International Journal of Hybrid Information Technology
Vol. 10, No. 6 (2017), pp.33-46
http//dx.doi.org/10.14257/ijhit.2017.10.6.04

Personalized Semantic Query Expansion Based on Dynamic User Query Profile and Spreading Activation Model

Shengrong Zhu[1], Xiangguang Meng[2], Feixiang Chen[3] and Xuan Tian*

[1,3,]School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China*

[2]School of Distance and Continuing Education, Communication University of China, Beijing 100024, China

[1]zhusr_bjfu@163.com, [2]mengxg@cuc.edu.cn，[3]fxchen@126.com，
**tianxuan@bjfu.edu.cn*

Abstract

Semantic query expansion is a widely used method to resolve the query problems of synonym and polysemy in the information retrieval field. However, it does not make users more satisfied with the search results because too much noise unfit to users' needs is introduced in the process. In this paper a new framework combining personalization with semantic query expansion is proposed to overcome the noise problem brought by semantic query expansion. In the proposed framework, firstly, instead of using traditional hierarchical expansion strategy, the spreading activation model (SAM) is used for enhancing the selection of expansion terms to reduce the noise. Secondly, to get more accurate expansion terms for individual search, dynamic user query profile is built to capture individual variable query needs and is integrated into the semantic expansion process. The proposed expansion process is described by four steps: building dynamic user query profile, concepts mapping, personalized semantic query expansion and determining the final expansion terms. Four groups of experiments were designed to verify the validity of the proposed method. The experiment results show that the proposed method outperforms both traditional hierarchical expansion and keyword-based query, which manifests that building dynamic user query profile is important for depicting user query needs in semantic query expansion and it is more rational to improve query expansion based on spreading activation model. Moreover, personalized semantic query expansion based on dynamic user query profile and spreading activation model can reduce noise of semantic query expansion and improve the search effectiveness.

***Keyword**: semantic query expansion, personalized information retrieval, dynamic user query profile, spreading activation model*

1. Introduction

In the information retrieval field, there exists some serious semantic problem because of the following two facts, (1) the document indexers and the users' queries may be not the same words but with closely related senses; (2) queries and documents terms can have multiple senses which may lead to sense ambiguity problem [1,2]. It is known as the vocabulary problems, compounded by synonym (different words with similar meaning) and polysemy (a word with different meanings)[3]. Semantic query expansion is a broadly explored method to resolve the problems of synonym and polysemy. It is often to expand the user query with the new terms that meaning are closely related to the input keywords. Such relationships are usually extracted from large scale thesauri or semantic supporter, such as WordNet, domain ontology, linked data [5-8] and semantic expansion strategy [9-15], in which various sets of synonyms, hyponyms, etc. are often predefined.

ISSN: 1738-9968 IJHIT

However, semantic query expansion approaches may result in topic drift problem, namely too many noise terms unfit to users' needs are introduced in the process. For example, "program" have multiple paraphrases: television show, plan or computer program. Assuming user tends to search something about computer program, the retrieval without considering semantic problem may get the results only if it contains "program" and including all sense of "program", while an automatic expansion of queries containing "program" with "computer" might work well in this case. That means individual needs can help to exclude noise terms in semantic query expansion.

In fact, there have been many researches on semantic query expansion combined with personalization [7,8,16-19]. These researches often study user profiles and personalized search algorithm to enhance the semantic query expansion. However, how to enhance the selection of expansion terms on large scale thesauri or semantic supporter to reduce the noise is often ignored. And another fact is that a search user's needs are also dynamically changed with time, and which is often neglected in the aforementioned personalized semantic query expansion related researches.

In this paper, a new framework is proposed to combine personalization with semantic query expansion. A spreading activation model which simulates the thinking process of human brain more vividly is utilized as the expansion strategy to enhance the selection of expansion terms. And a kind of dynamic user query profile is built to capture individual variable query interest to reduce the noise terms unfit to users' needs. The implementation of the framework includes four steps**.** Firstly, building dynamic user query profile on the basis of query logs; secondly, mapping query terms to appropriate concepts on WordNet; thirdly, exploring spreading activation model to implement personalized semantic query expansion; finally, determining the selection of final expansion terms. In addition, variant systems are built for comparison, like keyword-based searching with default Lucene performance (NE), general hierarchical query expansion without mapping filtration (EHNF), and hierarchical query expansion with mapping filtration (EHF). The evaluation is performed using MAP and the degree of relevance (DOR). The result shows that our proposed method performs well and is better than compared ones.

In the rest of this paper, Section 2 reviews related work on two aspects: semantic query expansion and personalized search. Section 3 describes the proposed framework and detailed implementation. Section 4 presents experiment data, experiment design and evaluation result. Section 5 concludes the paper.

2. Related Work

This paper brings together two IR areas: semantic query expansion and personalized search. A bundle of research study both domains. However, not much has been done specifically aiming at combing them in good way. In this section we first introduce some approaches for semantic query expansion and personalized search separately and then discuss personalized semantic query expansion.

Semantic Query expansion is essential to overcome the difficulty in using one single term to represent an abstract concept and also natural language ambiguity. Query expanding with related words increases the performance of search engines, while finding and using them is really an open problem [20]. Semantics considered as one kind of relevant relations also confront this issue. Semantic query expansion methods may differ in semantic supporter and semantic expansion strategy. Many studies utilize ontology to express semantic relation and implement semantic query expansion on ontology[5-7]. M. Shabanzadeh et al. extracted semantically related words from WordNet[5]. D. Pal extracted candidate expansion terms from a set of pseudo-relevant documents and then filtered terms according to the semantical relation to query based on WordNet[6]. In domain retrieval, domain ontology was built and query expansion was implemented with different strategies[7-12], L. Khan et al selected all successor concepts of disambiguated concepts of query for query expansion[9], and G. Zou et al only selected an instance set, direct descendant concept set and direct grandfather concept set[10], M. C. Lee et al employed concept hierarchy of domain ontology for expansion with a pruning threshold [11]. N. A. Segura et al considered concepts and relations together in the expansion process, concepts with meaningful relations were given a high priority for expansion[12]. There are also other semantic supporters like Linked data[21] and so on. In addition, other strategies can be introduced into semantics query expansion, L. Li et al. found out terms both semantically and co-occurrence related to query on massive web page set and search engine performance evaluation data[22]. Although using co-occurrence to complement semantic query expansion is quite a good idea, it may lead to topic drift easily especially when the document is not quite related to query.

Most query expansion approaches mentioned above have inherent limitations when trying to understand user intentions from a short query and introducing too much noise may lead to topic drift [7]. Therefore, for more precise query expansion, semantic query expansion is significant to be combined with personalization. Personalized search comprises two major components: User profiles and the retrieval algorithm[16]. This paper focus on User profiles, which considers what the factor is to express personalization. K. Sugiyama et al. analyzed surfing behavior to generated user profiles as features of the visited pages and then ranked search results based on the similarity between each URL and the user profile[17]. F. Cai et al. and R. W. White et al. not only considered user query behavior, but also consulted other users with similar interests[23, 24], T. T. Vu et al. went a step further of similar user to study dynamic group formation[25]. A. Chirita et al. utilized desktop terms describing user’s interests[18]. P. N. Bennett et al. considered a broad variety of context including the users’ characteristics, interests, recent activity, focus and so on[26].

This paper study personalized search combines with semantic query expansion. Many studies combined semantics with personalization by the way of interactive query refinement [13], relevance feedback[8,14] or even search results clustering[15]. These methods made certain improvements in retrieval precision but demanded user’s interaction to certain extend. Generally, automatically search should be better. G. J. Hahm et al studied engineering document retrieval, learnt user query profile to obtain user interests, built domain ontology, subsequently implemented query expansion on the domain ontology according to user query profile[7] and it obtained notable improvement. O. C. Cure et al. applied expansion in medical field, while achieving personalization requiring individuals’ interaction. N. Ahmadian et al. utilized context to pick expansion terms, but the problem of sense ambiguity is not taken into account[19]. X. Tian et al. expanded economic concepts based on spreading activation model, while the weight of concepts were static given and the expanded concepts might be beyond user’s usual search range[27].

In this paper, spreading activation model which simulates the thinking process of human brain more vividly is utilized as the expansion strategy to reduce noise and dynamic user query profile is built and integrated to capture individual variable query needs. The detailed framework working process can be shown in Figure 1. For user retrieval, remove stop words based on stop terms list first and get the preprocessed terms. After terms preprocessing, on the one hand, new input is treated as query record to update the user query profile, on the other hand, mapping terms to concepts in WordNet. Then mapped concepts are treated as objects to experience the personalized semantic query expansion process on the foundation of spreading activation model and integrating dynamic user query profile. Finally, determining the final expansion terms from the expansion concepts set.

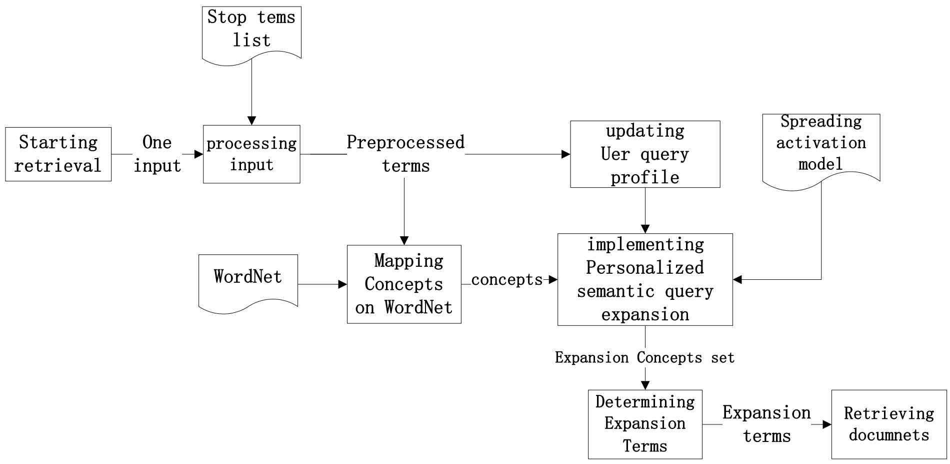

Figure 1. The Framework Working Process

3. Modeling

To overcome the problem of general semantic query expansion, SAM is utilized for enhancing the selection of expansion terms to reduce the noise. Moreover, to get more accurate expansion terms for individual search, dynamic user query profile is built to capture individual variable query needs and is integrated into the semantic expansion process. Naturally, building dynamic user query profile and the integration process occupy the main modeling focus. This section is described according to the proposed framework working process.

3.1. Building Dynamic User Query Profile (DUQP)

Definition 1(Dynamic User Query Profile). Dynamic user query profile describes the query state of terms in individual retrieval over time, it can be formalized as triples like $DUQP=List(qw,TW,lqt)$, where qw represents the retained query term after preprocessing, lqt is the last time querying qw, and TW describes the final term query weight at time lqt, which reflects the user query interest and changes over time. The bigger TW is, the more likely is user to query qw.

The basic idea of building DUQP is to analyze user query considering both query time and term frequency, which is based on the following assumption. Firstly, query itself ought to be the most informative source for digging out current user query needs. Secondly, query history reflects individual interest to certain extend, and the interest are more likely to represent users' query need It is more likely for user to query terms he or she retrieved more before. Therefore, query is set as research object, query time and term frequency are considered as important index in DUQP. Before r introducing the detailed idea, related concepts are defined firstly.

Definition 2(One Input). One input represents one valid input, corresponding to one record in query log.

Definition 3(One Query/Query Session). One query includes one or several input with the identical or similar query intention, also called one query session, corresponding to several adjacent records in query log.

Individual inclines to reformulate query when they are not satisfied with the current results or intend to obtain more related results, the success of information retrieval depends on how he or she formulates queries [28]. Therefore, we divide query session, and regard the time of the last input in one session as the session time. Generally, the time interval of each input in one session won't be long, we divide sessions according to the time interval and set it to twenty minutes.

Assuming there are three input in one query session, and each input is preprocessed into terms, $q_1 = \{(w_1, w_2, w_3), t_1\}$, $q_2 = \{(w_1, w_3), t_2\}$, $q_3 = \{(w_1, w_4), t_3\}$. The current input is q_3 and t_3 is the current time, the weight of w_1, w_4 in current input are set as 1,the decline parameters of weight is $\beta(0<\beta<1)$, so the weight of w_1 and w_3 in q_2 are β^2. By analogy the weight of w_1, w_2, w_3, w_4 in this query session respectively is $(1+\beta+\beta^2)$,1,$(\beta+\beta^2)$, β^2. Moreover, the weight of the term not appearing in the session is 0.

In summary, for one query session, supposing there are n input and the query terms are gathered as set ses, the decline parameters of weight between input is $\beta(0<\beta<1)$, the last input is viewed as current input, then the weight of term qw in this session is calculated as follows:

$$w = \begin{cases} 0, & (qw \notin ses) \\ \sum_{i=1}^{n} \beta^{(n-i)}, & otherwise. \end{cases} \tag{1}$$

Where i represents the $i\text{-}th$ input. For term qw, considering the whole query history, the query weight can be obtained through the equation as follows:

$$TW_t = TW_{t-1} * \eta + w \tag{2}$$

Where TW_t and TW_{t-1} describe the weight by the end of the $t\text{-}th$ and $(t\text{-}1)\text{-}th$ query of term qw separately, η expresses attenuation effect over time, w means the weight of term qw in the $t\text{-}th$ query, which can be calculated through formula(1). Referred to the forgetting rule of memory proposed by H.Ebbinghaus[29], we utilize the classical damping function Eq.(3) to express the attenuation effect of the query state of terms over time, as a rule to reflect the variability of user's interest in query terms.

$$\eta = e^{-\delta(t_t - t_{t-1})} \tag{3}$$

Where δ is the attenuation factor. Given that multiple queries slow down the attenuation effect, we set $\delta = \mu / CCnt_t$, where $CCnt_t$ refers to the total count of term qw being searched in the $t\text{-}th$ query. The more certain terms searched, the more slowly interest declined correspondingly. μ is a regulatory factor. t_t and t_{t-1} represent the time of $t\text{-}th$ and $(t\text{-}1)\text{-}th$ query of qw respectively. Assuming that the current time is T, the last time for querying qw is at time t_l, then the weight for term qw can be expressed as below:

$$TW = TW_l * e^{-\frac{\mu}{CCnt_t}(T - t_l)} + w \tag{4}$$

Where TW_l expresses the weight by the end of time t_l, and w refers to the current weight of qw, which can be calculated via Eq.(1). When term qw doesn't exist in current query, the current weight is 0, otherwise, it is calculated as $\sum_{i=1}^{n} \beta^{(n-i)}$. Then the more comprehensive equation can be as follows:

$$TW = \begin{cases} TW_l * e^{-\frac{\mu}{CCnt_t}(T - t_l)}, (qw \notin ses) \\ TW_l * e^{-\frac{\mu}{CCnt_t}(T - t_l)} + \sum_{i=1}^{n} \beta^{(n-i)}, otherwise. \end{cases} \tag{5}$$

In addition, considering the problem of data sparsity, for each term which doesn't occur in query history is weighted 1, and for the other terms, they are weighted $TW+1$. Finally, we get DUQP.

3.2. Mapping Concepts

Mapping concepts is the procedure of mapping preprocessed query terms to concepts, it is significant since the query expansion is based on concept, moreover, most terms own more than one paraphrase, and one term can match with several concepts, although user may be just interested in one of them, then, appropriate concepts should be picked if possible.

WordNet is the supporter for mapping here. In WordNet, terms are grouped into synsets and synsets are interlinked by means of conceptual-semantic and lexical relations. One synset corresponds to one concept. The ones contain certain query term and own higher similarity with query are chosen preferentially. Given the fact that query is often short, we take other input in current query session as portion to calculate the similarity between term and concept. The mapping procedure can be disassembled into two steps: finding concepts containing the term and picking the appropriate concepts according to the similarity.

Related symbols are defined first. For query, the current input being preprocessed is $cin = \{i_1, i_2, \dots, i_c\}$ and the others in the same session is $coi = \{oi_1, oi_2, \dots, oi_c\}$. For term i_c in cin, we assemble a new set $ic = \{i_c, oi_1, oi_2, \dots, oi_o\}$. In addition, the synset in WordNet can be formulized as $syn = \{t_1, t_2, \dots t_s\}$. The term weight in syn and cin are set as 1, and the term weight in coi are calculated as formula (1). The similarity between syn and i_c determines which concept term i_c mapped to, which can be described as below:

$$sim(syn, ic) = \sum_{t \in (syn \cap ic)} w_{ic}^{t} \tag{6}$$

Where t represents the common term in both syn and ic, w_{ic}^{t} refers to the weight of t in ic.

It should be mentioned here that the terms in cin are preprocessed noun in current input, and we do not expand the terms which cannot find appropriate mapping concept.

3.3. PSQE (Personalized Semantic Query Expansion)

The main focus of this paper is building DUQP and integrating it into the semantic expansion process based on spreading activation model. This section is mainly focused on the latter. It is called the process of personalized semantic query expansion (PSQE). The strategies of general semantic query expansion are often based on hierarchical relationship and the expansion level is determined empirically just like selecting all successor concepts, instance set, direct descendant concept set and direct grandfather concept set[9-12]. Here we borrow the thought of spreading activation model to enhance the selection of expansion terms to reduce noise and combine it with dynamic user query profile to get more accurate expansion terms.

Spreading activation model was utilized in retrieval many years ago[30]. It is mainly determined by two parts, the presentation network and the spreading strategy. In the network, when a concept is activated, the activation will spread through the links to its directly connected concepts and to the concepts related to those concepts. As the activation becomes too weaker to active others it stopped and when to stop is determined by the spreading strategy. The abstract network structure graph can be toughly expressed as Figure 2.

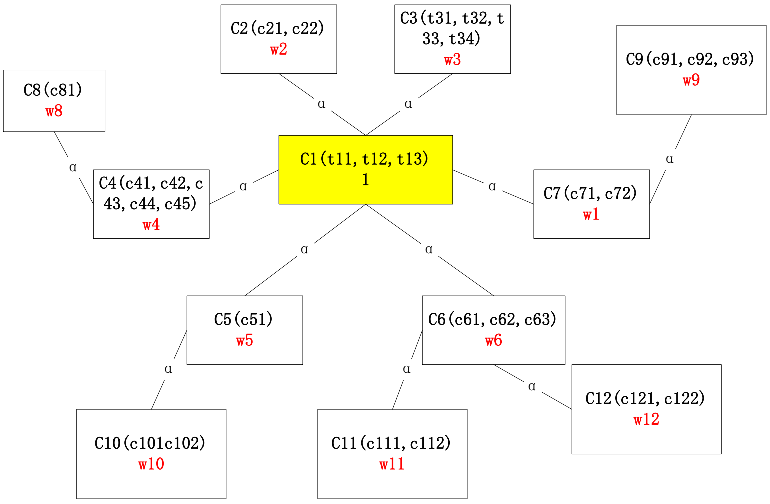

Figure 2. Abstract Network Structure Graph of PSQE

WordNet is the presentation network here and the impact of relation distinction between concepts is weaken in WordNet. For mapped concepts, they are regarded as expansion centers. In Figure.2, c_1 is the expansion center, the concept from c_2 to c_{12} is directly or indirectly related to c_1 and w in the rectangle is the weight of corresponding concept. Spreading strategy, also the constraint of picking expansion concepts, is based on both semantic and personalization. Semantics relies on the conceptual-semantic and lexical relations of WordNet itself, we use parameter α to express the distance impact on the relation between concepts. The personalization relies on DUQP. To describe the detail process, several related concepts are defined firstly.

Definition 3(WOC). $\mathrm{WOC}(c_i)$ is the weight of concept c_i in WordNet, c_i consists of one or several terms with similar sense, the largest term weight is employed to describe individual interest of the concept. Then WOC equals to the largest term weight in the concept and the term weight is from DUQP.

Definition 4(DOCC). $\mathrm{DOCC}(c_i)$ is the degree of concept c_i related to center concept. It not only indicates the semantic relationship between concept c_i and center concept c_c, but also reflects user's interest in the concept, which can be formulated as equation: $DOCC(c_i) = WOC(ci) * \alpha^{d(c_i,c_c)}$, where $d(c_i, c_c)$ refers to the distance between c_i and c_c.

Definition 5(Expanded Concept Set (ECS)). ECS is the set consists of expanded concepts, which are the ones whose DOCC value is larger than threshold γ. The distance between concepts and center concept $d(c_i, c_c)$ is less than the maximum distance $MaxD = \lfloor d(c_i, c_c) \rfloor$, where $\alpha^{d(c_i,c_c)} >= \gamma$.

The primary expansion procedure is as follows. Firstly the threshold γ is set and the maximum distance MaxD can be figured out, then Start from the expansion center concepts, traverse every directly related concept in WordNet by width first search and put unvisited ones into temporary set. Pick out the concepts whose DOCC value is larger than γ and put into ESC. Then view the concepts in temporary set as center concepts, traverse the directly related concept of them, repeat the process until there is no more unvisited directly related concepts or the distance is bigger than MaxD. Algorithm 1 illustrates the approach. During the expansion procedure, we can retain the concepts c_i with higher WOC and relatively larger $d(c_i, c_r)$, which may be abandoned in paper [27].

Algorithm 1 the Algorithm of PSQE

```
Input: DUQP, threshold γ of DOCC value, parameter α,center concept c_c
Output: ECS
Begin
    MaxD ← calMaxDis( γ , α )   //calculate max expanding distance
    Dim Layer As Integer; Dim tempSet1,tempSet2 As List
    Layer ←0     //current expansion layer
    tempSet1.push_back(c_c);
    while  isNotEmpty(tempSet1) & Layer<= MaxD
        Foreach c in tempSet1
            IF DOCC(c)>= γ
                ECS.push_back(c)
            End IF
        End Foreach
        Layer++

        Foreach c in tempSet1
            tempSet2 ← GetDirectlyRelatedCs(c)
            Foreach t in tempSet2
                IF  Visited(t)
                    tempSet2.remove(t)
                EndIF
            EndForeach
            tempSet1.remove(c)
        EndForeach

        tempSet1 ←  tempSet2
    EndWhile
End
```

All the center concepts are carried out with the same procedure, and duplicate concepts are merged by retaining the biggest weight, then we get the final ESC.

3.4. Determining Expansion Terms

This section describes the procedure of term selection and weight normalization. Expansion terms are derived from ECS. All terms are gathered together to obtain the final expansion terms. Terms occur in one concept own the same weight equaling to the value of DOCC, and the biggest weight of term t_i is retained when it occurred in different concepts, which can be expressed as Eq.(7). Sort all the terms by weight in descending order and choose k highest ones as the final expansion terms.

$$w_{t_i} = \max_{t_i \in c_j}(w_{c_j}) \tag{7}$$

The step before retrieval is weight normalization. We comply with the principle that original input should have the highest weight and they are given 1, the expanded ones ought to be no more than 1, we normalized it as below:

$$w'_{t_i} = \begin{cases} w_{t_i} & ,(w_{t_i} < 1) \\ (w_{t_i} - 1) * \frac{0.99-\alpha}{max(w_t)-1} + \alpha, & (w_{t_i} \geq 1) \end{cases} \tag{8}$$

Where w_{t_i} describes the original expanded term weight calculated in Eq.(7),

w'_{t_i} refers to the final expanded term weight, α is the parameter introduced in PSQE, $max(w_t)$. is the maximum weight value of all terms. When the original term weight is bigger than 1, we normalized them to the range from α to 0.99. In this way, we emphasis the importance of bigger weight terms and following the principle as well.

4. Experiments

In this section we briefly describe the experiment related design and the results we have obtained through experimentation. The experiment is based on the following assumption. For one ambiguous term, user tends to attach more attention on the paraphrase which was queried more in history. Thus, we judge the relevance between return results and query based on the relevance between results and user's preferred paraphrase implied in query history.

Experiments data and design

Document collection: We download web pages related to query from the top pages result of search engine "being" and also some other pages irrelevant to our queries or with certain query terms of other paraphrases to be treated as noise. For each query we have at least 10 relevant pages and 5 noisy pages.

User query log: Aol2006 query log data is utilized as user query history, it contains query records from March 1th to May 31th in 2006. We pick users who queried the ambiguous terms we choose more in logs.

Query: To demonstrate the effective of our approach, ambiguous terms such as: bat, program, hardware were chosen as our input test, and the query time is set June 1th in 2006 to simulate the daily search. In addition, the WordNet version is 2.1.

Method Compare: To evaluate the proposed expansion approach we proposed four variant systems. Firstly, the system, **NE**, no expansion, is the baseline keyword-based system, in which Lucene with the default function is used. Secondly, the system, **EHNF**, expand without mapping filtration, each query term maps to all concepts containing them in WordNet and employs traditional hierarchy expansion method giving the expansion layer three. Thirdly, the system, **EHF**, expand with mapping filtration when other related input, called preceding input, exists in current query, has the identical process with EHNF except for the term mapping filtration before expansion, namely, each query term maps to certain concepts considering the preceding related input. Fourthly, the system of our approach, **PSQE**.

Design and evaluation: As shown in Table1, two group experiments are desighed. In group one, the input has no preceding input, while group two has related ones. We evaluate result in two aspects, relevance and the degree of relevance (DOR) between related results and user query history. Relevance demonstrates the general effective of proposed method, which is based on mean average pre@n and we assign n=5,10,15,20. Two relevant documents may own different DOR, which can be assessed through the order of return results.

Table 1. The Experiment Design

input	preceding input(Group one)	preceding input(Group two)
bat	-	mammal chiropteran
program	-	computer interface
hardware	-	computer memory
mouse	-	device computer
…	…	…

Experiment results

Parameters: During the experiments, we adjust all parameters mentioned in previous sections to optimize the proposed method, the value are given in Table2. Besides, given the query sparsity in query history, the time unit in damping function η is set one month.

Table 2. Parameter Value

parameter	meaning	value
β	decline parameter between session input	0.8
α	decline parameter between directly related concepts in WordNet	0.85
γ	expansion threshold	0.6
μ	regulatory factor in damping function	0.2

Figure 3 depicts the MAP result comparison of four methods without preceding input, which illustrates that PSQE outperforms the other three method in the terms of MAP of methods without preceding input. EHNF and EHF have the same performance, while a bit bad than NE. Figure 4 describes the MAP result comparison of four methods with preceding input, which depicts that PSQE has the best performance, EHF, NE has performance in descending order with preceding input and EHNF has the worst performance.

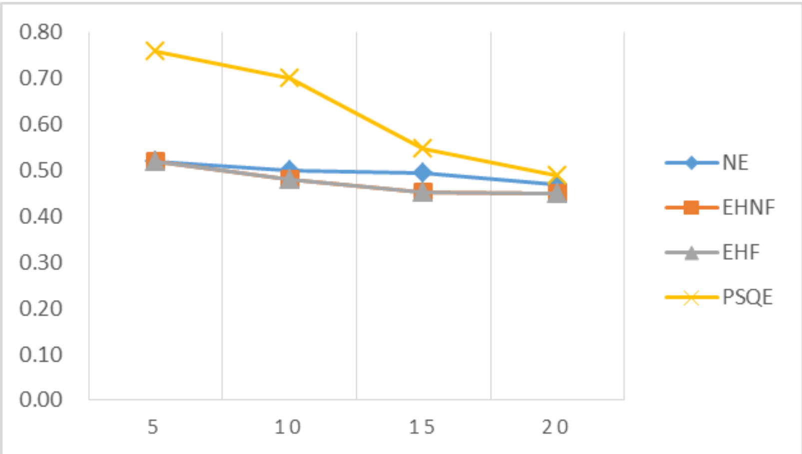

Figure 3. MAP of Methods without Preceding Input

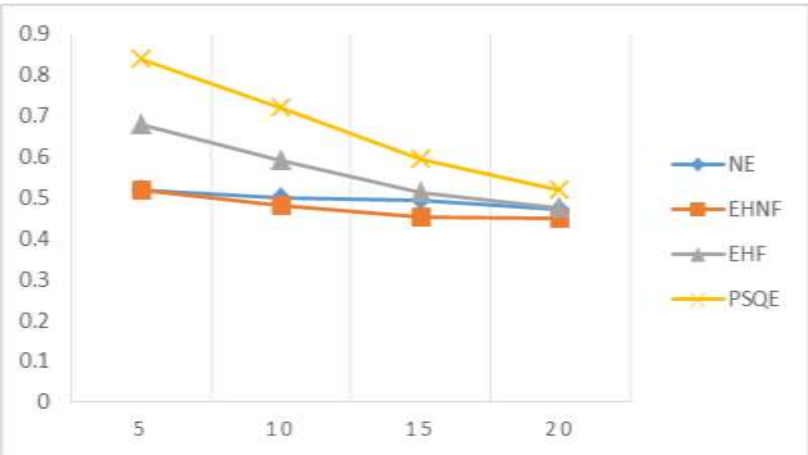

Figure 4. MAP of Methods with Preceding Input

PSQE outperforms other methods in both situation, especially when preceding input does not exist. When preceding input does not exist, the expanded term weight difference between PSQE and EHF is totally decided by user query history. Once individual's query history was inclined to certain paraphrases, the expanded terms of corresponding meaning will have greater weight. If preceding input exist, irrelevant concepts can be filtered first during concepts mapping, user query interests can be better captured and return results can be more personalized. Therefore, PSQE and EHF both have better performance in Figure 4 than Figure 3. EHNF and EHF own the same procedure when there is no preceding input, so they have the identical performance. Here we see it is even worse than NE. During the experiment, we noticed that EHNF has unstable performance, when the term has less paraphrase, it returned with better result, on the contrary, it performed worse. Actually we also noticed that the higher DOR documents ranked higher in PSQE than EHF in return results. The document rank distinction between EHF and PSQE implies the influence of individual query history to query result, which also demonstrates that PSQE performs better than EHF.

In conclusion, the MAP result comparison between EHNF and EHF manifests that user query profile is important for improving retrieval effectiveness. The comparison between PSQE and EHF indicates that spreading activation model is more rational than hierarchical model for expansion. What's more, the performance of PSQE shows that personalized semantic query expansion based on dynamic user query profile and spreading activation model can reduce noise of semantic query expansion and improve the search effectiveness.

5. Conclusion

In this paper a new framework combining personalization with semantic query expansion is proposed to enhance the retrieval effectiveness. In the framework, additional terms semantically related both to the query itself and to user's interests are automatically extracted based on WordNet and dynamic user query profile. While different from existing researches, in order to overcome the noise problem brought by traditional semantic query expansion, spreading activation model which simulates the thinking process of human brain more vividly is utilized as the semantic expansion strategy to enhance the selection of expansion terms. And dynamic user query profile based on user query history and timeline is built to capture individual variable query interest to reduce the noise terms unfit to users' needs. As a result, only the terms with higher comprehensive score of semantic relation and personalized satisfaction are included in the expansion output. The experiment results show that the proposed method outperforms both traditional hierarchical expansion and keyword-based query on both semantic relevance and search precision. In conclusion, the proposed framework suggest that combining dynamic user query profile with the semantic query expansion utilizing spreading activation model can help to reduce noise of semantic query expansion and improve the search effectiveness. In the future, domain ontology will be used to enhance the proposed framework to resolve the noise problem better, and furthermore, more experimental data of dynamic user query profile will be studied in this framework.

Acknowledgments

This paper is supported by the financial support of the Fundamental Research Funds for the Central Universities (NO. TD2014 -02) and the Fundamental Research Funds for the Central Universities (NO. YX2014-19).

References

[1] B. Elayeb, I. Bounhas, O. B. Khiroun, and N. B. B. Saoud, "Combining Semantic Query Disambiguation and Expansion to Improve Intelligent Information Retrieval," in Agents and Artificial Intelligence: Springer, **(2014)**, pp. 280-295.

[2] C. Carpineto and G. Romano, "A survey of automatic query expansion in information retrieval," ACM Computing Surveys (CSUR), vol. 44, **(2012)**, p. 1.

[3] M. G. Buey, Á. L. Garrido and S. Ilarri, "An approach for automatic query expansion based on NLP and semantics," in Database and Expert Systems Applications, **(2014)**, pp. 349-356.

[4] J. Gonzalo, H. Li, A. Moschitti, and J. Xu, "SIGIR 2014 workshop on semantic matching in information retrieval," in International ACM SIGIR Conference on Research & Development in Information Retrieval, **(2014)**, p. 1296-1296.

[5] M. Shabanzadeh, M. A. Nematbakhsh and N. Nematbakhsh, "A Semantic based query expansion to search," in Intelligent Control and Information Processing (ICICIP), 2010 International Conference on, **(2010)**, pp. 523-528.

[6] D. Pal, M. Mitra and K. Datta, "Improving query expansion using WordNet," Journal of the Association for Information Science and Technology, vol. 65, **(2014)**, pp. 2469-2478.

[7] G. J. Hahm, M. Y. Yi, J. H. Lee, and H. W. Suh, "A personalized query expansion approach for engineering document retrieval," Advanced Engineering Informatics, vol. 28**, (2014)**, pp. 344-359,.

[8] O. C. Cure, H. Maurer, N. H. Shah, and P. Le Pendu, "A formal concept analysis and semantic query expansion cooperation to refine health outcomes of interest," BMC Med Inform Decis Mak, vol. 15 Suppl 1, p. S8, **(2015).**

[9] L. Khan, D. McLeod and E. Hovy, "Retrieval effectiveness of an ontology-based model for information selection," The VLDB Journal—The International Journal on Very Large Data Bases, vol. 13, **(2004)**, pp. 71-85.

[10] G. Zou, B. Zhang, Y. Gan, and J. Zhang, "An ontology-based methodology for semantic expansion search," in Fuzzy Systems and Knowledge Discovery, 2008. FSKD'08. Fifth International Conference on, **(2008)**, pp. 453-457.

[11] M. C. Lee, K. H. Tsai and T. I. Wang, "A practical ontology query expansion algorithm for semantic-aware learning objects retrieval," Computers & Education, vol. 50, **(2008)**, pp. 1240-1257.

[12] N. A. Segura, Salvador-Sánchez, E. García-Barriocanal, and M. Prieto, "An empirical analysis of ontology-based query expansion for learning resource searches using MERLOT and the Gene ontology," Knowledge-Based Systems, vol. 24, **(2011)**, pp. 119-133.

[13] Ribeiro-Neto and Berthier, Modern information retrieval, Addison Wesley, **(2011)**.

[14] C. D. Manning, P. Raghavan, Sch, and H. Tze, Introduction to Information Retrieval, Cambridge University Press, **(2008)**.

[15] H. Otukesh, S. Otukesh, M. Mojtahedzadeh, R. Hoseini, S. M. Fereshtehnejad, A. R. Fard, N. Sadigh, A. H. Behzadi, R. Javadi, and N. Hooman, "A Survey of Web Clustering Engines," Acm Computing Surveys, vol. 41, **(2009)**, pp. 75-79.

[16] P. A. Chirita, C. S. Firan and W. Nejdl, "Personalized query expansion for the web," Proceedings of Sigir', **(2007)**,pp. 7-14.

[17] K. Sugiyama, K. Hatano and M. Yoshikawa, "Adaptive web search based on user profile constructed without any effort from users," in Proceedings of the 13th international conference on World Wide Web, WWW 2004, New York, NY, USA, May 17-20, **(2004)**, pp. 675-684.

[18] P. A. Chirita, C. S. Firan and W. Nejdl, "Summarizing local context to personalize global web search," in Proceedings of the 2006 ACM CIKM International Conference on Information and Knowledge Management, Arlington, Virginia, USA, November 6-11, **(2006)**, pp. 287-296.

[19] N. Ahmadian, M. A. Nematbakhsh and H. Vahdat-Nejad, "A context aware approach to semantic query expansion," in Innovations in Information Technology (IIT), 2011 International Conference on, **(2011)**, pp. 57-60.

[20] J. Bhogal, A. Macfarlane and P. Smith, "A review of ontology based query expansion," Information processing & management, vol. 43, **(2007)**, pp. 866-886.

[21] S. Shekarpour, K. Hoffner, J. Lehmann, and S. Auer, "Keyword Query Expansion on Linked Data Using Linguistic and Semantic Features," in 2013 IEEE Seventh International Conference on Semantic Computing, **(2013)**, pp. 191-197.

[22] L. Li and H. Wang, "Multi-Strategy Query Expansion Method Based on Semantics," Journal of Digital Information Management, vol. 12, **(2014)**.

[23] F. Cai, S. Liang and M. De Rijke, "Personalized document re-ranking based on Bayesian probabilistic matrix factorization," in International ACM SIGIR Conference on Research & Development in Information Retrieval, **(2014)**, pp. 835-838.

[24] R. W. White, W. Chu, A. Hassan, X. He, Y. Song, and H. Wang, "Enhancing personalized search by mining and modeling task behavior," in Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro, Brazil, **(2013)**, pp. 1411-1420.

[25] T. T. Vu, D. Song, A. Willis, S. N. Tran, and J. Li, "Improving search personalisation with dynamic group formation," in The International ACM SIGIR Conference, **(2014)**, pp. 951-954.

[26] P. N. Bennett, "Search from Personal to Social Context: Progress and Challenges," p.2, **(2015)**.
[27] X. Tian, X. Du, H. Hu, and H. Li, "Modeling individual cognitive structure in contextual information retrieval," Computers & Mathematics with Applications, vol. 57, **(2009)**, pp. 1048-1056.
[28] K. Kinley, D. W. Tjondronegoro, H. L. Partridge, and S. L. Edwards, "Human-Computer Interaction: The Impact of users' Congitive Styles on Query Reformulation Behaviour during Web searching," in OZCHI 2012 The Annual Conference of the Australian Computer-Human Interaction Special Interest Group, **(2012),** pp. 299-307.
[29] E. H, "Memory: A Contribution to Experimental Psychology.," Annals of Neurosciences, vol. 20, **(1913)**, pp. 155-156.
[30] F. Crestani, "Application of Spreading Activation Techniques in Information Retrieval," Artificial Intelligence Review, vol. 11, **(1997)**, pp. 453-482.

Authors

Shengrong Zhu, she is from Hubei Province of China, and a master at Information Science and Technology Dept in Beijing Forestry University, interested in intelligent information processing and 3D visualization.

Xuan Tian, she is from Shandong Province of China, and an associated professor at Information Science and Technology Dept in Beijing Forestry University, interested in intelligent information processing and intelligent information retrieval. E-mail: tianxuan@bjfu.edu.cn.