

A Re-examination of Query Expansion Using Lexical Resources

Hui Fang

Department of Computer Science and Engineering
The Ohio State University
Columbus, OH, 43210
hfang@cse.ohio-state.edu

Abstract

Query expansion is an effective technique to improve the performance of information retrieval systems. Although hand-crafted lexical resources, such as WordNet, could provide more reliable related terms, previous studies showed that query expansion using only WordNet leads to very limited performance improvement. One of the main challenges is how to assign appropriate weights to expanded terms. In this paper, we re-examine this problem using recently proposed axiomatic approaches and find that, with appropriate term weighting strategy, we are able to exploit the information from lexical resources to significantly improve the retrieval performance. Our empirical results on six TREC collections show that query expansion using only hand-crafted lexical resources leads to significant performance improvement. The performance can be further improved if the proposed method is combined with query expansion using co-occurrence-based resources.

1 Introduction

Most information retrieval models (Salton et al., 1975; Fuhr, 1992; Ponte and Croft, 1998; Fang and Zhai, 2005) compute relevance scores based on matching of terms in queries and documents. Since various terms can be used to describe a same concept, it is unlikely for a user to use a query term that is exactly the same term as used in relevant documents. Clearly, such vocabulary gaps make the retrieval performance non-optimal. Query expansion (Voorhees, 1994; Mandala et al., 1999a; Fang and

Zhai, 2006; Qiu and Frei, 1993; Bai et al., 2005; Cao et al., 2005) is a commonly used strategy to bridge the vocabulary gaps by expanding original queries with related terms. Expanded terms are often selected from either co-occurrence-based thesauri (Qiu and Frei, 1993; Bai et al., 2005; Jing and Croft, 1994; Peat and Willett, 1991; Smeaton and van Rijsbergen, 1983; Fang and Zhai, 2006) or hand-crafted thesauri (Voorhees, 1994; Liu et al., 2004) or both (Cao et al., 2005; Mandala et al., 1999b).

Intuitively, compared with co-occurrence-based thesauri, hand-crafted thesauri, such as WordNet, could provide more reliable terms for query expansion. However, previous studies failed to show any significant gain in retrieval performance when queries are expanded with terms selected from WordNet (Voorhees, 1994; Stairmand, 1997). Although some researchers have shown that combining terms from both types of resources is effective, the benefit of query expansion using only manually created lexical resources remains unclear. The main challenge is how to assign appropriate weights to the expanded terms.

In this paper, we re-examine the problem of query expansion using lexical resources with the recently proposed axiomatic approaches (Fang and Zhai, 2006). The major advantage of axiomatic approaches in query expansion is to provide guidance on how to weight related terms based on a given term similarity function. In our previous study, a co-occurrence-based term similarity function was proposed and studied. In this paper, we study several term similarity functions that exploit various information from two lexical resources, i.e., WordNet

and dependency-thesaurus constructed by Lin (Lin, 1998), and then incorporate these similarity functions into the axiomatic retrieval framework. We conduct empirical experiments over several TREC standard collections to systematically evaluate the effectiveness of query expansion based on these similarity functions. Experiment results show that all the similarity functions improve the retrieval performance, although the performance improvement varies for different functions. We find that the most effective way to utilize the information from WordNet is to compute the term similarity based on the overlap of synset definitions. Using this similarity function in query expansion can significantly improve the retrieval performance. According to the retrieval performance, the proposed similarity function is significantly better than simple mutual information based similarity function, while it is comparable to the function proposed in (Fang and Zhai, 2006). Furthermore, we show that the retrieval performance can be further improved if the proposed similarity function is combined with the similarity function derived from co-occurrence-based resources.

The main contribution of this paper is to re-examine the problem of query expansion using lexical resources with a new approach. Unlike previous studies, we are able to show that query expansion using only manually created lexical resources can significantly improve the retrieval performance.

The rest of the paper is organized as follows. We discuss the related work in Section 2, and briefly review the studies of query expansion using axiomatic approaches in Section 3. We then present our study of using lexical resources, such as WordNet, for query expansion in Section 4, and discuss experiment results in Section 5. Finally, we conclude in Section 6.

2 Related Work

Although the use of WordNet in query expansion has been studied by various researchers, the improvement of retrieval performance is often limited. Voorhees (Voorhees, 1994) expanded queries using a combination of synonyms, hypernyms and hyponyms *manually* selected from WordNet, and achieved limited improvement (i.e., around -2% to

$+2\%$) on short verbose queries. Stairmand (Stairmand, 1997) used WordNet for query expansion, but they concluded that the improvement was restricted by the coverage of the WordNet and no empirical results were reported.

More recent studies focused on combining the information from both co-occurrence-based and hand-crafted thesauri. Mandala et. al. (Mandala et al., 1999a; Mandala et al., 1999b) studied the problem in vector space model, and Cao et. al. (Cao et al., 2005) focused on extending language models. Although they were able to improve the performance, it remains unclear whether using only information from hand-crafted thesauri would help to improve the retrieval performance.

Another way to improve retrieval performance using WordNet is to disambiguate word senses. Voorhees (Voorhees, 1993) showed that using WordNet for word sense disambiguation degrade the retrieval performance. Liu et. al. (Liu et al., 2004) used WordNet for both sense disambiguation and query expansion and achieved reasonable performance improvement. However, the computational cost is high and the benefit of query expansion using only WordNet is unclear. Ruch et. al. (Ruch et al., 2006) studied the problem in the domain of biology literature and proposed an argumentative feedback approach, where expanded terms are selected from only sentences classified into one of four disjunct argumentative categories.

The goal of this paper is to study whether query expansion using only manually created lexical resources could lead to the performance improvement. The main contribution of our work is to show query expansion using only hand-crafted lexical resources is effective in the recently proposed axiomatic framework, which has not been shown in the previous studies.

3 Query Expansion in Axiomatic Retrieval Model

Axiomatic approaches have recently been proposed and studied to develop retrieval functions (Fang and Zhai, 2005; Fang and Zhai, 2006). The main idea is to search for a retrieval function that satisfies all the desirable retrieval constraints, i.e., axioms. The underlying assumption is that a retrieval function sat-

isfying all the constraints would perform well empirically. Unlike other retrieval models, axiomatic retrieval models directly model the relevance with term level retrieval constraints.

In (Fang and Zhai, 2005), several axiomatic retrieval functions have been derived based on a set of basic formalized retrieval constraints and an inductive definition of the retrieval function space. The derived retrieval functions are shown to perform as well as the existing retrieval functions with less parameter sensitivity. One of the components in the inductive definition is primitive weighting function, which assigns the retrieval score to a single term document $\{d\}$ for a single term query $\{q\}$ based on

$$S(\{q\}, \{d\}) = \begin{cases} \omega(q) & q = d \\ 0 & q \neq d \end{cases} \quad (1)$$

where $\omega(q)$ is a term weighting function of q . A limitation of the primitive weighting function described in Equation 1 is that it can not bridge vocabulary gaps between documents and queries.

To overcome this limitation, in (Fang and Zhai, 2006), we proposed a set of semantic term matching constraints and modified the previously derived axiomatic functions to make them satisfy these additional constraints. In particular, the primitive weighting function is generalized as

$$S(\{q\}, \{d\}) = \omega(q) \times f(s(q, d)),$$

where $s(q, d)$ is a semantic similarity function between two terms q and d , and f is a monotonically increasing function defined as

$$f(s(q, d)) = \begin{cases} 1 & q = d \\ \frac{s(q, d)}{s(q, q)} \times \beta & q \neq d \end{cases} \quad (2)$$

where β is a parameter that regulates the weighting of the original query terms and the semantically similar terms. We have shown that the proposed generalization can be implemented as a query expansion method. Specifically, the expanded terms are selected based on a term similarity function s and the weight of an expanded term t is determined by its term similarity with a query term q , i.e., $s(q, t)$, as well as the weight of the query term, i.e., $\omega(q)$. Note that the weight of an expanded term t is $\omega(t)$ in traditional query expansion methods.

In our previous study (Fang and Zhai, 2006), term similarity function s is derived based on the mutual information of terms over collections that are constructed under the guidance of a set of term semantic similarity constraints. The focus of this paper is to study and compare several term similarity functions exploiting the information from lexical resources, and evaluate their effectiveness in the axiomatic retrieval models.

4 Term Similarity based on Lexical Resources

In this section, we discuss a set of term similarity functions that exploit the information stored in two lexical resources: WordNet (Miller, 1990) and dependency-based thesaurus (Lin, 1998).

The most commonly used lexical resource is WordNet (Miller, 1990), which is a hand-crafted lexical system developed at Princeton University. Words are organized into four taxonomies based on different parts of speech. Every node in the WordNet is a synset, i.e., a set of synonyms. The definition of a synset, which is referred to as *gloss*, is also provided. For a query term, all the synsets in which the term appears can be returned, along with the definition of the synsets. We now discuss six possible term similarity functions based on the information provided by WordNet.

Since the definition provides valuable information about the semantic meaning of a term, we can use the definitions of the terms to measure their semantic similarity. The more common words the definitions of two terms have, the more similar these terms are (Banerjee and Pedersen, 2005). Thus, we can compute the term semantic similarity based on synset definitions in the following way:

$$s_{def}(t_1, t_2) = \frac{|D(t_1) \cap D(t_2)|}{|D(t_1) \cup D(t_2)|},$$

where $D(t)$ is the concatenation of the definitions for all the synsets containing term t and $|D|$ is the number of words of the set D .

Within a taxonomy, synsets are organized by their lexical relations. Thus, given a term, related terms can be found in the synsets related to the synsets containing the term. In this paper, we consider the following five word relations.

- **Synonym(Syn)**: X and Y are synonyms if they are interchangeable in some context.
- **Hypernym(Hyper)**: Y is a hypernym of X if X is a (kind of) Y.
- **Hyponym(Hypo)**: X is a hyponym of Y if X is a (kind of) Y.
- **Holonym(Holo)**: Y is a holonym of X if X is a part of Y.
- **Meronym(Mero)**: X is a meronym of Y if X is a part of Y.

Since these relations are binary, we define the term similarity functions based on these relations in the following way.

$$s_R(t_1, t_2) = \begin{cases} \alpha_R & t_1 \in T_R(t_2) \\ 0 & t_1 \notin T_R(t_2) \end{cases}$$

where $R \in \{syn, hyper, hypo, holo, mero\}$, $T_R(t)$ is a set of words that are related to term t based on the relation R , and α s are non-zero parameters to control the similarity between terms based on different relations. However, since the similarity values for all term pairs are same, the values of these parameters can be ignored when we use Equation 2 in query expansion.

Another lexical resource we study in the paper is the dependency-based thesaurus provided by Lin¹ (Lin, 1998). The thesaurus provides term similarities that are automatically computed based on dependency relationships extracted from a parsed corpus. We define a similarity function that can utilize this thesaurus as follows:

$$s_{Lin}(t_1, t_2) = \begin{cases} L(t_1, t_2) & (t_1, t_2) \in TP_{Lin} \\ 0 & (t_1, t_2) \notin TP_{Lin} \end{cases}$$

where $L(t_1, t_2)$ is the similarity of terms stored in the dependency-based thesaurus and TP_{Lin} is a set of all the term pairs stored in the thesaurus. The similarity of two terms would be assigned to zero if we can not find the term pair in the thesaurus.

Since all the similarity functions discussed above capture different perspectives of term relations, we

propose a simple strategy to combine these similarity functions so that the similarity of a term pair is the highest similarity value of these two terms of all the above similarity functions, which is shown as follows.

$$s_{combined}(t_1, t_2) = \max_{R \in Rset}(s_R(t_1, t_2)),$$

where

$$Rset = \{def, syn, hyper, hypo, holo, mero, Lin\}.$$

In summary, we have discussed eight possible similarity functions that exploit the information from the lexical resources. We then incorporate these similarity functions into the axiomatic retrieval models based on Equation 2, and perform query expansion based on the procedure described in Section 3. The empirical results are reported in Section 5.

5 Experiments

In this section, we experimentally evaluate the effectiveness of query expansion with the term similarity functions discussed in Section 4 in the axiomatic framework. Experiment results show that the similarity function based on synset definitions is most effective. By incorporating this similarity function into the axiomatic retrieval models, we show that query expansion using the information from only WordNet can lead to significant improvement of retrieval performance, which has not been shown in the previous studies (Voorhees, 1994; Stairmand, 1997).

5.1 Experiment Design

We conduct three sets of experiments. First, we compare the effectiveness of term similarity functions discussed in Section 4 in the context of query expansion. Second, we compare the best one with the term similarity functions derived from co-occurrence-based resources. Finally, we study whether the combination of term similarity functions from different resources can further improve the performance.

All experiments are conducted over six TREC collections: ap88-89, doe, fr88-89, wt2g, trec7 and trec8. Table 1 shows some statistics of the collections, including the description, the collection size,

¹Available at <http://www.cs.ualberta.ca/~lindek/downloads.htm>

Table 1: Statistics of Test Collections

Collection	Description	Size	# Voc.	# Doc.	#query
ap88-89	news articles	491MB	361K	165K	150
doe	technical reports	184MB	163K	226K	35
fr88-89	government documents	469MB	204K	204K	42
trec7	ad hoc data	2GB	908K	528K	50
trec8	ad hoc data	2GB	908K	528K	50
wt2g	web collections	2GB	1968K	247K	50

the vocabulary size, the number of documents and the number of queries. The preprocessing only involves stemming with Porter’s stemmer.

We use WordNet 3.0 ², Lemur Toolkit ³ and TrecWN library ⁴ in experiments. The results are evaluated with both MAP (mean average precision) and gMAP (geometric mean average precision) (Voorhees, 2005), which emphasizes the performance of difficulty queries.

There is one parameter β in the query expansion method presented in Section 3. We tune the value of β and report the best performance. The parameter sensitivity is similar to the observations described in (Fang and Zhai, 2006) and will not be discussed in this paper. In all the result tables, ‡ and † indicate that the performance difference is statistically significant according to Wilcoxon signed rank test at the level of 0.05 and 0.1 respectively.

We now explain the notations of different methods. *BL* is the baseline method without query expansion. In this paper, we use the best performing function derived in axiomatic retrieval models, i.e, F2-EXP in (Fang and Zhai, 2005) with a fixed parameter value ($b = 0.5$). QE_X is the query expansion method with term similarity function s_X , where X could be *Def.*, *Syn.*, *Hyper.*, *Hypo.*, *Mero.*, *Holo.*, *Lin* and *Combined*.

Furthermore, we examine the query expansion method using co-occurrence-based resources. In particular, we evaluate the retrieval performance using the following two similarity functions: s_{MIBL} and s_{MIImp} . Both functions are based on the mutual information of terms in a set of documents. s_{MIBL} uses the collection itself to compute the mutual information, while s_{MIImp} uses the working sets con-

structed based on several constraints (Fang and Zhai, 2006). The mutual information of two terms t_1 and t_2 in collection C is computed as follow (van Rijsbergen, 1979):

$$I(X_{t_1}, X_{t_2}) = \sum p(X_{t_1}, X_{t_2}) \log \frac{p(X_{t_1}, X_{t_2})}{p(X_{t_1})p(X_{t_2})}$$

X_{t_i} is a binary random variable corresponding to the presence/absence of term t_i in each document of collection C .

5.2 Effectiveness of Lexical Resources

We first compare the retrieval performance of query expansion with different similarity functions using short keyword (i.e., title-only) queries, because query expansion techniques are often more effective for shorter queries (Voorhees, 1994; Fang and Zhai, 2006). The results are presented in Table 2. It is clear that query expansion with these functions can improve the retrieval performance, although the performance gains achieved by different functions vary a lot. In particular, we make the following observations.

First, the similarity function based on synset definitions is the most effective one. QE_{def} significantly improves the retrieval performance for all the data sets. For example, in trec7, it improves the performance from 0.186 to 0.216. As far as we know, none of the previous studies showed such significant performance improvement by using only WordNet as query expansion resource.

Second, the similarity functions based on term relations are less effective compared with definition-based similarity function. We think that the worse performance is related to the following two reasons: (1) The similarity functions based on relations are binary, which is not a good way to model term similarities. (2) The relations are limited by the part

²<http://wordnet.princeton.edu/>

³<http://www.lemurproject.org/>

⁴<http://l2r.cs.uiuc.edu/cogcomp/software.php>

Table 2: Performance of query expansion using lexical resources (short keyword queries)

	trec7		trec8		wt2g	
	MAP	gMAP	MAP	gMAP	MAP	gMAP
<i>BL</i>	0.186	0.083	0.250	0.147	0.282	0.188
QE_{def}	0.216\ddagger	0.105\ddagger	0.266\ddagger	0.164\ddagger	0.301\ddagger	0.210\ddagger
	(+16%)	(+27%)	(+6.4%)	(+12%)	(+6.7%)	(+12%)
QE_{syn}	0.194	0.085 \ddagger	0.252 \ddagger	0.150 \ddagger	0.287 \ddagger	0.194 \ddagger
	(+4.3%)	(+2.4%)	(+0.8%)	(+2.0%)	(+1.8%)	(+3.2%)
QE_{hyper}	0.186	0.086	0.250	0.152	0.286 \ddagger	0.192 \ddagger
	(0%)	(+3.6%)	(0%)	(+3.4%)	(+1.4%)	(+2.1%)
QE_{hypo}	0.186 \ddagger	0.085 \ddagger	0.250	0.147	0.282 \ddagger	0.190
	(0%)	(+2.4%)	(0%)	(0%)	(0%)	(+1.1%)
QE_{mero}	0.187 \ddagger	0.084 \ddagger	0.250	0.147	0.282	0.189
	(+0.5%)	(+1.2%)	(0%)	(0%)	(0%)	(+0.5%)
QE_{holo}	0.191 \ddagger	0.085 \ddagger	0.250	0.147	0.282	0.188
	(+2.7%)	(+2.4%)	(0%)	(0%)	(0%)	(0%)
QE_{Lin}	0.193 \ddagger	0.092 \ddagger	0.256 \ddagger	0.156 \ddagger	0.290 \ddagger	0.200 \ddagger
	(+3.7%)	(+11%)	(+2.4%)	(+6.1%)	(+2.8%)	(+6.4%)
$QE_{Combined}$	0.214 \ddagger	0.104 \ddagger	0.267\ddagger	0.165\ddagger	0.300 \ddagger	0.208 \ddagger
	(+15%)	(+25%)	(+6.8%)	(+12%)	(+6.4%)	(+10.5%)
	ap88-89		doe		fr88-89	
	MAP	gMAP	MAP	gMAP	MAP	gMAP
<i>BL</i>	0.220	0.074	0.174	0.069	0.222	0.062
QE_{def}	0.254\ddagger	0.088\ddagger	0.181\ddagger	0.075\ddagger	0.225\ddagger	0.067\ddagger
	(+15%)	(+19%)	(+4%)	(+10%)	(+1.4%)	(+8.1%)
QE_{syn}	0.222 \ddagger	0.077 \ddagger	0.174	0.074	0.222	0.065
	(+0.9%)	(+4.1%)	(0%)	(+7.3%)	(0%)	(+4.8%)
QE_{hyper}	0.222 \ddagger	0.074	0.175	0.070	0.222	0.062
	(+0.9%)	(0%)	(+0.5%)	(+1.5%)	(0%)	(0%)
QE_{hypo}	0.222 \ddagger	0.076 \ddagger	0.176 \ddagger	0.073 \ddagger	0.222	0.062
	(+0.9%)	(+2.7%)	(+1.1%)	(+5.8%)	(0%)	(0%)
QE_{mero}	0.221	0.074 \ddagger	0.174 \ddagger	0.070 \ddagger	0.222	0.062
	(+0.45%)	(0%)	(0%)	(+1.5%)	(0%)	(0%)
QE_{holo}	0.221	0.076	0.177 \ddagger	0.073	0.222	0.062
	(+0.45%)	(+2.7%)	(+1.7%)	(+5.8%)	(0%)	(0%)
QE_{Lin}	0.245 \ddagger	0.082 \ddagger	0.178	0.073	0.222	0.067 \ddagger
	(+11%)	(+11%)	(+2.3%)	(+5.8%)	(0%)	(+8.1%)
$QE_{Combined}$	0.254\ddagger	0.085 \ddagger	0.179 \ddagger	0.074 \ddagger	0.223 \ddagger	0.065
	(+15%)	(+12%)	(+2.9%)	(+7.3%)	(+0.5%)	(+4.3%)

Table 3: Performance comparison of hand-crafted and co-occurrence-based thesauri (short keyword queries)

Data	MAP			gMAP		
	QE_{def}	QE_{MIBL}	QE_{MIImp}	QE_{def}	QE_{MIBL}	QE_{MIImp}
ap88-89	0.254	0.233‡	0.265‡	0.088	0.081‡	0.089‡
doe	0.181	0.175†	0.183	0.075	0.071†	0.078
fr88-89	0.225	0.222‡	0.227†	0.067	0.063	0.071‡
trec7	0.216	0.195‡	0.236‡	0.105	0.089‡	0.097
trec8	0.266	0.250‡	0.278	0.164	0.148‡	0.172
wt2g	0.301	0.311	0.320‡	0.210	0.218	0.219‡

of speech of the terms, because two terms in WordNet are related only when they have the same part of speech tags. However, definition-based similarity function does not have such a limitation.

Third, the similarity function based on Lin’s thesaurus is more effective than those based on term relations from the WordNet, while it is less effective compared with the definition-based similarity function, which might be caused by its smaller coverage.

Finally, combining different WordNet-based similarity functions does not help, which may indicate that the expanded terms selected by different functions are overlapped.

5.3 Comparison with Co-occurrence-based Resources

As shown in Table 2, the similarity function based on synset definitions, i.e., s_{def} , is most effective. We now compare the retrieval performance of using this similarity function with that of using the mutual information based functions, i.e., s_{MIBL} and s_{MIImp} . The experiments are conducted over two types of queries, i.e. short keyword (keyword title) and short verbose (one sentence description) queries.

The results for short keyword queries are shown in Table 3. The retrieval performance of query expansion based on s_{def} is significantly better than that based on s_{MIBL} on almost all the data sets, while it is slightly worse than that based on s_{MIImp} on some data sets. We can make the similar observation from the results for short verbose queries as shown in Table 4. One advantage of s_{def} over s_{MIImp} is the computational cost, because s_{def} can be computed offline in advance while s_{MIImp} has to be computed online from query-dependent working sets which takes much more time. The low computa-

tional cost and high retrieval performance make s_{def} more attractive in the real world applications.

5.4 Additive Effect

Since both types of similarity functions are able to improve retrieval performance, we now study whether combining them could lead to better performance. Table 5 shows the retrieval performance of combining both types of similarity functions for short keyword queries. The results for short verbose queries are similar. Clearly, combining the similarity functions from different resources could further improve the performance.

6 Conclusions

Query expansion is an effective technique in information retrieval to improve the retrieval performance, because it often can bridge the vocabulary gaps between queries and documents. Intuitively, hand-crafted thesaurus could provide reliable related terms, which would help improve the performance. However, none of the previous studies is able to show significant performance improvement through query expansion using information only from manually created lexical resources.

In this paper, we re-examine the problem of query expansion using lexical resources in recently proposed axiomatic framework and find that we are able to significantly improve retrieval performance through query expansion using only hand-crafted lexical resources. In particular, we first study a few term similarity functions exploiting the information from two lexical resources: WordNet and dependency-based thesaurus created by Lin. We then incorporate the similarity functions with the query expansion method in the axiomatic retrieval

Table 4: Performance Comparison (MAP, short verbose queries)

Data	BL	QE_{def}	QE_{MIBL}	QE_{MIImp}
ap88-89	0.181	0.220 \ddagger (21.5%)	0.205 \ddagger (13.3%)	0.230 \ddagger (27.1%)
doe	0.109	0.121 \ddagger (11%)	0.119 (9.17%)	0.117 (7.34%)
fr88-89	0.146	0.164 \ddagger (12.3%)	0.162 \ddagger (11%)	0.164 \ddagger (12.3%)
trec7	0.184	0.209 \ddagger (13.6%)	0.196 (6.52%)	0.224 \ddagger (21.7%)
trec8	0.234	0.238 \ddagger (1.71%)	0.235 (0.4%)	0.243 \ddagger (3.85%)
wt2g	0.266	0.276 (3.76%)	0.276 \ddagger (3.76%)	0.282 \ddagger (6.02%)

Table 5: Additive Effect (MAP, short keyword queries)

	ap88-89	doe	fr88-89	trec7	trec8	wt2g
QE_{MIBL}	0.233	0.175	0.222	0.195	0.250	0.311
$QE_{def+MIBL}$	0.257 \ddagger	0.183 \ddagger	0.225 \ddagger	0.217 \ddagger	0.267 \ddagger	0.320 \ddagger
QE_{MIImp}	0.265	0.183	0.227	0.236	0.278	0.320
$QE_{def+MIImp}$	0.269\ddagger	0.187	0.232\ddagger	0.237\ddagger	0.280\ddagger	0.322\ddagger

models. Systematical experiments have been conducted over six standard TREC collections and show promising results. All the proposed similarity functions improve the retrieval performance, although the degree of improvement varies for different similarity functions. Among all the functions, the one based on synset definition is most effective and is able to significantly and consistently improve retrieval performance for all the data sets. This similarity function is also compared with some similarity functions using mutual information. Furthermore, experiment results show that combining similarity functions from different resources could further improve the performance.

Unlike previous studies, we are able to show that query expansion using only manually created thesauri can lead to significant performance improvement. The main reason is that the axiomatic approach provides guidance on how to appropriately assign weights to expanded terms.

There are many interesting future research directions based on this work. First, we will study the same problem in some specialized domain, such as biology literature, to see whether the proposed approach could be generalized to the new domain. Second, the fact that using axiomatic approaches to incorporate linguistic information can improve retrieval performance is encouraging. We plan to extend the axiomatic approach to incorporate more linguistic information, such as phrases and word

senses, into retrieval models to further improve the performance.

Acknowledgments

We thank ChengXiang Zhai, Dan Roth, Rodrigo de Salvo Braz for valuable discussions. We also thank three anonymous reviewers for their useful comments.

References

- J. Bai, D. Song, P. Bruza, J. Nie, and G. Cao. 2005. Query expansion using term relationships in language models for information retrieval. In *Fourteenth International Conference on Information and Knowledge Management (CIKM 2005)*.
- S. Banerjee and T. Pedersen. 2005. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*.
- G. Cao, J. Nie, and J. Bai. 2005. Integrating word relationships into language models. In *Proceedings of the 2005 ACM SIGIR Conference on Research and Development in Information Retrieval*.
- H. Fang and C. Zhai. 2005. An exploration of axiomatic approaches to information retrieval. In *Proceedings of the 2005 ACM SIGIR Conference on Research and Development in Information Retrieval*.
- H. Fang and C. Zhai. 2006. Semantic term matching in axiomatic approaches to information retrieval. In *Proceedings of the 2006 ACM SIGIR Conference on Research and Development in Information Retrieval*.

- N. Fuhr. 1992. Probabilistic models in information retrieval. *The Computer Journal*, 35(3):243–255.
- Y. Jing and W. Bruce Croft. 1994. An association thesaurus for information retrieval. In *Proceedings of RIAO*.
- D. Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of International Conference on Machine Learning (ICML)*.
- S. Liu, F. Liu, C. Yu, and W. Meng. 2004. An effective approach to document retrieval via utilizing wordnet and recognizing phrases. In *Proceedings of the 2004 ACM SIGIR Conference on Research and Development in Information Retrieval*.
- R. Mandala, T. Tokunaga, and H. Tanaka. 1999a. Ad hoc retrieval experiments using wordnet and automatically constructed thesauri. In *Proceedings of the seventh Text REtrieval Conference (TREC7)*.
- R. Mandala, T. Tokunaga, and H. Tanaka. 1999b. Combining multiple evidence from different types of thesaurus for query expansion. In *Proceedings of the 1999 ACM SIGIR Conference on Research and Development in Information Retrieval*.
- G. Miller. 1990. Wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4).
- H. J. Peat and P. Willett. 1991. The limitations of term co-occurrence data for query expansion in document retrieval systems. *Journal of the American Society for Information Science*, 42(5):378–383.
- J. Ponte and W. B. Croft. 1998. A language modeling approach to information retrieval. In *Proceedings of the ACM SIGIR'98*, pages 275–281.
- Y. Qiu and H.P. Frei. 1993. Concept based query expansion. In *Proceedings of the 1993 ACM SIGIR Conference on Research and Development in Information Retrieval*.
- P. Ruch, I. Tbahriti, J. Gobeill, and A. R. Aronson. 2006. Argumentative feedback: A linguistically-motivated term expansion for information retrieval. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 675–682.
- G. Salton, C. S. Yang, and C. T. Yu. 1975. A theory of term importance in automatic text analysis. *Journal of the American Society for Information Science*, 26(1):33–44, Jan-Feb.
- A. F. Smeaton and C. J. van Rijsbergen. 1983. The retrieval effects of query expansion on a feedback document retrieval system. *The Computer Journal*, 26(3):239–246.
- M. A. Stairmand. 1997. Textual context analysis for information retrieval. In *Proceedings of the 1997 ACM SIGIR Conference on Research and Development in Information Retrieval*.
- C. J. van Rijsbergen. 1979. *Information Retrieval*. Butterworths.
- E. M. Voorhees. 1993. Using wordnet to disambiguate word sense for text retrieval. In *Proceedings of the 1993 ACM SIGIR Conference on Research and Development in Information Retrieval*.
- E. M. Voorhees. 1994. Query expansion using lexical-semantic relations. In *Proceedings of the 1994 ACM SIGIR Conference on Research and Development in Information Retrieval*.
- E. M. Voorhees. 2005. Overview of the trec 2005 robust retrieval track. In *Notebook of the Thirteenth Text REtrieval Conference (TREC2005)*.