Correspondence

Fuzzy Query Processing for Document Retrieval Based on Extended Fuzzy Concept Networks

Shyi-Ming Chen and Yih-Jen Horng

***Abstract*— In this paper, we present a new method for fuzzy query processing for document retrieval based on extended fuzzy concept networks. In an extended fuzzy concept network, there are four kinds of fuzzy relationships between concepts, i.e., fuzzy positive association, fuzzy negative association, fuzzy generalization, and fuzzy specialization. An extended fuzzy concept network can be modeled by a relation matrix and a relevance matrix, where the elements in a relation matrix represent the fuzzy relationships between concepts, and the elements in a relevance matrix indicate the degrees of relevance between concepts. The implicit fuzzy relationships between concepts can be inferred by the transitive closure of the relation matrix. The implicit degrees of relevance between concepts also can be inferred by the transitive closure of the relevance matrix. The proposed method is more flexible than the ones presented in [8] and [17] due to the fact that it allows the users to perform positive queries, negative queries, generalization queries, and specialization queries. The proposed method allows the users to perform fuzzy queries in a more flexible and more intelligent manner.**

***Index Terms*— Document retrieval, extended fuzzy concept networks, fuzzy query processing, relation matrix, relevance matrix.**

I. Introduction

In [24], Salton *et al*. pointed out that an information retrieval system is a system which is used to store items of information that need to be processed, searched, retrieved, and disseminated to various user populations. The primary purpose of establishing an information retrieval system is to assist the users to efficiently acquire information [8]. Most commercial information retrieval systems currently still adopt the Boolean logic model. These information retrieval systems are based on the assumption that documents can be precisely described by sets of index terms and that information needed by the users can be represented by Boolean search requests. However, the information retrieval systems based on the Boolean logic model are rather restricted in applications due to the fact that these systems are unable to represent uncertain information. If there is uncertain information, the query processing of these systems is not handled properly [8]. In recent years, several fuzzy information retrieval methods based on fuzzy set theory [27] have been proposed for improving the disadvantage of the Boolean logic model, such as [8], [9], [12], [17]–[21], [25], and [28].

In [8], we presented a knowledge-based fuzzy information retrieval method to deal with document retrieval, where concept matrices are used for knowledge representation, and simple queries, weighted queries, interval queries, and weighted-interval queries are allowed for document retrieval. In [9], Ke *et al*. presented a fuzzy information retrieval system model for document retrieval. In [12], Kamel *et al*. presented a fuzzy query processing method using clustering techniques. In [17], Lucarella *et al*. proposed an information retrieval method based on fuzzy concept networks. In [18], Murai *et al*. presented a fuzzy document retrieval method based on two-valued indexing. In [19], Miyamoto presented a fuzzy information retrieval method based on fuzzy associations. In [20], Radechi presented a mathematical model of information retrieval system based on the theory of fuzzy sets. In [21], Radechi presented a fuzzy set theoretical approach to document retrieval. In [25], Tahuni presented a fuzzy model of document retrieval system. In [28], Zemankova presented a fuzzy intelligent information system FIIS. However, either efficiency or effectiveness of these methods are not satisfied. Thus, there is an increasing demand to develop a more powerful fuzzy information retrieval method to deal with document retrieval.

In [8], we have presented a method to deal with document retrieval based on concept networks [17], where concept matrices are used for modeling concept networks. The method presented in [8] is more flexible than the ones presented in [9] and [17] due to the fact that it has the capability to deal with interval queries and weighted-interval queries. However, there is only one kind of fuzzy relationship between concepts in the concept networks presented in [8] and [17], i.e., fuzzy positive association relation. If we can provide more kinds of fuzzy relationships between concepts in a concept network, then there is room for more flexibility. In [14], Kracker has presented a fuzzy concept network model which has four kinds of fuzzy relationships between concepts, (i.e., fuzzy positive association, fuzzy negative association, fuzzy generalization, and. fuzzy specialization) for supporting database queries. In this paper, we generalize the definitions of fuzzy concept networks presented in [8], [11], and [17] to propose the concept of extended fuzzy concept networks based on [14]. We also present a new method for document retrieval based on the extended fuzzy concept networks. In an extended fuzzy concept network, there are four kinds of fuzzy relationships between concepts, i.e., fuzzy positive association, fuzzy negative association, fuzzy generalization, and fuzzy specialization. An extended fuzzy concept network can be modeled by a relation matrix and a relevance matrix, where the elements in a relation matrix represent the fuzzy relationships between concepts, and the elements in a relevance matrix indicate the degrees of relevance between concepts. The implicit fuzzy relationships between concepts can be inferred by the transitive closure of the relation matrix. The implicit degrees of relevance between concepts also can be inferred by the transitive closure of the relevance matrix. The proposed method is more flexible than the ones presented in [8] and [17] due to the fact that it allows the users to perform positive queries, negative queries, generalization queries, and specialization queries. The proposed method allows the users to perform fuzzy queries in a more flexible and more intelligent manner.

The rest of this paper is organized as follows. In Section II, we briefly review the definitions of concept networks from [8] and [17]. In Section III, we present the definitions of extended fuzzy concept networks. In Section IV, we use relation matrices and relevance matrices to model extended fuzzy concept networks. In Section V, we propose a new method for document retrieval based on extended fuzzy concept networks. The conclusions are discussed in Section VI.

Manuscript received January 6, 1997; revised August 31, 1997. This work was supported in part by the National Science Council, Republic of China, under Grant NSC 86-2213-E-009-018.

S.-M. Chen is with the Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C.

Y.-J. Horng is with the Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

Publisher Item Identifier S 1083-4419(99)00907-3.

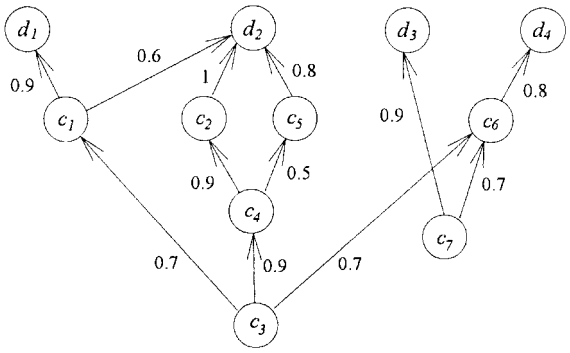

Fig. 1. A concept network.

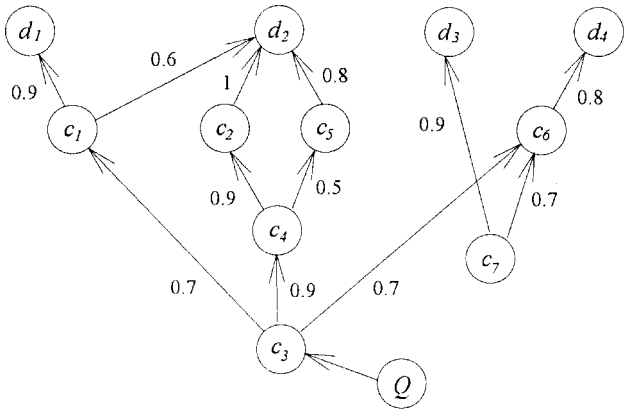

Fig. 2. A concept network of Example 2.1.

II. Concept Networks

In [17], Lucarella *et al*. have proposed concept networks for fuzzy information retrieval. A concept network includes nodes and directed links, where each node represents a concept or a document; each directed link connects two concepts or directs from one concept C_i to one document d_j and is labeled with a real value between zero and one. If $C_i \xrightarrow{\mu} C_j$, then it indicates that the degree of relevance from concept C_i to concept C_j is μ, where $\mu \in [0,1]$. If $C_i \xrightarrow{\mu} d_j$, then it indicates that the degree of relevance of document d_j with respect to concept C_i is μ, where $\mu \in [0,1]$. For Example, Fig. 1 shows a concept network, where $C_1, C_2, \cdots$, and C_7 are concepts; d_1, d_2, d_3, and d_4 are documents.

From Fig. 1, we can see that document d_2 can be expressed as a fuzzy subset of concepts, where

$$d_2 = \{(C_1, 0.6), (C_2, 1), (C_5, 0.8)\}.$$

A concept network is assumed to consist of n nodes and some directed links between concepts. Let C be a set of concepts, $C = \{C_1, C_2, \cdots, C_n\}$, and let the value associated with the directed link form concept C_i to concept C_j be denoted by $F(C_i, C_j)$, where F is a mapping function, $F\colon C \times C \to [0,1]$, and $F(C_i, C_j) \in [0,1]$. If the relevance value from concept C_i to concept C_j is $F(C_i, C_j)$, and if the relevance value from concept C_j to concept C_k is $F(C_j, C_k)$, then the relevance value from concept C_i to concept C_k can be evaluated by the following expression:

$$F(C_i, C_k) = \min\left(F(C_i, C_j), F(C_j, C_k)\right). \tag{1}$$

Similarly, if $F(C_1, C_2), F(C_2, C_3), \cdots$, and $F(C_n, C_{n-1})$ are known, then we can get

$$F(C_1, C_n) = \min\left(F(C_1, C_2), F(C_2, C_3), \cdots, F(C_{n-1}, C_n)\right). \tag{2}$$

In a concept network, each document has a different relevance value with respect to each concept. The document descriptor [8] for the document d_j is defined as a fuzzy subset of the collection of concepts by the following expression:

$$d_j = \{(C_i, f_{d_j}(C_i)) | (C_i \in C\}$$

where $f_{d_j}(C_i), f_{d_j}\colon C \to [0,1]$, represents the degree of relevance of document d_j with respect to concept C_i. Each user's query can be represented by a query descriptor Q expressed as a fuzzy subset of the collection of concepts by the following expression:

$$Q = \{(C_i, f_Q(C_i)) | C_i \in C\}$$

where $f_Q(C_i), f_Q\colon C \to [0,1]$, represents the relevance value of the query descriptor Q with respect to the concept C_i.

Example 2.1: Assume that the concept network shown in Fig. 2 consists of four documents d_1, d_2, d_3, d_4, and seven concepts $C_1, C_2, \cdots, C_7$.

If the query descriptor Q is

$$Q = \{(C_3, 1.0)\}$$

where 1.0 represents the relevance value of the query descriptor Q with respect to the concept C_3, then the relevance value of document d_2 with respect to concept C_3 can be evaluated. From Fig. 2, we can see that there are three different routes which can be applied for determining the relevance value of document d_2 with respect to the concept C_3.

1) The first route is $C_3 \to C_1 \to d_2$.
   Based on [17], the relevance value of document d_2 with respect to concept C_3 can be determined as follows:
   $$\min(0.7, 0.6) = 0.6.$$
2) The second route is $C_3 \to C_4 \to C_2 \to d_2$.
   Based on [17], the relevance value of document d_2 with respect to concept C_3 can be determined as follows:
   $$\min(0.9, 0.9, 1) = 0.9.$$
3) The third route is: $C_3 \to C_4 \to C_5 \to d_2$.
   Based on [17], the relevance value of document d_2 with respect to concept C_3 can be determined as follows:
   $$\min(0.9, 0.5, 0.8) = 0.5.$$

Then, based on [17], we can see that the relevance value of the document d_2 with respect to the concept C_3 is

$$\max(0.6, 0.9, 0.5) = 0.9.$$

The reasoning procedure should be repeated n times if there are n documents. However, there is only one kind of fuzzy relationship between concepts in the concept networks presented in [8] and [17], i.e., the fuzzy positive association relation. If we can provide more kinds of relationships between concepts in a concept network, then there is room for more flexibility. In Section III, we will generalize the concepts of concept networks to propose the concepts of extended fuzzy concept networks which allows four kinds of fuzzy relationships between concepts, i.e., fuzzy positive association, fuzzy negative association, fuzzy generalization, and fuzzy specialization. More powerful knowledge representation capability is consequently provided for.

III. Extended Fuzzy Concept Networks

In this section, we propose the definitions of extended fuzzy concept networks based on [14]. The extended fuzzy concept networks are more general than the fuzzy concept networks presented in [8], [11], [14], and [17]. There are four kinds of fuzzy relationships between concepts in an extended fuzzy concept network, i.e., fuzzy positive association, fuzzy negative association, fuzzy generalization, and fuzzy specialization. The fuzzy relationships between concepts are described as follows.

1) *Fuzzy positive association* relates concepts which in some contexts have a fuzzy similar meaning.
2) *Fuzzy negative association* relates concepts which are fuzzy complementary, fuzzy incompatible or fuzzy antonyms.
3) *Fuzzy generalization* is a concept that is regarded as a fuzzy generalization of another concept if it includes that concept in an analytic or partitive sense.
4) *Fuzzy specialization* is the inverse of the fuzzy generalization relationship.

The fuzzy relationships between concepts introduced above are described formally as follows.

Definition 3.1: Let C be a set of concepts, then

1) fuzzy positive association P is a fuzzy relation, P: $C \times C \rightarrow [0,1]$, which is reflexive, symmetric, and max-*-transitive;
2) fuzzy negative association N is a fuzzy relation, N: $C \times C \rightarrow [0,1]$, which is anti-reflexive, symmetric, and max-*-nontransitive;
3) fuzzy generalization G is a fuzzy relation, G: $C \times C \rightarrow [0,1]$, which is anti-reflexive, antisymmetric, and max-*-transitive;
4) fuzzy specialization S is a fuzzy relation, S: $C \times C \rightarrow [0,1]$, which is anti-reflexive, antisymmetric, and max-*-transitive.

Furthermore, the following restrictions hold [14].

1) $P(c_i,c_j) \neq 0 \rightarrow N(c_i,c_j) = 0$ and $G(c_i,c_j) = 0$ and $S(c_i,c_j) = 0$ and $P(c_j,c_i) = P(c_i,c_j)$;
2) $N(c_i,c_j) \neq 0 \rightarrow P(c_i,c_j) = 0$ and $G(c_i,c_j) = 0$ and $S(c_i,c_j) = 0$ and $N(c_j,c_i) = N(c_i,c_j)$;
3) $G(c_i,c_j) \neq 0 \rightarrow P(c_i,c_j) = 0$ and $N(c_i,c_j) = 0$ and $S(c_i,c_j) = 0$ and $S(c_j,c_i) = G(c_i,c_j)$;
4) $S(c_i,c_j) \neq 0 \rightarrow P(c_i,c_j) = 0$ and $N(c_i,c_j) = 0$ and $G(c_i,c_j) = 0$ and $G(c_j,c_i) = S(c_i,c_j)$;

for every $c_i, c_j \in C$.

In the following, we present the definition of extended fuzzy concept networks.

Definition 3.2: An extended fuzzy concept network consists of nodes and directed links. Each node represents a concept or a document. Each directed link connects two concepts or connects from a concept c_i to a document d_j. If

1) $c_i \xrightarrow{(\mu,P)} c_j$, then there is a positive association relationship between concept c_i and concept c_j, and the relevance degree is μ, where $\mu \in [0,1]$.
2) $c_i \xrightarrow{(\mu,N)} c_j$, then there is a negative association relationship between concept c_i and concept c_j, and the relevance degree is μ, where $\mu \in [0,1]$.
3) $c_i \xrightarrow{(\mu,G)} c_j$, then concept c_i is more general than concept c_j, and the degree of generalization is μ, where $\mu \in [0,1]$.
4) $c_i \xrightarrow{(\mu,S)} c_j$, then concept c_i is more special than concept c_j, and the degree of specialization is μ, where $\mu \in [0,1]$.
5) $c_i \xrightarrow{(\mu,P)} d_j$, then there is a positive association relationship between concept c_i and document d_j, and the relevance degree is μ, where $\mu \in [0,1]$ (i.e., document d_j possesses concept c_i with the degree $\mu \times 100\%$).

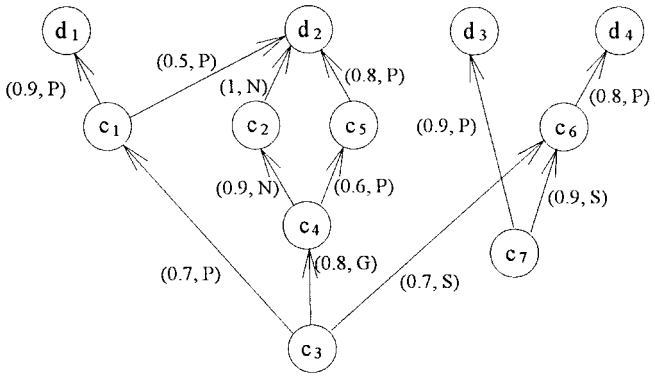

Fig. 3. An extended fuzzy concept network.

6) $c_i \xrightarrow{(\mu,N)} d_j$, then there is a negative association relationship between concept c_i and document d_j, and the relevance degree is μ, where $\mu \in [0,1]$ (i.e., document d_j possesses the concept which is $\mu \times 100\%$ complementary with the concept c_i).
7) $c_i \xrightarrow{(\mu,G)} d_j$, then there is a generalization relationship between concept c_i and document d_j, and the relevance degree is μ, where $\mu \in [0,1]$ and concept c_i is more general than the concept possessed by document d_j with the degree of $\mu \times 100\%$.
8) $c_i \xrightarrow{(\mu,S)} d_j$, then there is a specialization relationship between concept c_i and document d_j, and the relevance degree is μ, where $\mu \in [0,1]$ and concept c_i is more special than the concept possessed by document d_j with the degree of $\mu \times 100\%$.

Every directed link in an extended fuzzy concept network is labeled with a pair of values (μ, FR), where μ denotes the degree of relevance, $\mu \in [0,1]$, and FR denotes the fuzzy relationship between concept c_i and concept c_j or between concept c_i and document d_j, where $FR \in \{P, N, G, S\}$.

Example 3.1: Assume that an extended fuzzy concept network as shown in Fig. 3, where $c_1, c_2, \cdots, c_7$ are concepts, and d_1, d_2, d_3, and d_4 are documents, then we can see that document d_2 possesses 50% of concept c_1, 80% of concept c_5, and document d_2 possesses the concept which is 100% complementary with the concept c_2.

In an extended fuzzy concept network, if the relevance degree between concept c_i and concept c_j is μ_{ij}, where $\mu_{ij} \in [0,1]$, and if the relevance degree between concept c_j and concept c_k is μ_{jk}, where $\mu_{jk} \in [0,1]$, then the relevance degree μ_{ik} between concept c_i and concept c_k can be calculated as follows:

$$\mu_{ik} = \min(\mu_{ij}, \mu_{jk}) \tag{3}$$

where $\mu_{ik} \in [0,1]$. Furthermore, if the relevance degree between concept c_1 and concept c_2 is μ_{12}, the relevance degree between concept c_2 and concept c_3 is $\mu_{23}, \cdots$, and the relevance degree between concept c_{n-1} and concept c_n is $\mu_{(n-1)n}$, where $\mu_{12} \in [0,1], \mu_{23} \in [0,1], \cdots$, and $\mu_{(n-1)n} \in [0,1]$, then the relevance degree between concept c_1 and concept c_n is μ_{1n}, where $\mu_{1n} \in [0,1]$ and

$$\mu_{1n} = \min(\mu_{12}, \mu_{23}, \cdots, \mu_{(n-1)n}). \tag{4}$$

In an extended fuzzy concept network, if the fuzzy relationship between concept c_i and concept c_j is FR_{ij}, and if the fuzzy relationship between concept c_j and concept c_k is FR_{jk}, then the fuzzy relation FR_{ik} between concept c_i and concept c_k can be obtained by Table I, where P, N, G, and S stand for fuzzy positive association, fuzzy negative association, fuzzy generalization, and

TABLE I
THE COMBINATION OF FUZZY RELATIONSHIPS

	P	N	G	S
P	P	N	G	S
N	N	P	N	N
G	G	N	G	P
S	S	N	P	S

TABLE II
THE COMBINATION OF FUZZY RELATIONSHIPS IN RELATION MATRICES

	P	N	G	S	Z
P	P	N	G	S	P
N	N	P	N	N	N
G	G	N	G	P	G
S	S	N	P	S	S
Z	P	N	G	S	Z

fuzzy specialization, respectively. In Table I, the first row shows the four possible fuzzy relationships of FR_{ij}, and the first column shows the four possible fuzzy relationships of FR_{jk}. The other elements in the table are the combination of FR_{ij} and FR_{jk}. From Table I, we can see that the combination of two relationships of the same type results in a relationship of this type except for negative associations (N) which get positive associations (P). In Table I, we let these four kinds of fuzzy relationships have different priorities, i.e., the negative association (N) has the highest priority, generalization (G) and specialization (S) have lower priority, and the priority of the positive association (P) is the lowest. In Table I, the combination of the high priority relationship and the low priority relationship results in a relationship of high priority except the combination of generalization (G) and specialization (S) which results in positive association (P).

In order to describe the different relevance degrees and fuzzy relationships between documents and concepts, we can represent the documents by extended fuzzy sets which are fuzzy subsets of the set of concepts, where extended fuzzy sets are the generalization of fuzzy sets [27]. For example, let C be a set of concepts. Then, a document d_j can be represented as follows:

$$d_j = \{(c_i, \langle v_{d_j}(c_i), r_{d_j}(c_i)\rangle)|c_i \in C\}$$

where $v_{d_j}(c_i)$ represents the relevance degree between document d_j and concept c_i, v_{d_j}: $C \rightarrow [0,1]$, and $r_{d_j}(c_i)$ stands for the fuzzy relationship between the document d_j and the concept c_i, r_{d_j}: $C \rightarrow \{\mathrm{P},\mathrm{N},\mathrm{G},\mathrm{S}\}$.

A user's queries Q also can be represented by an extended fuzzy set shown as follows:

$$Q = \{(c_i, \langle v_Q(c_i), r_Q(c_i)\rangle)|c_i \in C\}$$

where $v_Q(c_i)$ represents the relevance degree between the query Q and concept c_i, v_Q: $C \rightarrow [0,1]$, and $r_Q(c_i)$ stands for the fuzzy relationship between the query Q and concept c_i, r_Q: $C \rightarrow \{\mathrm{P},\mathrm{N},\mathrm{G},\mathrm{S}\}$.

IV. RELATION MATRICES AND RELEVANCE MATRICES

In this section, we present the definitions of relation matrices and relevance matrices which can be used to model the extended fuzzy concept networks. The definitions of the transitive closure of relation matrices and the transitive closure of relevance matrices are also presented.

Definition 4.1: A relevance matrix V is a fuzzy matrix [13], where the element $V(c_i, c_j)$ represents the relevance degree between concepts c_i and c_j, and $V(c_i, c_j) \in [0,1]$. If $V(c_i, c_j) = 0$, then the relevance degree between concept c_i and concept c_j is not defined explicitly by the experts.

Definition 4.2: Assume that V is a relevance matrix

$$V = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1n} \\ V_{21} & V_{22} & \cdots & V_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ V_{n1} & V_{n2} & \cdots & V_{nn} \end{bmatrix}$$

where n is the number of concepts, $v_{ij} \in [0,1], 1 \leq i \leq n$, and $1 \leq j \leq n$. See (5), at the bottom of the page, where "$\vee$" is the maximum operator and "$\wedge$" is the minimum operator. Then, there exists a positive integer p, where $p \leq n-1$, such that $V^p = V^{p+1} = V^{p+2} = \cdots$. Let $T = V^p$, then T is called the transitive closure [13] of the relevance matrix V.

Definition 4.3: The relation matrix R is a fuzzy matrix, where the element $R(c_i, c_j)$ represents the fuzzy relationship between concept c_i and concept c_j, where $R(c_i, c_j) \in \{\mathrm{P},\mathrm{N},\mathrm{G},\mathrm{S},\mathrm{Z}\}$ and P, N, G, S stand for fuzzy positive association, fuzzy negative association, fuzzy generalization, and fuzzy specialization, respectively. If $R(c_i, c_j) = Z$, then the fuzzy relationship between concept c_i and concept c_j is not defined explicitly by the experts.

Let R be a relation matrix

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

where n is the number of concepts, $r_{ij} \in \{\mathrm{P},\mathrm{N},\mathrm{G},\mathrm{S},\mathrm{Z}\}, 1 \leq i \leq n$, and $1 \leq j \leq n$. See (6), at the bottom of the next page, where "$\boxed{\vee}$" is the operation of choosing the highest priority fuzzy relationship and "$\boxed{\wedge}$" is the operation of choosing the combination of two relationships according to Table II, where Table II is similar to Table I except that we add character "Z" to represent the relationship between concepts which is not explicitly defined by the experts. From Table II, we can see that the combination of two relationships of the same type results in a relationship of this type except for negative

$$V^2 = V \otimes V = \begin{bmatrix} \bigvee_{i=1,\cdots,n} (v_{1i} \wedge v_{i1}) & \bigvee_{i=1,\cdots,n} (v_{1i} \wedge v_{i2}) & \cdots & \bigvee_{i=1,\cdots,n} (v_{1i} \wedge v_{in}) \\ \bigvee_{i=1,\cdots,n} (v_{2i} \wedge v_{i1}) & \bigvee_{i=1,\cdots,n} (v_{2i} \wedge v_{i2}) & \cdots & \bigvee_{i=1,\cdots,n} (v_{2i} \wedge v_{in}) \\ \vdots & \vdots & \vdots & \vdots \\ \bigvee_{i=1,\cdots,n} (v_{ni} \wedge v_{i1}) & \bigvee_{i=1,\cdots,n} (v_{ni} \wedge v_{i2}) & \cdots & \bigvee_{i=1,\cdots,n} (v_{ni} \wedge v_{in}) \end{bmatrix} \quad (5)$$

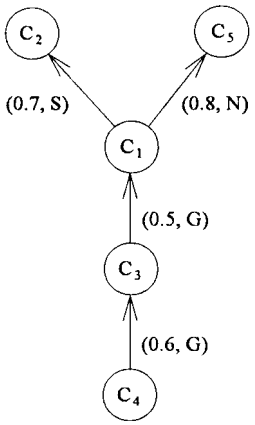

Fig. 4. An extended fuzzy concept network of Example 4.1.

associations (N) which get positive associations (P). Furthermore, in Table II, we let these five kinds of fuzzy relationships have different priorities, i.e., the negative association (N) has the highest priority, generalization (G) and specialization (S) have the second highest priority, the priority of the positive association (P) is lower, and the relationships (Z) not explicitly defined by the experts have the lowest priority. In Table II, the combination of the high priority relationship and the low priority relationship results in a relationship of high priority except the combination of generalization (G) and specialization (S) which results in positive association (P). Then, there exists a positive integer p, where $p \leq n-1$, such that $R^p = R^{p+1} = R^{p+2} = \cdots$. Let $L = R^p$, then L is called the transitive closure of relation matrix R.

Example 4.1: Assume that there is an extended fuzzy concept network as shown in Fig. 4, then, we can model this extended fuzzy concept network by the relevance matrix V and relation matrix R shown as follows:

$$V = \begin{bmatrix} 1 & 0.7 & 0.5 & 0 & 0.8 \\ 0.7 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 1 & 0.6 & 0 \\ 0 & 0 & 0.6 & 1 & 0 \\ 0.8 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{Z} & \mathrm{N} \\ \mathrm{G} & \mathrm{P} & \mathrm{Z} & \mathrm{Z} & \mathrm{Z} \\ \mathrm{G} & \mathrm{Z} & \mathrm{P} & \mathrm{S} & \mathrm{Z} \\ \mathrm{Z} & \mathrm{Z} & \mathrm{G} & \mathrm{P} & \mathrm{Z} \\ \mathrm{N} & \mathrm{Z} & \mathrm{Z} & \mathrm{Z} & \mathrm{P} \end{bmatrix}.$$

Then, based on the previous discussion, we can obtain the transitive closure T of the relevance matrix V and the transitive closure L of the relation matrix R as follows:

$$T = \begin{bmatrix} 1 & 0.7 & 0.5 & 0.5 & 0.8 \\ 0.7 & 1 & 0.5 & 0.5 & 0.7 \\ 0.5 & 0.5 & 1 & 0.6 & 0.5 \\ 0.5 & 0.5 & 0.6 & 1 & 0.5 \\ 0.8 & 0.7 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$$L = \begin{bmatrix} \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{G} & \mathrm{P} & \mathrm{P} & \mathrm{S} & \mathrm{N} \\ \mathrm{G} & \mathrm{P} & \mathrm{P} & \mathrm{S} & \mathrm{N} \\ \mathrm{G} & \mathrm{G} & \mathrm{G} & \mathrm{P} & \mathrm{N} \\ \mathrm{N} & \mathrm{N} & \mathrm{N} & \mathrm{N} & \mathrm{P} \end{bmatrix}.$$

V. Fuzzy Query Processing for Document Retrieval Based on Extended Fuzzy Concept Networks

In Section III, we have introduced that a document can be represented by an extended fuzzy set, where each concept represents a topic or an attribute. In this section, we use document descriptor relevance matrices and document descriptor relation matrices to represent documents, where the document descriptor relevance matrix is used to represent the relevance degrees between concepts and documents, and the document descriptor relation matrix is used to represent the fuzzy relationships between concepts and documents. The definitions of document descriptor relevance matrices and document descriptor relation matrices are presented as follows.

Definition 5.1: Let P be a set of documents, $P = \{d_1, d_2, \cdots, d_m\}$, and let C be a set of concepts, $C = \{c_1, c_2, \cdots, c_n\}$. The document descriptor relevance matrix D is shown as follows:

$$D = \begin{array}{c} \\ d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_m \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_n \end{array} \\ \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{bmatrix} \end{array}$$

where m is the number of documents, n is the number of concepts, v_{ij} stands for the relevance degree between document d_i and concept c_j, $v_{ij} \in [0,1], 1 \leq i \leq m$, and $1 \leq j \leq n$.

Definition 5.2: Let P be a set of documents, $P = \{d_1, d_2, \cdots, d_m\}$, and C be a set of concepts, $C = \{c_1, c_2, \cdots, c_n\}$. The document descriptor relation matrix M is shown as follows:

$$M = \begin{array}{c} \\ d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_m \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_n \end{array} \\ \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \end{array}$$

$$R^2 = R \otimes R = \begin{bmatrix} \bigvee_{i=1,\cdots,n} (r_{1i} \wedge r_{i1}) & \bigvee_{i=1,\cdots,n} (r_{1i} \wedge r_{i2}) & \cdots & \bigvee_{i=1,\cdots,n} (r_{1i} \wedge r_{in}) \\ \bigvee_{i=1,\cdots,n} (r_{2i} \wedge r_{i1}) & \bigvee_{i=1,\cdots,n} (r_{2i} \wedge r_{i2}) & \cdots & \bigvee_{i=1,\cdots,n} (r_{2i} \wedge r_{in}) \\ \vdots & \vdots & \vdots & \vdots \\ \bigvee_{i=1,\cdots,n} (r_{ni} \wedge r_{i1}) & \bigvee_{i=1,\cdots,n} (r_{ni} \wedge r_{i2}) & \cdots & \bigvee_{i=1,\cdots,n} (r_{ni} \wedge r_{in}) \end{bmatrix} \quad (6)$$

where m is the number of documents, n is the number of concepts, r_{ij} stands for the fuzzy relationship between document d_i and concept $c_j, r_{ij} \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}, \mathrm{Z}\}, 1 \le i \le m$, and $1 \le j \le n$.

In a document descriptor relevance matrix D and a document descriptor relation matrix M, the relevance degrees and fuzzy relationship between concepts and documents are given subjectively by experts. However, the experts may somehow forget to set some relevance degrees and fuzzy relationship between concepts and documents. In this case, we can obtain the implicit relevance degrees and fuzzy relationships between concepts and documents by means of the transitive closure T of the relation matrix V and the transitive closure L of the relevance matrix R, respectively. Let $D^* = D \otimes T$, then D^* is the document descriptor relevance matrix containing implied relevance degrees between concepts and documents. Let $M^* = M \otimes L$, then M^* is the document descriptor relation matrix containing implied fuzzy relationships between concepts and documents. The matrices D^* and M^* are used as a basis for similarity measures between queries and documents described later.

The user's query Q can be represented by a query descriptor relevance vector $\overline{qv}$ and a query descriptor relation vector $\overline{qr}$. In this case, if the user's query is shown as follows:

$$Q = \{(c_1, \langle x_1, y_1\rangle), (c_2, \langle x_2, y_2\rangle), \cdots, (c_n, \langle x_n, y_n\rangle)\}$$

then

$$\overline{qv} = \langle x_1, x_2, \cdots, x_n\rangle$$
$$\overline{qr} = \langle y_1, y_2, \cdots, y_n\rangle$$

where $x_i \in [0, 1]$ indicates the desired relevance degree of the document with respect to concept c_i, and $y_i \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}\}$ indicates the desired fuzzy relationship of the document with respect to concept c_i, and $1 \le i \le n$. In a query descriptor relevance vector $\overline{qv}$, if $x_i = 0$, then it indicates that documents desired by the user don't possess concept c_i. If $x_i =$ "-," then it indicates that the relevance degree of the desired documents with respect to concept c_i can be neglected. In a query descriptor relation vector $\overline{qr}$, if $y_i =$ "-," then it indicates that the fuzzy relationships of the desired documents with respect to concept c_i can be neglected. If $y_i =$ "N," then the user wants to perform a negative query, i.e., there is a negative relationship between the desired documents and concept c_i. If $y_i =$ "G," then the user wants to perform a generalization query, i.e., there is a generalization relationship between the desired documents and concept c_i. If $y_i =$ "S," then the user wants to perform a specialization query, i.e., there is a specialization relationship between the desired documents and concept c_i.

Let $\langle x, s\rangle$ and $\langle y, t\rangle$ be two pairs of values where $x \in [0, 1], y \in [0, 1], s \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}\}$, and $t \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}\}$, then the degree of similarity between $\langle x, s\rangle$ and $\langle y, t\rangle$ can be evaluated by the function T,

$$T(\langle x, s\rangle, \langle y, t\rangle) = \begin{cases} 0 & \text{if } s \ne t \\ 1 - |x - y| & \text{if } s = t \end{cases} \tag{7}$$

where $T(\langle x, s\rangle, \langle y, t\rangle) \in [0, 1]$. The larger the value of $T(\langle x, s\rangle, \langle y, t\rangle)$, the more the similarity between $\langle x, s\rangle$ and $\langle y, t\rangle$. Assume that the document descriptor relevance vector $\overline{dv_i}$ (i.e., the ith row of the document descriptor relevance matrix D^*), the document descriptor relation vector $\overline{dr_i}$ (i.e., the ith row of the document descriptor relation matrix M^*), the query descriptor relevance vector $\overline{qv}$ and the query descriptor relation vector $\overline{qr}$ are represented as follows:

$$\overline{dv_i} = \langle s_{i1}, s_{i2}, \cdots, s_{in}\rangle$$
$$\overline{dr_i} = \langle t_{i1}, t_{i2}, \cdots, t_{in}\rangle$$
$$\overline{qv} = \langle x_1, x_2, \cdots, x_n\rangle$$
$$\overline{qr} = \langle y_1, y_2, \cdots, y_n\rangle$$

where $s_{ij} \in [0, 1], x_i \in [0, 1], t_{ij} \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}, \mathrm{Z}\}, y_i \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}, \mathrm{Z}\}, 1 \le j \le n, 1 \le i \le m, n$ is the number of concepts, and m is the number of documents. Let $\overline{qv(j)}$ and $\overline{qr(j)}$ be the jth element of the query descriptor relevance vector $\overline{qv}$ and the jth element of the query descriptor relation vector $\overline{qr}$, respectively. If $\overline{qv(j)} =$ "-" or $\overline{qr(j)} =$ "-," then it indicates that concept c_j is neglected by the user's query. The degree of satisfaction that document d_i satisfies the user's query Q can be evaluated by

$$RS(d_i) = \frac{\displaystyle\sum_{\overline{qv(j)} \ne \text{"-"} \text{ and } \overline{qr(j)} \ne \text{"-"} \text{ and } j=1,\cdots,n} T(\langle s_{ij}, t_{ij}\rangle, \langle x_j, y_j\rangle)}{k} \tag{8}$$

where $RS(d_i) \in [0, 1], 1 \le i \le m$, and k is the number of concepts not neglected by the user's query. The larger the value of $RS(d_i)$, the more the degree of satisfaction that the document d_i satisfies the user's query. In a fuzzy information retrieval system, we also can set up a retrieval threshold value λ, where $\lambda \in [0, 1]$. If $RS(d_i) \ge \lambda$, which indicates that document d_i satisfies the user's query. The information retrieval system would display every document having a retrieval status value greater than the threshold value λ, where $\lambda \in [0, 1]$, in a sequential order from the document with the highest retrieval status value to that with the lowest one.

Example 5.1: Consider the extended fuzzy concept network shown in Example 4.1, where the extended fuzzy concept network has been modeled by the relevance matrix V and the relation matrix R as shown in Example 4.1, we can see that the transitive closure T of the relevance matrix R and the transitive closure L of the relation matrix R follows:

$$T = \begin{bmatrix} 1 & 0.7 & 0.5 & 0.5 & 0.8 \\ 0.7 & 1 & 0.5 & 0.5 & 0.7 \\ 0.5 & 0.5 & 1 & 0.6 & 0.5 \\ 0.5 & 0.5 & 0.6 & 1 & 0.5 \\ 0.8 & 0.7 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$$L = \begin{bmatrix} \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{G} & \mathrm{P} & \mathrm{P} & \mathrm{S} & \mathrm{N} \\ \mathrm{G} & \mathrm{P} & \mathrm{P} & \mathrm{S} & \mathrm{N} \\ \mathrm{G} & \mathrm{G} & \mathrm{G} & \mathrm{P} & \mathrm{N} \\ \mathrm{N} & \mathrm{N} & \mathrm{N} & \mathrm{N} & \mathrm{P} \end{bmatrix}.$$

Assume that there are five documents in a fuzzy information retrieval system, and the document descriptor relevance matrix D and the document descriptor relation matrix M are shown as follows:

$$D = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0.5 & 1 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.6 & 0 \\ 0.8 & 1 & 1 & 1 & 0 \\ 0.4 & 0.9 & 0 & 0 & 1 \end{bmatrix}$$

$$M = \begin{bmatrix} \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{Z} & \mathrm{Z} \\ \mathrm{G} & \mathrm{P} & \mathrm{Z} & \mathrm{S} & \mathrm{Z} \\ \mathrm{Z} & \mathrm{Z} & \mathrm{Z} & \mathrm{S} & \mathrm{Z} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{Z} \\ \mathrm{P} & \mathrm{S} & \mathrm{Z} & \mathrm{Z} & \mathrm{N} \end{bmatrix}.$$

Then, based on the previous discussions, the document descriptor relevance matrix D^* and the document descriptor relation matrix M^* can be obtained as follows:

$$D^* = D \otimes T = \begin{bmatrix} 1 & 1 & 1 & 0.6 & 0.8 \\ 0.7 & 1 & 0.6 & 0.7 & 0.7 \\ 0.5 & 0.5 & 0.6 & 0.6 & 0.5 \\ 0.8 & 0.7 & 1 & 0.6 & 0.8 \\ 0.8 & 0.9 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$$M^* = M \otimes Q = \begin{bmatrix} \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{G} & \mathrm{P} & \mathrm{P} & \mathrm{S} & \mathrm{N} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \end{bmatrix}.$$

If the user's query represented by the query descriptor relevance vector $\overline{qv}$ and the query descriptor relation vector $\overline{qr}$ are as follows:

$$\overline{qv} = \langle 0.6, 1, 0.8, \text{-}, 0.7 \rangle$$
$$\overline{qr} = \langle \mathrm{P}, \mathrm{S}, \mathrm{G}, \text{-}, \mathrm{N} \rangle$$

then based on (7) and (8), the degree of satisfaction that each document satisfies the user's query can be evaluated shown as follows:

$$RS(d_1) = 0.625$$
$$RS(d_2) = 0.25$$
$$RS(d_3) = 0.55$$
$$RS(d_4) = 0.6$$
$$RS(d_5) = 0.6.$$

If the retrieval threshold given by the user is $\lambda = 0.5$, then we can see that document d_2 is not suitable to the user's query due to the fact that the retrieval status value of the document d_2 is less than the retrieval status value λ (where $\lambda = 0.5$). Furthermore, we also can see that the documents which satisfy the user's query are d_1, d_4, d_5, d_3. In this case, document d1 is the best choice for the user's query, due to the fact that it has the largest retrieval status value.

Consider the following OR-connected query Q

$$Q = Q_1 \text{ or } Q_2$$

where

$$Q_1 = \{(c_1, \langle x_{11}, y_{11} \rangle), (c_2, \langle x_{12}, y_{12} \rangle), \cdots (c_n, \langle x_{1n}, y_{1n} \rangle)\},$$
$$Q_2 = \{(c_1, \langle x_{21}, y_{21} \rangle), (c_2, \langle x_{22}, y_{22} \rangle), \cdots (c_n, \langle x_{2n}, y_{2n} \rangle)\}$$

then the sub-query Q_1 can be represented by a query descriptor relevance vector $\overline{qv_1}$ and a query descriptor relation vector $\overline{qr_1}$; the sub-query Q_2 can be represented by a query descriptor relevance vector $\overline{qv_2}$ and a query descriptor relation vector $\overline{qr_2}$, where

$$\overline{qv_1} = \langle x_{11}, x_{12}, \cdots, x_{1n} \rangle$$
$$\overline{qr_1} = \langle y_{11}, y_{12}, \cdots, y_{1n} \rangle$$
$$\overline{qv_2} = \langle x_{21}, x_{22}, \cdots, x_{2n} \rangle$$
$$\overline{qr_2} = \langle y_{21}, y_{22}, \cdots, y_{2n} \rangle$$

where $x_{tj} \in [0,1], y_{tj} \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}\}, 1 \le t \le 2$, and $1 \le j \le n$.

Assume that the document descriptor relevance vector $\overline{dv_i}$ (i.e., the ith row of the document relevance matrix D^*) and the document descriptor relation vector $\overline{dr_i}$ (i.e., the ith row of the document relation matrix M^*) are as follows:

$$\overline{dv_i} = \langle s_{i1}, s_{i2}, \cdots, s_{in} \rangle$$
$$\overline{dr_i} = \langle t_{i1}, t_{i2}, \cdots, t_{in} \rangle$$

where $s_{ij} \in [0,1], t_{ij} \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}, \mathrm{Z}\}, 1 \le i \le m$, and $1 \le j \le n$. Then, based on formulas (7) and (8), the degree of similarity between the sub-query Q_i and the documents can be expressed by a $1 \le m$ matrix RS_i, where m is the number of document and $1 \le i \le 2$. In this case, the degree of similarity between the user's query Q and the documents can be calculated as follows:

$$RS^*(d_i) = \max\left(RS_1(d_i), RS_2(d_i)\right) \tag{9}$$

where

$$RS_1(d_i) = \frac{\sum_{\overline{qv_1(j)} \neq \text{“-”} \text{ and } \overline{qr_i(j)} \neq \text{“-”} \text{ and } j=1,\cdots,n} T(\langle s_{ij}, t_{ij} \rangle, \langle x_{1j}, y_{1j} \rangle)}{k_1} \tag{10}$$

$$RS_2(d_i) = \frac{\sum_{\overline{qv_2(j)} \neq \text{“-”} \text{ and } \overline{qr_2(j)} \neq \text{“-”} \text{ and } j=1,\cdots,n} T(\langle s_{ij}, t_{ij} \rangle, \langle x_{2j}, y_{2j} \rangle)}{k_2} \tag{11}$$

where k_1 is the number of concepts not neglected by the sub-query Q_1, k_2 is the number of concepts not neglected by the sub-query $Q_2, RS_1(d_i) \in [0,1], RS_2(d_i) \in [0,1]$, and $1 \le i \le m$. $RS_1(d_i)$ represents the degree of similarity between the sub-query Q_1 and document d_i; $RS_2(d_i)$ represents the degree of similarity between the sub-query Q_2 and document d_i; the retrieval status value $RS^*(d_i)$ represents the degree of similarity of the user's query Q with respect to document d_i, and $1 \le i \le m$. The fuzzy information retrieval system would display every document having a retrieval status value greater than the threshold value λ in a sequential order from the document with the highest degree of retrieval status value to that with the lowest one, where $\lambda \in [0,1]$.

Example 5.2: Same assumption as in Example 5.1, where the retrieval status value λ given by the user is 0.5 (i.e., $\lambda = 0.5$), and the document descriptor relevance matrix D^* and document descriptor relation matrix M^* are as follows:

$$D^* = \begin{bmatrix} 1 & 1 & 1 & 0.6 & 0.8 \\ 0.7 & 1 & 0.6 & 0.7 & 0.7 \\ 0.5 & 0.5 & 0.6 & 0.6 & 0.5 \\ 0.8 & 0.7 & 1 & 0.6 & 0.8 \\ 0.8 & 0.9 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$$M^* = \begin{bmatrix} \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{G} & \mathrm{P} & \mathrm{P} & \mathrm{S} & \mathrm{N} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \end{bmatrix}.$$

Assume that the user's query Q is as follows:

$$Q = Q_1 \,\mathrm{OR}\, Q_2$$

where the sub-query Q_1 can be represented by the query descriptor relevance vector $\overline{qv_1}$ and the query descriptor relation vector $\overline{qr_1}$ shown as follows:

$$\overline{qv_1} = \langle 0.6, 1, 0.8, \text{-}, 0.7 \rangle$$
$$\overline{qr_1} = \langle \mathrm{P}, \mathrm{S}, \mathrm{G}, \text{-}, \mathrm{N} \rangle$$

and the sub-query Q_2 can be represented by the query descriptor relevance vector $\overline{qv_2}$ and the query descriptor relation vector $\overline{qr_2}$ shown as follows:

$$\overline{qv_2} = \langle 0.9, \text{-}, \text{-}, \text{-}, \text{-} \rangle$$
$$\overline{qr_2} = \langle \mathrm{P}, \text{-}, \text{-}, \text{-}, \text{-} \rangle.$$

Then, based on formula (10), we can get

$$RS_1(d_1) = 0.625$$
$$RS_1(d_2) = 0.25$$
$$RS_1(d_3) = 0.55$$
$$RS_1(d_4) = 0.6$$
$$RS_1(d_5) = 0.6.$$

Based on formula (11), we can get

$$RS_2(d_1) = 0.95$$
$$RS_2(d_2) = 0$$
$$RS_2(d_3) = 0.55$$
$$RS_2(d_4) = 0.85$$
$$RS_2(d_5) = 0.85.$$

Furthermore, based on (9), we can get

$$RS^*(d_1) = \max(0.625, 0.95) = 0.95$$
$$RS^*(d_2) = \max(0.25, 0) = 0.25$$
$$RS^*(d_3) = \max(0.55, 0.55) = 0.55$$
$$RS^*(d_4) = \max(0.6, 0.85) = 0.85$$
$$RS^*(d_5) = \max(0.6, 0.85) = 0.85.$$

Because the retrieval status value λ given by the user is 0.5 (i.e., $\lambda = 0.5$), we can see that the document d_2 is not suitable to the user's query due to the fact that the retrieval status value of the document d_2 is less than the retrieval threshold value λ (where $\lambda = 0.5$). In this case, the documents which satisfy the user's query are d_1, d_3, d_4, and d_5. Furthermore, we also can see that the document d_1 is the best choice for the user's query due to the fact that it has the largest retrieval status value.

Weighted queries can also be processed by our method. Assume that there are n concepts in a fuzzy information retrieval system, and assume that the weight of the concept c_j given by the user is w_j, where $w_j \in [0,1]$, and $\Sigma_{j=1}^{n} w_j = 1$. Furthermore, assume that the user's query is shown as follows:

$$Q = \{(c_1, \langle x_1, y_1\rangle), (c_2, \langle x_2, y_2\rangle), \cdots, (c_n, \langle x_n, y_n\rangle)\}$$

where $x_i \in [0,1]$, which indicates the desired relevance degree of the document with respect to concept c_i, $y_i \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}\}$ indicates the desired fuzzy relationship of the document with respect to concept c_i, and $1 \le i \le n$. Based on the previous discussions, the user's query Q can be represented by a query descriptor relevance vector $\overline{qv}$ and a query descriptor relation vector $\overline{qr}$, where

$$\overline{qv} = \langle x_1, x_2, \cdots, x_n\rangle$$
$$\overline{qr} = \langle y_1, y_2, \cdots, y_n\rangle.$$

Assume that the ith row of the document descriptor relevance matrix D^* be $\langle s_{i1}, s_{i2}, \cdots, s_{in}\rangle$ and assume that the ith row of the document descriptor relation matrix M^* be $\langle t_{i1}, t_{i2}, \cdots, t_{in}\rangle$, where $s_{ij} \in [0,1], t_{ij} \in \{\mathrm{P}, \mathrm{N}, \mathrm{G}, \mathrm{S}, \mathrm{Z}\}, 1 \le i \le m$, and $1 \le j \le n$. Then, the degree of similarity between the user's query Q and the document d_i can be calculated as follows:

$$RS_w^*(d_i) = \sum_{\overline{qv(j)} \neq \text{"-"} \text{ and } \overline{qr(j)} \neq \text{"-"} \text{ and } j=1,\cdots,n} \cdot T(\langle s_{ij}, t_{ij}\rangle, \langle x_j, y_j\rangle) \times w_j \quad (12)$$

where the retrieval status value $RS_w^*(d_i)$ indicates the degree of similarity between the user's query Q and the document d_i, $RS_w^*(d_i) \in [0,1]$, and $1 \le i \le m$. The system would display every document having a retrieval status value greater than the threshold value λ in a sequential order from the document with the highest degree of retrieval status value to that with the lowest one, where $\lambda \in [0,1]$.

Example 5.3: Same assumption as in Example 5.1, where the retrieval status value λ given by the user is 0.5 (i.e., $\lambda = 0.5$), and the document descriptor relevance matrix D^* and the document descriptor relation matrix M^* are as follows:

$$D^* = \begin{bmatrix} 1 & 1 & 1 & 0.6 & 0.8 \\ 0.7 & 1 & 0.6 & 0.7 & 0.7 \\ 0.5 & 0.5 & 0.6 & 0.6 & 0.5 \\ 0.8 & 0.7 & 1 & 0.6 & 0.8 \\ 0.8 & 0.9 & 0.5 & 0.5 & 1 \end{bmatrix}$$

$$M^* = \begin{bmatrix} \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{G} & \mathrm{P} & \mathrm{P} & \mathrm{S} & \mathrm{N} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \\ \mathrm{P} & \mathrm{S} & \mathrm{S} & \mathrm{S} & \mathrm{N} \end{bmatrix}.$$

Assume that the user's query represented by the query descriptor relevance vector $\overline{qv}$ and the query descriptor relation vector $\overline{qr}$ are as follows:

$$\overline{qv} = \langle 0.6, 1, 0.8, \text{-}, 0.7\rangle$$
$$\overline{qr} = \langle \mathrm{P}, \mathrm{S}, \mathrm{G}, \text{-}, \mathrm{N}\rangle$$

and assume that the weights of the concepts c_1, c_2, c_3, and c_5 given by the user are 0.4, 0.4, 0.1, and 0.1, respectively, then based on formula (12), we can get

$$RS_w^*(d_1) = 0.6 * 0.4 + 1 * 0.4 + 0 * 0.1 + 0.9 * 0.1 = 0.73,$$
$$RS_w^*(d_2) = 0 * 0.4 + 0 * 0.4 + 0 * 0.1 + 1 * 0.1 = 0.1,$$
$$RS_w^*(d_3) = 0.9 * 0.4 + 0.5 * 0.4 + 0 * 0.1 + 0.8 * 0.1 = 0.64,$$
$$RS^*w(d_4) = 0.8 * 0.4 + 0.7 * 0.4 + 0 * 0.1 + 0.9 * 0.1 = 0.69,$$
$$RS_w^*(d_5) = 0.8 * 0.4 + 0.9 * 0.4 + 0 * 0.1 + 0.7 * 0.1 = 0.75.$$

Because the retrieval status value λ given by the user is 0.5 (i.e., $\lambda = 0.5$), we can see that the documents which satisfy the user's query are d_1, d_3, d_4, and d_5, where the document d_2 is not suitable to the user's query due to the fact that the retrieval status value of the document d_2 is less than the retrieval status value λ (where $\lambda = 0.5$). In this case, document d_5 is the best choice for the user's query due to the fact that it has the largest retrieval status value.

VI. Conclusions

In this paper, we have presented the concepts of extended fuzzy concept networks, where there are four kinds of fuzzy relationships between concepts in an extended fuzzy concept network, i.e., fuzzy positive association, fuzzy negative association, fuzzy generalization, and fuzzy specialization. We also presented a fuzzy information retrieval method based on the extended fuzzy concept networks for document retrieval. The proposed method is more flexible and more intelligent than the ones presented in [8] and [17] due to the fact that it allows the users to perform positive queries, negative queries, generalization queries, and specialization queries. The proposed method allows the users to perform fuzzy queries in a more flexible and more intelligent manner.

REFERENCES

[1] S. M. Chen, "A new approach to handling fuzzy decision making problems," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, pp. 1012–1016, 1988.

[2] ____, "An improved algorithm for inexact reasoning based on extended fuzzy production rules," *Cybern. Syst.: Int. J.*, vol. 23, no. 5, pp. 463–481, 1992.

[3] ____, "A new approach to inexact reasoning for rule-based systems," *Cybern. Syst.: An Int. J.*, vol. 23, no. 6, pp. 561–582, 1992.

[4] ____, "Techniques for handling multicriteria fuzzy decision-making problems," in *Proc. 4th Int. Symp. Computer Information Sciences*, Cesme, Turkey, vol. 2, pp. 919–925, 1989.

[5] S. M. Chen, J. S. Ke, and J. F. Chang, "Knowledge representation using fuzzy Petri nets," *IEEE Trans. Knowl. Data Eng.*, vol. 2, pp. 311–319, 1990.

[6] ____, "An inexact reasoning algorithm for dealing with inexact knowledge," *Int. J. Softw. Eng. Knowl. Eng.*, vol. 1, no. 3, pp. 227–244, 1991.

[7] S. M. Chen and Y. J. Horng, "Finding inheritance hierarchies in interval-valued fuzzy concept-networks," *Fuzzy Sets Syst.*, vol. 84, no. 1, pp. 75–83, 1996.

[8] S. M. Chen and J. Y. Wang, "Document retrieval using knowledge-based fuzzy information retrieval techniques," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, pp. 793–803, May 1995.

[9] G. T. Her and J. S. Ke, "A fuzzy information retrieval system model," in *Proc. 1983 National Computer Symp.*, Taiwan, R.O.C., 1983, pp. 147–151.

[10] Y. J. Horng and S. M. Chen, "Document retrieval based on extended fuzzy concept networks," in *Proc. 4th Nat. Conf. Defense Management*, Taipei, Taiwan, R.O.C., 1996, vol. 2, pp. 1039–1050.

[11] I. Itzkovich and L. W. Hawkes, "Fuzzy extension of inheritance hierarchies," *Fuzzy Sets Syst.*, vol. 62, no. 2, pp. 143–153, 1994.

[12] M. Kamel, B. Hadfield, and M. Ismail, "Fuzzy query processing using clustering techniques," *Inf. Process. Manage.*, vol. 26, no. 2, pp. 279–293, 1990.

[13] A. Kandel, *Fuzzy Mathematical Techniques with Applications*. Reading, MA: Addison-Wesley, 1986.

[14] M. Kracker, "A fuzzy concept network model and its applications," in *Proc. 1st IEEE Int. Conf. Fuzzy Systems*, 1992, pp. 761–768.

[15] D. H. Kraft and D. A. Buell, "Fuzzy sets and generalized Boolean retrieval systems," *Int. J. Man-Mach. Stud.*, vol. 19, no. 1, pp. 45–56, 1983.

[16] C. G. Looney, "Fuzzy Petri nets for rule-based decision making," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, pp. 178–183, 1988.

[17] D. Lucarella and R. Morara, "FIRST: Fuzzy information retrieval system," *J. Inf. Sci.*, vol. 17, pp. 81–91, 1991.

[18] T. Murai, M. Miyakoshi, and M. Shimbo, "A fuzzy document retrieval method based on two-valued indexing," *Fuzzy Sets Syst.*, vol. 30, pp. 103–120, 1989.

[19] S. Miyamoto, "Information retrieval based on fuzzy associations," *Fuzzy Sets Syst.*, vol. 38, pp. 191–205, 1990.

[20] T. Radechi, "Mathematical model of time effective information retrieval system based on the theory of fuzzy set," *Inf. Process. Manage.*, vol. 13, pp. 109–116, 1977.

[21] T. Radechi, "Fuzzy set theoretical approach to document retrieval," *Inf. Process. Manage.*, vol. 15, vol. 247–259, 1979.

[22] ____, "Generalized Boolean methods of information retrieval," *Int. J. Man-Mach. Stud.*, vol. 18, no. 5, pp. 409–439, 1983.

[23] R. Rousseau, "On relative indexing in fuzzy retrieval systems," *Inf. Process. Manage.*, vol. 21, no. 5, pp. 415–417, 1985.

[24] G. Salton and M. J. Mcgill, *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1983.

[25] V. Tahani, "A fuzzy model of document retrieval system," *Inf. Process. Manage.*, vol. 12, pp. 177–187, 1976.

[26] J. Y. Wang and S. M. Chen, "A knowledge-based method for fuzzy information retrieval," in *Proc. 1st Asian Fuzzy Systems Symp.*, Singapore, 1993.

[27] L. A. Zadeh, "Fuzzy sets," *Inf. Contr.*, vol. 8, pp. 338–353, 1965.

[28] M. Zemankova, "FIIS: A fuzzy intelligent information system," *Data Eng.*, vol. 12, no. 2, pp. 11–20, 1989.

[29] R. Zwick, E. Carlstein, and D. V. Budescu, "Measures of similarity among fuzzy concepts: A comparative analysis," *Int. J. Approx. Reas.*, vol. 1, pp. 221–242, 1987.

Process and Data Nets: The Conceptual Model of the M^*-OBJECT Methodology

Giuseppe Berio, Antonio di Leva,
Piercarlo Giolito, and Francois Vernadat

***Abstract*—The paper describes a specification model, called the Process and Data Net (PDN) model, used as the modeling tool for the M^*-OBJECT information system design methodology. The model integrates the representation of static, dynamic, and behavioral aspects of a database application. PDN consists of two components: an object-oriented data model that describes static and behavioral aspects of objects of the system under analysis, and a process model that specifies a way organization activities must be coordinated. The major features of the proposed approach are: 1) the system representation captures all relevant properties from the end-user viewpoint without unnecessary details concerning implementation, 2) complex data structures and data manipulations can be specified, and 3) specifications are executable for rapid prototyping.**

I. INTRODUCTION

To design and develop an information system which supports modern (and complex) applications, such as office automation (OA), computer-integrated manufacturing (CIM), or software engineering, a sound conceptual design methodology is required. This problem has been recognized to be of strategic importance for CIM by several international projects such as CIMOSA [1] and the Purdue Enterprise Reference Architecture [2].

A methodology consists of *models* used to describe a real-world system under consideration (e.g., an enterprise), and *methods*, i.e., design strategies to elaborate the real-world description. First approaches devoted to conceptual modeling were mostly concerned with the representation of static properties, i.e., the modeling of *data structures* (to represent components of the object system) and *integrity constraints* on data (i.e., rules that data must satisfy). Significant models of this type, such as the Entity-Relationship model [3] and the semantic data model [4], have been developed.

The need to integrate the specification of **dynamic properties** in the system representation was soon recognized. Dynamic properties refer to *processes*. Processes are made of *activities* (linked by causal relationships) that must be specified to describe the system organization, and *dynamic constraints*, i.e., rules which must be satisfied by processes under analysis [5]. Several approaches have been proposed to take into account dynamic properties. Some of them (e.g., REMORA [6], TAXIS [7], or TEMPORA [8]) are based on the concept of an *event*, which can be seen as a control mechanism

Manuscript received March 18, 1995; revised December 15, 1995 and January 6, 1997.

G. Berio was with the Dipartimento di Informatica, Universitá di Torino, Corso Svizzera 185, I-10149 Torino, Italy. He is now with LGIPM, Université de Metz, F-57012 Metz, France.

A. DiLeva and P. Giolito are with the Dipartimento di Informatica, Universitá di Torino, I-10149 Torino, Italy.

F. Vernadat is with LGIPM, Université de Metz, F-57012 Metz, France.

Publisher Item Identifier S 1083-4419(99)00777-3.