given query is made possible. One should emphasize that due to the time required for document retrieval the process of assigning a response to a given query should take place on the basis of the λ^*-level meanings of the descriptors $t \in T$. Thus a matter of special significance here is the correct choice of the value of λ^*. The value of this parameter should depend directly

on the size of the set of documents in a given information system. With regard to the difficulty in the analytical determination of a threshold value λ^* the choice of this value for each actual information system should be carried out experimentally.

In the document retrieval method described it has been assumed that the grades of importance of descriptors occurring in complex descriptors describing particular queries are identical. However, it is also possible to take into account the differentiation of the importance of particular descriptors in query search patterns by introducing—into the syntax of the document retrieval language—a set o of special operators and the definitions of the meanings of the expressions created from descriptors and these special operators. Generally, if the λ^*-level meaning $M_{t(\lambda^*)}$ of a descriptor $t \in T$ is given in the form:

$$M_{t(\lambda^*)} = \{\langle d, \mu_{M_{t(\lambda^*)}}(d)\rangle | d \in M_t(\lambda^*)\},$$

then the λ^*-level meaning of an expression $o_i(t) \in T^*$, where $t \in T$, $o_i \in o$, can be written as follows:

$$M_{o_i(t)(\lambda^*)} = \{\langle d, \mu_{M_{o_i(t)(\lambda^*)}}(d) = f_{o_i}[\mu_{M_{t(\lambda^*)}}(d)]\rangle | d \in M_{o_i(t)}(\lambda^*)\}.$$

The task of the functions f_{o_i} corresponding to particular special operators, is to differentiate the importance of descriptors in particular query search patterns. The grades of membership of particular information system documents in λ^*-level fuzzy sets being the λ^*-level meanings of descriptors which are particulary important from the point of view of the content of the queries under consideration should be increased, whereas the grades of membership of these documents in λ^*-level fuzzy sets being the λ^*-level meanings of less important descriptors should be reduced. Below we present this type of functions with the reservation that this is not the ultimate choice. As an example, the function corresponding to the special operator $o_1 \in o$ causing an increase in the importance of the descriptor $t \in T$ in the search pattern of a given query can be defined as follows:

$$f_{o_1}[\mu_{M_{t(\lambda^*)}}(d)] = 1 - [1 - \mu_{M_{t(\lambda^*)}}(d)]^\alpha, \alpha > 1.$$

Thus, the λ^*-level meaning of expression $o_1(t)$ is a λ^*-level fuzzy set in the form:

$$M_{o_1(t)(\lambda^*)} = \{\langle d, \mu_{M_{o_1(t)(\lambda^*)}}(d) = 1 - [1 - \mu_{M_{t(\lambda^*)}}(d)]^\alpha\rangle | 1 - [1 - \mu_{M_{t(\lambda^*)}}(d)]^\alpha \geq \lambda^*, d \in D\}.$$

On the other hand, a function corresponding to the special operator $o_2 \in o$ causing a reduction in the importance of the descriptor $t \in T$ in the search pattern of a given query can be defined as follows:

$$f_{o_2}[\mu_{M_{t(\lambda^*)}}(d)] = [\mu_{M_{t(\lambda^*)}}(d)]^\beta, \beta > 1.$$

That is, the λ^*-level meaning of expression $o_2(t)$ is a λ^*-level fuzzy set in the form:

$$M_{o_2(t)(\lambda^*)} = \{\langle d, \mu_{M_{o_2(t)(\lambda^*)}}(d) = [\mu_{M_{t(\lambda^*)}}(d)]^\beta\rangle | [\mu_{M_{t(\lambda^*)}}(d)]^\beta \geq \lambda^*, d \in D\}.$$

The values of parameters α and β are dependent on the importance of a given descriptor $t \in T$ in a given query search pattern.

Acknowledgements—The author would like to express his gratitude to Dr. Czesław Daniłowicz of the Technical University of Wrocław for discussing the text of this paper. He would also like to thank Prof. Manfred Kochen of the University of Michigan and Prof. Michael Lynch of the University of Sheffield for fruitful discussions on the application of the fuzzy set theory to the analysis of document retrieval systems.

REFERENCES

[1] G. Salton (Ed.), *The SMART Retrieval System—Experiments in Automatic Document Processing*. Prentice-Hall, Englewood Cliffs, New Jersey (1971).

[2] M. DABROWSKI, A general model of distribution of objects in information retrieval systems. *Inform. Systems* 1975, 1, 4.
[3] D. HSIAO and F. HARARY, A formal system for information retrieval from files. *Commun. ACM* 1970, 13, 2.
[4] K. LAUS and M. DABROWSKI, A model of information retrieval process for hierarchical set of descriptors. *Inform. Stor. Ret.* 1974, 10, 7/8.
[5] W. LIPSKI and W. MAREK, O sformalizowanej teorii systemow informacyjnych. *Podstawy Sterowania* 1976, 6, 1.
[6] W. MAREK and Z. PAWLAK, Information storage and retrieval systems: mathematical foundations. *Theore. Comp. Sci.* 1976, 1, 4.
[7] G. SALTON, *Automatic Information Organization and Retrieval.* McGraw-Hill, New York (1968).
[8] G. SALTON, *Dynamic Information and Library Processing.* Prentice-Hall, Englewood Cliffs, New Jersey (1975).
[9] W. M. TURSKI, On a model of information retrieval system based on thesaurus. *Inform. Stor. Retr.* 1971, 7.
[10] C. J. VAN RIJSBERGEN, *Information Retrieval.* Butterworths, London (1975).
[11] E. WONG and T. C. CHIANG, Canonical structure in attribute based file organization. *Commun. ACM* 1971, 14, 9.
[12] L. A. ZADEH, Fuzzy sets. *Inform. Control* 1965, 8, 3.
[13] L. A. ZADEH, Similarity relations and fuzzy orderings. *Inform. Sci.* 1971, 3, 1/2.
[14] L. A. ZADEH, Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. Systems, Man and Cybernetics* 1973, SMC-3, 1.
[15] L. A. ZADEH, The concept of a linguistic variable and its application to approximate reasoning—I. *Inform. Sci.* 1975, 8, 3.
[16] T. RADECKI, Mathematical model of information retrieval strategy based on the theory of fuzzy sets. Communications of the Main Library and Scientific Information Centre of the Technical University of Wrocław, Poland, Ser. D, No 26, 1975.
[17] T. RADECKI, Mathematical model of information retrieval system based on the concept of fuzzy thesaurus. *Inform. Proc. Management* 1976, 12, 5.
[18] T. RADECKI, Mathematical model of time—effective information retrieval system based on the theory of fuzzy sets. *Inform. Proc. Management* 1977, 13, 2.
[19] T. RADECKI, Application of fuzzy sets theory to the description of information retrieval process. Communications of the Main Library and Scientific Information Centre of the Technical University of Wrocław, Poland, Ser. A, No 56, 1976.
[20] T. RADECKI, Semantic normal forms and their application to documents retrieval system description. Communications of the Main Library and Scientific Information Centre of the Technical University of Wrocław, Poland, Ser. A, No 62, 1976.
[21] C. V. NEGOITA, On the application of the fuzzy sets separation theorem for automatic classification in information retrieval systems. *Inform. Sci.* 1973, 5.
[22] C. V. NEGOITA, On the decision process in information retrieval. *Studii si Cercetari de Documentare* 1973, 15, 3.
[23] C. V. NEGOITA, On the notion of relevance in information retrieval. *Kybernetes* 1973, 2, 3.
[24] C. V. NEGOITA, On the search strategies in automatic information systems. *Acta Proc. of the 6th Int. Cong. on Cybernetics*, Namur, 7–11 Sept. 1970.
[25] M. P. REBROVA, Formalnaja aglomerativnaja ierarchičeskaja klassifikackija dokumentov i strategija poiska v nej. *Naučno—techničeskaja informacija* 1975, Ser. 2, No 10.
[26] W. M. SACHS, An approach to associative retrieval trough the theory of fuzzy sets. *J. ASIS* 1976, 27, 2.
[27] V. TAHANI, A fuzzy model of document retrieval systems. *Inform. Proc. Management* 1976, 12, 3.
[28] T. RADECKI, Level fuzzy sets. *J. Cybernetics* 1977, 7, 3/4.