[10] D. R. Hush and B. G. Horne, "Progress in supervised neural networks," *IEEE Signal Processing Mag.*, pp. 8–39, Jan. 1993.
[11] R. C. Lacher, S. I. Hruska, and D. C. Kuncicky, "Back-propagation learning in expert networks," *IEEE Trans. Neural Networks*, vol. 3, pp. 62–72, Jan. 1992.
[12] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representation by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, vol. 1. Cambridge, MA: MIT Press, 1986.
[13] L. G. Valiant, "A theory of the learnable," *Commun. ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.
[14] V. N. Vapnik and A. Ya. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theor. Probab. Applicat.*, vol. 16, no. 2, pp. 264–280, 1971.
[15] Y. Xie and M. A. Jabri, "Analysis of the effects of quantization in multilayer neural networks using a statistical model," *IEEE Trans. Neural Networks*, vol. 3, pp. 334–338, Mar. 1992.

Supporting Conceptual and Neighborhood Queries on the World Wide Web

Chih-Shyang Chang and Arbee L. P. Chen

***Abstract*—A document retrieval system mainly consists of three components: document representation, user queries, and document evaluation. Each component may involve some uncertainties. Fuzzy set theory is a natural approach to coping with the representation of documents, queries, and the relevance of documents to a given query. We propose a fuzzy document retrieval model on the World Wide Web (WWW) environment to support *conceptual queries*. A flexible query expression is proposed to support different semantics of the queries. A *concept network* is adopted as the knowledge base to represent the relevance of the concepts. The concept network is explored from the WWW. Moreover, we also support *neighborhood queries*, which retrieve documents relevant to a document specified by a user. A system is currently being implemented to achieve these functions.**

***Index Terms*—Concept network, fuzzy set, information retrieval, World Wide Web (WWW).**

I. Introduction

The World Wide Web (WWW) provides a resource repository. To discover the needed documents on the WWW, we can navigate the WWW from links by a web browser or a search engine provided by the web developers [1]. Currently, there exist many search engines, such as Aliweb, Alta Vista, InfoSeek, Lycos, Yahoo, WebCrawler, and WWWW. Most of these search engines are *word-oriented systems*. When a query is submitted, a word-oriented system tracks down all documents containing some of the words in the query phrase. For example, given the request "intellectual property rights," the system will return documents that contain one or more of the words in the request. There may be other documents about intellectual property rights that do not contain these words. These documents will not be retrieved by the system.

There exist three sources [7] of uncertainty in the document retrieval system [3], [14].

1) Uncertainty of the user request: the user may be looking for documents whose specifications are not precisely known. Therefore, it is unlikely that the actual user needs be exactly reflected in the submitted query.
2) Uncertainty of the document representation: the representation of a document is difficult to build to reflect the complete content of the document. We can only provide an uncertain representation of the document content.
3) Uncertainty of the relationship between the user request and the request result: a result document should be considered as a degree of relevance to the actual need of the user. Hence, there exists an uncertain relationship between the request and the request result.

Fuzzy set theory [17], [20] provides a sound mathematical framework to deal with the uncertainty of document representation, query specification, and the document retrieval process. The approach, starting from the initial works of Tahani [15] and Radecki [12], [13], has been widely investigated [16], [18], [19]. Tahani [15] addressed an organization of document files and a strategy for the document retrieval by using basic notions and operations of the theory of fuzzy set. Miyamoto [8], [9] proposed a formulation of information retrieval based on the fuzzy thesaurus [10] and the citation analysis [5]. In [6] and [7], Lucarella *et al.* proposed a document retrieval technique based on fuzzy reasoning. This work differs from previous ones in that a knowledge-base approach is proposed. The knowledge-base can be achieved by the *concept network* [7], [11], such that queries and documents can be uniformly represented by a set of concepts with a *relevance degree*. A strategy to reduce the search space is also proposed. Chen and Wang [2] represented the concept network as a *concept matrix*. The concept matrix and its transitive closure are used to deal with document retrieval. The proposed query expression allows negative conditions. However, there exist problems in query processing.

In this paper, we propose a framework of document retrieval on the WWW based on fuzzy sets. We introduce a flexible query expression by combining *range query* and *point query* to solve the problems of Chen and Wang's work. To support the conceptual query, we adopt the concept network to specify the relationships among the concepts. Moreover, the concept network can be constructed by exploring the concept hierarchies from the WWW. The *neighborhood query*, which can retrieve the documents that are relevant to a given document, is also supported. In order to measure the similarity of the documents, an evaluation model is proposed. Consequently, a system designed to support these features is being implemented.

The paper is organized as follows. The basic concepts of fuzzy sets and a fuzzy document retrieval model are introduced in Section II. The conceptual query, the concept network, the semantics of query expressions, and the flexible query expression and processing are addressed in Section III. To support neighborhood queries, we measure the *degree of relevance between two documents* using the document contents and/or the link relationships of the documents in Section IV. In Section V, the exploration of the concept network is considered. A system, WORDS, presented in Section VI, is being implemented to achieve these functions.

Manuscript received November 24, 1996; revised October 3, 1997. This work was supported in part by the National Science Council of the Republic of China under Contract NSC 87-2213-E-007-029.

C.-S. Chang is with Trilogy Technologies, Inc., Taipei 300, Taiwan, R.O.C. (e-mail: dr808304@cs.nthu.edu.tw).

A. L. P. Chen is with the Department of Computer Science, National Tsing Hua University, Hsinchu 30043, Taiwan, R. O. C. (e-mail: alpchen@cs.nthu.edu.tw).

Publisher Item Identifier S 1094-6977(98)02569-3.

II. Fuzzy Document Retrieval System

A *fuzzy set* [20] consists of data items and their corresponding grades of membership in the set. The *grade of membership* of a data item in the fuzzy set is given by a subjectively defined membership function, and its value can range from zero to one, where the value of one denotes full membership.

Definition 1: Let U be the universe of discourse. A *fuzzy subset* S of U is characterized by a *membership function* $\mu_S : U \rightarrow [0,1]$, which associates with each element $u \in U$, $\mu_S(u)$ representing the *grade of membership* of u in S. S is denoted by $\{(\mu_S(u), u) \mid u \in U\}$.

Other widely used notations are

$$S = \int_U \mu_S(u)/u$$

when U is a continuum and

$$S = \{\mu_S(u)/u \mid u \in U\}$$

when U is a finite or countable set.

The basic operations that can be performed on the fuzzy set and used in the query processing are shown as follow.

Definition 2: Let A and B be two fuzzy subsets of a universe of discourse U, characterized by the membership functions μ_A and μ_B, respectively. The *union* of A and B is denoted by $A \vee B$ and is defined by

$$A \vee B = \{\max(\mu_A(u), \mu_B(u))/u \mid u \in U\}.$$

Definition 3: Let A and B be two fuzzy subsets of a universe of discourse U, characterized by the membership functions μ_A and μ_B, respectively. The *intersection* of A and B is denoted by $A \wedge B$ and is defined by

$$A \wedge B = \{\min(\mu_A(u), \mu_B(u))/u \mid u \in U\}.$$

Definition 4: Let A be a fuzzy subset of a universe of discourse U, characterized by the membership function μ_A. The *complement* of A is denoted by $\neg A$ and is defined by

$$\neg A = \{(1 - \mu_A(u))/u \mid u \in U\}.$$

Definition 5: A *binary fuzzy relation* r from U_1 to U_2 is a fuzzy subset of $U_1 \times U_2$, where U_1 and U_2 are two universes of discourse, characterized by a membership function, as follows:

$$\mu_r : U_1 \times U_2 \rightarrow [0,1].$$

We define a fuzzy document retrieval system $\mathcal{S}$ as follows.

Definition 6: A model of a fuzzy document retrieval system $\mathcal{S}$ is given by

$$\langle \mathcal{H}, \mathcal{C}, \mathcal{Q}, I, \mathcal{K}, \phi, \psi \rangle$$

where $\mathcal{H}$ is a set of documents, $\mathcal{C}$ is a set of concepts, $\mathcal{Q}$ is a set of queries, I is a binary fuzzy *indexing* relation from $\mathcal{H}$ to $\mathcal{C}$, $\mathcal{K}$ is a knowledge base, $\phi : \mathcal{Q} \times \mathcal{H} \rightarrow [0,1]$ is a retrieval function, and $\psi : \mathcal{H} \times \mathcal{H} \rightarrow [0,1]$ is a relevance function. For each pair $(q,h), q \in \mathcal{Q}, h \in \mathcal{H}$, $\phi(q,h) \in [0,1]$ is called the *retrieval status value*. For each pair $(h_1,h_2), h_1, h_2 \in \mathcal{H}$, $\psi(h_1,h_2) \in [0,1]$ is called the *degree of relevance between h_1 and h_2* or *relevance degree between h_1 and h_2*.

The binary fuzzy indexing relation I is represented as the form

$$I = \{\mu_I(h,c)/(h,c) \mid h \in \mathcal{H}, c \in \mathcal{C}\}$$

with a membership function $\mu_I : \mathcal{H} \times \mathcal{C} \rightarrow [0,1]$, indicating for each pair (h,c) to what degree the concept c is relevant to the document h.

Definition 7: For each document $h \in \mathcal{H}$, on the basis of the binary indexing relation I, the *document descriptor* I_h of h is a fuzzy subset of $\mathcal{C}$ defined as follows:

$$I_h = \left\{\mu_{I_h}(c)/c \mid c \in \mathcal{C}, \mu_{I_h}(c) = \mu_I(h,c)\right\}$$

where $\mu_{I_h}(c)$ is the *degree of relevance* (or *relevance degree*) of document h with respect to concept c.

Notice that, in the document descriptor, the pair $(\mu_{I_h}(c), c)$ is not stored when $\mu_{I_h}(c) = 0$. Let C_h be denoted by the set of concepts, in which all of the degrees of relevance of the document h, with respect to those concepts, are larger than zero.

To illustrate, let $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$. $\mu_{h_2}(c_3) = 0.9$ means that the document h_2 contains the concept c_3 with the relevance degree 0.9, and $I_{h_2} = \{0.5/c_1, 0.9/c_3\}$ means that the document h_2 contains both concepts c_1 and c_3 with the relevance degrees 0.5 and 0.9, respectively. Then, we have $C_{h_2} = \{c_1, c_3\}$.

The relevance degrees between the documents and the concepts can be represented by a matrix D called the *document descriptor matrix* [2]. Let $\mathcal{H} = \{h_1, h_2, \ldots, h_m\}$ and $\mathcal{C} = \{c_1, c_2, \ldots, c_n\}$. Then, D can be shown as follows:

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix}$$

where $d_{ij} = I_{h_i}(c_j), 1 \leq i \leq m, 1 \leq j \leq n$.

III. Conceptual Query

A. Concept Network

To achieve the conceptual query for document retrieval, a knowledge-base must be supported. The *keyword connection matrix* [11] and *concept network* [7] have been used to specify the relationships among keywords and concepts, respectively. In this paper, we adopt the concept network. A concept network consists of nodes and directed links associated with a real number ranging from zero to one (except zero). Each node denotes a concept or a document. Each directed link connects two concepts or directs from one concept to a document. If a directed link, associated with a real number μ, directs from concept c_i to concept c_j, it means that the degree of relevance from concept c_i to concept c_j is equal to μ. If a directed link, associated with a real number μ, directs from concept c_i to document h, it means that the document h contains the concept c_i associated with the relevance degree μ [7].

Definition 8 ([2]): Let $\mathcal{C} = \{c_1, c_2, \ldots, c_n\}$ be a set of concepts. A *concept matrix* K is a matrix in which $K_{ij} \in [0,1]$. The (i,j) element of K represents the degree of relevance from concept c_i to concept c_j. Let $K^2 = K \otimes K$ be the *multiplication* of the concept matrix and $K^2_{ij} = \bigvee_{l=1}^{n}(K_{il} \wedge K_{lj}), 1 \leq i,j \leq n$, where $\vee$ and $\wedge$ represent the $\max$ operation and the $\min$ operation, respectively. Then, there exists an integer $\rho \leq n-1$, such that $K^\rho = K^{\rho+1} = K^{\rho+2} = \cdots$. Let $K^* = K^\rho$. K^* is called the *transitive closure* of the concept matrix K.

Referring to [2], the relevance degree of each document, with respect to a specific concept, can be improved by computing the multiplication of the document descriptor matrix D and the transitive closure of the concept matrix K, as follows:

$$D^* = D \otimes K^*.$$

D^* is called the *expanded document descriptor matrix*.

B. Semantics of Query Expression

From past works on fuzzy document retrieval [2], [6], [7], [12], [13], [15], we observe that there exist two semantics on the queries, the *range query* and the *point query*. When a positive (negative) range query is submitted, it means that the documents containing the concepts with a relevance degree *more than* those specified in the query are more (less) desired. When a positive (negative) point query is submitted, it means that the documents containing the concepts with the relevance degree *near* those specified in the query are more (less) desired. Consider the following examples: a user submits a query $\{0.8/c\}$ (*positive query*), where c is a concept. If it is a range query, the documents containing the concept c with the relevance degree more than 0.8 are more desired; if it is a point query, the documents containing the concept c with the relevance degree near 0.8 are more desired. Consider another example: a user submits a query $\neg\{0.8/c\}$ (*negative query*). If it is a range query, the documents containing the concept c with the relevance degree more than 0.8 are less desired; if it is a point query, the documents containing the concept c with the relevance degree near 0.8 are less desired. A flexible query expression to support the range query and the point query is presented.

C. Query Expression and Processing

A query q is specified as a disjunctive normal form, i.e., $q = Q_1 \vee Q_2 \vee \cdots \vee Q_m$, where each *subquery* Q_i is a conjunction of a *positive query component* $X_i^{\theta_i}$ and a *negative query component* $\bar{X}_i^{\bar{\theta}_i}$. Each query component can be either a range query or a point query, where r and p are used to denote the range query and the point query, respectively. Hence, let Q_i be denoted by

$$Q_i = X_i^{\theta_i} \wedge \neg \bar{X}_i^{\bar{\theta}_i}$$

where $\theta_i, \bar{\theta}_i \in \{r, p\}$ and X_i and $\bar{X}_i$ are the *query vectors* defined as follows.

Definition 9: Let $\mathcal{C} = \{c_1, c_2, \ldots, c_n\}$ be a set of concepts. A *query vector* X is represented as $\langle x_1, x_2, \ldots, x_n \rangle$, where $x_i \in [0,1] \cup \{-, \epsilon\}$. If $x_i \in [0,1]$, it indicates that the query vector contains the concept c_i with the relevance degree x_i, i.e., $\mu_X(c_i) = x_i$; if $x_i =$ "$-$," it indicates that the concept c_i in the query vector is *neglectable*; if $x_i =$ "ϵ," it indicates that the query vector contains the concept c_i with the relevance degree near zero.

Notice that ϵ is used to support the combination of range query and point query, so that an overflow operation in the query processing does not occur.

Let X be a query vector. The C_X denotes the set of concepts concerned by X and is defined as follows:

$$C_X = \{c_i \mid X = \langle x_1, x_2, \ldots, x_n \rangle \wedge x_i \neq \text{“}-\text{”}, 1 \leq i \leq n\}.$$

Each query vector can be either r (range query) or p (point query). Let $\phi^r(X,h)$ [respectively, $\phi^p(X,h)$] be denoted by the retrieval function, with respect to the range query (respectively, point query) to return the relevance degree between the query vector X and the document h. $\phi^r(X,h)$ and $\phi^p(X,h)$ are defined as follows:

- *Range query*

$$\phi^r(X,h) = \frac{\sum_{c \in C_X} \min(\mu_h(c), \mu_X(c))}{\sum_{c \in C_X} \mu_X(c)}.$$

- *Point query*

$$\phi^p(X,h) = \frac{\sum_{c \in C_X} S(\mu_h(c), \mu_X(c))}{|C_X|}$$

where $S(x,y) \geq 0$ is a function proportional to $-|x-y|$, e.g., $S(x,y)$ can be equal to

$$\exp(-\beta|x-y|), \quad \beta > 0 \tag{1}$$

or

$$1 - |x-y|. \tag{2}$$

In this paper, we consider (2) for convenience.

Notice that, when the retrieval functions ϕ^r and ϕ^p are evaluated, the operations on ϵ will be considered. We define minimum (min), addition $(+)$, and difference $(-)$ operations on ϵ with the real number $x \in [0,1]$, as follows:

$$\min(x, \epsilon) = \begin{cases} 0, & \text{if } x = 0 \\ \epsilon, & \text{otherwise} \end{cases}$$

$$|x \pm \epsilon| = \begin{cases} \epsilon, & \text{if } x = 0 \\ x, & \text{otherwise.} \end{cases}$$

For a query $q = Q_1 \vee Q_2 \vee \cdots \vee Q_n$, the *retrieval response* F_q is a fuzzy subset of $\mathcal{H}$, as defined as follows:

$$F_q = F_1 \vee F_2 \vee \cdots \vee F_n$$

where F_i, the retrieval response of the subquery Q_i, is a fuzzy subset of $\mathcal{H}$, which is defined as follows:

$$F_i = F_i^+ \wedge F_i^-$$

where F_i^+ and F_i^- are the retrieval responses of the positive query component of Q_i and the negative query component of Q_i, respectively, as defined as follows:

- Suppose a subquery Q_i of q consists of the positive query component $X_i^{\theta_i}$ and the negative query component $\bar{X}_i^{\bar{\theta}_i}$, the retrieval response F_i^+ of $X_i^{\theta_i}$ is a fuzzy subset of $\mathcal{H}$, whose membership function is given by

$$\mu_{F_i^+}(h) = \begin{cases} \phi^r(X_i, h), & \text{if } \theta_i = r \\ \phi^p(X_i, h), & \text{if } \theta_i = p. \end{cases}$$

- Similarly, the retrieval response F_i^- of $\bar{X}_i^{\bar{\theta}_i}$ is a fuzzy subset of $\mathcal{H}$, whose membership function is given by

$$\mu_{F_i^-}(h) = \begin{cases} 1 - \phi^r(\bar{X}_i, h), & \text{if } \bar{\theta}_i = r \\ 1 - \phi^p(\bar{X}_i, h), & \text{if } \bar{\theta}_i = p. \end{cases}$$

Consider the following example.

Example 1: Let $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$ be a set of concepts and $\mathcal{H} = \{h_1, h_2\}$ be a set of documents. Consider an expanded document descriptor matrix D^*, as follows:

$$D^* = \begin{matrix} h_1 \\ h_2 \end{matrix} \begin{bmatrix} 1 & 0.9 & 0 & 0.8 \\ 0.7 & 1 & 0.6 & 0.4 \end{bmatrix}.$$

Four queries that consist of only one subquery are shown below.

1) $q_1 = \langle 0.6, -, -, 0.8 \rangle^r \wedge \neg \langle -, -, \epsilon, - \rangle^r = X^r \wedge \neg \bar{X}^r$.
Semantics: The documents that contain the concepts c_1 and c_4 with the relevance degrees more than 0.6 and 0.8, respectively, and do not contain the concept c_3, are more desired.
Processing: For each document $h_i, i = 1, 2$, we compute the degree of relevance, with respect to the query q_1. We have

$$\phi^r(X, h_1) = \frac{0.6 + 0.8}{0.6 + 0.8} = 1$$
$$\phi^r(X, h_2) = \frac{0.6 + 0.4}{0.6 + 0.8} = 0.71$$

and

$$\phi^r(\bar{X}, h_1) = \frac{0}{\epsilon} = 0, \quad \phi^r(\bar{X}, h_2) = \frac{\epsilon}{\epsilon} = 1.$$

We obtain $F^+ = \{1/h_1, 0.71/h_2\}$ and $F^- = \{(1-0)/h_1, (1-1)/h_2\} = \{1/h_1, 0/h_2\}$. Consequently, the retrieval response of q_1 is

$$F_{q_1} = F^+ \wedge F^- = \{1/h_1, 0/h_2\} = \{1/h_1\}.$$

2) $q_2 = \langle 0.6, -, -, 0.8\rangle^r \wedge \neg\langle -, -, \epsilon, -\rangle^p = X^r \wedge \neg\bar{X}^p$.
Semantics: The documents that contain the concepts c_1 and c_4 with the relevance degrees more than 0.6 and 0.8, respectively, are more desired, except for those documents that do not contain the concept c_3.
Processing: For each document $h_i, i = 1, 2$, we compute the degree of relevance, with respect to the query q_2. We have

$$\phi^r(X, h_1) = \frac{0.6 + 0.8}{0.6 + 0.8} = 1$$
$$\phi^r(X, h_2) = \frac{0.6 + 0.4}{0.6 + 0.8} = 0.71$$

and

$$\phi^p(\bar{X}, h_1) = \frac{1}{1} = 1, \quad \phi^p(\bar{X}, h_2) = \frac{0.4}{1} = 0.4.$$

We obtain $F^+ = \{1/h_1, 0.71/h_2\}$ and $F^- = \{(1-1)/h_1, (1-0.4)/h_2\} = \{0/h_1, 0.6/h_2\}$. Consequently, the retrieval response of q_2 is

$$F_{q_2} = F^+ \wedge F^- = \{0/h_1, 0.6/h_2\} = \{0.6/h_2\}.$$

3) $q_3 = \langle 0.6, -, -, 0.8\rangle^p \wedge \neg\langle -, -, \epsilon, -\rangle^r = X^p \wedge \neg\bar{X}^r$.
Semantics: The documents that contain the concepts c_1 and c_4 with the relevance degrees near to 0.6 and 0.8, respectively, and do not contain the concept c_3 are more desired.
Processing: For each document $h_i, i = 1, 2$, we compute the degree of relevance, with respect to the query q_3. We have

$$\phi^p(X, h_1) = \frac{0.6 + 1}{2} = 0.8$$
$$\phi^p(X, h_2) = \frac{0.9 + 0.6}{2} = 0.75$$

and

$$\phi^r(\bar{X}, h_1) = \frac{0}{\epsilon} = 0, \quad \phi^r(\bar{X}, h_2) = \frac{\epsilon}{\epsilon} = 1.$$

We obtain $F^+ = \{0.8/h_1, 0.75/h_2\}$ and $F^- = \{(1-0)/h_1, (1-1)/h_2\} = \{1/h_1, 0/h_2\}$. Consequently, the retrieval response of q_3 is

$$F_{q_3} = F^+ \wedge F^- = \{0.8/h_1, 0/h_2\} = \{0.8/h_1\}.$$

4) $q_4 = \langle 0.6, -, -, 0.8\rangle^p \wedge \neg\langle -, -, \epsilon, -\rangle^p = X^p \wedge \neg\bar{X}^p$.
Semantics: The documents that contain the concepts c_1 and c_4 with the relevance degrees near to 0.6 and 0.8, respectively, are more desired, except for those documents that do not contain the concept c_3.
Processing: For each document $h_i, i = 1, 2$, we compute the degree of relevance, with respect to the query q_4. We have

$$\phi^p(X, h_1) = \frac{0.6 + 1}{2} = 0.8$$
$$\phi^p(X, h_2) = \frac{0.9 + 0.6}{2} = 0.75$$

and

$$\phi^p(\bar{X}, h_1) = \frac{1}{1} = 1, \quad \phi^p(\bar{X}, h_2) = \frac{0.4}{1} = 0.4.$$

We obtain $F^+ = \{0.8/h_1, 0.75/h_2\}$ and $F^- = \{(1-1)/h_1, (1-0.4)/h_2\} = \{0/h_1, 0.6/h_2\}$. Consequently, the retrieval response of q_4 is

$$F_{q_4} = F^+ \wedge F^- = \{0/h_1, 0.6/h_2\} = \{0.6/h_2\}.$$

Fig. 1. No links between h_i and h_j.

IV. Neighborhood Query

By supporting the neighborhood query, the documents that are relevant to a given document can be retrieved. A mechanism must be given to measure the degree of relevance between two documents. To do this, the *prior information*, including the description of the documents and the link relationship between the documents, must be given. According to the availability of the prior information, we divide the work into three parts: 1) the document descriptors are given and the document link relationships are not given (Section IV-A), 2) the former are not given and the latter given (Section IV-B), and 3) the former and the latter are both given (Section IV-C). Recall that the document descriptor of the document h is denoted by I_h. Let $\boldsymbol{l}(h_i, h_j)$ denote the relevance degree from a document h_i to another document h_j, where a link l of h_i points to h_j.

We introduce some terminologies for the subsequent discussion.

- The *s-step-directed relevance degree from h_i to h_j* represents the degree of relevance from h_i through s links to h_j, denoted as

$$\boldsymbol{\psi}^s(h_i, h_j), \quad s = 1, 2, \ldots$$

- The *s-step-indirected relevance degree between h_i and h_j* represents the degree of relevance between h_i and h_j that they are at intervals of s links, denoted as

$$\psi^s(h_i, h_j) = \max\{\boldsymbol{\psi}^s(h_i, h_j), \boldsymbol{\psi}^s(h_j, h_i)\}, \quad s = 1, 2, \ldots$$

- The *relevance degree between h_i and h_j* represents the degree of relevance between h_i and h_j.

$$\psi(h_i, h_j) = \psi(h_j, h_i) = \max_{s=1}^{\infty} \psi^s(h_i, h_j).$$

A. I_h Given and $\boldsymbol{l}(h_i, h_j)$ Not Given

When the document descriptor of the documents are given, the degree of relevance between two documents can be obtained by computing the similarity for each concept in those documents.

Let I_{h_i} and I_{h_j} be two document descriptors of the document h_i and the document h_j, respectively. The concept sets of h_i and h_j are C_{h_i} and C_{h_j}, respectively. Then the degree of relevance between h_i and h_j can be computed as follows:

$$\psi(h_i, h_j) = \frac{\sum_{c \in C_{h_i} \cup C_{h_j}} S\big(\mu_{h_i}(c), \mu_{h_j}(c)\big)}{\big|C_{h_i} \cup C_{h_j}\big|} = \Delta_{ij} \tag{3}$$

where $S(x, y)$ is defined as in (2) and Δ_{ij} is a notation that will be used in the following discussion.

B. I_h Not Given and $\boldsymbol{l}(h_i, h_j)$ Given

When the *direct relevance degree* of all document pairs are given, the degree of relevance between two documents can be obtained by computing various primitive link relationships, including no links, direct link, transitive link, more than one direct link, and circuits.

1) No Links: If no links exist between h_i and h_j, as shown in Fig. 1, in this case, the relevance degree between h_i and h_j is equal to zero

$$\boldsymbol{\psi}^s(h_i, h_j) = \boldsymbol{\psi}^s(h_j, h_i) = 0, \quad s = 1, 2, \ldots.$$

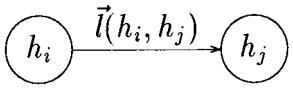


Fig. 2. One direct link from h_i to h_j.

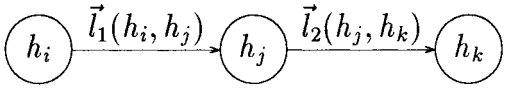


Fig. 3. Transitive direct link from h_i through h_j to h_k.

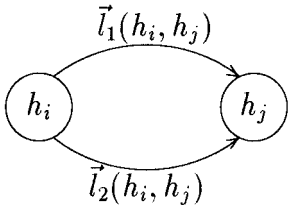


Fig. 4. More direct links from h_i to h_j.

2) Direct Link: If h_i contains a link l points to h_j, as shown in Fig. 2, the one-step-directed relevance degree from h_i to h_j is $\boldsymbol{l}(h_i, h_j)$. That is

$$\begin{aligned}\boldsymbol{\psi}^1(h_i, h_j) &= \boldsymbol{l}(h_i, h_j)\\ \boldsymbol{\psi}^s(h_i, h_j) &= 0, \quad s = 2, 3, \ldots.\\ \boldsymbol{\psi}^s(h_j, h_i) &= 0, \quad s = 1, 2, \ldots.\end{aligned}$$

3) Transitive Link: If there exists a link l_1 from h_i to h_j associated with the relevance degree $\boldsymbol{l}_1(h_i, h_j)$ and a link l_2 from h_j to h_k associated with the relevance degree $\boldsymbol{l}_2(h_j, h_k)$, as shown in Fig. 3, the two-step-directed relevance degree from h_i to h_k is defined as the minimum of $\boldsymbol{l}_1(h_i, h_j)$ and $\boldsymbol{l}_2(h_j, h_k)$. That is

$$\begin{aligned}\boldsymbol{\psi}^2(h_i, h_k) &= \min\{\boldsymbol{l}_1(h_i, h_j), \boldsymbol{l}_2(h_j, h_k)\}\\ \boldsymbol{\psi}^s(h_i, h_k) &= 0, \quad s = 1, 3, 4, \ldots.\\ \boldsymbol{\psi}^s(h_k, h_i) &= 0, \quad s = 1, 2, \ldots.\end{aligned}$$

4) More than One Direct Link: If there exists two links l_1 and l_2, both from h_i to h_j, associated with the relevance degrees $\boldsymbol{l}_1(h_i, h_j)$ and $\boldsymbol{l}_2(h_i, h_j)$, respectively, as shown in Fig. 4, the one-step-directed relevance degree from h_i to h_j is defined as the maximum of $\boldsymbol{l}_1(h_i, h_j)$ and $\boldsymbol{l}_2(h_i, h_j)$. That is

$$\begin{aligned}\boldsymbol{\psi}^1(h_i, h_j) &= \max\{\boldsymbol{l}_1(h_i, h_j), \boldsymbol{l}_2(h_i, h_j)\}\\ \boldsymbol{\psi}^s(h_i, h_j) &= 0, \quad s = 2, 3, 4, \ldots.\\ \boldsymbol{\psi}^s(h_j, h_i) &= 0, \quad s = 1, 2, \ldots.\end{aligned}$$

5) With Circuits: If there exists a link from h_i to h_j associated with the relevance degree $\boldsymbol{l}_1(h_i, h_j)$ and a link from h_j to h_i associated with the relevance degree $\boldsymbol{l}_2(h_j, h_i)$, as shown in Fig. 5,

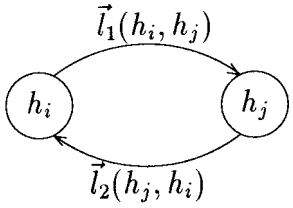


Fig. 5. More direct links with a circuit between h_i and h_j.

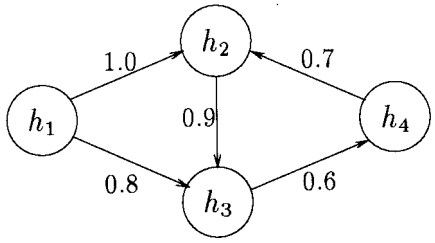


Fig. 6. Example for I_h not given and $\boldsymbol{l}(h_i, h_j)$ given.

the s-step-directed relevance degree for $s = 1, 2, \ldots$ are shown as follows:

$$\begin{aligned}\boldsymbol{\psi}^1(h_i, h_j) &= \boldsymbol{l}_1(h_i, h_j)\\ \boldsymbol{\psi}^1(h_j, h_i) &= \boldsymbol{l}_2(h_j, h_i)\end{aligned}$$

$$\boldsymbol{\psi}^{2s}(h_i, h_j) = \boldsymbol{\psi}^{2s}(h_j, h_i) = 0, \quad s = 1, 2, \ldots$$

$$\begin{aligned}\boldsymbol{\psi}^{2s+1}(h_i, h_j) &= \boldsymbol{\psi}^{2s+1}(h_j, h_i)\\ &= \min\{\boldsymbol{l}_1(h_i, h_j), \boldsymbol{l}_2(h_j, h_i)\}, \quad s = 1, 2, \ldots\end{aligned}$$

The one-step-directed relevance degree for all document pairs can be represented by a matrix $\boldsymbol{M}^1$, called *one-step-directed relevance degree matrix*, which is defined as follows.

Definition 10: Let $\mathcal{H}$ be a set of documents where $\mathcal{H} = \{h_1, h_2, \ldots, h_n\}$. A one-step-directed relevance degree matrix $\boldsymbol{M}^1$ is a matrix where $\boldsymbol{M}^1_{ij}$, the (i, j) element of $\boldsymbol{M}^1$, is the one-step-directed relevance degree from h_i to h_j, as denoted by

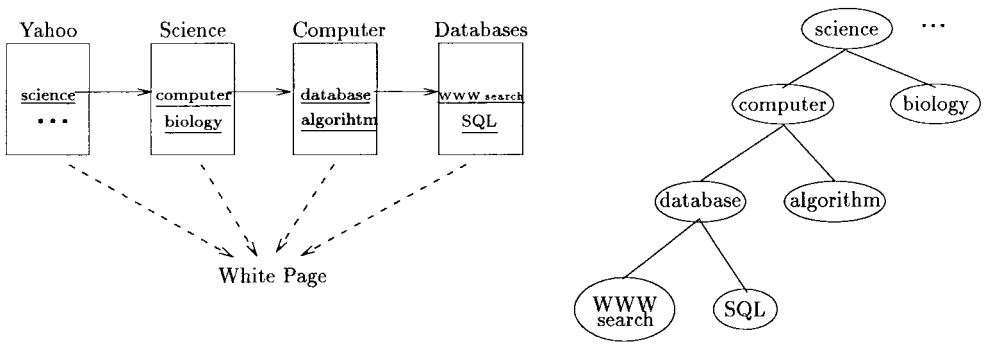
$$\boldsymbol{M}^1_{ij} = \begin{cases}\boldsymbol{\psi}^1(h_i, h_j), & \text{if } i \neq j, \quad 1 \leq i, \quad j \leq n\\ 1, & \text{if } i = j.\end{cases}$$

Recall that, if there exists more than one direct link from h_i to h_j., $\boldsymbol{\psi}^1(h_i, h_j)$ can be obtained from the discussion in Section IV–B4. Therefore, we can use $\boldsymbol{M}^1$ to obtain the relevance degree of all document pairs by computing the transitive closure of $\boldsymbol{M}^1$.

Property 1: Let $\mathcal{H}$ be a set of documents $\mathcal{H} = \{h_1, h_2, \ldots, h_n\}$. Let $\boldsymbol{M}^1$ be a one-step-directed relevance degree matrix, as shown as follows:

$$\boldsymbol{M}^1 = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}$$

where $m_{ij} = \boldsymbol{M}^1_{ij}$ is defined in Definition 10. Equation (4) computes the *two-step-directed relevance degree matrix* $\boldsymbol{M}^2$, where "$\vee$" represents the $\max$ operation and "$\wedge$" represents the $\min$ operation. Then, there exists an integer ρ ($\rho \leq n - 1$), such that $\boldsymbol{M}^\rho = \boldsymbol{M}^{\rho+1} = \boldsymbol{M}^{\rho+2} = \cdots$. The (i, j) element of the *relevance degree matrix* M^*,



(a) (b)

Fig. 7. (a) The white pages and their link relationships beginning from Yahoo homepage. (b) The concept hierarchy from (a).

denoted M_{ij}^*, which represents the relevance degree between h_i and h_j, can be computed by the maximum of $\boldsymbol{M}_{ij}^{\rho}$ and $\boldsymbol{M}_{ji}^{\rho}$. That is

$$M_{ij}^* = M_{ji}^* = \max\{\boldsymbol{M}_{ij}^{\rho}, \boldsymbol{M}_{ji}^{\rho}\}$$

$$\boldsymbol{M}^2 = \begin{bmatrix} \bigvee_{i=1}^n (m_{1i} \wedge m_{i1}) & \bigvee_{i=1}^n (m_{1i} \wedge m_{i2}) & \cdots & \bigvee_{i=1}^n (m_{1i} \wedge m_{in}) \\ \bigvee_{i=1}^n (m_{2i} \wedge m_{i1}) & \bigvee_{i=1}^n (m_{2i} \wedge m_{i2}) & \cdots & \bigvee_{i=1}^n (m_{2i} \wedge m_{in}) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \bigvee_{i=1}^n (m_{ni} \wedge m_{i1}) & \bigvee_{i=1}^n (m_{ni} \wedge m_{i2}) & \cdots & \bigvee_{i=1}^n (m_{ni} \wedge m_{in}) \end{bmatrix}. \quad (4)$$

Example 2: Consider an example for a set of documents $\mathcal{H} = \{h_1, h_2, h_3, h_4\}$, in which the graph of the link relationships and the one-step-directed relevance degrees are shown in Fig. 6. The one-step-directed relevance degree matrix $\boldsymbol{M}^1$ is shown as follows:

$$\boldsymbol{M}^1 = \begin{bmatrix} 1 & 1 & 0.8 & 0 \\ 0 & 1 & 0.9 & 0 \\ 0 & 0 & 1 & 0.6 \\ 0 & 0.7 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{M}^{\rho} = \begin{bmatrix} 1 & 1 & 0.9 & 0.6 \\ 0 & 1 & 0.9 & 0.6 \\ 0 & 0.6 & 1 & 0.6 \\ 0 & 0.7 & 0.7 & 1 \end{bmatrix}, \quad \rho \geq 2.$$

Consequently, we obtain the relevance degree matrix M^* as follows:

$$M^* = \begin{bmatrix} 1 & 1 & 0.9 & 0.6 \\ 1 & 1 & 0.9 & 0.7 \\ 0.9 & 0.9 & 1 & 0.7 \\ 0.6 & 0.7 & 0.7 & 1 \end{bmatrix}.$$

For example, from M^*, we can obtain the degree of relevance between h_1 and h_4, which is 0.6 and so on.

C. I_h Given and $\boldsymbol{l}(h_i, h_j)$ Given

If I_h and $\boldsymbol{l}(h_i, h_j)$ are given for each $h \in \mathcal{H}$, the relevance degree can be computed by combining the results from Section IV-A and B.

For a set of documents $\mathcal{H} = \{h_1, h_2, \ldots, h_n\}$, we first compute Δ_{ij} from (3) and the relevance degree matrix M^* from Section IV-B. Then, the relevance degree between h_i and h_j can be defined as the maximum of Δ_{ij} and M_{ij}^*. That is

$$\psi(h_i, h_j) = \max\{\Delta_{ij}, M_{ij}^*\}, \quad 1 \leq i, \quad j \leq n.$$

D. Assignment of $\boldsymbol{l}(h_i, h_j)$

In this section, we consider the assignment of $\boldsymbol{l}(h_i, h_j)$. When the document h_i is considered as a technical report or a paper, the link l, which points to h_j, can be considered as the citation of h_i. In general, the technical reports or the papers can be combined into five parts: abstract, introduction, technical description, related work, and conclusion. According to the position of the citation appearing in the document, we can assign $\boldsymbol{l}(h_i, h_j)$ with an appropriate value. For example, if the citation h_j is cited in the abstract part of the document h_i, we may assign $\boldsymbol{l}(h_i, h_j)$ as the largest value comparing to those cited in the other parts. In the following, a subjective assignment of $\boldsymbol{l}(h_i, h_j)$ is listed:

- h_j is cited in the abstract part of h_i: $\boldsymbol{l}(h_i, h_j) = 0.9$;
- h_j is cited in the introduction part of h_i: $\boldsymbol{l}(h_i, h_j) = 0.5$;
- h_j is cited in the related-work part of h_i: $\boldsymbol{l}(h_i, h_j) = 0.6$;
- h_j is cited in the technical description part of h_i: $\boldsymbol{l}(h_i, h_j) = 0.7$; and
- h_j is cited in the conclusion part of h_i: $\boldsymbol{l}(h_i, h_j) = 0.8$.

V. Concept Network Exploration

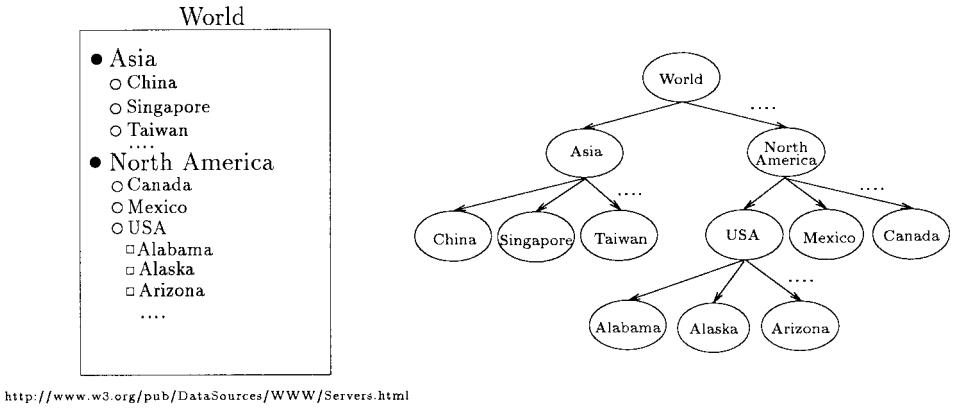
In this section, the exploration of the concept network is addressed. We use the easily available information from the WWW, including homepage link relationships and HTML (HyperText Markup Language) structures, to construct the concept network.

A. Using WWW Link Relationships

A *white page* is a homepage on the WWW. It contains a list of concepts. Each concept in a white page may link to another homepage, which may be a white page or not. Suppose that we click a concept t in a white page and the link navigates to another white page H. We may obtain the concept t as a generalization of all concepts in H. For example, in Fig. 7(a), four white pages are linked in a chain from the Yahoo homepage to the other three homepages that are all white pages. The concept "science" links to the homepage that contains two concepts: "computer" and "biology." We may say that the concept "science" is a generalization of the concept "computer" and the concept "biology." Similarly, we may say that the concept "computer" is a generalization of the concept "databases" and the concept "algorithm." Then, we can construct a *concept hierarchy* with the root concept "science," as shown in Fig. 7(b).

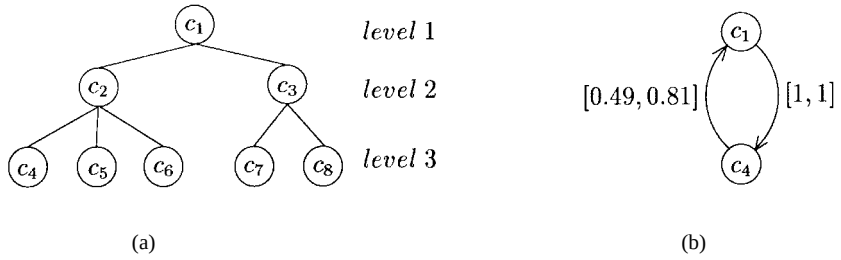
B. Using HTML Structures

HTML is the language used to define the structure of the hypertext documents on the WWW. In general, we always list the same category of the concepts in the same homepage and put the *specialized*



(a) (b)

Fig. 8. (a) Homepage as url: http://www.w3.org/pub/DataSources/WWW/Servers.html. (b) The concept hierarchy from (a).



(a) (b)

Fig. 9. (a) Concept hierarchy. (b) Concept-pair and the assignment of the relevance degrees.

concepts behind the corresponding *generalized* concept with an indentation. To illustrate, consider Fig. 8. In Fig. 8(a), a homepage (url: http://www.w3.org/pub/DataSources/WWW/Servers.html) contains the regions, countries, and counties on earth (in this paper, we only show a part of the items). For example, Asia includes the countries of China, Singapore, Taiwan, etc.; North America includes the countries of Canada, Mexico, the United States (including the states of Alabama, Alaska, Arizona, etc.), and so on. The concept hierarchy is shown in Fig. 8(b). After a set of concept hierarchies is collected, a strategy is proposed to a construct concept network in the following.

C. Construction

Definition 11: Let W be a concept hierarchy. Each node in W represents a concept. The *concept set* of W, denoted C_W, is the set of concepts in W. Let c be a concept in C_W and *level*(c) be the level of c in W. A *concept pair* is a pair of two nodes associated with a *level difference* δ, denoted by (c_i, c_j, δ), where $c_i, c_j \in C_W$ and $\delta =$ *level*$(c_j) -$ *level*(c_i), and c_i is the *ancestor* of c_j. Let $\Pi(W)$ be denoted as all possible concept pairs in W.

Example 3 Consider the concept hierarchy W shown in Fig. 9. The concept set of W, C_W is $\{c_1, c_2, \ldots, c_8\}$. There exist 12 concept pairs in W. That is

$$\begin{aligned}\Pi(W) = \{&(c_1, c_2, 1), (c_1, c_3, 1), (c_1, c_4, 2), (c_1, c_5, 2)\\ &(c_1, c_6, 2), (c_1, c_7, 2), (c_1, c_8, 2), (c_2, c_4, 1)\\ &(c_2, c_5, 1), (c_2, c_6, 1), (c_3, c_7, 1), (c_3, c_8, 1)\}.\end{aligned}$$

For each concept pair, the relevance degree will be assigned according to its level difference. Let $[a, b],\ 0 \le a \le b \le 1$ be a *base-interval* for the assignment of the relevance degree. The relevance degree of the concept pair (c_i, c_j, δ) is considered as $[a^\delta, b^\delta]$. It means that the degree of relevance from the concept c_j to the concept c_i is $[a^\delta, b^\delta]$. Implicitly, the degree of relevance from c_i to c_j is $[1, 1],\ i < j$. To illustrate, consider the concept pair $(c_1, c_4, 2)$ in Example 3, and let [0.7, 0.9] be the base interval. The degree of relevance from c_4 to c_1 is $[0.7^2, 0.9^2] = [0.49, 0.81]$, and the relevance degree from c_1 to c_4 is [1, 1], as shown in Fig. 9(b).

Let $\mathcal{W} = \{W_1, W_2, \ldots, W_n\}$ be a set of concept hierarchies and $\mathcal{C}$ be a set of concepts. The degree of relevance from the concept c_i to the concept c_j can be computed by accumulating the average of the occurrence of the concept pair. An algorithm is shown to reflect the above descriptions.

Algorithm 1 (Construction):

Input: a set of concept hierarchies $\mathcal{W} = \{W_1, W_2, \ldots, W_n\}$ and a base interval $[a, b],\ 0 \le a \le b \le 1$;

Output: a concept matrix $[K_{ij}]$;

Comment: Z_{ij} and $K'_{ij},\ 1 \le i,\ j \le |\mathcal{C}|$ are temporary variables and are initiated to zero.

The *lower*(x) and *upper*(x) denote the lower part and the upper part of an interval value x, respectively

```
0:  begin
1:      C ← ⋃_{l=1}^{n} C_{W_l}
2:      for each W_l ∈ W
3:      begin
4:          Π(W_l) ← {(c_i, c_j, level(c_j) − level(c_i)) | c_i is the
              ancestor of c_j in W_l ∧ c_i, c_j ∈ C}
5:          for each (c_i, c_j, δ) ∈ Π(W_l)
6:          begin
7:              lower(K'_ij) ← lower(K'_ij) + 1
8:              upper(K'_ij) ← upper(K'_ij) + 1
9:              lower(K'_ji) ← lower(K'_ji) + a^δ
```

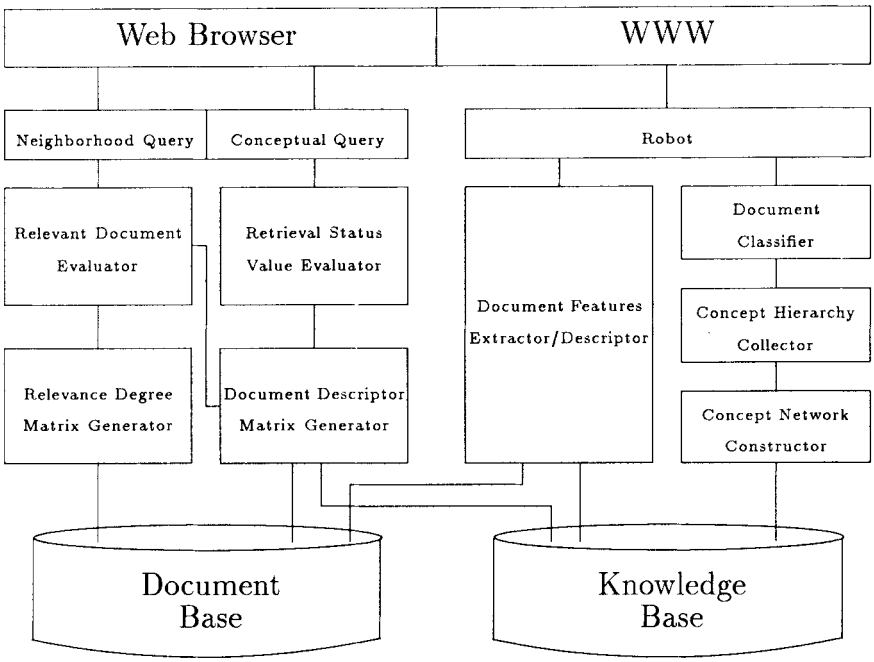


Fig. 10. Architecture of WORDS.

10: $upper(K'_{ji}) \leftarrow upper(K'_{ji}) + b^{\delta}$
11: $Z_{ij} \leftarrow Z_{ij} + 1$
12: $Z_{ji} \leftarrow Z_{ji} + 1$
13: **end**
14: **end**
15: $K_{ij} \leftarrow \frac{lower(K'_{ij}) + upper(K'_{ij})}{2 \times Z_{ij}}, 1 \leq i, j \leq |\mathcal{C}|$
16: **end**.

VI. WORDS

The Web knOwledge and Resource Discovery System (WORDS), providing the functions addressed in the previous sections, is being implemented using Java programming language [4]. The architecture of WORDS is shown in Fig. 10. WORDS mainly consists of two parts: the resource discovery part and the knowledge discovery part. In the resource discovery part, shown on the left-hand side of Fig. 10, users can submit conceptual queries or neighborhood queries by using the web browser, e.g., Netscape, Internet Explorer, etc. To achieve the conceptual query, the expanded document descriptor matrix D^* (discussed in Section III-A) is generated to compute the retrieval status values ϕ (discussed in Section III-C). To achieve the neighborhood query, the relevance degree matrix M^* (discussed in Section IV) is generated to compute the degree of relevance of the relevant documents for a given document.

In the knowledge discovery part, a robot is built to retrieve documents from the WWW. By employing the document classifier, an incoming document is determined to be a white page or not. If a chain of incoming documents are all white pages, we can obtain a concept hierarchy by using the method discussed in Section V (concept hierarchy collector). A strategy is proposed to construct the concept network from a collection of the concept hierarchies (concept network constructor). Consequently, the concept network is stored in the knowledge base. In the document feature extraction or description, we adopt the IR techniques or the knowledge of the experts to assign the metadata, which includes concepts and associated relevance degrees, for the incoming documents. The metadata of the incoming documents are then stored in the document base.

Currently, the system is being implemented by using Java programming language [4]. We have completed the robot part, concept hierarchy collector part, and the concept network constructor part. Regarding the document classifier part, the mechanism of determining whether an incoming homepage is a white page is manual.

VII. Conclusion

In this paper, we propose a framework of document retrieval on the WWW based on fuzzy sets. We support a flexible conceptual query specification, which allows the users to specify the range query and the point query within a query, and the neighborhood query, which allows the users to allocate documents by a given document. The construction of a concept network is through exploring the concept hierarchies from the WWW environment.

Since the number of concepts and documents are huge, the matrix calculation (the transitive closure of the concept matrix and the relevance degree matrix) requires a great amount of computation. The computations of the matrix can be set off-line, such that the response time of the on-line query (conceptual query and neighborhood query) can be reduced.

References

[1] T. Berners-Lee, R. Cailliau, A. Luotonen, H. F. Nielsen, and A. Secret, "The World-Wide Web," *Commun. ACM*, vol. 37, no. 8, 1994.
[2] 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.
[3] D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. New York: Plenum, 1986.
[4] D. Flanagan, *Java In A Nutshell*. Sebastopol, CA: O'Reilly & Associates, Inc., 1996.
[5] E. Garfield, *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. New York: Wiley, 1979.

[6] G. Guardalben and D. Lucarella, "Information retrieval based on fuzzy reasoning," *Data Knowl. Eng.*, vol. 10, pp. 29–44, 1993.
[7] D. Lucarella and R. Morara, "FIRST: Fuzzy information retrieval system," *J. Inf. Sci.*, vol. 17, no. 2, pp. 81–91, 1991.
[8] S. Miyamoto, "Two approaches for information retrieval through fuzzy associations," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, pp. 123–130, Jan. 1989.
[9] ——, "Information retrieval based on fuzzy associations," *Fuzzy Sets Syst.*, vol. 38, no. 2, pp. 191–205, 1990.
[10] S. Miyamoto, T. Miyake, and K. Nakayama, "Generation of a pseudothesaurus for information retrieval based on cooccurrences and fuzzy set operations," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 62–70, Jan. 1983.
[11] Y. Ogawa, T. Morita, and K. Kobayashi, "A fuzzy document retrieval system using the keyword connection matrix and a learning method," *Fuzzy Sets Syst.*, pp. 163–179, 1991.
[12] T. Radecki, "Mathematical model of time-effective information retrieval system based on the theory of fuzzy sets," *Inf. Process. Manage.*, vol. 13, pp. 109–116, 1977.
[13] ——, "Fuzzy set theoretical approach to document retrieval," *Inf. Process. Manage.*, vol. 15, pp. 247–259, 1979.
[14] G. Salton and M. J. Mcgill. *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1983.
[15] V. Tahani, "A fuzzy model of document retrieval systems," *Inf. Process. Manage.*, vol. 12, pp. 177–187, 1976.
[16] G. Triantafyllos, S. Vassiliadis, and G. G. Pechanek, "A fuzzy information retrieval system," in *Proc. IEEE Conf. Fuzzy Syst.*, 1994, pp. 150–155.
[17] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 1, no. 1, pp. 3–28, 1978.
[18] M. Zemankova and A. Kandel, "Implementing imprecision in information systems," *Inf. Sci.: Int. J.*, vol. 37, pp. 107–141, 1985.
[19] R. B. R. C. Zenner and M. M. D. Caluwe, "A new approach to information retrieval systems using fuzzy expressions," *Fuzzy Sets Syst.*, vol. 17, pp. 9–22, 1985.
[20] H. J. Zimmermann, *Fuzzy Set Theory—and Its Applications*, 2nd ed. London, U.K.: Kluwer, 1991.