Effect of log-based Query Term Expansion on Retrieval Effectiveness in Patent Searching

Wolfgang Tannebaum, Parvaz Madabi and Andreas Rauber

Institute of Software Technology and Interactive Systems,
Vienna University of Technology, Austria
http://www.ifs.tuwien.ac.at
{tannebaum, rauber}@ifs.tuwien.ac.at

Abstract. In this paper we study the impact of query term expansion (*QTE*) using synonyms on patent document retrieval. We use an automatically generated lexical database from USPTO query logs, called *PatNet*, which provides synonyms and equivalents for a query term. Our experiments on the CLEF-IP 2010 benchmark dataset show that automatic query expansion using *PatNet* tends to decrease or only slightly improve the retrieval effectiveness, with no significant improvement. An analysis of the retrieval results shows that *PatNet* does not have generally a negative effect on the retrieval effectiveness. Recall is drastically improved for query topics, where the baseline queries achieve, on average, only low recall values. But we have not detected any commonality that allows us to characterize these queries. So we recommend using *PatNet* for semi-automatic *QTE* in Boolean retrieval, where expanding query terms with synonyms and equivalents with the aim of expanding the query scope is a common practice.

Keywords: Patent Searching, Query Term Expansion, Query Log Analysis;

1 Introduction

Patent search is the task of finding relevant patent information in patent databases to judge the validity of an applied or granted patent based on novelty and inventiveness. This task is usually performed by examiners in a patent office and patent searchers in private companies. For searching the patent databases the patent searchers commonly formulate complex Boolean queries, which are easy for patent experts to manipulate and which provide a record of what documents were searched. Query terms are expanded with synonyms or equivalents, co-occurring terms and keyword phrases [1,2]. Especially the expansion of the query terms with synonyms is particular common in Boolean patent retrieval, as shown in [7]. In this paper, we want to measure the effect of query expansion using synonyms on retrieval effectiveness in patent searching, in particular when used in a fully automatic manner. We use the lexical database *PatNet* automatically generated from USPTO query logs, which provides synonyms and equivalents for a query term [8].

2 Related Work

Currently, fully automatic query expansion in patent search is mostly based on computing co-occurring terms in a patent corpus [1,4]. Additional query terms are extracted automatically from the query documents, the feedback documents or from the cited documents, for example based on statistical measures, such as term frequencies (tf) and a combination of term frequencies and inverted document frequencies (tfidf) [4,10]. Also, whole documents or whole sections of the query documents are used for query generation and query expansion [9]. One approach, which uses synonyms for automatic query expansion in the patent domain is described in [4]. The standard dictionary *WordNet* and a lexical database extracted from a European Patent Office (*EPO*) patent collection, called *SynSet*, was used to improve the retrieval effectiveness. Experiments show better retrieval precision over the baseline queries, but not for recall. So for recall-oriented patent searching, this result is negative. Contrary to the usage of synonyms from standard dictionaries and extracted from patent documents, as indicated in [4], we propose to use query logs as presented in [6] and in particular query logs of patent examiners as suggested in [7,8] for automatic query expansion. This allows us to use specific terms for query expansion, in particular the query and expansion terms to the patent applications used by the patent examiners for searching.

3 Experiments

For our experiments, we use the data set of the CLEF-IP initiative[1], namely the CLEF-IP 2010 data set. To evaluate the performance of IR approaches we use the test set of the CLEF-IP data set, which is based on 1,348 English patent topics, and the metrics Precision, Recall, Average Precision (AP) and Mean Average Precision (MAP). In addition, we use the Patent Retrieval Evaluation Score (PRES), which combines recall and the user's search effort in one single score [3].

3.1 Baseline runs

The first query model (*QS-BL*) estimates the importance of each term according to a weighted log-likelihood based approach comparing the foreground (query patent) and background (collection) language models. Terms with high similarity to the foreground language model and low similarity to the background language model are used as query terms representing the specific terminology of the query patent. Top k terms with higher weights are selected as query terms from this query model. All fields of the query patents are considered in the query estimation process and k is experimentally set to 100. Initial query *QS-BL* is expanded using the information available in the citations of the query patent. Two different weighting algorithms are

---

[1] http://ifs.tuwien.ac.at/ clef-ip/

used for calculating query weights while taking into account the citation information. The first approach (*QS-PR*) uses PageRank scores to identify influential documents in the citation graph of a query patent and then uses those documents for drawing expansion terms. The second approach (*QS-TPR*) uses a time-aware decay function to give importance to newer documents in the citation graph and penalize older documents. Further explanations on baseline queries can be found in [5]. We also combined all the query lists above and we refer to this combined list as (*QS-MQL*). Table 1 shows the evaluation results using the CLEF-IP 2010 corpora in terms of MAP, Recall, and PRES at cut-off value of 1000.

Table 1. Retrieval Results of the baseline runs.

Query Model	MAP		Recall		PRES	
	value	*change*	*value*	*change*	*value*	*change*
QS-BL	0.1368	NA	**0.6215**	NA	**0.5067**	NA
QS-PR	**0.1392**	+1.7%	0.6302	+1.4%	0.5121	+2.1%
QS-TPR	0.1391	+1.7%	0.6305	+1.4%	0.5123	+1.1%
QS-MQL	**0.0815**	-40%	**0.6761**	+8.8%	**0.5484**	+8.2%

The results show that the expanded query sets obtain better performance compared to the baseline query set in view of recall and PRES. Further, the citation query model, which is based on citation information together with the publication dates, and the citation query model using Page Rank scores achieve similar performance in view of recall and PRES. Compared to the previous approaches tested on CLEF-IP 2010, the runs can be considered as the second best methods in terms of recall and PRES.[2]

3.2 Using *PatNet* for Query Term Expansion

We use *PatNet* to expand the query and expansion terms of the baseline query sets with synonymous expansion terms (*ET*). We use the most likely *ETs*, which are commonly used by patent examiners of the USPTO for *QTE*. For each expansion we use the highest ranked, in particular the most frequent, *ET* provided by *PatNet* for a query and expansion term. Specifically, we replace the terms in the baseline query sets with synonyms for which *PatNet* suggests *ETs*. So we generate four additional query sets for the query topics. Table 2 shows the evaluation results in terms of MAP, Recall, and PRES at cut-off value of 1000 when using *PatNet* for *QTE*.

Table 2. Retrieval Results of the expanded query sets.

Query Model	MAP		Recall		PRES	
	value	*change*	*value*	*change*	*value*	*change*
QS-BLE	0.0848	-38%	0.4983	**-19%**	0.3835	-24%
QS-PRE	**0.1390**	-0.1%	**0.6307**	+0.1%	**0.5123**	+0.1%
QS-TPRE	**0.0066**	-95%	**0.1871**	-70%	**0.1238**	-76%
QS-MQLE	0.0132	-84%	0.2033	-70%	0.1478	-73%

The results show that, when querying *QS-BL* in combination with *QS-BLE*, the retrieval performance drastically decreases. In particular, recall goes down (-19%) from 62% to 50%. Further, PRES decreases by 24% from 51% to 38% and MAP decreases from 14% to 8% (-38%). Further, recall and precision can be slightly improved, while precision decreases slightly (-0.1%), when using *PatNet* with *QS-PR*. In combination with the second and third expansion approach the retrieval performance drastically decreases.

3 Analysis of the retrieval results

The experiments show that there was no significant improvement in the retrieval effectiveness, when using *PatNet* for fully-automatic *QTE*. We now analyze the results per topic (1348 topics) to validate whether there are certain characteristics that indicate when the approach comes in useful. Table 3 shows the percentage of topics for which the retrieval performance is improved, remains unchanged, or is degraded.

Table 3. Improved, unchanged, and degraded query topics.

QS-BLE	Recall	MAP	PRES
improved	**13.96%**	23.46%	24.20%
unchanged	36.90%	**2.52%**	**2.52%**
degraded	**49.15%**	**74.02%**	**73.27%**

Table 3 shows that expanding the baseline queries with synonyms results in improving the recall of 14% of query topics. For about 49% of the topics recall decreases. As expected through the expansion of the query scope, precision decreases for large number of topics (74%). Through the expansion a lot of additional non-relevant documents are retrieved. But Table 4 shows that recall can be significantly improved (+34%) for query topics, which achieve, on average, only low recall (44%). Otherwise, recall drastically degrades (from 64% to 35%) when queries, which still provide good recall measures, are expanded with *PatNet* (initially retrieved relevant documents are lost from the rank list).

Table 4. Recall achieved for improved, unchanged and degraded query topics.

Recall	QS-BL	QS-PR	QS-TPR	QS-BLE
Avg.	0.6215	0.6302	0.6305	0.4983
improved	**0.4407**	**0.5263**	**0.5298**	**0.5911**
unchanged	0.6635	**0.6726**	**0.6724**	0.6635
degraded	**0.6411**	0.6289	0.6285	**0.3478**

Further, *PatNet* significantly outperform the related expansion approaches *QS-PR* and *QS-TPR,* which achieve for these query topics compared to their avg. recall performance and values achieved for the unchanged and degraded query topics, on average, only moderate recall. The retrieval performance of these query topics are apparently difficult to improve with the related expansion approaches.

To characterize for which queries the expansion performs better, we now try to detect commonalities. At first, we consider the patent classifications of the query topics and the cited documents, and the classes *PatNet* was extracted from. We measure the overlap of the classes based on the queries for which the retrieval performance is improved, remains unchanged, or is degraded. The analysis shows for the query topics as well as for the citations that in each case (for improved, unchanged or degraded topics) about half of the query patents and citations are classified in the same classes as *PatNet* was extract from. So the patent classification is no criteria to detect queries for which the expansion performs better. Next, we evaluate whether the performance of the lexical database depends on the number of provided *ETs (n)*, the query topic length *(l)* or on the number of retrieved relevant documents *(c)*.

Table 5. Query topic, query and citation characteristics.

QS-BL		Avg.	Max.	Min.
	n	51	74	28
improved	l	14,959	133,762	1,280
	c	13	76	0
	n	56	75	26
unchanged	l	11,174	110,506	1,513
	c	11	57	0
	n	56	77	19
degraded	l	12,370	102,371	1,509
	c	16	85	1

Table 5 shows that the performance of *PatNet* is independent from the number of provided *ETs* and from the query topic length. We consider the number of character strings of each query patent. For improved, unchanged or degraded query topics virtually the same number of *ETs* are used for query expansion. Further, query topics have, on average, virtually equivalent topic lengths showing that *PatNet* can be used both for shorter topics and for longer query topics. Also the number of retrieved relevant documents is no criteria to detect when *PatNet* comes in useful. Table 6 shows the rank positions of the relevant documents provided by the baseline query set to detect whether it is an issue of being to generic or not found via the given query terms. The latter would argue for extending the query scope using synonyms.

Table 6. Rank positions of the retrieved relevant documents provided by *QS-BL*.

QS-BL	1 - 250	251 - 500	501 - 750	751 - 1000
improved	65%	16%	11%	8%
unchanged	79%	12%	6%	3%
degraded	63%	18%	11%	8%

As shown more than two-thirds of the retrieved relevant documents appear in the rank lists under the top 250 documents. Less than 8% appear in the last 250 documents. These distributions of the documents speaks for extending the query scope using synonyms. But the experiments indicate just the opposite. Finally, we consider the patent conventions and countries the relevant documents have been filed to detect,

whether it is an issue that *PatNet* was extracted only from US patents. In each case about half of the relevant documents are *EP* or *WO* patents and about one third of the topics are US patents. There is no increase of US patents for improved query topics.

5 Conclusions and future work

In this paper used the lexical database *PatNet* for automatic *QTE* in patent searching. The experiments show that the retrieval performance of the query generation and expansion models presented in this work is decreased or only marginally improved, when using *PatNet* for *QTE*. No significant improvement is recognized. But the analysis of the retrieval results shows that the query log-based *QTE* method does not have generally a negative effect on the retrieval effectiveness. Recall is drastically improved for query topics, where the baseline queries achieve, on average, only low recall values. But we have not detected any commonality that allows us to characterize these queries. So we recommend to use *PatNet* as a lexical resource for semi-automatic *QTE* in Boolean patent retrieval, where synonym expansion is particularly common to improve recall. In our future work we will focus on semi-automatic *QTE* to assist patent searchers in assembling complex Boolean queries.

References

1. Jochim, C., Lioma, C., Schütze, H. 2011. Expanding queries with term and phrase translations in patent retrieval. In Proc. of the Second Int. Conf. on Multidisciplinary Inf. Retrieval Facility (IRFC 2011), Vienna, Austria, pp. 16-29.
2. Kim, Y., Seo, J., Croft, W.B. 2011. Automatic Boolean query suggestion for professional search. In Proc. of the 34th int. ACM SIGIR Conf. on Research and development in Inf. Retrieval (SIGIR 2011), Beijing, China, pp. 825-834.
3. Magdy, W., Jones, G.J.F. 2010. PRES: A score metric for evaluating recall-oriented information retrieval applications. In Proc. of the 33rd int. ACM SIGIR Conf. on Research and development in Inf. Retrieval (SIGIR 2010), Geneva, Switzerland, pp. 611-618.
4. Magdy, W., Jones, G.J.F. 2011. A Study of Query Expansion Methods for Patent Retrieval. In Proc. of PaIR 2011, Glasgow, Scotland, pp. 19-24.
5. Mahdabi P., Crestani, F. 2014. Patent Query Formulation by Synthesizing Multiple Sources of Relevance Evidence. In Trans. on Inf. Systems, Volume 32, Issue 4, Article No. 4.
6. Silvestri, F. 2010. Mining Query Logs: Turning Search Usage Data into Knowledge. In Foundations and Trends in Information Retrieval, Volume 4, Issue 1-2, pp. 1-174.
7. Tannebaum, W., Rauber, A. 2013. Mining Query Logs of USPTO Patent Examiners. In Proc. of 4th Int. Conf. of the CLEF Initiative (CLEF 2013), Valencia, Spain, pp. 136-142.
8. Tannebaum W., Rauber, A. 2015. *PatNet*: A lexical database for the patent domain. In Proc. of the 37th Europ. Conf. on Inf. Retrieval (ECIR 2015), Vienna, Austria, pp. 550-555.
9. Xue, X., Croft, W. 2009. Transforming patents into prior-art queries. In Proc. of the 32nd Int. ACM SIGIR Conf. on Research and Development in Inf. Retrieval, Boston, USA, 808-80.
10. Xue, X., Croft, W. 2009. Automatic query generation for patent search. In Proc. of CIKM 2009, Hong Kong, China, pp. 2037-2040.