THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

TECHNICAL REPORT OF IEICE

A Semantic-based Personalized Retrieval System Using Individual Cognitive Structures

XiaoYong DU [†,‡] HaiHua LI [†] Xuan TIAN[†] Zhe SU[†] and Xijun LUO [†]

[†] Department of Computer Science, School of Information, Renmin University of China

[‡] Key Labs of Data Engineering and Knowledge Engineering, Ministry of Education, China

E-mail: duyong@ruc.edu.cn

Abstract "What you get is what you want" is undoubtedly an ideal objective of modern IR systems. Personalization is a kind of powerful mechanism to approach this objective. However, most current solutions for personalization focus on descriptions of user preferences from syntax level. In this paper, we propose a novel ICS (Individual Cognitive Structure)-based user's personalization model and a semantic-based personalized information retrieval system by using ontology technologies. Experimental results show the effectiveness of our proposed model and system by comparisons of different approaches.

Keyword Personalization, Individual Cognitive Structure, domain ontology, Information Retrieval.

1. Introduction

"What you get is what you want" is an ideal objective of current IR systems. However, the final retrieved results of current IR systems often meet the user's information requirements insufficiently, in particular for those "high-level" users. In essence, the above situation is caused primarily by the two following reasons. On one hand, the lack of reasonable organizations for existing information resources hinders the performance of information retrieval. On the other hand, the lack of the complete description for user's information requirements also weakens the precision of the returned results in IR system.

The Semantic Web (SW), the Knowledge Grid (KG), Web Service (WS) are the key supporting techniques for solving the first problem. As described by Tim Berners-Lee, the Semantic Web has spurred an intense activity in industry and academia [1]. The core idea of SW is to represent concepts and semantic relationships between these concepts using (domain) ontologies for the sake of the inter-operability between machines and machines or machines and people. Therefore, we can reasonably image that these technologies will be utilized as management of the stacks of information resources in current IR.

The solutions for the second problem focus on the modeling of user's personalized information needs by different ways, such as, to use a list of keywords to represent the user's interests, to mine the user's behaviors from browser logs of users. But, few research concerns the influence of individual cognitive structure [18]. Intuitively, the individual cognitive structure of a user is a great important factor to order the retrieved results. For example, a professor and a student have clearly different judgment for the retrieved results of the same query expression, say "database".

In this paper, we propose an ICS (individual cognitive structures)-based user's personalization model and a semantic-based personalized retrieval approach, particularly for those "high level" users. Here, we use "high level" to indicate a requirement for high quality of the retrieved results. The core idea of this approach is to describe static cognitive structures of users using the knowledge of users themselves. The production of this idea originates from an intuition, i.e., the ways of user's retrieval are generally implemented based on his/her existed knowledge. We will focus on the implementation of this approach in the semantic-based retrieval system.

The paper is organized as follows. In section 2 we define the individual cognitive structures of users. Section 3 introduces the system architecture. Section 4 provides the definition of the association degrees between keywords and concepts. Section 5 implements the ICS-based query refinement approach. Section 6 illustrates the experimental evaluation and Section 7 concludes this paper.

2. ICS (Individual Cognitive Structure)

Individual knowledge background of a user is often one of the key factors to measure the user's information requirements. Specially, a mining of such individual knowledge can help others to ascertain the meaning of user requests farther. Hence, we compute ICS in terms of individual knowledge background aiming at obtaining the deep meaning of user's requirements. ICS can be categorized into two types: one is static cognitive structure, and the other is users' behaviors related dynamic cognitive structure [2]. We focus on the discussion of static cognitive structure in this paper, since ICS is generally more stable and user behaviors' impact on it is very small.

Using DOSAM (Domain-Ontology Spreading-Activation Model) approach, we can deduce ICS of users pertaining to a specific domain. The origination of DOSAM is from SAM[15-17] model in Cognitive Psychology and its details are described in [3] and we omit the details of this model here for lack of space.

In the following discussions, suppose $DO = \{C, R\}$ represents a specific domain ontology where C represents the set of concepts and R represents the relationships between concepts of this domain, that is, $R = \{(c_i, c_j) \mid c_i, c_j \in C\}$. Then, we present some primary definitions of ICS involved.

Definition 1 (Cognitive Center Concepts). The concepts by which user u describes his/her basic interests or cognitions on DO are called cognitive center concepts. Let V_u be the set of cognitive center concepts for u.

Definition 2 (DOC_u, short for the Degree of Cognition). For concept c_i in DO, $DOC_u(c_i)$ is given for u to describe the extent of his cognition for c_i where $0 <= DOC_u(c_i) <= 1$.

It's obvious that it is impracticable for a user to provide the *DOC* of each concept in *DO*. However, it's possible for a user to manually provide the *DOC* of each cognitive center concept, since the number of cognitive center concepts is often very small. Hence, the definition of $DOC_u(c_i)$ is described as the following equation 1.

$$DOCu(ci) = \begin{cases} \lambda_i, & c_i \in V_u \\ MAX_{c_j \in V_u}\{DOC_u(c_j) * DOA(c_i, c_j)\}, & c_i \in C - V_u \end{cases} \quad (1)$$

where $DOA(c_i, c_j)$ represents the semantic association degree between concept c_i and c_j and λ_i is given by u . Also, you can obtain the computation method of $DOA(c_i, c_j)$ in [3] if necessary.

Definition 3 (ICS). The definition of *ICS* for u is described as $O_u = \{C', R'\}$ where:

$$C' = \{c_i \mid DOC_u(c_i) >= \theta, c_i \in C\}$$
$$R' = \{(c_i, c_j) \mid c_i, c_j \in C' \text{ and } (c_i, c_j) \in R\}.$$

Where $0 <= \theta <= 1$, is a threshold given by u. We can regard *ICS* as a subnet of *DO* if domain ontology is regarded as a network. In addition, each concept in this subnet is annotated by the degree of user cognition.

3. System Architecture

Following the architecture of SW, the information resources in our system are organized using domain ontologies. Here documents are sorted in the database together with domain ontologies (we called it ontology repository). Documents are annotated by the concepts in domain ontologies. On user's side, the knowledge background of a user is described by his/her cognitive structures in order to refine any future retrieval. The detailed system architecture is shown in Figure 1.

This prototype system consists of three key modules: (1) Semantic Annotator: the functionality of this module is to annotate information resources using domain ontologies. (2) DOSAM: the main role of it is to construct the ICSs for users. (3) Personalized Retrieval: the key functionality of this module is to return the required information resources for the users by utilizing the combination of the user requirements and their ICSs as refined search conditions. The box in dot line is the scope of this paper。

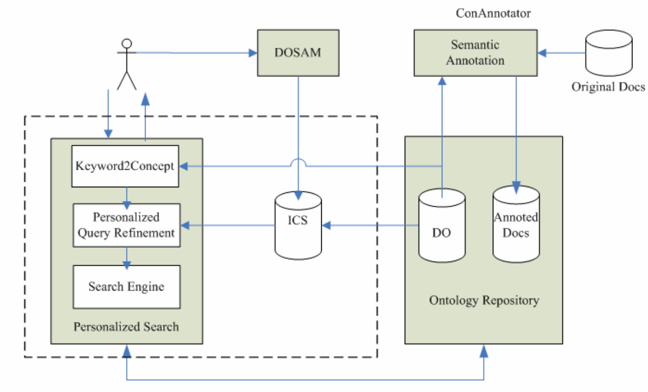

Figure 1: System Architecture.

3.1 Semantic Annotator

In our prototype system, information resources are modeled in form of documents. These documents are pre-processed as follows. First, we categorize the domain of each document. Second, they are annotated automatically first and then manually using this domain ontologies. Finally, these documents are organized by reverted indexes utilizing their annotations. In other words, each document can be regarded as the instances of some concepts in domain ontology. The details of the module is beyond the scope of this paper, and readers can find them in[19].

3.2 DOSAM

In module DOSAM, a user's ICS is constructed from the domain ontology and user's cognitive center concepts. The user firstly chooses several cognitive center concepts from the domain ontology. Accordingly, our system constructs the user's ICS from the domain ontology as following steps:

a. User *u* chooses a set of specific interested concepts as the cognitive center concepts from the domain ontology. A value of DOC is assigned to each concept manually.

b. ICS is constructed according to the DOSAM algorithm. It is a sub-network extracted from the corresponding domain ontology. By computing the semantic distance between the user's cognitive center concepts and other concepts, a set of most relevant concepts in domain ontology are activated and added into the ICS. Detailed algorithm please refer to [3].

3.3 Personalized Retrieval

It accepts a set of keywords and then returns a set of results from the annotated document repository. Firstly it transforms the input keywords into the corresponding concepts of domain ontologies and then return the final results.

For this purpose, two core techniques are required. One is to compute the association degrees between keywords and concepts. The other is to integrate the ICSs into the personalized retrieval. We will present details of these two techniques in section 4 and Section 5 respectively.

4. Association Degrees between Keywords and Concepts

The computing of the association degrees between retrieval keywords and concepts of domain ontologies is implemented using the statistic data of a large corpus, which is widely accepted by computer science society such as in [13]. In the following part, suppose C is the set of concepts in DO, $N = \|C\|$ represents the number of concepts in C, D is the set of documents annotated by DO, and $M = \|D\|$ represents the number of documents in D.

Below, we define two vectors **O-KCRV** and **TF-KCRV** to measure the association degrees between keywords and concepts respectively.

O-KCRV (Occurrence-based Keyword-Concept-Relevance Vector). This vector is used to describe the frequency of co-occurrence of a pair of a keyword and a concept in a given corpus. It is defined as $(kc_1, kc_2, \ldots, kc_N)^T$ obtained in a given size window W. Furthermore, kc_i is weighted by the following equation (2).

$$kc_i = count(k, c_i; W; D) / Max_{j \in N}\{count(k, c_j; W; D)\} \quad (2)$$

where $count(k, c_p; W; D)$ represents the frequency of co-occurrence of keyword k and concept c_p in window W.

In special case, if the keyword is just a concept, we directly set the corresponding element of it with itself in O-KCRV as 1.

TF-KCRV (Term-Frequency-based Keyword-Concept-Relevance Vector). TF-KCRV is also used to measure the association degrees between keywords and concepts. Furthermore, this vector is defined as $(kdc_1, kdc_2, \ldots, kdc_N)^T$ where kdc_i represents the occurrences of keyword k in documents which is annotated by concept c_i.

The relationships among keywords, documents and concepts are illustrated in Figure 2. A keyword may occur in several documents and each document may be annotated by several concepts. Therefore, we can utilize the statistical occurrences of keywords in documents annotated by concepts to measure the association degrees between keywords and concepts. Compared with traditional approaches, our approach is at the word level but not at the whole-document level. Based on word granularity, more precise statistic can be obtained. For example, if the occurrences of two keywords identified at the whole-document level are equal, we can distinguish their importance by their occurrences at the word level since they are often unequal. Therefore, the higher of the occurrences of keyword k in document d, the closer the association degrees between k and concepts annotating d, intuitionally.

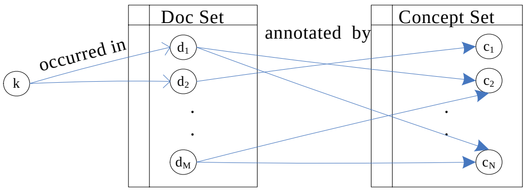

Figure 2: The relationships between keywords, documents and concepts.

Let $d_m \propto c_i$ represent that document d_m is an instance of concept c_i (if a document d is annotated by concept c, then we define that d is an instance of c.), $D_i = \{d_m \mid d_m \in D \wedge d_m \propto c_i\}$, $count(k\ ;\ d_m)$ represent the occurrences of keywords k in d_m, $c(k\ ,\ c_i) = \Sigma_{dm \in Di} count(k\ ;\ d_m)$ represent the occurrences of k in all instances of c_i. Hence, we describe the definition of the association degree kdc_i as shown in the equation (3).

$$kdc_i = c(k, c_i) / Max_{j \in N}\{c(k, c_j)\} \quad (3)$$

To sum up, O-KCRV measures the association degrees between keywords and concepts from the co-occurrences of keywords and concepts in entire corpus, while TF-KCRV does it from the occurrences of keywords in documents annotated by concepts of domain ontologies. In this work, we assume that the impact of two above measurements on the association degrees between keywords and concepts are equal. In addition, we adopt the inner product of the vector O-KCRV and TF-KCRV to measure the association degrees between keywords and concepts, since that can enhance the contributions of two above measurements each other. Therefore, we construct an integrated vector I-KCRV (Integrated Keyword-Concept-Relevance Vector) using O-KCRV and TF-KCRV metrics to measure the final association degrees between a keyword and a concept, that is,

$$\text{I-KCRV} = (kcr_1, \ldots, kcr_N) = (kc_1 \cdot kdc_1, \ldots, kc_N \cdot kdc_N)$$

The objective of computing the association degrees between keywords and concepts is to utilize the related concepts to re-define keywords. Also, we employ the top-k method in this re-definition. Since, for a given keyword, a majority of concepts often relate very loosely with this keyword. In the following section, we shall present the ICS-based query refinement approach.

5. ICS-based Query Refinement

To improve the retrieval precision, the traditional approaches taken in both some commercial IR engines and many IR researches are to refine and/or expand the original query automatically based on the documents retrieved by the original query [9,12]. Query expansion is to extend user queries without re-computing the weights of the query items, while query refinement is to re-compute the weights of query items[14]. In this paper, we chiefly discuss the implementation of query refinements.

Section 4 transforms keyword-based queries into concept-based ones. In this section, we focus on the ICS-based query refinement using the concept-based queries obtained in Section 4.

Given query $Q = (k_1, k_2, \ldots, k_n)$, the matrix $QCR_{n \times N}$ (Query-Concept-Relevance) represents the association degrees between keywords in Q and concepts in domain ontology DO, defined as:

$$QCR_{n\times N} = (I - KCV_1, \ldots, I - KCV_n)^T = (r_{ij})_{n\times N}$$

where r_{ij} represents the association degree between keyword k_i and concept c_j in DO.

After transforming keyword-based queries into concept-based ones, we obtain the vector QV (Query Vector) as the refined query , that is, $QV = (cw_1, cw_2, \ldots, cw_N)^T$ where $cw_i = \sum_{j=1\ldots n} r_{ij}$ and $i \in (1..N)$.

According to the definition of ICS in Section 2, we can obtain the vector CS (Cognitive Structure Vector) that represents the user's cognition degree for all concepts in DO, which can be defined as:

$$CS = (doc_1, doc_2, \ldots, doc_N)^T$$

where doc_i represents user u's cognition degree of concept c_i and is defined as equation (4).

$$doc_i = \begin{cases} DOC_u(c_i), & if\ DOC_u(c_i) >= \theta \\ 0, & others \end{cases} \quad (4)$$

where θ is a threshold given by u and $0 <= \theta <= 1$.

The vector QV' is the refinement of QV by CS (representing the information of ICS), that is,

$$QV' = \alpha * QV + (1 - \alpha) * CS$$

where α is the adjusting parameter given by our system and $0 <= \alpha <= 1$, as depicted by Figure 4. In other words, the higher the association degrees between keywords and concepts and the deeper the user's cognition of concepts, the larger is the weights in QV'.

$$\begin{array}{ccccccc} & QV & & & CS & & QV' \\ \alpha * & \begin{pmatrix} cw_1 \\ \cdot \\ \cdot \\ cw_N \end{pmatrix} & + & (1-\alpha) * & \begin{pmatrix} doc_1 \\ \cdot \\ \cdot \\ doc_N \end{pmatrix} & \Rightarrow & \begin{pmatrix} w_1 \\ \cdot \\ \cdot \\ w_N \end{pmatrix} \end{array}$$

Figure 4: The computation of QV'.

Since, the retrieval document set is very large, we index this set by concepts in order to improve the retrieval performance. Hence, we curtail the scope of retrieval document set before we match retrieval queries with document set. Subsequently, we describe the curtailing of retrieval document set algorithm formally as Algorithm 1 where $DC_{M \times N}$ represents the relationships between documents and concepts and the definition of dc_{ij} is described as follows.

$$dc_{ij} = \begin{cases} 1, \textit{if document } d_i \textit{ is annotated by concept } c_j \\ 0, \textit{others} \end{cases}.$$

As a result, the curtailed retrieval document set is very small compared with the original one. In the following section, our experimental comparisons of different approaches will be given.

Algorithm 1: Curtailing of Document Set.

Input:
$QV' = (w_i)^T$,
$DC_{M \times N} = (dc_{ij})$,
φ // φ represents the threshold given by our system
Output :
D' // the curtailed documents set
Begin

```
D'←NULL;
For each w_i in QV' Do {
    If (w_i >= φ) {
        For each j in M Do {
            If (dc_ij = = 1)
        // d_j represents the document j
                D'= D' ∪ { d_j };
        } //end for
    } // end if
} //end for
Return D';
```

End.

---

6. Experimental Evaluation

6.1 Experimental Setup

In our system, we employed two domain ontologies: *EO* (Economics Ontology)[4] in economics domain and *ACMCCS98* [6] ontology in computer science domain. There are 9740 classes and 15,222 properties in *EO*. This ontology almost covers most of concepts and their relationships in area of economics. *ACMCCS98* is the classification system recommended by *ACM* in the area of Computer Science. The statistic data about two domain ontologies are illustrated in Table 1, where *SER* denotes the semantic equivalence relationships, *SPR* denotes the semantic parent-child relationships and *SAR* denotes the semantic association relationships.

Table1 : Statistics of Domain Ontologies.

Items	EO	ACMCCS98
# of Concepts	9470	1472
# of SER	1516	0
# of SPR	5460	1461
# of SAR	4368	132

We used DBLP [7] and DLPers datasets to test our experiments. DLPers is the set of document resources in DLPers V2.0 (a sub-system of our university digital library system) which have been annotated by *EO*. DBLP are the resources from DBLP database. The details of DLPers and DBLP are shown as Table 2.

Table2 : Statistics of the Document Sets.

Document Sets	# of Docs.	Language
DLPers	785,426	Chinese
DBLP	19,229	English

The size of window, as used for computing the co-occurrences of keywords and concepts, was set as [-8, 9] recommended in [5]. The θ was set as the median of $QV' = (w_i)^T$, by considering that the median can not be influenced by irregular values. Accordingly, we can averagely reduce the search space to half by the selection of such θ and also that may lead to little impact on the recall.

Our experiments were performed on Windows 2000, Java, Oracle9.2, where Oracle9.2 was used for our domain ontologies, ICSs information and document sets. Moreover, the DB Server was on Intel 2.8 GHz CPU with 2G RAM.

To comparison, we provide three search ways using different implementation strategies in our prototype system: (1) Traditional keyword-based Search(KS), to treat the keywords as themselves; (2) Concept-based Search(CS), to return the required documents in terms of the concepts inputted by the user; (3) Personalized Search(PS), to transform the inputted keywords into relevant concepts of domain ontologies and then return results using user's ICS. The comparisons of the above three approaches are depicted as Table 3.

Table 3: The Comparisons of different approaches.

Search Approaches	1	2	3
TS	N	N	N
CS	Y	Y	N
PS	Y	Y	Y

where :

- Item 1: whether documents are annotated by concepts.
- Item 2: whether retrieval query is refined.
- Item 3: whether retrieval query is refined by user's ICS.

We use the topics to describe user's query requirement. Note that, we limited these topics in areas of economics and computer science domain currently. The format of topics is similar with TREC [8]: { No., Title, Desp (Description), Narr (narration)}. Table 4 gives an example for the description of a topic. In our experiments, 8 users provided 6 topics of economics domain and 15 topics of computer science domain together.

Table 4: The Example of a Topic.

No.	02
Title	The Postfix Expression
Desc	The computation of the Postfix Expression
Narr	To describe the computation approach and algorithm for the postfix expression and infix expression is not included

At the same time, the above users provided their cognitive center concepts and the corresponding degree of cognition for each concept. We manually constructed three queries for each topic for the three search approaches respectively.

6.2 Experimental Analysis

In this section, Precision@n and AP@k are used as measures. Precision@n is measured by the precision at a cut-off point of the top-n retrieved documents.

$$\Pr ecision@n = \frac{\#relevant\ docs\ \text{in}\ top-n\ retrieved}{n} \quad (5)$$

AP@k is measured by the average precision of top-k documents. Its computation is as equation (6), where k represents a cut-off point of the top-k retrieved documents, r represents the number of relevant documents before top-k, $rank_j$ represents the rank of the jth documents.

$$AP@k = \frac{1}{r}\sum_{j<=k}\frac{rank_j}{j} \quad (6)$$

In our experiments, both n and k are set as 20 since users often pay more attention to the top-20 retrieved documents. The experimental results of Precision@20 and AP@20 are depicted as Figure 6 and Figure 7 respectively.

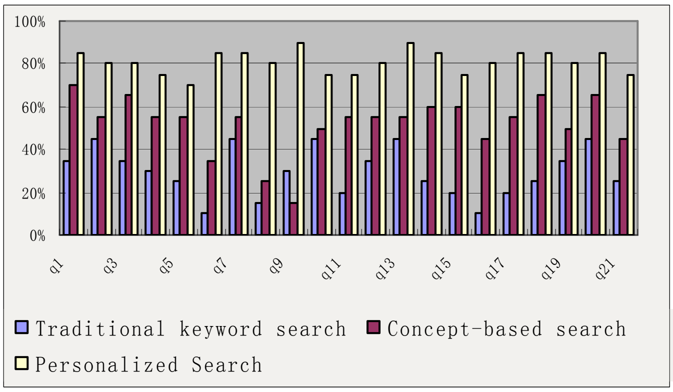

Figure 6 : Comparisons on Precison@20

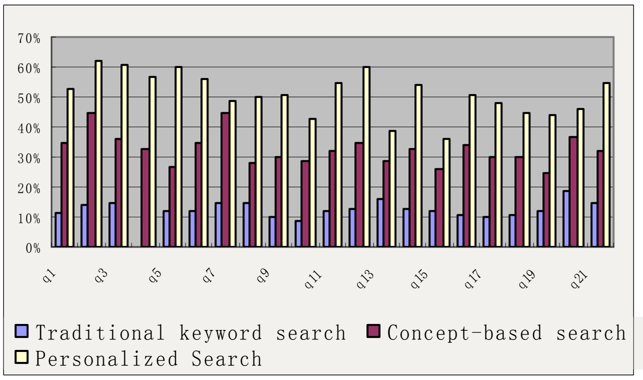

Figure 7 : Comparisons on AP@20.

From Figure 6, we can see that as far as Precision@20, PS > CS > TS, which is as our anticipation because ICS-based query refinement could enhance the weight of the concepts relevant to user's query needs. Figure 7 shows the similar result for the measure AP@20, that is, PS> CS > TS. Because concept-based search is benefit from concept relevance and therefore the semantic relevant terms are reinforced in query concept vector. On Precison@20, the average value of PS is 81%, which is 29% larger than that of CS, 51% larger than that of TS. On AP@20, the average value of PS is 51%, which is 18% larger than that of CS, 39% larger than that of TS.

6.3 Complexity of Time

As we have seen above, the three search engines use the similar way to find relevant documents. The key factors to affect the performance of search engine are the modules of vector matches and query refinement. In the following discussions, suppose the average length of keyword-based vectors is x, the average length of concept-based vectors is y and the number of documents is M. For traditional keyword search, the complexity of time is $O(M \times x^2)$, since there is no query refinement in this search and accordingly the execution time is chiefly consumed by vector matches.

For concept-based search, the complexity of time is $O(N' \times y^2)$ where N' represents the number of documents in the curtailed document set, since the query refinement focus on the vector additions essentially on which the consumed time is linear and accordingly the execution time of this approach is also mainly spent on vector matches.

For personalized search, although the computation of I-KCRVs takes time largely, we put it in the pre-processing phase so that we can ignore the execution time spent on it. Similar with concept-based search, the execution time of this approach is also consumed by vector matches and its complexity of time is $O(N' \times y^2)$.

The comparisons of these three approaches are shown in Table 6. Compared with two other approaches, the execution time of personalized search is impaired slightly while the precision of it is improved highly since additional computation of I-KCRV is pre-processed.

Table 6: Comparisons of Time Complexity.

Search Approaches	Query Refinement	Vector Matches	Total
TSE	—	O(N*x2)	O(N*x2)
CSE	O(1)	O(N'*y2)	O(N'*y2)
PSE	O(N*len*(y＋x)/2)	O(N'*y2)	O(N'*y2)

7. Conclusions and Further Work

In this paper, we propose a personalized retrieval approach using ICS and construct a prototype personalized search system. Our novel features are: first, we model user's preference by his/her static ICS, and use them to refine user's queries. Second, a novel method of computing the association degrees between keywords and concepts is proposed by utilizing the frequency of co-occurrence between keywords and concepts in a corpus and the occurrences of keywords in documents annotated by concepts. Third, experimental results show that our approach outperforms the keyword-based search and concept-based search as a whole, in particular from the precision point of view.

There are many interesting future research topics. For example, we can use user's dynamic cognitive behaviors to track the transition of cognitive structure; our method can be improved by distinguishing the concepts occurring section in documents; the relevance between concepts can be utilized in semantic-based search.

Acknowledgements

This work is partly supported by the National Natural Science Foundation of China under Grant NO. 60496325 and 60573092, and Project 985 of MOE.

References

[1] Berners-Lee T, Hendler J, Lassila O. The Semantic Web - a New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution

of New Possibilities. Scientific American. 2001,284(5):34-43

[2] Medin D, Ross B. Cognitive Psychology. Harcourt Brace College Publishers ,1997

[3] Tian X, Du X Y. Modeling User's Cognitive Structure in Contextual Information Retrieval. IEA/AIE2006, Submmited. 2006

[4] Li M, Wang D, Du X Y, Wang S.:Ontology construction for semantic web:A role-based collaborative development method. Lecture Notes in Computer Science (Proc. of APWeb2005), 2005, 3399: 609-619

[5] Lu S, Bai, S. Quantitative Analysis of Context Field in Natural Language Processing. Chinese Journal of Computers, 2001(7) :742-748

[6] ACMCSS98. http://www.acm.org/class/1998/

[7] DBLP. http://www.informatik.uni-trier.de/~ley/db/

[8] TREC. http://www.trec.org/

[9] Salton G, Buckley C. Improving retrieval performance by relevance feedback, Journal of the American Society for Information Science, 1990, 41(4) : 288-297

[10] Loh S, Wives L, Oliveira P. Concept-based knowledge discovery in texts extracted from the Web. ACM SIGKDD Explorations Newsletter. 2000 , 2(1) : 29-40

[11] Qiu Y, Frei H. Concept based query expansion. Proc. ACM SIGIR1998, ACM Press, 1998, 160-169

[12] Xu J, Croft W, Query expansion using local and global document analysis, Proc. ACM SIGIR 1996, ACM Press, 1996, 4-11

[13] Fang H, Zhai C X. Semantic term matching in axiomatic approaches to information retrieval, Proc. ACM SIGIR 2006, ACM Press, 2006,115-122

[14] Salton G, McGill M. Introduction to Modern Information Retrieval. New York: McGraw-Hill; 1983

[15] Collins A M, Loftus E F. A spreading-activation theory of semantic processing Psychological Review, 1975, 82(6) : 407-428..

[16] Salton G, Buckley C. On The Use Of Spreading Activation Methods In Automatic Information. Proc. ACM SIGIR1988. ACM Press, 1988, 147-160.

[17] Crestani F. Application of Spreading Activation Techniques in Information Retrieval. Artifical Intelligent Review. 1997, 11(6/December): 453-482

[18] Belew R K. Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW. Cambridge University Press, 2001

[19] Hu H, Du X Y. ConAnnotator: Ontology- aided collaborative Annotation System. Proc. CSCWD2006, IEEE Press, 2006, Vol.Ⅱ:850-855