



## Open Research Online

### Citation

Pan, Dazhao; Zhang, Peng; Li, Jingfei; Song, Dawei; Wen, Jirong; Hou, Yuexian; Hu, Bin; Jia, Yuan and De Roeck, Anne (2013). Using Dempster-Shafer's evidence theory for query expansion based on freebase knowledge. In: Ninth Asia Information Retrieval Societies Conference, 9-11 Dec 2013, Singapore.

### URL

<https://oro.open.ac.uk/38257/>

### License

None Specified

### Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from [Open Research Online \(ORO\) Policies](#)

### Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding

# Using Dempster-Shafer’s Evidence Theory for Query Expansion Based on Freebase Knowledge

Dazhao Pan<sup>1</sup>, Peng Zhang<sup>1</sup>, Jingfei Li<sup>1</sup>, Dawei Song<sup>1,2</sup>, Ji-Rong Wen<sup>3</sup>, Yuexian Hou<sup>1</sup>, Bin Hu<sup>4</sup>, Yuan Jia<sup>5</sup>, and Anne De Roeck<sup>2</sup>

<sup>1</sup> Tianjin University, Tianjin, China

<sup>2</sup> The Open University, Milton Keynes, UK

<sup>3</sup> Microsoft Research Asia

<sup>4</sup> Lanzhou University, Lanzhou, China

<sup>5</sup> Institute of Linguistics, Chinese Academy of Social Sciences, China

{dazhaopan, darcyzzj, dawei.song2010, krete1941}@gmail.com

tjucsljf@foxmail.com, jrwen@microsoft.com

bh@lzu.edu.cn, summeryuan\_2003@126.com, anne.deroeck@open.ac.uk

**Abstract.** Query expansion is generally a useful technique in improving search performance. However, some expanded query terms obtained by traditional statistical methods (e.g., pseudo-relevance feedback) may not be relevant to the user’s information need, while some relevant terms may not be contained in the feedback documents at all. Recent studies utilize external resources to detect terms that are related to the query, and then adopt these terms in query expansion. In this paper, we present a study in the use of Freebase [6], which is an open source general-purpose ontology, as a source for deriving expansion terms. FreeBase provides a graph-based model of human knowledge, from which a rich and multi-step structure of instances related to the query concept can be extracted, as a complement to the traditional statistical approaches to query expansion. We propose a novel method, based on the well-principled Dempster-Shafer’s (D-S) evidence theory, to measure the certainty of expansion terms from the Freebase structure. The expanded query model is then combined with a state of the art statistical query expansion model – the Relevance Model (RM3). Experiments show that the proposed method achieves significant improvements over RM3.

**Keywords:** Query Expansion, Freebase, Dempster-Shafer theory

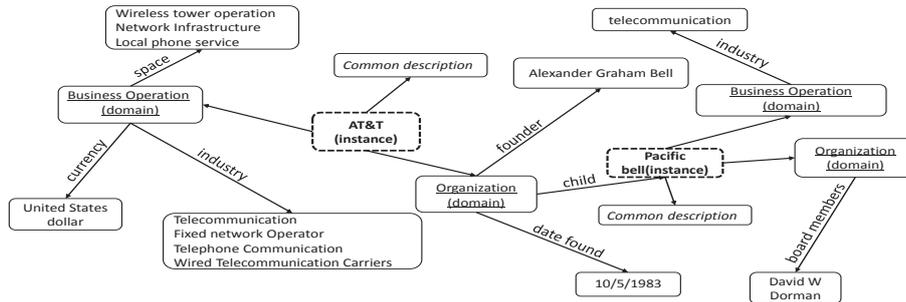
## 1 Introduction

The information needs of the searchers are often formulated as keyword queries. However, searchers tend to use short and incomplete queries. Consequently the original query input by the user may miss some important information, resulting in a negative impact on the retrieval effectiveness. To address this problem, query expansion (QE) has been widely used [5] [7] [15].

Pseudo-relevance feedback (PRF) is a typical query expansion method [8] [7] [19] [33]. It selects expansion terms from top-ranked documents in the first-round retrieval as feedback documents, by assuming that feedback documents are relevant and contain the key concepts related to the query. However, in

practice the top ranked documents may not all be truly relevant, so that many expansion terms derived may be irrelevant. The uncertainty of the relevance of expansion terms may cause the query drifting problem [8].

To tackle this problem, recent research attempts to make use of external knowledge, such as click data, user log, Wikipedia or simple ontology data WordNet. The analysis of the user search history (e.g., click data) may avoid the query drifting problem, but the coverage of such knowledge is limited[10]. The emergence of the Web and large-scale human edited knowledge (e.g., Wikipedia and Freebase) provides access to new sources of high quality term associations for query expansion [34].



**Fig. 1.** Fragment of the entity-relationship graph of Freebase

In this paper, we focus on studying the use of an ontology-structured knowledge source - Freebase, as the source of obtaining expansion terms. Freebase employs the entity-relation model to describe the data in the knowledge base. It has over 37 million topics about real-world entities like people, places and events. We choose Freebase for three reasons. First, it contains a large amount of knowledge, covering different aspects of the world, such as location, people, event, military, music, film, literature and business. As an illustration, we searched the TREC queries in the Freebase, and found that 90% of the queries (we used in our experiments) have matches of some relevant knowledge. Second, unlike Wikipedia that describes human knowledge with a long detailed article, Freebase describes the human knowledge using an ontological structure, and describes each property with keywords. Therefore, the relevant concepts underneath a query can be quickly located. Third, different from the WordNet which mainly contains synonymy relations, Freebase not only describes entities, but also contains different properties and entity-relationships. Thus one can walk through the entity-relation graph to find deeper relationships.

A fragment of the entity-relation graph in Freebase is shown in Figure 1. We can see that all the information about an instance (e.g., AT&T) is contained in its properties (e.g., space, industry, currency). We can also find other instances that are related to the current one. For example, starting from AT&T and walking through the Organization Domain, we can find AT&T’s successor node - “Pacific bell”, which is also an instance. For a query, we can find the matched

instances as well as their parent and child instances up to a certain step, and treat them as different aspects that describe the query. Then it is key to fuse the relevant information from different aspects. In this paper, we propose to solve the fusion problem by using the Dempster-Shafer’s (D-S) evidence theory [23]. D-S theory allows one to combine evidence from different sources and arrive at a degree of belief that takes into account all the available evidences. Treating different instances as different evidences for determining query expansion terms from Freebase, we redefine the Basic Probability Assignment (BPA) and Belief functions in the D-S theory, and define a new measure, called *Certainty*, to represent a term’s importance in the knowledge structure. Under the framework of D-S theory, we also define the transitivity of the nodes in the knowledge structure. Finally, we integrate the above parameters in a single model (see Section 4.3).

The knowledge terms are then combined with the terms given by the relevance model (RM3) [17] [18]. In the combination process, we balance the knowledge term’s Certainty values by D-S theory and the original weights by RM3, in order to compute the final term weights in the query expansion model. Our experiments show that the selected knowledge terms are complementary to the initial query (Table 2). Combining the knowledge terms and the RM3 terms can significantly improve the retrieval effectiveness.

## 2 Related Work

Pseudo relevance feedback (PRF) is an effective query expansion method by reformulating the original query using terms derived from the pseudo-relevance documents [8] [7] [19] [33]. Based on the assumption that top-ranked documents are relevant to the query, several PRF algorithms are proposed, e.g., Okapi [22], relevance model [15] and mixture model [35].

Despite the large number of studies in this area, a crucial question is that the expansion terms extracted from the pseudo-relevant documents are not all useful [8]. It was found that one of reasons for the failure of query expansion is the lack of relevant documents in local relevance feedback collection. Recently, some researches attempt to capture and utilize lexico-semantic relationships, based on the association hypothesis formulated by van Rijsbergen [25]. Early global expansion techniques [11] [12] [33] aim to determine the strength of semantic relatedness between terms based on the term co-occurrence statistics. The emergency of hand-crafted general purpose or domain-specific ontologies also provide access to high quality term associations for query expansion. Several approaches have been proposed to utilize external resources, such as query logs [10], WordNet [9], Wikipedia [34], ConceptNet [13], etc. Yin et al. proposed an expansion method based on random walking through URL graph [12]. Voorhees [32] experimentally determined an upper bound of the WordNet based on query expansion.

Furthermore, there has been an emerging line of research on ontology-based query expansion. Vallet et al. [3] used an ontology query language to search for relevant documents. Nagypal et al. [20] combined the use of ontology with the vector space model. Braga et al. [2] used ontologies for information extraction

and filtering across multiple domains. Ozcan et al. [4] adopted ontologies to represent concepts which are expanded by using WordNet. Meij et al. [16] showed that discriminative semantic annotations of documents using domain-specific ontologies, such as MeSH, can effectively improve retrieval effectiveness.

Several methods which combine multiple resources have also been proposed. Mandala et al. [31] proposed a method to combine three different thesaurus types for query expansion. Bodner et al. [27] combined WordNet and co-occurrence based thesauri for query expansion. In the work of Collins-Thompson and Callan [30], a Markov chain based framework were proposed for query expansion by combining multiple sources of knowledge on term associations. In Cao et al. [28], term relationships from co-occurrence statistics and WordNet were used to smooth the document language model.

Dempster-Shafer (D-S) theory has been applied to IR in several previous studies. In [21], D-S theory has been used to combine content and link information in Web IR, and in [24], it is used to combine textual and visual evidences for image retrieval. Theophylactou and Lalmas [14] also used D-S theory to model a compound term as a set of single terms. In our work, we employ the D-S theory to measure the scores of the terms in Freebase that are matched against a query, and to determine the weights of the expanded terms based on such scores.

### 3 Query Matching in Freebase

Given a query, in order to find matched instances in Freebase, we first split a query into sequences of words. For example, the query “commercial satellite launches” can be split into three phrases: “commercial satellite launches”, “commercial satellite” and “satellite launches”. We search these three phrases in Freebase and get relevant instances. Note that we do not split the query to single terms, because we think that single terms will more likely match non-relevant instances, due to lack of context. For example, the query “Black Monday” describes the largest one-day percentage decline in recorded stock market history. If we split it to “Black” and “Monday”, the two separate terms, individually, are irrelevant to the query. In our experiment, the minimum splitting unit is two sequential terms.

After getting the phrases from the query, we then adopt a two-round matching process to find candidate instances. In the first round, we will find whether a phrase extracted from query is recorded in the Freebase using exact match. If the phrase is recorded, we will get the corresponding instance and extract the notable domain from the instance. in the second round, we employ the Freebase’s API to search for (partially) matched instances within the notable domain. For example, for the phrase “Black Monday”, in the first round matching, we get the exactly matched instance “black Monday” and extract its notable domain “event”. In the second round, we only search the instances belonging to the “event” domain. Through the two-round matching, we get the matched instances. The Match Score provided by the Freebase’s API, is used to measure the relevance degree between a query phrase and an instance.

As shown in Fig.1, one instance’s property may be another instance. Therefore, in addition to the two round matching strategy described above, we go

deeper in the Freebase graph by walking through the path as follows:

$$instance_j \dashrightarrow property_{jk} \dashrightarrow instance_k \dashrightarrow property_{kw}$$

where  $property_{jk}$  is the  $k$ -th property of  $instance_j$ . We use the  $property_{jk}$  to find the child  $instance_k$  and finally we can get the property of  $instance_k$ . In our paper, we call the instances directly from the two round matching process as the **first level instances** and their successor instances as **second level instances**.

## 4 RANKING KNOWLEDGE TERMS BASED ON DEMPSTER-SHAFER THEORY

Table 1. Notations

Parameters	Definition
Q	Query
$q_i$	term in Query
$I_i$	first level Instance
$FS(I_i)$	instance $I_i$ 's Freebase score provided by Freebase
$Ip_i$	second level Instance
$tf_k(t_i)$	term $t_i$ 's term frequency in instance $I_k$
$m_k(t_i)$	basic probability assignment (BPA) in instance $I_k$
$Bel_k(t_i)$	belief of term in instance $I_k$
$Bel_k(Ip_i)$	belief of second level instance in instance $I_k$
$Pl_k(t_i)$	plausibility of term in instance $I_k$
$Pl_k(Ip_j)$	plausibility of second level instance in instance $I_k$
$Certainty(I_k)$	initial $I_k$ 's relevance degree with query
$Certainty(Ip_i)$	the measure of certainty of the second level instance
$Certainty(t_i)$	the measure of certainty of the single term

### 4.1 An Introduction to Dempster-Shafer Theory

The Dempster-Shafer (D-S) theory is a mathematical theory of evidence [23]. It allows one to combine evidence from different sources and arrive at a degree of belief that takes into account all the available evidence. It is developed to account for the uncertainty. Specifically, let  $\theta$  represent all possible states (or elements) under consideration. Function  $m: 2^\theta \rightarrow [0, 1]$  is called a basic probability assignment (BPA), which assigns a probability mass to each element. It has two properties:

First, the empty set is assigned the value 0:

$$m(\emptyset) = 0$$

Second, the sum of the probabilities assigned to all elements of  $\theta$  is 1:

$$\sum_{X \in 2^\theta} m(X) = 1$$

Based on the mass assignments, the upper and lower bounds of a probability interval can be defined. This interval contains the precise probabilities of a set of elements (denoted as  $A$ ), and is bounded by two non-additive continuous measures called belief (or support) and plausibility:

$$Bel(A) \leq P(A) \leq Pl(A)$$

The belief  $Bel(A)$  for a set  $A$  is defined as the sum of all the masses of subsets of  $A$ :

$$Bel(A) = \sum_{B:B \subseteq A} m(B)$$

That is,  $Bel(A)$  gathers all the evidence directly in support of  $A$ . The plausibility  $Pl(A)$  is the sum of all the masses of the sets  $B$  that intersect  $A$ :

$$Pl(A) = \sum_{B:B \cap A \neq \emptyset} m(B) = 1 - Bel(\bar{A})$$

It defines the upper bounds of  $A$ .

Dempster-Shafer theory corresponds to the traditional probability theory when  $m$  assigns non-zero probabilities only to individual elements. The Dempster's rule of combination provides a method to fuse the evidences for an element from multiple independent sources and calculate an overall belief for the element. In our paper, we use the Dempster's rule to measure the Certainty of a term from different matched instances, detailed in the next subsection.

#### 4.2 Using D-S Theory to Rank the Candidate Terms

**Table 2.** Top ranked terms extracted from Freebase combined (by D-S theory) with expanded terms from Feedback documents (by RM3) on AP collection.

Query	Top Knowledge Terms(AP)
Black Monday	finance stock market crash industrial
Diversify of Pacific Telesis	pactel bell commun telecom wireless
Conflict in Horn of Africa	east somalia eritrea ethiopia djibouti
Nuclear Prolifer	weapon india treati north korea
McDonnell Douglas Contracts Military Aircraft	germani navi radar bomber jet

We treat different instances as different sources of evidence. Our objective is to measure the certainty of a term that is contained in multiple instances. We assume the hypothesis space is composed of all the candidate terms extracted from Freebase, e.g.,  $instance \rightarrow \{a, b, c, d, e, \{d, e\}\}$ , where  $d, e$  are second level instances. We use  $P$  to represent the hypothesis space.

We measure the basic probability assignment ( $BPA$ ) of a single term  $t_i$  in  $P$  as follows:

$$m_k(t_i) \propto tf_k(t_i) Certainty(I_k) \quad (1)$$

where  $tf_k(t_i)$  represents  $t_i$ 's term frequency in the Instance  $I_k$ , and we define the initial  $Certainty(I_k)$  as the Match score of the Instance  $I_k$  with respect to

the query. We adopt different strategies to measure the Certainty for instances got from the two round matching process. Intuitively, if an instance contains more terms from query, the instance is more relevant to the query. Therefore we use the length of the first round matched instance to measure  $I_1$ 's Certainty.

$$Certainty(I_1) = \frac{length(I_1)}{length(Q)} \quad (2)$$

For the instances got from the second round retrieval, we initialize their Certainty according to the comparison with  $I_1$ :

$$Certainty(I_i) = \frac{FS(I_i)}{FS(I_1)} Certainty(I_1) \quad (3)$$

where  $FS(I_i)$  represents the *Match Score* provided by Freebase and  $i \neq 1$ . We use the  $FS(I_1)$  as a standard value, and use  $\frac{FS(I_i)}{FS(I_1)}$  to represent the ratio between  $FS(I_i)$  and  $FS(I_1)$ , where  $FS(I_i) < FS(I_1)$ .

For the single term, we calculate the  $Bel(t_i)$  as follows:

$$Bel_k(t_i) \propto tf_k(t_i) Certainty(I_k) \quad (4)$$

For the second level instance  $Ip_i$ , we calculate the  $Bel(Ip_i)$  as follows:

$$Bel_k(Ip_i) \propto \sum_{t_i: t_i \in Ip_i} tf_k(t_i) Certainty(I_i) \quad (5)$$

According to the Dempster-Shafer evidence theory, the plausibility of the second level instance calculated as follows:

$$Pl_k(Ip_i) \propto Certainty(I_k) - Bel_k(\overline{Ip_i}) \quad (6)$$

As mentioned in Section 4.2, the plausibility of the single term is the same as  $Bel$ . So,  $Pl(t_i) = Bel(t_i)$ .

**Fusion** Recall that, we treat different instances as different sources of evidence, and Dempster's rule of combination that aggregates two or more bodies of evidence (defined within the same frame of discernment) into one body of evidence. Let  $m_j$  and  $m_k$  be two different *BPA* calculated from  $I_j$  and  $I_k$ .  $A$  as the common terms. we calculate the combined *BPA* as follows :

$$\begin{aligned} m(A) &\propto (m_j \oplus m_k)(A) \\ &= \frac{\sum_{B \cap C = A \neq \emptyset} (tf_j(B) Certainty(I_j)) \times (tf_k(C) Certainty(I_k))}{(1-K)} \end{aligned} \quad (7)$$

$$K \propto \sum_{B \cap C = \emptyset} (tf_j(B) Certainty(I_j)) \times (tf_k(C) Certainty(I_k)) \quad (8)$$

Where  $B$  and  $C$  are two hypothesis space from two different Instances. According to the Dempster's rule of combination, we fuse the term's *BPA* from different instances. We calculate the  $Bel$  using the uniformed *BPA*.

**Transitivity** In order to measure the second level instance’s properties, we need to measure the *Certainty* of the second level instance, and according to the *Bel* and *Pl*, we measure the second level instance as follows:

$$Certainty(Ip_i) \propto Bel_{Ip_i} + \frac{|Ip_i|}{|P|} (Pl_k(Ip_i) - Bel_k(Ip_i)) \quad (9)$$

where  $|Ip_i|$  indicates the length of the instance property, and  $|P|$  indicates the number of the elements in  $P$ . Therefore we get the *Certainty* of the second level instance which is passed through its precursor’s instance. After that, we can use such method to calculate *Certainty* of second level instance’s property. We use *Certainty* to represent term’s final score in the Freebase ( Considering  $Bel = Pl$  for single term. So,  $Certainty(t_i) = Bel(t_i)$  ).

### 4.3 Extended Relevance Model

In the framework of Relevance Model(RM), we estimate the probability distribution  $P(w|R)$ , where  $w$  is an arbitrary word and  $R$  is the unknown underlying relevance model, which is usually extracted by RM based on the top-ranked documents of the initial retrieval.

Although the candidate expansion words extracted from knowledge are relevant to the query in Freebase, we do not know whether such knowledge words are suitable to the local corpus. In order to measure the relevance of the knowledge words in the local corpus, we measure two features of knowledge candidate words. First, we will count the co-occurrence of the candidate expanded knowledge term and query term in a Window ( we segment the documents to small passages ). It is formulated as follows:

$$co - occurrence_{(q_i, t_i)} = \#win_n(t_i, q_i) \quad (10)$$

We filter the knowledge terms using the co-occurrence feature, by removing the terms that don’t co-occur with the query terms. Second, based on the assumption, if a knowledge candidate term is relevant to the query, it will occur in the top-ranked feedback documents. We calculate two term distributions: 1) the term frequency in the top-ranked feedback documents, 2) the term frequency in the whole corpus. Using the method mentioned in the work of Parapar and Barreiro [29]. We assume that the term with a big divergence is important. Combing with the *Certainty* from knowledge, we measure the term score as follows:

$$Socre(t_i) = Certainty(t_i) \log\left(\frac{tf(t_i|D)}{tf(t_i|C)}\right) IDF(t_i|D) \quad (11)$$

$$IDF(t_i) = \log\left(\frac{|C|}{|j : t_i \in d_j|}\right) \quad (12)$$

where  $D$  is the top-ranked documents,  $C$  is the whole corpus and  $Certainty(t_i)$  is the measure of term’s *Certainty* in the Freebase. We use the  $\log$  operation to smooth the distribution gap.  $IDF$  is used to represent the importance of

term in the corpus.  $|C|$  represents the total amount of documents in the corpus.  $|j : t_i \in d_j|$  represents the amount of documents that contain the term  $t_i$ . We think that the Freebase term with a higher  $IDF$  and a bigger divergence of term distribution is more relevant to the query. We use the score computed by Eq. 11 to measure the importance of the terms. Table 2 shows the top ranked knowledge terms.

We will combine our knowledge terms with the Relevance Model (RM3). Specifically, we formulate the new query  $Q_{new}$  as follows:

$$Q_{new} \propto \lambda Q_{ori} + (1 - \lambda)(Q_{exp} + Q_{kb}) \quad (13)$$

where  $Q_{ori}$  is the original query,  $Q_{exp}$  is the expanded query by RM, and  $Q_{kb}$  represents the terms extracted from Freebase. We do not involve extra parameters in Eq. 13, where  $\lambda$  is the parameter of RM3 (i.e.,  $\lambda Q_{ori} + (1 - \lambda)Q_{exp}$ ).

## 5 Experiment

### 5.1 Experimental Settings

We use three standard TREC collections: AP8889 with topics.101-150; WSJ with topics.101-150; ROBUST04 with topics.301-450. Table 3 shows the detail of the collections. We use mean average precision(MAP) to evaluate the retrieval effectiveness. Lemur4.12 is used for indexing and retrieval. Documents and queries are stemmed with the Porter stemmer and the stopwords are removed through the Stopwords list.

**Table 3.** Overview of TREC collections and topics

Corpus	Size	# of Doc	Topics
AP8889	0.6G	164,597	101-150
WSJ	0.5G	173,252	101-150
ROBUST04	1.8G	528,155	301-450

### 5.2 Baseline Models

In our experiments, we select two baselines. One is the query-likelihood language model (QL), and the other is the relevance Model (RM3). RM3 is a strong baseline which is widely used, and we choose the parameters of RM3 to get its optimal result on each collection. RM3 + FB1 and RM3 + FB are the results that we didn't apply D-S. RM3 + FB1 is the result that we just use the first level of the Freebase, RM3 + FB is the result that we use both level of the Freebase.

### 5.3 Parameter Settings

In our experiment, we adopt the Dirichlet smoothing method and set the Dirichlet prior as the default 1000. RM3 + KB is our model that combines the RM3 terms and Freebase terms using D-S theory. Both RM3 and RM3 + KB methods have parameters  $N$ ,  $k$ ,  $\lambda$ . For number of feedback documents, we tested  $N$  from 10, 20, 30, 40, 50. We evaluated the  $\lambda$  for different values: 0.1, 0.2, 0.3, 0.4, 0.5. And for the feedback terms number  $k$ , we tested :10, 20, 30, 40, 50. When  $N = 10$ ,  $k = 50$  and  $\lambda = 0.1$ , RM3 gets the optimal result. In Eq. 13, we combine

**Table 4.** Performance comparisons on all the topics.

Method	Topics.101-150(AP8889)	Topics.101-150(WSJ)	Topics.301-450(ROBUST04)
QL	0.2423	0.2629	0.2480
RM3	0.3077	0.2941	0.2550
RM3 + FB1	0.3105(0.90%)	0.2970(0.986%)	0.2554(0.157%)
RM3 + FB	0.3142(2.10%)	0.2921(-0.680%)	0.2558(0.317%)
RM3 + KB	*0.3197(3.90%)	*0.3008(2.28%)	0.2598(1.18%)

\* indicates statistically significant improvements (at level 0.05) over RM3.

knowledge candidate terms and RM3 terms, so we denote the new query model in Eq. 13 as RM3+KB. RM3+KB adopts the same parameter (i.e.,  $\lambda$ ) as RM3 model.

#### 5.4 Experiment Results

**Table 5.** Performance comparisons on the topics for which related knowledge instances can be found in Freebase.

Method	Topics.101-150(AP8889)	Topics.101-150(WSJ)	Topics.301-450(ROBUST04)
QL	0.2446	0.2532	0.2351
RM3	0.3045	0.2858	0.2406
RM3 + KB	*0.3207(5.32%)	*0.2945(3.04%)	0.2459(2.20%)

\* indicates statistically significant improvements (at level 0.05) over RM3.

As can be seen from the Table 4, RM3 largely outperforms QL, which demonstrates RM3 is an effective query expansion method on all collections. For RM3+KB, On AP8889 and WSJ, it has a significant performance improvement over the RM3 model. For ROBUST04, RM3+KB also improves the performance over RM3. RM3+KB also outperforms RM3 + FB1 and RM3 + FB in which the D-S theory is not involved. We can see that RM3+KB performs the best on all collections.

As we have mentioned, not all of queries have the relevant knowledge in Freebase, e.g., for Topics.101-150, about 90% queries can find relevant instance in Freebase. We then report the results of the topics for which the related instances can be found in Freebase. From the result showed by Table 5, we can see that on AP8889 collection, the performance is improved by RM3 + KB about 5% over RM3, and on WSJ collection, performance is improved about 3% over RM3. For ROBUST04, the performance is also improved about 2% over RM3. These results indicate that the terms extracted from the Freebase are relevant to the query and RM3 with knowledge terms works better than using RM3 terms only.

## 6 Conclusions

In this paper, we have explored the use of Freebase as a resource for query expansion. In our work, we adopt a two round query match in Freebase to

find knowledge instances related to query. We propose to apply the Dempster-Shafer evidence theory to rank relevant terms for the query. We then combine the extracted terms (from Freebase) with the expanded terms (from feedback documents) by RM3. The experiment result shows that terms extracted from Freebase have a significant improvement performance over RM3, which is a state-of-the-art query expansion method.

## 7 Acknowledgments

This work is funded in part by the Chinese National Program on Key Basic Research Project (973 Program, grant no. 2013CB329304 and 2014CB744604), the Natural Science Foundation of China (grant no. 61272265), and the European Union Framework 7 Marie-Curie International Research Staff Exchange Programme (grant no. 247590).

## References

1. Ciorascu, C., Ciorascu, I., Stoffel, K.: knowler-ontological support for information retrieval systems. In: *SIGIR 2003*, volume 2, (2003).
2. Braga, Regina MM, Cludia Maria Lima Werner, and Marta Mattoso.: Using ontologies for domain information retrieval. In: *Database and Expert Systems Applications, 2000*, pp. 836-840 (2000).
3. Vallet, D., Fernandez, M., Castells, P.: An ontology-based information retrieval model. In: *The Semantic Web: Research and Applications*, pp. 455-470 (2005).
4. Ozcan, R., Aslangodan, Y. A.: Concept based information access using ontologies and latent semantic analysis. In: *Dept. of Computer Science and Engineering*, 8:2004, (2004).
5. Bendersky, M., Croft, W. B.: Discovering key concepts in verbose queries. In: *SIGIR 2008*, pp. 491-498 (2008).
6. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: *SIGMOD 2008*, pp. 1247-1250 (2008).
7. Buckley, C., Salton, G., Allan, J., Singhal, A.: Automatic query expansion using smart: Trec 3. In: *NIST SPECIAL PUBLICATION SP*, pp. 69-69 (1995).
8. Cao, G., Nie, J. Y., Gao, J., Robertson, S.: Selecting good expansion terms for pseudo-relevance feedback. In: *SIGIR 2008*, pp. 243-250 (2008).
9. Collins-Thompson, K., Callan, J.: Query expansion using random walk models. In: *ACM international conference on Information and knowledge management*, pp. 704-711 (2005).
10. Cui, H., Wen, J. R., Nie, J. Y., Ma, W. Y.: Query expansion by mining user logs. In: *Knowledge and Data Engineering, IEEE Transactions on*, pp. 829-839 (2003).
11. Gauch, S., Wang, J.: A corpus analysis approach for automatic query expansion. In: *Proceedings of the sixth international conference on Information and knowledge management*, pp. 278-284 (1997).
12. Jing, Y., Croft, W. B.: An association thesaurus for information retrieval. In: *Proceedings of RIAO 1994*, volume 94, pp. 146-160 (1994).
13. Kotov, A., Zhai, C.: Tapping into knowledge base for concept feedback: leveraging conceptnet to improve search results for difficult queries. In: *ACM international conference on Web search and data mining*, pp. 403-412 (2012).

14. Lalmas, M., Ruthven, I.: Representing and retrieving structured documents using the dempster-shafer theory of evidence: Modelling and evaluation. In: *Journal of Documentation*, pp. 529-565, (1998).
15. Lavrenko, V., Croft, W. B.: Relevance based language models. In: *SIGIR 2001*, pp. 120-127 (2001).
16. Meij, E., Trieschnigg, D., De Rijke, M., Kraaij, W.: Conceptual language models for domain-specific retrieval. In: *Information Processing & Management*, pp. 448-469 (2010).
17. Lavrenko, V., Croft, W. B.: Relevance based language models. In: *ACM SIGIR* , pp. 120-127 (2001).
18. Lv, Y., Zhai, C.: A comparative study of methods for estimating query language models with pseudo feedback. In: *ACM conference on Information and knowledge management*, pp. 1895-1898 (2009).
19. Metzler, D., Croft, W. B.: Latent concept expansion using markov random fields. In: *SIGIR 2007*, pp. 311-318 (2007).
20. Nagypál, G.: Improving information retrieval effectiveness by using domain knowledge stored in ontologies. In: *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops*, pp. 780-789 (2005).
21. Plachouras, V., Ounis, I.: Dempster-shafer theory for a query-biased combination of evidence on the web. In: *Information Retrieval*, pp. 197-218 (2005).
22. Robertson, S. E., Walker, S., Beaulieu, M. M., Gatford, M., Payne, A.: Okapi at trec-4. In: *text retrieval conference*, pp. 73-97 (1996).
23. Shafer, G.: *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, (1976).
24. Urban, J., Jose, J. M., Van Rijsbergen, C. J.: An adaptive technique for content-based image retrieval. In: *Multimedia Tools and Applications*, pp. 1-28 (2006).
25. Van Rijsbergen, C. J.: A non-classical logic for information retrieval. In: *The computer journal*, pp. 481-485 (1986).
26. Bai, J., Song, D., Bruza, P., Nie, J. Y., Cao, G.: Query expansion using term relationships in language models for information retrieval. In: *ACM international conference on Information and knowledge management* , pp. 688-695 (2005).
27. Bodner, R. C., Song, F.: In: *Knowledge-based approaches to query expansion in information retrieval*. Springer, (1996).
28. Cao, G., Nie, J. Y., Bai, J.: Integrating word relationships into language models. In: *SIGIR 2005* , pp. 298-305 (2005).
29. Parapar, J., Barreiro, Á.: Promoting divergent terms in the estimation of relevance models. In: *Advances in Information Retrieval Theory*, pp. 77-88 (2011).
30. Collins-Thompson, K., Callan, J.: Query expansion using random walk models. In: *ACM international conference on Information and knowledge management*, pp. 704-711 (2005).
31. Mandala, R., Tokunaga, T., Tanaka, H.: Combining multiple evidence from different types of thesaurus for query expansion. In: *SIGIR 1999*, pp. 191-197 (1999).
32. Voorhees, E. M.: Query expansion using lexical-semantic relations. In: *SIGIR 1994*, pp. 61-69 (1994).
33. Xu, J., Croft, W. B. : Query expansion using local and global document analysis. In: *SIGIR 1996*, pp. 4-11 (1996).
34. Xu, Y., Jones, G. J., Wang, B.: Query dependent pseudo-relevance feedback based on wikipedia. In: *SIGIR 2009*, pp. 59-66 (2009).
35. Zhai, C., Lafferty, J.: Model-based feedback in the language modeling approach to information retrieval. In: *the tenth international conference on Information and knowledge management*, pp. 403-410 (2001).