A Study on Query Expansion Methods for Patent Retrieval

Walid Magdy and Gareth J. F. Jones
Centre for Next Generation Localization
School of Computing, Dublin City University, Dublin 9, Ireland
{wmagdy, gjones}@computing.dcu.ie

ABSTRACT
Patent retrieval is a recall-oriented search task where the objective is to find all possible relevant documents. Queries in patent retrieval are typically very long since they take the form of a patent claim or even a full patent application in the case of prior-art patent search. Nevertheless, there is generally a significant mismatch between the query and the relevant documents, often leading to low retrieval effectiveness. Some previous work has tried to address this mismatch through the application of query expansion (QE) techniques which have generally showed effectiveness for many other retrieval tasks. However, results of QE on patent search have been found to be very disappointing. We present a review of previous investigations of QE in patent retrieval, and explore some of these techniques on a prior-art patent search task. In addition, a novel method for QE using automatically generated synonyms set is presented. While previous QE techniques fail to improve over baseline retrieval, our new approach show statistically better retrieval precision over the baseline, although not for recall. In addition, it proves to be significantly more efficient than existing techniques. An extensive analysis to the results is presented which seeks to better understand situations where these QE techniques succeed or fail.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms
Algorithms, Experimentation

Keywords
Patent retrieval, Query expansion, SynSet

1. INTRODUCTION
One of the key differences between patent search and other search tasks such as web search is that the query in patent search is generally much longer. For example, in patent invalidity search task, the query consists of a patent claim which comprises several sentences, and in the case of prior-art search task, the query can be a full patent application. Despite the length of the query, there is often a significant term mismatch problem between topics and relevant documents. In [10], an analysis of the matching between patent topics and their relevant documents was presented for the CLEF-IP 2009 prior-art search task. This analysis showed the seriousness of the mismatch problem, where 12% of the relevant patents do not share any terms in common with the patent topics after filtering out stop words. This result highlights the challenge of patent search, since not only are long queries often very ambiguous with respect to the information need, but they frequently share few terms with the target documents.

In order to achieve higher recall, which is the main objective in patent search, there needs to be an overlap between the topic and the relevant documents. In this paper, different query expansion (QE) methods are investigated and a novel QE approach is introduced for patent retrieval. The main focus here is the exploration of methods which seek to improve the retrieval effectiveness of a well formulated long query. The hypothesis here assumes that expanding the query with additional terms will lead to increased possibilities for term matching between query and relevant documents, which can potentially lead to improved retrieval results. However at the same time, it can lead to further false matches between the query and non-relevant documents leading to degradation in retrieval effectiveness.

Several studies have sought to improve retrieval effectiveness for patent search tasks through QE with relevance feedback techniques [5,4,15,6]. Unfortunately, none of these studies succeeded in achieving a stable significant improvement in retrieval effectiveness. Here, we re-examine this existing work and apply additional QE techniques through expanding queries using WordNet. In addition, we introduce a novel approach that automatically generates candidate synonyms sets (SynSet) for terms, and use it as a source of expansion terms. These QE techniques were applied to the CLEF-IP 2010 prior-art patent search task. None of these approaches were able to achieve a significant improvement over the baseline. However, some of them are shown to improve retrieval effectiveness for some of the queries. Of the approaches tested, the novel QE approach introduced in here achieved the best results. It also proves to be the most efficient of the techniques examined. This success indicates its potential, while can be used immediately on demand by users, further research is needed to better understand whether it can be applied completely automatically.

The remainder of this paper is organized as follows: Section 2 gives background on different QE techniques in general and for patent search in particular; Section 3 tests the current standard QE techniques on the prior-art patent search task; Section 4 presents the approach used for generating SynSet and tests its impact on retrieval effectiveness; Section 5 analyses the results, and finally, Section 6 concludes the paper.

2. BACKGROUND
2.1 Query Expansion Techniques
Many expansion techniques for queries have been introduced in the field of information retrieval (IR) with the objective of improving retrieval effectiveness. In general these operate by

PaIR'11, October 24, 2011, Glasgow, Scotland, UK.

providing additional descriptive detail of the user's information need by adding additional terms to the original user's query. Many approaches have been proposed and explored for the selection of these additional terms and how they are weighted in combination with the original query terms. The expansion terms can be selected from a feedback process [2,14], or from external sources such as Wikipedia, dictionaries, or query logs. Expansion can be per term such as using WordNet [16,8] or per query as in the case of relevance feedback.

The main assumption when expanding the original query with additional terms is that the added terms increase the probability of matching of the query with relevant documents, with the objective of improving the retrieval effectiveness. However, this assumption is not always valid, since the expansion terms can also lead to promote retrieval of irrelevant documents. Thus research in QE typically focuses on seeking to expand queries with "good" terms that lead to overall improvement in the effectiveness of the IR system [2].

2.2 Query Expansion for Patent Retrieval

The main objective of QE is to overcome mismatch between search queries and relevant documents. This is typical for situations where queries are short and do not describe the user's information need well. However, while for patents the query is typically very long, there is still often a significant mismatch between queries and relevant items [10,13]. This has led researchers to investigate QE techniques for patent search. However, reported work on QE for patent search has never demonstrated consistent effectiveness.

Some of the initial trials for utilizing pseudo relevance feedback (PRF) for QE in patent search are described in [5]. PRF is a standard techniques used to enrich a search query with additional terms from the top ranked documents from an initial retrieval run under the assumption that these documents are relevant [14]. In this work a novel mechanism for PRF specifically designed for patent search was introduced and compared to the standard Rocchio method. Experiments on the NTCIR-3 patent retrieval task did not produce any significant improvement in retrieval results. The author commented that the reason for this may be that all words from the documents assumed to be relevant were used without any selection process. In NTCIR-4, there was another attempt at utilizing QE through PRF to improve the retrieval effectiveness [4]. However, it was found that while retrieval effectiveness was improved for a few topics, it was degraded for many others. The authors did not provide a clear analysis of possible reasons.

In the patent invalidity search task in NTCIR-5, another group tried to use a PRF algorithm using a different approach by only reweighting the terms of the query based on comparing the hierarchical structure of the patent classification of the initial retrieved documents to that of the query [15]. Again, this technique did not lead to any significant improvements in the final search results. In the same year, another QE method was applied to the patent queries in the invalidity search task from the patent topic itself instead of the collection [6]. The technique attempted to expand the patent claim query using additional explanation text of the claim from the description section of the patent. The challenge for this method was to locate the part of the description section that describes the given claim. They used morphological analysis and pattern matching techniques to extract these relevant parts and appended them to the query claim. Their technique achieved some significant improvement. However, the main disadvantage of this method was that it was specifically designed for patent invalidity search and cannot be generalized to other tasks such as prior-art search, where the query is the full patent application including the claims and the description sections.

Another investigation explored use of PRF to improve the patent retrievability in patent search rather than improving the retrieval effectiveness directly [1]. The problem addressed in this research was that some patents have a low chance of being retrieved or sometimes cannot be retrieved by any query. The objective for this research was to enrich the patent queries with additional terms using the PRF method to improve the retrievability score for patents in the collection. They succeeded in significantly improving the Gini coefficient, which is used to measure the retrievability. However, they did not test how this would affect the retrieval effectiveness for a patent search task.

3. TESTING STANDARD QUERY EXPANSION TECHNIQUES WITH PATENT RETRIEVAL

3.1 Experimental Data and Baseline Run

All experiments described in this paper were performed on the CLEF-IP 2010 prior-art search task [13] where the objective is to find relevant patents for a set of patent applications. The collection consists of 1.35M patents in three different languages of which 69% are English and 31% are French and German. The French and German patents are provided with their title, abstract, and claims sections manually translated into English. The English text of all patents was indexed to create a multilingual English index. A set of 1,348 English patent topics was provided for this task. Each topic is a full patent application which can be tens of pages in length. For our experiments we used a simple state-of-the-art query formulation technique presented in [12]. The applied search query was constructed from terms in the description section of the patent topic by filtering out terms that appeared only once, and including term bigrams appearing in the title and abstract sections of the query patent more than once. The Indri[1] toolkit was used for the retrieval process. It has been shown that citation extraction techniques for the prior-art patent search task improve the results significantly [13,12]. However, in our experiments, we do not apply any of these extraction techniques, since our focus is the retrieval algorithm itself. MAP and PRES are used for evaluating the results with more emphasize on PRES since it is specifically designed for the recall-oriented patent search task [11]. The scores for the baseline run were PRES = 0.486 and MAP = 0.1399.

The baseline formulated queries were then expanded using standard QE techniques as described in the next sections.

3.2 Psuedo Relevance Feedback

Although the existing work showed that PRF is not effective for improving retrieval effectiveness for patent search, we apply it on our data set to see if this finding is replicated for our task. This is important since the reported results are for a patent invalidity search task, not the prior-art search task investigated here. In this experiment, the PRF implemented in the Indri search toolkit is used with different numbers of assumed relevant documents and expansion terms for the feedback process. Indri's PRF mechanism is an adaptation of Lavrenko's relevance models [7]. The default weighting between the original query and expansion terms in Indri

[1] http://www.lemurproject.org/

is 1:1. The numbers of documents tested for the PRF process were {5, 10, 20}, and the number of expansion terms {10, 20, 30, 50}.

The results in Table 1 show that the best PRF run led to degradation in retrieval effectiveness of 48%, indicating its unsuitability as a QE method for this task. This poor result did not motivate us to try different weightings of the expansion terms to the original query, since the expansion terms appeared to be very destructive to the query and the best result was likey to be achieved by assigning a weighting of zero to the expansion terms. A possible explanation for this result is that the initial performance of the baseline is relatively low, which means that the top ranked documents used for PRF are mostly non relevant, meaning that QE is likely to add noise terms leading to degradation in the retrieval effectiveness.

These results are much worse than those reported in previous research [5,4,15]. This is because the previous work evaluated a patent invalidity search task, where the query is a patent claim (one or few sentences). The task tested here is a prior-art patent search task where the query is much longer since it is a full document. Our findings confirm that PRF is not an effective QE algorithm for standard patent search tasks.

Table 1: Effect of PRF with varying numbers of feedback documents and expansion terms on prior-art patent search

	Docs \ Terms	**10**	**20**	**30**	**50**
MAP BL = 0.1399	**5**	0.037	0.053	0.062	**0.072**
	10	0.031	0.046	0.053	0.061
	20	0.026	0.036	0.042	0.049
PRES BL = 0.486	**5**	0.196	0.234	0.247	**0.265**
	10	0.190	0.222	0.235	0.251
	20	0.178	0.205	0.216	0.232

3.3 Expanding Queries using WordNet

WordNet has been utilized in several IR research investigations to expand queries to achieve improved retrieval effectiveness [8,16]. To the best of our knowledge, it has not previously been tested on patent search tasks. We explore the potential for use of WordNet to expand patent queries to improve the retrieval effectiveness. Each term in a query is expanded with its WordNet synonyms and hyponyms. It was observed that expanding patent queries using WordNet slows down the retrieval process dramatically, especially when using many expansions since the number of expansion terms is very large. For initial experiments, only the first 100 topics from the English topics set were selected as a pilot run to select the best expansion set of terms from WordNet. Four test runs using noun/verb synonyms/hyponyms for expanding the meaning of each term were carried out as follows: NS (each term is expanded with its noun synonyms), NS+VS (noun and verb synonyms), NS+NH (noun synonyms and hyponyms), and NS+VS+NH+VH (synonyms and hyponyms for nouns and verbs). The "#syn" operator in Indri query language was used to enable the presence of synonyms of terms in the query[2].

Table 2 reports results of the four runs. For the 100 pilot topics, it was found that expanding the query terms with WordNet leads to a slight improvement in MAP, but significant degradation in PRES. This result means that a few of the relevant documents are being promoted to higher positions in the ranked list, but that a greater number of relevant documents are moved lower in the list or even lost from the ranked list completely. For a patent task, this outcome is considered a negative result. In addition, it was found that the retrieval speed when expanding terms with WordNet was massively slowed down, especially when more expansion terms were used (for the NS+VS+NH+VH run, speed of retrieval was more than 50 times slower). Although the results were not positive, from further analysis, we found that retrieval for some of the topics was improved. In order to perform deeper analysis and get a conclusive result, we applied the expansion using NS (noun synonyms) only to the full topics set since it achieved the best results and was the fastest among all the WordNet runs. Table 3 compares the retrieval effectiveness for the CLEF-IP full English topics with and without expansion using WordNet. Unlike the pilot run, both the MAP and PRES for the QE run were lower than the baseline. This result confirms that WordNet is not an effective method for QE for patent search. This is in addition to the slow search speed and the language dependency of this method.

Table 2: Effect of using WordNet for QE on the retrieval effectiveness of 100 pilot patent queries

	MAP		**PRES**	
	value	%change	value	%change
Baseline	0.1668	NA	0.584	NA
NS	0.1680	+0.7%	0.562	-3.7%
NS+NH	0.1680	+0.7%	0.561	-3.8%
NS+VS	0.1677	+0.5%	0.551	-5.6%
NS+NH+VS+VH	0.1540	-7.6%	0.544	-6.8%

Table 3: Effect of using the "NS" in WordNet for QE on the retrieval effectiveness for the English topics in CLEF-IP 2010

	MAP		**PRES**	
	value	%change	value	%change
Baseline	0.1399	NA	0.486	NA
WordNet (NS)	0.1364	-2.5%	0.484	-1.0%

4. QUERY EXPANSION USING AUTOMATICALLY GENERATED SYNSET

The previous section found that QE using PRF and WordNet is not effective for patent search. PRF is characterized by its general applicability and language independency, but it showed a highly negative effect on the retrieval effectiveness. The WordNet approach showed insignificant change to the overall retrieval effectiveness, but a degree of improvement for some topics.

An alternative WordNet type expansion technique is proposed here based on an automatic method for generating synonyms or related words. The idea for automatically generating the synonyms set (SynSet) originates from the characteristics of the CLEF-IP patent collection, where some of the sections in some patents are translated into three languages (English, French, and German). The idea is to use these parallel manual translations to create possible synonyms sets. Although the idea was based on the presence of this data, this approach can potentially be applied to other kinds of IR applications when parallel multilingual corpora of a domain close to the data collection are available.

4.1 Generating the SynSet

Related work for automatically building a SynSet from a word-to-word translation model was presented in [17], where automatically generated synonyms were used in conjunction with WordNet and translation models to enhance cross language IR. In our approach, a word-to-word translation model is used to create a SynSet for QE in monolingual search. The main idea from using

[2] http://www.lemurproject.org/lemur/IndriQueryLanguage.php

parallel corpora is generating synonym sets from word translations. For a word in one language f which has possible translations to a set of words in another language $\{e_1, e_2 \dots e_n\}$, this set of words can be considered as synonyms or at least related to each other. The probability of e_1 to be a synonym of word e_2 can be computed using Equation 1.

$$p(e_1|e_2) = \sum_{i=1}^{n} p(f_i|e_2) \cdot p(e_1|f_i) \quad \textbf{(1)}$$

where $p(e_1|e_2)$ is the probability that e_1 is a synonym of e_2, $\{f_1, f_2 \dots f_n\}$ are possible translations for word e_2, $p(f_i|e_2)$ is the probability that f_i is a translation of e_2, and $p(e_1|f_i)$ is the probability that e_1 is a translation of f_i.

Automatic SynSets were created as follows:

- English and French translations for the 1.35M patents title and claims sections were extracted and aligned by sentences. Long claims where split at punctuation points to produce shortened aligned sentences. A set of 8M parallel sentences was extracted using this approach.
- Stopword removal was applied for the both languages[3].
- Words in both languages were stemmed using Snowball[4].
- GIZA++[5] was used for cross-language word alignment creating English to French and French to English dictionaries.
- Equation 1 was used to produce the SynSet for English terms.

The resulting SynSet contains a set of synonyms (related terms) for each term including the original term. Subjective analysis showed the SynSet to be reasonable, although containing some noisy terms with low probabilities. In order to reduce the number of noisy synonyms, pruning was applied removing all terms with low probability (less than 0.1), and adding their probabilities to the original term (Equation 2). This step was found to improve the retrieval effectiveness when using the SynSet for QE.

$$p(e_x|e_x)|_{pruned} = p(e_x|e_x)|_{original} + \sum_{\forall P(e_i|e_x)<0.1} p(e_i|e_x) \quad \textbf{(2)}$$

Applying Equation 2 led to many terms not having any synonyms other than themselves (i.e. $p(e_x|e_x) = 1$), which means that these terms has no expansion terms added when they appear in a query. A further pruning step was applied which removed SynSet entries for all terms that appeared less than 20 times in the 8M sentences training set, since these terms could not have enough training instances to produce a reliable SynSet. Some samples of the produced SynSets are shown in Table 4 (note that terms are in their stemmed form).

The generated SynSet was then used to expand the 1,348 queries from the CLEF-IP 2010 task and the resulting IR effectiveness observed.

Table 4: Sample of SynSet. Probabilities are between brackets

Term	SynSet
Motor	motor (0.63), engin (0.37)
weight	weight (0.86), wt (0.14)
Travel	travel (0.67), move (0.19), displac (0.14)
Color	color (0.56), colour (0.25), dye (0.19)
Cloth	fabric (0.36), cloth (0.3), garment (0.2), tissu (0.14)
Tube	tube (0.88), pipe (0.12)
Area	area (0.4), zone (0.23), region (0.2), surfac (0.17)

[3] http://members.unine.ch/jacques.savoy/clef/index.html
[4] http://snowball.tartarus.org/
[5] http://code.google.com/p/giza-pp/

4.2 Effect of SynSet on Retrieval Effectiveness

In order to test the effect of using the automatically generated SynSet on the retrieval effectiveness when used for patent QE, two experiments were conducted. The first one used the probability associated with the SynSet entries as a weight for each expanded term in the query (Wsynset). Therefore, each term was replaced with its SynSet entries with the probability of each item in the SynSet acting as a weight to the term within the query. The "#wsyn" operator in Indri query language was used to enable of the presence of weighted synonyms for terms in the query[2]. The second one neglected this associated probability and used uniform weighting for all synonyms of a given term (Usynset), this strategy is similar to adding synonyms from WordNet where no probability is assigned. Table 5 reports the retrieval results.

Table 5: Effect of using the SynSet for QE on the retrieval effectiveness for the English topics in CLEF-IP 2010

	MAP		PRES	
	value	%change	value	%change
Baseline	0.1399	NA	0.486	NA
Wsynset	0.1440	+2.9%	0.485	-0.7%
Usynset	0.1402	+0.2%	0.480	-1.7%

The results show the superiority of our new QE technique over PRF and WordNet. However, the impact of the SynSet technique was overall still not superior to the baseline. The results achieved when using the weighted SynSet method (Wsynset) were statistically better than the baseline when compared using MAP, but statistically worse than the baseline when compared using PRES. This result means that this technique achieved the opposite of what it was intended for, where it improved the precision and degraded the recall. For a recall-oriented task such as patent search, this result is considered a negative outcome. Nevertheless, this small benefit overall shows that there are topics which are improved, as well as others which are degraded or not changed.

Understanding situations where each of the QE techniques improves or degrades retrieval effectiveness is important if they are to be applied to improve patent search reliably.

5. ANALYSIS OF RESULTS

Unfortunately none of the QE techniques for prior-art patent search examined in this study task achieved overall superior results to the baseline. Here, the expansion techniques and their results are analyzed to seek to understand the circumstances in which they work and those where they fail. The goal is to find possible noticeable features that could be extracted to help in improving the results through understanding the reasons for success and failure. In this analysis, only the WordNet and the SynSets were analyzed since PRF introduced a very large degradation in retrieval effectiveness, whereas the other techniques led to insignificant average changes in the retrieval results with some instances of success and failure.

5.1 Number of Expansion Terms

Table 6 shows the average number of expansion terms added per query using the WordNet and SynSet methods. In addition, the average increase in the query size is reported for each run by calculating the ratio of query size after expansion to the original query size. Table 6 shows the percentage of expanded terms per query ranges from 57% of the terms in case of using the SynSet to 85% when using all expansion terms of WordNet (nouns and verbs, synonyms and hyponyms). This shows that most of the terms in the queries are enriched with additional terms that are related in meaning. Regarding the average increase in query size, the difference between SynSet and the runs of WordNet are shown clearly. For the SynSet method, there was an average increase in the query size of 60%, which does not slow down the retrieval process markedly. On average, only one term is added as an expansion synonym for 57% of the terms in queries. For WordNet, the number of terms added to the query was very large, where the size of query was increased 5 times when only the noun synonyms (NS) of the terms were considered for expansion, and when using all the noun/verb synonyms/hyponyms, the query size reached 34 times its original size. This remarkable increase in the query size led to a very large reduction in the speed of the retrieval process of more than 50 times compared to the baseline.

The conclusion from Table 6 is that WordNet is not an efficient method for QE for patent search making it unsuitable for consideration to potentially enhance retrieval effectiveness. Moreover, SynSet is the most efficient method for QE in patents among the methods explored here, namely WordNet and PRF.

Table 6: Statistics of number of terms added per query when using WordNet and SynSet QE methods for patent search

	$\frac{Expanded\ terms}{Total\ query\ terms}$	$\frac{Query\ size\|_{expansion}}{Query\ size\|_{original}}$
NS	76%	4.9
NS+VS	84%	9.5
NS+NH	80%	21
NS+NH+VS+VH	85%	33.9
SynSet	57%	1.6

5.2 Success and Failure per Topics

There was no significant improvement or degradation in the retrieval effectiveness when using WordNet (NS) or SynSet for QE. The results were reported for the full English topics set. Here, the numbers of topics that are improved, degraded, or unchanged are counted for each method. Since MAP is a much more sensitive metric than PRES, we assume that a change in the score of 5% and 1% to be classified as a noticeable change for MAP and PRES respectively. Figure 1 shows the number of topics which were improved, degraded, or unchanged for QE techniques using WordNet (NS), Usynset, and Wsynset.

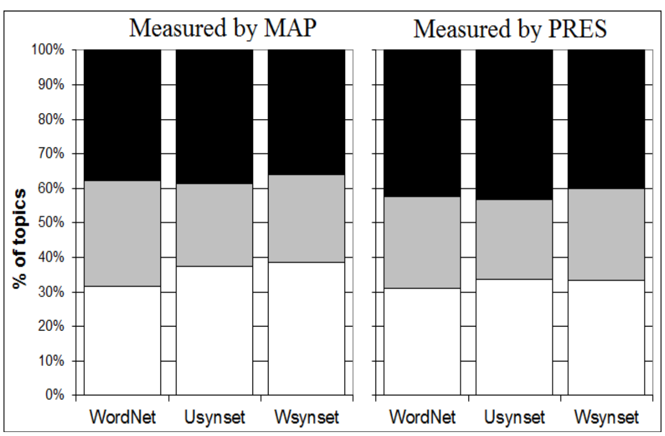

Figure 1: The number of topics which improved (white), degraded (black), or unchanged (gray) when applying different QE techniques to patent search measured using MAP and PRES

From Figure 1, it can be seen that the Wsynset approach led to the highest number of topics improved and correspondingly the lowest number of topics degraded compared to the Usynset and WordNet approaches. Nonetheless, the number of topics improved is always less than the number of topics degraded when compared using PRES. These results indicate that methods have potential. However, the key challenge for further work is to explore when to apply one of the techniques, even if this selection has to be done manually. This issue of when to apply an unreliable QE method is a well-known challenge of research into PRF for other search tasks [2,3,9].

5.3 Features of Expansion Success or Failure

In this section, an experiment is reported which extracts features for the topics which improved or degraded for later use as evidence to decide automatically when to apply QE.

There is some reported work on training classifiers to decide when and when not to apply QE expansion to a topic [3,9]. Most of this work is in the area of query performance prediction, which seeks to predict the initial performance of a user's query. Using this approach, different query processing is applied depending on the prediction of performance for the query [3]. Applying this approach to patent topics, we noticed that the QE behaviour has no correlation to the performance of the initial baseline. Table 7 shows some sample topics which were improved or degraded when comparing the PRES value before and after expansion using Wsynset as the expansion method. As shown clearly, there is no correlation between the initial value of the baseline and the expected performance of the QE. Hence, query performance prediction of this type cannot be used here.

Since initial query performance does not provide a useful indication of the effectiveness of the QE methods, different features need to be extracted as an alternative to explore possible combinations of features to determine whether the QE will improve or degrade retrieval performance for a query. For this investigation features were extracted based on: length of query before and after expansion, ratio between query length before and after expansion, average document frequency (DF) of terms in the query, and percentage of change in average DF of terms before and after expansion. These sets of features were calculated twice: based on all query terms and based only on unique query terms. A total of 15 features were extracted from examples of the experimental topic set through selecting the topics which were improved and degraded (Figure 1) as positive and negative examples respectively. The distributions of each of the features were plotted for positive and negative examples. This analysis was not able to identify features which were discriminative enough, since the distributions of features for the positive and negative examples were almost entirely overlapping. To further investigate these features, a support vector machine (SVM) was trained using 80% of the topic set and tested on the remainder. The output of the SVM is a binary decision, where the objective was to predict whether or not to apply the QE technique for the query. Only the topics of the Wsynset run, of which some change occurred for their PRES values, were used for training and testing. Unfortunately, the SVM was also not able to reliably predict when to apply the QE processes. Actually some of the runs could not complete the SVM training, since the positive and negative examples for some of the features were inseparable. These results illustrate the challenge of determining when to apply QE in patent search, since there is no easily available set of features to predict the success of QE for individual queries.

Table 7: Sample topics which were improved (LHS) or degraded (RHS) by expansion using Wsynsets based on change in PRES

Topic ID	Baseline	Wsynset	%change
PAC-1704	0.000	0.174	+∞
PAC-195	0.000	0.215	+∞
PAC-1225	0.105	0.532	+408%
PAC-1670	0.124	0.637	+415%
PAC-954	0.514	0.763	+48%
PAC-122	0.590	0.944	+60%
PAC-579	0.630	0.902	+43%
PAC-1113	0.669	0.880	+32%

Topic ID	Baseline	Wsynset	%change
PAC-1510	0.030	0.012	-60%
PAC-210	0.160	0.000	-100%
PAC-220	0.201	0.000	-100%
PAC-56	0.263	0.040	-85%
PAC-784	0.323	0.027	-92%
PAC-42	0.459	0.216	-53%
PAC-906	0.571	0.214	-63%
PAC-1498	0.662	0.307	-54%

5.4 Recommendation for Usage

Although the overall retrieval of Wsynset was statistically better when compared by MAP, it was statistically worse when compared by PRES, which is against our main objective in patent search. The previous analysis showed that Wsynset is not a fully reliable approach for expanding patent queries in order to achieve overall better retrieval results. However, the experiments showed it to be the most effective of the QE techniques investigated in this study, and the most efficient one. Furthermore, the SynSet approach is general and language independent, and can thus be applied to any language pair as long as a suitable parallel corpus is available. Our analysis showed that the technique works for a good portion of the patent topics, however, our trials failed to be able to automatically enable/disable the application of the expansion for the cases when it is likely to be effective.

Our recommendation for the usage of QE using SynSets is to apply it on demand by the user (patent examiner), since it can improve the retrieval effectiveness for some topics, even if the initial retrieval was good. However, this does not eliminate the importance of further investigation of how QE might be made more effective automatically for larger numbers of patent topics. In addition, SynSets may be usefully exploited as a lexical resource for use directly by a patent examiner to suggest possible related terms when they are manually formulating a search query.

6. Conclusion

This paper has presented a study of three approaches to QE for the prior-art patent search task. We confirmed previous results that PRF is not effective in patent search tasks. We investigated the use of two resources of synonyms for QE. The first used WordNet, and the other used a set of automatically generated synonyms (SynSet). Unfortunately, neither of these techniques led to a superior overall improvement in retrieval effectiveness. Query by query analysis was not able to identify situations where QE would succeed or fail. Nonetheless, we showed than the SynSet method is the most effective of QE approaches investigated. Moreover, it is the most efficient of the approaches examined and is language independent, since it can be applied automatically as long as a parallel corpus is available.

For future investigation, additional analysis of the success and failure of SynSet should be applied. Also, further pruning methods in SynSet creation could be explored, since there may be some terms that degrade retrieval effectiveness when used for expansion which could be eliminated using alternative pruning methods. Finally, this approach should be applied on real patent examiners queries that are formulated manually, which to the best of our knowledge are not available yet for research.

7. ACKNOWLEDGEMENT

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (CNGL) project.

8. REFERENCES

1. Bashir S. and A. Rauber. Improving Retrievability of Patents in Prior-Art Search. In *Proceedings of ECIR 2010*, (2010)
2. Cao G., J.-Y. Nie, J. Gao, and S. Robertson. Selecting Good Expansion Terms for Pseudo-Relevance Feedback. In *Proceedings of SIGIR'08*, (2008)
3. Cronen-Townsend S., Y. Zhou, and W. B. Croft. Predicting Query Performance. In *Proceedings of SIGIR'0*2, (2002)
4. Itoh H. NTCIR-4 Patent Retrieval Experiments at RICOH. In *Proceedings of NTCIR-4*, (2004)
5. Kishida K. Experiments on Psuedo Relevance Feedback Method Using Taylor Formula at NTCIR-3 Patent Retrieval Task. In *Proceedings of NTCIR-3*, (2003)
6. Konishi K. Query Terms Extraction from Patent Document for Invalidity Search. In *Proceedings of NTCIR-5*, (2005)
7. Lavrenko, V. and Croft, W.B. Relevance-Based Language Models. In *Proceedings of SIGIR '01*, (2001)
8. Liu S., F. Liu, C. Yu, and W. Meng. An Effective Approach to Document Retrieval via Utilizing WordNet and Recognizing Phrases. In *Proceedings of SIGIR'04*, (2004)
9. Lv Y. and C. Zhai. Adaptive Relevance Feedback in Information Retrieval. In *Proceedings of CIKM'09*, (2009)
10. Magdy W., J. Leveling, and G. J. F. Jones. Exploring Structured Documents and Query Formulation Techniques for Patent Retrieval. In *Proceedings of CLEF 2009*, (2009)
11. Magdy W. and G. J. F. Jones. PRES: A Score Metric for Evaluating Recall-Oriented Information Retrieval Applications. In *Proceedings of SIGIR'09*, (2010)
12. Magdy W., P. Lopez, and G. J. F. Jones. Simple vs. Sophisticated Approaches for Patent Prior-Art Search. In *Proceedings of ECIR 2011*, (2011)
13. Piroi F. CLEF-IP 2010: Retrieval Experiments in the Intellectual Property Domain. In *Proceedings of the CLEF 2010*, (2010)
14. Rocchio J. The SMART retrieval system: Experiments in Automatic Document Processing. In *Relevance Feedback in Iinformation Retrieval*, (1971)
15. Takeuchi H., N. Uramoto, and K. Takeda. Experiments on Patent Retrieval at NTCIR-5 Workshop. In *Proceedings of NTCIR-5*, (2005)
16. Voorhees E. M. Using WordNet for Text Retrieval. In *WordNet, an Electronic Lexical Database, MIT Press*, (1998)
17. Wang J., D. W. Oard. Combining Bidirectional Translation and Synonymy for Cross-Language Information Retrieval. In *Proceedings of SIGIR'06*, (2006)