

Using query logs of USPTO patent examiners for automatic query expansion in patent searching

Wolfgang Tannebaum · Andreas Rauber

Received: 2 May 2013 / Accepted: 11 February 2014 / Published online: 22 February 2014
© Springer Science+Business Media New York 2014

Abstract In the patent domain significant efforts are invested to assist researchers in formulating better queries, preferably via automated query expansion. Currently, automatic query expansion in patent search is mostly limited to computing co-occurring terms for the searchable features of the invention. Additional query terms are extracted automatically from patent documents based on entropy measures. Learning synonyms in the patent domain for automatic query expansion has been a difficult task. No dedicated sources providing synonyms for the patent domain, such as patent domain specific lexica or thesauri, are available. In this paper we focus on the highly professional search setting of patent examiners. In particular, we use query logs to learn synonyms for the patent domain. For automatic query expansion, we create term networks based on the query logs specifically for several USPTO patent classes. Experiments show good performance in automatic query expansion using these automatically generated term networks. Specifically, with a larger number of query logs for a specific patent US class available the performance of the learned term networks increases.

Keywords Patent searching · Query expansion · Query log analysis

1 Introduction

In preparing a patent application or judging the validity of a patent applied for based on novelty and inventiveness, an essential task is searching patent databases for related patents

W. Tannebaum (✉) · A. Rauber
Institute of Software Technology and Interactive Systems, Vienna University of Technology,
Vienna, Austria
e-mail: tannebaum@ifs.tuwien.ac.at
URL: <http://www.ifs.tuwien.ac.at>

A. Rauber
e-mail: rauber@ifs.tuwien.ac.at

that may invalidate the patent. Patent searching is usually performed by examiners in a patent office and patent searchers in private companies (Alberts et al. 2011). For searching patent databases the patent searchers follow a strict scheme. They compartmentalize the invention into searchable features and expand the features using synonyms, equivalents and co-occurring terms, using the concept of an invention diagram. To narrow the search topic, query terms are specialized into keyword phrases (Hunt and Nyugen 2007). Table 1 shows an example of a diagram including the features of the invention completed with expansion terms as they are used for query generation by the patent searchers.

The first column includes the searchable features of the invention selected from the source document, particularly from a patent document or an invention report. The second column provides the corresponding expansion terms. The terms are synonyms or equivalents, such as “screen” for “display”, co-occur in the source document, for example “signal” with “transmitter”, or limit a feature of the invention to a keyword phrase, such as “control module” for “module”. Particularly for finding synonyms to the searchable features of the invention, there is an increasing need to assist patent searchers as this process is very time-intensive and the probability to miss relevant expansion terms is high (Azzopardi et al. 2010; Fujita 2007; Hunt and Nyugen 2007). Yet, no sources providing synonyms, such as patent domain specific lexica or thesauri, are available.

The goal of this paper is to learn synonyms for the patent domain to enhance patent searchers in query generation, particularly in query expansion. Particularly, patent searchers shall be allowed to formulate queries and to evaluate patent applications in the same way as patent examiners do it. To this end, we learn domain-specific term networks from the query logs created by patent examiners as part of the application validation procedure.

The remainder of the paper is organized as follows. We first review related work on automatic query expansion in patent searching and enhancing query generation using query logs. In Sect. 3 we present the structure and characteristics of query logs of USPTO patent examiners. We then present our approach to learn synonyms from the query logs and the term networks each learned for a specific US patent class in Sect. 4. Experiments based on query expansion done by patent examiners are provided in Sect. 5, followed by conclusions and an outlook on future work in Sect. 6.

2 Related work

2.1 Automatic query expansion in patent searching

Currently, automatic query expansion in patent search is mostly limited to computing co-occurring terms for the searchable features of the invention (Konishi 2005; Mahdabi et al. 2011; Magdy and Jones 2011). Additional query terms are extracted automatically from the query documents, the feedback documents or from the cited documents based on statistical measures, such as term frequencies (tf) and a combination of term frequencies and inverted document frequencies (tfidf), or from the translations of the claim sections (Jochim et al. 2011; Konishi 2005; Mahdabi and Crestani 2011; Magdy and Jones 2011; Russo 2011; Xue and Croft 2009).

Also whole documents or whole sections of the query documents, like the title, abstract, description or the claim section are used for query generation and query expansion (Xue and Croft 2009).

Table 1 Invention diagram

Features	Expansion terms
Module	Control, terminal, computer, station, ...
Display	Screen, keypad, touchpad, interface, ...
Transmitter	Radio, infrared, ultrasonic, transmit, send, signal
Receiver	Radio, infrared, ultrasonic, receive, reception, ...

For conceptual search (searching by meanings rather than literal strings) in the patent domain, the international patent classification (IPC), in particular their categories and short descriptions, or standard dictionaries, such as *WordNet*, or lexica, like *Wikipedia*, are used for query refinement (Bashar and Myaeng 2011; Herbert et al. 2009). Especially, to provide synonyms for conceptual search, the related experiments for automatic query expansion commonly rely on the usage of the standard dictionaries and lexica (Bashar and Myaeng 2011). We assume, that these standard dictionaries will achieve only fair performance in the patent domain. In Sect. 5 we will reassess this assumption.

One approach computing synonyms from European Patent Office (EPO) patent collection is described in Jochim et al. (2010, 2011). The claim sections in English, German and French are aligned to extract translation relations for each language pair. Based on the language pairs having the same translation terms, synonyms are learned in English, French and German. We refined the approach to learn synonyms on specific claim terms, identified as the subject features of the inventions (Tannebaum and Rauber 2012b). We used the reference signs appearing next to these specific claim terms in the granted patents of the EPO. We assembled bigrams learned from the claim sections to learn translation relations from the claim sections. From these we learned synonyms based on the same translation terms as shown in (Tannebaum and Rauber 2012b) but for specific claim terms. Further, our approach avoids mistakes through incorrect word alignment. The limited collection of documents available for learning patent class-specific term networks prove to be disadvantage. Since 2003, about forty thousand granted patent documents including the translations of the claim sections have been published by the EPO each year. Hence, in view of more than 600 subclasses of the cooperative patent classification (CPC) classification scheme of the EPO, the document collection will be too small and finally inappropriate for learning patent class specific term networks.

2.2 Enhancing query generation based on query logs

Query logs are being intensively studied in many information retrieval settings, specifically for web searchers (Silvestri 2010). The main focus is on the analysis of the queries to enhance searches (Amitay and Broder 2008; Clough and Berendt 2009; Kato et al. 2013). In previous research for learning lexical term networks using query logs, terms are extracted directly from the query log collection. Relations, specifically synonym relations, are generally learned by using external sources, such as lexica, glossaries or databases like *WordNet* (Sekine and Suzuki 2007; Zhang et al. 2006). Further relations are retrieved by analyzing the retrieved, particularly the clicked, documents. If two queries are related with the same document, these two queries are associated with each other (Hang et al. 2002; Kunpeng et al. 2009). All approaches to find synonyms depend on external sources such as lexica or glossaries. These approaches do not utilize relations between the query terms in

the query logs. Yet, we need these and can use these for learning synonyms in the patent domain due to the radically different setting in which patent searches are performed. Contrary to conventional query logs, where searchers are usually one-step single query events, patent search sessions extend over many queries that are gradually refined, and that rely more heavily on the use of synonyms to ensure coverage, as patent applicants are permitted to be their own lexicographers, i.e. they can define their own terminology.

Finding query logs in the patent domain has been a difficult task due to the lack of publicly available logs (Jürgens et al. 2012). Private companies and searchers are not interested in making their logs that will usually include or hint at features of their inventions available. The USPTO is the only source known to us which publishes the query logs of patent examiners. In (De Marco 2011) a detailed analysis of individual the USPTO patent examiners query logs is presented to reveal search strategies and potential shortcomings/limitations. In (Tannebaum and Rauber 2012a, b) we analyzed the basic characteristics of USPTO patent examiners' query logs. We manually downloaded a limited set of query logs (346 log files) of USPTO patent examiners, particularly for one specific patent domain from the US Patent and Trademark Office Portal PAIR. Initial results indicated that specialized term networks can be extracted directly from query logs to complement resources for standard English using the domain specificity of patents and the extensive classification scheme they are structured in. This has positive effects on automated query expansion in patent searching.

In this paper we radically extended these initial studies, proving a method to create valuable domain-specific term networks to assist in query expansion in this highly professional search setting. We collected and preprocessed a significantly larger corpus of patent query logs, facilitating more in-depth studies and providing reasonably comprehensive term networks.

3 Query logs of the USPTO

The query logs of USPTO patent examiners called “Examiner’s search strategy and results” are published for most patent applications since 2003 by the US Patent and Trademark Office Portal PAIR (Patent Application Information Retrieval).¹ The download is limited by the USPTO. For each patent application a verification code has to be entered. Google has begun crawling the USPTO’s public PAIR sites and provides free download of the patent applications.² Since, April 2013 Reed Technology a contractor to the USPTO undertakes this task and hosts the data.³ For each patent application a single zip file is available containing several folders including information such as: Address and Attorney/Agent, Application Data, Continuity Data, Foreign Priority, Image File Wrapper, Patent Term Adjustments, Patent Term Extension History and Transaction History. The Image File Wrapper is of concern to us here. This folder can contain one or multiple query log files.

Each query log of the USPTO is a PDF file consisting of a series of queries. Figure 1 shows an example of such a query log. Each query has several elements. We focus on the search query element showing the query formulated by the patent examiner. Further elements are: reference, hits, database(s), default operator, plurals, and time stamp.

¹ <http://www.uspto.gov/>.

² <http://www.google.com/googlebooks/uspto-patents.html>.

³ <http://patents.reedtech.com/Public-PAIR.php>.

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S1	11759	(leadframe or (lead adj frame) or foil) with (plastic adj (film or layer))	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/02/16 20:11
S2	2952	(leadframe or (lead adj frame) or foil) with (diode or photodiode)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/02/16 20:13
S3	12	S1 and S2	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/02/16 20:13
S4	6	@ad <= "20030604" and S3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/02/16 20:49

Fig. 1 Example of a USPTO query log

There are several kinds of queries in the search query element as shown in Fig. 1. Text queries, such as queries *S1* and *S2*, are used for querying whole documents (fulltext search) or only sections of patent documents, such as the title section (title search) using query terms. Non-text queries are used for searching patent document numbers, classifications, or application and publications dates. For example, the non-text query “@ad ≤ 20030604” is used for searching patent documents applied before 4th June of 2003, as shown in query *S4*. A further kind of query is the reference query, such as query *S3*, which is a combination of earlier queries, in particular of query *S1* and *S2*, i.e. re-using the terms of a previous query and expanding it with further elements, thus avoiding to have to re-type an earlier query. Text queries include search operators between the query terms. The types of search operators are (1) Boolean operators, such as “AND or OR” and (2) Proximity operators, like “SAME, ADJ(acent), NEAR, or WITH”. Furthermore, Truncation Limiters, such as “\$”, are used for query formulation. If the search operators are added manually, they are shown between the query terms in the text query element, else they are indicated by the default operator element. For our purposes we are specifically interested in the queries including the Boolean operator “OR” to learn synonyms.

4 Learning term networks

4.1 Experiment set up

The USPTO published about 2.7 million patent applications, since 2003. The applications are classified into 473 classes each including hundreds or thousands of subclasses. Hence, on average, about 6,000 application documents are available for each class. Because patent searchers use the classification system to narrow the search, we selected fifteen classes for our experiments. At first we collected all application numbers of the published patent

applications for the fifteen classes and generated a list of download links for each class based on the download URL “http://storage.googleapis.com/uspto-pair/applications/APP_NUM.zip”, where we replaced “APP_NUM” in the URL with the application numbers. Secondly, we harvested the zip files via *Wget*⁴ a free software package for retrieving files from web servers, unzipped and filtered the files using the file name ending “SRNT.pdf” of the query log files. Following, we carried out OCR conversion using *ABCocr*⁵ a product to extract text from images on a Windows 7 platform and converted the PDF files to TXT files. Subsequently, all terms were fed into the extraction process.

Overall, we downloaded and preprocessed 103,896 query logs available for the fifteen US classes, making it the largest collection of query logs used for experiments in the patent IR domain. Table 2 shows the number and title of the selected classes and the number of downloaded query logs for each class.

As shown in Table 2 the number of query logs for the classes differs between 1,820 and 16,864 files. For our experiments, particularly for learning the class-specific term networks and for evaluating them, we conceptually grouped the classes according to their size: (*small*) having less than 4,000 query logs (*medium*) having up to 8,000 files and (*large*) having more than 8,000 logs. The grouping allows us to assess, on how far the performance of class specific term networks depend on the class size (number of query logs) and whether a minimum number of query logs is needed to achieve accurate performance in automatic query expansion. Furthermore, we selected some classes that are topically related (e.g. classes 384 and 148; or classes 128, 433 and 623 from the medical domain) as well as completely disjunct classes to evaluate in how far term networks can be learned on a more generic level, using the hierarchical relationship between classes.

4.2 Extracting term networks based on synonym detection

In patent searching the Boolean Operator “OR” is used to expand a query term with an expansion term, which has the same meaning, such as “drill” for “burr” or “tool” for “instrument” in the medical domain concerning dentistry equipment. We use that for automatically detecting synonyms in the query logs based on the Boolean operator “OR”, which indicates that two query terms are synonyms, or can at least be considered as equivalents. The process works as follows: We extract all text queries including the search operators between the query terms from the query log collection. We then filter all 3-g generated from the text queries in the form “X b Y”, where *b* is the Boolean operator “OR” and X and Y are query terms. In addition, to exclude mismatches and misspellings and for ranking of the extracted synonyms according to their support in the specific classes, in particular for suggesting initially the synonyms having the highest support followed by additional terms that have a lower support later-on, we utilize a confidence value *CV*. We measure the frequency of each synonym in the specific class. Table 3 shows for each class the number of extracted synonyms based on the confidence values CV_1 to CV_5 , i.e. that have a support greater than or equal to 1–5.

As shown in Table 3, in particular for CV_1 , for each class the number of unique synonyms extracted from the query logs increases with the size of the query log collection. The highest number of synonym relations (36,366 relations) could be extracted from the *large* class 422 for “Chemical apparatus and process disinfecting, preserving, or sterilizing”. Furthermore, using the CLAWS part of speech tagger for English terms (Garside and

⁴ <http://www.gnu.org/software/wget/>.

⁵ <http://www.websupergoo.com/abcocr-1.htm>.

Table 2 Experiment set up

Nr.	US class	Title	#Query logs
1	454	Ventilation	1,820
2	384	Bearings	1,901
3	126	Stoves and furnaces	2,720
4	148	Metal treatment	2,877
5	219	Electric heating	3,926
6	433	Dentistry	4,025
8	180	Motor vehicles	5,205
7	417	Pumps	5,423
9	398	Optical communications	6,028
10	280	Land vehicles	7,905
11	128	Surgery	8,757
12	379	Telephonic communications	9,897
13	422	Chemical apparatus and process disinfecting, preserving, or sterilizing	11,842
14	439	Electrical connectors	14,706
15	623	Prosthesis (i.e. artificial body members)	16,864
Σ	–	–	103,896

Table 3 Number of extracted synonyms based on confidence values CV_1 to CV_5

US class	CV_1	CV_2	CV_3	CV_4	CV_5
454	3,126	826	383	215	135
384	1,778	525	297	190	136
126	4,155	1,358	675	418	276
148	6,103	2,903	1,813	1,312	980
219	6,323	2,729	1,582	1,103	812
433	7,441	3,280	1,713	1,184	851
180	7,682	2,998	1,675	1,067	711
417	6,634	2,345	1,251	774	541
398	9,165	4,209	2,501	1,762	1,280
280	7,009	2,377	1,279	776	524
128	18,106	7,765	4,168	2,733	1,826
379	22,780	11,021	6,388	4,369	3,065
422	36,366	14,055	7,776	5,288	3,816
439	10,129	3,967	2,157	1,452	989
623	31,693	11,534	6,385	4,203	2,771
Σ	159,713	64,750	36,268	24,488	17,112

Total number of extracted synonyms based on CV_2 (64,750) set in bold

Smith 1997) we identified the synonyms based on the confidence value CV_1 w.r.t. part of speech and find out that more than half of the terms are nouns (69.61 %) followed by adjectives (15.53 %) and verbs (14.87 %).

As expected, we notice for all classes a considerable decrease of the number of synonyms having a support between 2 and 5. Because patent searching is a recall orientated

task, we consider those synonyms that were encountered at least two times as synonyms (CV_2) in the specific class to learn the term networks. This reduces spurious mismatches, but provides as many synonyms for automatic query term suggestion as possible. Table 4, shows the learned term networks which resemble thesauri of English concepts.

The thesauri provide English synonyms for each specific patent class. In each term network terms that have the same meaning are linked to each other. In total, the learned term networks provide 64,750 unique synonym relations based on 36,601 unique query terms. Finally, the learned term networks can be used, particularly in each specific patent US class, for (semi-) automated query suggestion, particularly query expansion.

5 Automatic query expansion

In this section we use the term networks for automatic query expansion in patent searching. For the evaluation we expand query terms of queries from real query sessions of patent examiners (gold standard). Based on the suggested terms from the term networks and the expansion terms used by the examiners, we calculate recall and precision scores to evaluate the performance of the learned term networks.

For each class we split the query log collection in a training set and a test set for evaluation. The training set is further divided into several sub-sets to learn multiple term networks for each class to evaluate size and time dependency characteristics. Specifically, having the query logs ordered by time of application of the patent, we use the first set of query logs of each class for training, with the test set being created from the chronologically last set of query logs in each class. For each class we generate up to five training sets for learning the class specific term networks (*TS1* to *TS5*). The size of these sub-sets depends on the class size. For the five classes having less than 4,000 query logs (grouped as *small*) we learn 17 term networks based on training sets having between 500 and 2,500 query logs in increments of 500. For the *medium* grouped classes we learn 20 term networks having between 3,000 and 5,000 query logs. Finally, for the *large* classes we generate 21 term networks based on training sets having between 6,000 and 10,000 log files. In total, for all classes we learn 59 term networks based on specific class and training set size. Table 5 shows the generated training sets used for learning the class specific term networks.

Furthermore, we learn a class-independent term network, which we call *LogNet*. For this we use the largest training sets of each specific class. Table 4 shows the selected training sets in bold. This network is still domain-specific in the sense that it is based on patent query logs, yet it stretches across class boundaries and is thus less specific.

To evaluate the lexical term networks based on recall and precision of the suggested expansion terms, we query the synonyms from the test sets. Particularly, to calculate the recall scores, we compare the suggested terms from the term networks with the synonym or equivalent terms from the test sets, which were used by the patent examiners for searching. To compute precision we compare the synonyms used by the examiners in the test sets with all expansion terms suggested by the term networks.

5.1 Query expansion based on training set size

First we use the specific term networks *TS1* to *TS5* of each class and the class-independent term network *LogNet* to evaluate the performance of the term networks based on the size of the training sets and for each class. For calculating the recall and precision measures we

Table 4 Learned term networks

Term networks	Unique relations	Unique terms
Ventilation	826	1,078
Bearings	525	692
Stoves and furnaces	1,358	1,508
Metal treatment	2,903	2,770
Electric heating	2,729	2,799
Dentistry	3,280	3,578
Motor vehicles	2,998	2,962
Pumps	2,345	2,435
Optical communications	4,209	3,499
Land vehicles	2,377	2,450
Surgery	7,765	7,633
Telephonic communications	11,021	7,386
Chemical apparatus and process disinfecting, preserving, or sterilizing	14,055	10,609
Electrical connectors	3,967	3,462
Prosthesis (i.e. artificial body members)	11,534	9,496
Σ	64,750	36,601

Table 5 Test and training sets, sets in bold were used for *LogNet*

US class	Test sets	TS1	TS2	TS3	TS4	TS5
454	500	500	1,000	–	–	–
384	500	500	1,000	–	–	–
126	500	500	1,000	1,500	2,000	–
148	500	500	1,000	1,500	2,000	–
433	500	500	1,000	1,500	2,000	2,500
219	500	3,000	3,500	–	–	–
180	500	3,000	3,500	4,000	4,500	–
417	500	3,000	3,500	4,000	4,500	–
398	500	3,000	3,500	4,000	4,500	5,000
280	500	3,000	3,500	4,000	4,500	5,000
128	500	6,000	7,000	8,000	–	–
379	500	6,000	7,000	8,000	9,000	–
422	500	6,000	7,000	8,000	9,000	10,000
439	500	6,000	7,000	8,000	9,000	10,000
623	500	6,000	7,000	8,000	9,000	10,000

Bold values indicate the best performing training sets for each class

excluded synonyms in the test sets which are out of the vocabulary of the term networks, i.e. terms that did not appear in any earlier query log. Table 5 shows the achieved recall and precision measures.

As shown in Table 6, for almost all classes the recall measures increase with the increase in training set size. Specifically, we can assume that the recall scores will further increase with even larger training sets. Because we excluded synonyms that are out of the

vocabulary, in some cases, particularly for classes 180 and 417, with larger training sets the recall scores go down. The reason for that is, with the larger training sets more synonyms and equivalent terms appear in the term networks (the terms are not out of vocabulary any more) but not necessarily as synonyms. Best recall measures are provided, on average, by the term networks learned from the training sets of the *large* US classes with a size larger than 6,000 query logs (with one exception, class 128). In particular, best recall is provided by the term Network *TS5* learned for the class 623. The term network *TS5* provides with a recall of 73.33 %, on average, 7 of 10 synonyms, which are used by the patent examiners for query expansion.

The precision values show, that with increasing training set sizes the achieved precision scores decrease as the number of suggested synonyms increases. The term network *TS3* learned for class 128 provides with a score of 44.17 % best precision. On average, 4 out of 10 terms that are suggested by the term network as synonyms were actually used by the examiners for query expansion. Considering the term network providing the best recall performance (*TS5* for class 623) on average only 2 out of 10 terms suggested are used by the patent examiners for query expansion. Note that the lower precision may not be a serious impediment for deployment of the query term expansion: as patent search is recall-oriented rather than precision-oriented, i.e. preferring a higher number of potentially irrelevant documents in a result set over a more limited result set missing relevant documents, especially when suggested by a system for manual deployment rather than performed in a fully autonomous manner, may be assistive rather than harmful. Furthermore, precision is likely to be under-estimated, as the fact that certain suggested expansion terms are incorrect rather than potentially useful but having been not thought of by the searcher would need to be confirmed by expert searchers. This would help to determine whether the terms suggested but not used in the original patent verification search are actually wrong, or whether the examiner performing the validation search simply did not think them.

Furthermore, we evaluate the class-independent term network *LogNet*. In almost all classes the recall measures of the class-specific networks *TS1* to *TS5* are further improved. In particular, best recall is provided for class 417 with a recall of 78 %. As explained above, in two cases the recall decreases because of the synonyms that appear in *LogNet*, but not as synonyms. Mirroring the trend observed with the class-specific term networks, *LogNet* achieves only weak precision measures across all classes, peaking at 7.14 % for class 128.

In addition, we measure coverage of the respective networks by determining the number of out-of-vocabulary words, i.e. expansion terms that were used later in time that the network could not learn. This provides an indication on the comprehensiveness of the expansion term suggested. Table 7 shows the coverage of the class-specific term networks and *LogNet*.

With the increasing class size the coverage of the networks obviously increases. Best coverage scores of the class-specific term networks are provided by the term networks learned from the *large* classes. In particular, the term network learned for class 379 provides 81.48 % of the query terms from the test set. For all classes, on average, the class-independent term network *LogNet* provides a coverage of 87.90 %.

The experiments show that for almost all classes the recall and coverage measures of the class-specific term networks rise with the class size and can be further improved using the class-independent term network *LogNet*. On the other hand, the class-specific term networks achieve much better precision scores than *LogNet*. In this case, query terms are expanded in a certain context. To provide term networks for automatic query term

Table 6 Query expansion based on training set size and class

<i>US class</i>	Recall						Precision					
	<i>TS1</i>	<i>TS2</i>	<i>TS3</i>	<i>TS4</i>	<i>TS5</i>	<i>LogNet</i>	<i>TS1</i>	<i>TS2</i>	<i>TS3</i>	<i>TS4</i>	<i>TS5</i>	<i>LogNet</i>
454	45.83	48.39	–	–	–	54.10	20.00	21.13	–	–	–	1.88
384	18.18	26.09	–	–	–	58.21	25.00	39.58	–	–	–	5.68
126	29.41	29.63	30.30	32.56	–	51.43	31.50	17.68	14.76	10.74	–	3.32
148	35.16	39.22	39.62	46.28	–	61.39	22.01	17.23	13.09	18.78	–	5.82
219	40.74	41.46	36.21	35.71	40.54	66.67	40.13	18.95	21.88	19.70	21.84	2.74
433	43.75	45.31	–	–	–	59.60	10.15	10.51	–	–	–	2.33
180	68.00	67.31	63.16	62.07	–	47.42	13.34	13.07	12.60	12.24	–	5.35
417	54.17	60.00	64.00	62.96	–	78.00	12.40	10.57	10.09	9.83	–	5.43
398	52.63	50.85	50.85	51.67	56.67	59.21	12.94	12.37	11.58	10.90	10.70	5.63
280	50.00	44.00	46.15	48.28	51.72	56.52	12.74	11.38	14.33	16.66	12.24	4.48
128	30.06	29.35	32.09	–	–	45.21	21.64	20.30	44.17	–	–	7.14
379	60.76	64.89	66.91	67.91	–	72.26	7.98	6.64	5.62	5.66	–	4.15
422	56.88	58.56	56.52	58.26	59.83	70.00	8.25	8.27	7.79	7.06	7.22	6.29
439	60.00	57.14	59.10	59.10	59.10	68.97	7.31	6.46	6.73	6.13	6.02	2.66
623	64.29	64.29	64.29	66.67	73.33	61.90	25.52	25.25	25.53	19.87	22.05	3.79

Bold values indicate the best performing training sets for each class

Table 7 Coverage provided by the class-specific term networks and *LogNet*

<i>US class</i>	Coverage					
	<i>TS1</i>	<i>TS2</i>	<i>TS3</i>	<i>TS4</i>	<i>TS5</i>	<i>LogNet</i>
454	42.45	52.83	–	–	–	90.57
384	23.76	44.55	–	–	–	96.04
126	39.82	48.67	55.75	64.60	–	95.58
148	34.62	45.45	49.65	56.64	–	80.42
219	27.08	43.75	55.21	61.46	62.50	92.08
433	65.88	66.47	–	–	–	90.59
180	51.47	53.43	56.37	56.86	–	82.35
417	62.89	65.98	67.01	68.04	–	91.75
398	65.49	60.90	67.61	69.01	69.72	85.21
280	51.04	54.17	56.26	61.46	61.46	88.54
128	62.86	65.15	66.60	–	–	82.57
379	78.31	78.84	80.42	81.48	–	84.13
422	75.41	77.46	78.69	78.69	79.51	85.25
439	65.57	68.85	72.13	72.13	72.13	85.25
623	73.95	73.95	74.79	75.63	75.63	88.24

Bold values indicate the best performing training sets for each class

suggestion achieving high recall/coverage and precision scores, either (1) the recall measures of the class-specific term networks or (2) the precision scores of *LogNet* have to be improved. We address this issue in the following experiments.

5.2 Using a confidence value for query expansion

To optimize precision values we re-run the experiments for the *large* classes which achieve highest recall and coverage scores but provide the lowest precision measures of the class-specific term networks utilizing the confidence values for the synonym relations. In this setting, we only considered those expansion terms that were encountered at least five times as synonyms in the training sets, i.e. that have a higher support for the respective mapping in the training set. The resulting scores are provided in Table 8.

While the recall achieved with this more limited term network obviously decreases considerably, we also observe a drastic increase in precision compared to the values provided in Table 6. For example, the term network *TS5* of class 422 provides with a recall of 59.83 %, on average, 6 of 10 synonyms, which are used by the patent examiners for query expansion. The corresponding limited term network provides with a value of 22.00 % a considerable lower recall score. But a drastic increase in precision can be observed. The limited term network achieving a precision score of 23.05 % drastically outperform the term network *TS5* providing a value of 7.22 %.

The experiments show, that we can use the confidence values to iteratively suggest an increasing number of expansion terms as the search evolves. This allows the system to strike a reasonable balance between increasingly higher recall/coverage by suggesting additional expansion terms that have a lower support in the training set at the cost of lower precision after having initially suggested the most likely, highest-precision expansion terms.

5.3 Query expansion across different US patent classes

In Sect. 5.2, we addressed the low precision performance of the term networks learned from *large* classes achieving highest recall and coverage scores. In this section we evaluate the performance of the class-specific term networks when used for patents from other classes. We assume that this will help to detect classes where cross-domain applications might be useful, in particular to improve recall and coverage of *small* classes providing accurate precision scores. We test the term networks *TS2* of the *small* and *medium* classes and the term networks *TS3* of the *large* classes across test sets from other classes without excluding out-of-vocabulary words. We have not considered out-of-vocabulary words, because our goal is to detect related class having the most common synonyms. Across different patent classes we cannot assume that expansion terms were used later in time. Table 9 shows the achieved recall measures.

As shown in Table 9 the learned class specific term networks achieve respectable recall measures across some class boundaries. In particular, the term network *TS3* learned for the US Class 128 called “Surgery” achieves a recall measure of 29.73 % for class 623 called “Prosthesis”. This hints at the fact that term networks learned from one class may, in fact, be applied across related classes. Yet, there is no guarantee that this will always work for each pair of related classes. For example the term network learned from query logs of class 148 called “Metal Treatment” achieves at 12.68 % a better recall measure for class 384 called “Bearings” than the corresponding term network learned from class 384 when applied to class 148 with a recall score of only 4.59 %. As known, “Bearings” are generally created from metal material. “Metal Treatment” is a common process for manufacturing bearings and popular in the class 384. Yet, “Bearings” may be irrelevant for treating metal and thus for the class 148 leading to a non-bijective relationship.

Table 8 Considering confidence values in the *large* classes

<i>US class</i>	Recall					Precision				
	<i>TS1</i>	<i>TS2</i>	<i>TS3</i>	<i>TS4</i>	<i>TS5</i>	<i>TS1</i>	<i>TS2</i>	<i>TS3</i>	<i>TS4</i>	<i>TS5</i>
128	10.74	12.36	14.57	–	–	13.55	9.87	25.66	–	–
379	21.57	24.11	25.89	26.39	–	23.94	19.87	16.77	16.86	–
422	13.76	16.22	17.40	21.70	22.00	45.65	41.22	32.80	25.55	23.05
439	20.51	20.51	23.08	23.08	23.08	33.06	19.42	11.69	7.27	15.29
623	22.21	23.56	23.56	25.31	31.55	34.33	35.56	32.43	23.31	18.99

Bold values indicate the best performing training sets for each class

Table 9 Recall measures achieved when using class-specific term network for other than the class they were based upon

Test sets	Training sets TS2				
	454	384	126	148	219
<i>Small</i>					
454	–	5.56	16.67	2.78	12.50
384	2.82	–	1.41	12.68	2.82
126	3.90	5.19	–	1.30	16.88
148	2.75	4.59	1.83	–	2.75
219	7.19	5.76	5.76	9.35	–
Test sets	Training sets TS2				
	433	180	417	398	280
<i>Medium</i>					
433	–	12.17	6.09	11.30	7.83
180	4.72	–	3.94	3.94	14.17
417	15.00	11.67	–	11.67	6.67
398	5.05	6.06	5.05	–	2.02
280	7.02	17.54	10.53	3.51	–
Test sets	Training sets TS3				
	128	379	422	439	623
<i>Large</i>					
128	–	8.96	19.36	2.89	9.54
379	11.24	–	8.28	7.69	4.14
422	20.12	2.44	–	3.66	7.93
439	23.08	2.56	15.38	–	17.95
623	29.73	2.70	18.92	2.70	–

Bold values indicate the best performing training sets for each class

Hence, the experiments show that for the classes marked in bold cross-domain applications seems promising. This can be used for expanding smaller sets of query logs of specific classes to learn improved term networks of related classes providing improved recall and coverage scores, i.e. for classes where few query logs are available.

5.4 Query expansion compared to *WordNet*

As mentioned in Sect. 2, most approaches use standard dictionaries for automatic query expansion, particularly for finding synonyms. In this section we evaluate the performance of our approach compared to the dictionary *WordNet* (Miller 1995).

In particular, we test the best performing class-specific term networks, the class-independent network *LogNet* and the dictionary *WordNet* based on the test sets generated for each specific class and used in the Sects. 5.1 and 5.2.

Again, we calculate the recall/coverage and precision scores based on the suggested terms from the various term networks and the expansion terms used by the examiners. For the expansion of the query terms we use all lexical relations included in the patent domain specific term networks and in *WordNet*. We will not consider the meaning of the query terms, as our main focus is on the recall score in automatic query term suggestion. Thus, *WordNet* should benefit from higher recall due to the large number of synonyms added without a potentially harmful limitation to specific word senses. We are aware, that the precision measures can be improved when considering the word senses. In spite of this rather defensive assumption, *LogNet* achieves better recall measures than the standard dictionary *WordNet* across all classes, as shown in Table 10.

For larger training sets the recall measures of the learned term network *LogNet* increases. For the large classes *LogNet* is learned from training sets having more than 6,000 query logs files and provides, on average, for all classes, best recall scores. Compared to *LogNet* and the best performing class-specific networks, *WordNet* achieves only low recall for all classes. A comparative performance is only achieved for class 454. Over all classes, *WordNet* provides, on average, only a recall of 22.06 %. Comparing the precision measures, *WordNet* achieves as expected, like *LogNet*, only weak precision across all classes, peaking at 5.49 % for class 398.

In addition, to see if the differences between the results of *WordNet* and the best performing query log based expansion method, in particular *LogNet*, were statistically relevant, we run a *t* test. The test allows us to conclude that there were a statistically significance ($p < 0.05$). With respect to the measures for each specific class achieved by *WordNet* and *LogNet*, the *t* test confirms that for all classes except class 280 (0.27), the differences are significant ($p < 0.05$).

Again, we measure coverage of the respective networks by determining the number of out-of-vocabulary words. Table 11 shows the vocabulary covered by the term networks. *WordNet* being the most comprehensive thesaurus provides best coverage followed by *LogNet*. Best coverage is provided by *WordNet* for the class 128 at 98.55 %. *LogNet* has the highest coverage for class 384 at 96.04 %.

Through the analysis of the failed synonym relations provided by *WordNet* we learn that (1) patent examiners expand class-specific query terms using general terms. For example, in the class 379 called “Telephonic communications” they expand the specific query term “cellphone” using the general expansion term “device”. A further example is the expansion of the class-specific term “camper” with the general term “vehicle”. Further (2) the examiners expand query terms w.r.t. part of speech, such as “burn” for “burning” or “coat” for “coating”. Furthermore (3) they relate terms, which have the same meaning in a

Table 10 Recall and precision values for query expansion based on *LogNet*, *WordNet* and the best performing class-specific network

US class	Recall			Precision		
	Class-specific Network	LogNet	WordNet	Class-specific network	LogNet	WordNet
454	48.39 (TS2)	54.10	43.06	21.13	1.88	1.60
384	26.09 (TS2)	58.21	22.54	39.58	5.68	4.49
126	32.56 (TS4)	51.43	27.27	10.74	3.32	1.96
148	46.28 (TS4)	61.39	27.52	18.78	5.82	1.96
219	40.54 (TS5)	66.67	32.37	21.84	2.74	1.91
433	45.31 (TS2)	59.60	19.13	10.51	2.33	1.12
180	68.00 (TS1)	47.42	19.69	13.34	5.35	1.00
417	64.00 (TS3)	78.00	16.67	10.09	5.43	2.89
398	56.67 (TS5)	59.21	16.16	10.70	5.63	5.49
280	51.72 (TS5)	56.52	33.33	12.24	4.48	1.18
128	32.09 (TS3)	45.21	26.30	44.17	7.14	1.80
379	67.91 (TS4)	72.26	15.98	5.66	4.15	1.71
422	59.83 (TS5)	70.00	15.85	7.22	6.29	2.50
439	59.10 (TS5)	68.97	30.77	6.02	2.66	2.67
623	73.33 (TS5)	61.90	17.57	22.05	3.79	1.75

Bold values indicate the best performing training sets for each class

Table 11 Coverage provided by the class-specific term networks, *LogNet* and *WordNet*

US class	Class-specific term networks	LogNet	WordNet
454	52.83 (TS2)	90.57	93.40
384	44.55 (TS2)	96.04	93.07
126	65.00 (TS4)	95.58	93.81
148	56.64 (TS4)	80.42	97.55
219	62.50 (TS5)	92.08	96.35
433	66.47 (TS2)	90.59	95.88
180	51.47 (TS1)	82.35	96.57
417	67.01 (TS3)	91.75	92.78
398	69.72 (TS5)	85.21	95.07
280	61.46 (TS5)	88.54	92.71
128	66.60 (TS3)	82.57	98.55
379	81.48 (TS4)	84.13	96.21
422	79.51 (TS5)	85.25	97.13
439	72.13 (TS5)	85.25	88.52
623	75.63 (TS5)	88.24	94.12

Bold values indicate the best performing training sets for each class

specific classes, such as “portable” for “handheld”, in particularly for the class 379 called “Telephonic communications”. Additionally, through analysis of the vocabulary, which is not covered by *WordNet*, we find out, that patent examiners (4) use popular trademarks, such as “iphone”, “ipad” or “blackberry” for query expansion. Further, (5) the patent

applicants are allowed to create their own terms, such as “pocketpc” for “notebook”, “watergas” for “steam” or “passcode” for “password”. Because of these highly specific expansions of query terms in the patent domain, standard dictionaries, such as *WordNet*, achieve only low performance. In these standard dictionaries, such patent domain specific vocabulary and relations are not included. But even these kinds of synonyms, equivalents and relations between the vocabulary are needed for automatic query expansion in the patent domain. Using our approach to learn term networks from the patent domain and directly from the query logs of patent examiners fulfills the requirements of this highly domain specific query expansion.

Finally, the experiments show, that the term networks learned directly from the patent domain, in particular *LogNet* and the best performing class-specific term networks, drastically outperform the general term network *WordNet*. The standard dictionary *WordNet* achieves for all US classes only low performance in recall. As expected, all term networks, the patent domain specific and the general term network *WordNet*, achieve low precision measures. The reason for that is, that we have not considered the meaning of the query terms. In Sects. 5.1 and 5.2, we show how to counter the low precision values of term networks when used for automatic query term suggestion by (1) suggesting class-specific expansion terms first, followed by additional expansion terms provided by the class-independent term network; or (2) by suggesting expansion terms provided by the class-independent term networks in the order of their support in the training set.

Hence, the experiments show that the query logs are valuable resources to learn term networks from the patent domain for query expansion achieving high recall and precision scores.

6 Conclusions and future work

In this paper we presented an approach to support query expansion in the domain of patent search. We used real query expansion sessions done by patent professionals to learn term networks for automatic query expansion. Several US class-specific term networks and a class-independent term network, referred to as *LogNet*, were learned from the query logs provided by the USPTO to capture synonym relations. We evaluated these term networks based on query expansion done by patent professionals in real sessions.

The experiments show for the class-specific term networks, that recall measures increase with the availability of a larger set of query logs. Best recall measures are provided, on average, by the term networks learned from training sets with a size larger than 6,000 query logs. The class-specific networks provide up to 7 out of 10 synonyms, which are used by the patent examiners for query expansion. We assume that the recall measures will further increase with the rise of the training set size. Because the USPTO publish new query logs regularly for each US class, the size of the collections rises automatically. This will be advantageous for our expansion approach.

Further, the experiments show that for almost all classes the recall measures of the class-specific term networks can be further improved using the class-independent term network *LogNet*. *LogNet* suggests, on average, up to 8 out of 10 synonyms, which are used by the examiners for query expansion. Expectedly the class-independent term network achieves lower precision scores.

In addition, we show how to strike reasonable balance between increasingly higher recall/coverage and lower precision. We learned that the class-specific term networks provide better precision scores than the class-independent term network *LogNet*. Query

terms are expanded in a certain context (patent class). On the other hand the class-independent term network *LogNet* achieves best recall and coverage scores. Hence, (1) expansion terms can be incrementally suggested initially from the class-specific term networks providing high precision scores, followed by more generic terms from the class-independent term network *LogNet* achieving higher recall measures later-on, or (2) expansion terms can be suggest in the order of their support in the training set. After having initially suggested the most likely and highest-precision terms (i.e. using expansion terms that were encountered most frequently as synonyms in the training sets), additional terms that have a lower support and lower precision (i.e. encountered at least one time) can be suggested. Furthermore, (3) related US classes allow the use of class-specific term networks across boundaries, providing valuable expansion opportunities, specifically for smaller classes, in example for classes where few query logs are available.

More importantly, the specific term networks drastically outperform general-purpose sources such as *WordNet*. The standard dictionary *WordNet* achieves for all US classes only low performance in recall. This may be attributed to the fact that patent searchers, (1) expand class-specific query terms using general terms, (2) expand query terms w.r.t. part of speech, (3) relate terms, which have the same meaning in a specific class, (4) use popular trademarks and (5) patent applicants are allowed to create their own terms for query expansion. These kinds of synonyms, equivalents and relations between the vocabulary are not included in standard dictionaries, such as *WordNet*, but are needed for automatic query expansion in the patent domain.

We show, that our approach to learn term networks from the patent domain, specifically directly from the query logs helps in meeting the requirements of this highly domain specific setting.

In future work we aim to improve the performance of the learned class-specific and class-independent term networks. To this end, we want to learn class-related term networks to improve recall and coverage of the class-specific term networks that achieved best precision measures. Further, we want to use a dynamic thresholding on the support scores for synonym relationships, in particular for the class-independent term network *LogNet* providing best recall and coverage scores but lowest precision measures. Thus, we will be able to sort the expansion terms in such a way as to optimize precision while keeping recall high. In addition, we want to consider the default operator element and the set operator (which can be set at “AND” or “OR”) to learn synonyms from co-occurring query terms, which are not linked to each other by the Boolean operator “OR” to expand our term networks.

Furthermore, we want to use the query log collections to learn further semantic relations that are needed for automatic query expansion in patent searching. In particular, we aim to learn term networks of keyword phrases, which we use for automatic query limitation in patent searching. To learn the keyword phrases we will use the proximity operators appearing in the text queries of the query logs.

References

- Alberts, D., Yang, C., Fobare-DePonio, D., Koubek, K., Robins, S., Rodgers, M., Simmons, E., & De Marco, D. (2011). Introduction to patent searching. In M. Lupu, K. Mayer, J. Tait, & A. J. Trippe (Eds.), *Current challenges in patent information retrieval. The information retrieval series* (Vol. 29, pp. 3–43). Springer.
- Amitay, E., & Broder, A. (2008). Introduction to special issue on query log analysis: Technology and ethics. In *ACM Trans. Web 2*, Article 18.

- Azzopardi, L., Vanderbauwhede, W., & Joho, H. (2010). Search system requirements of patent analysts. In *Proceeding of the 33rd international ACM SIGIR conference on research and development in information retrieval (SIGIR 2010)*. Geneva, Switzerland, pp. 775–776.
- Bashar, A., & Myaeng, S. (2011). Query phrase expansion using wikipedia in patent class search. In *Proceedings of the 7th Asia conference on information retrieval technology (AIRS'11)*. Dubai, United Arab Emirates, pp. 115–126.
- Clough, P., & Berendt, B. (2009). Report on the treble CLEF query log analysis workshop 2009. *SIGIR Forum*, 43, 71–77.
- De Marco, D. (2011). *Plumbing the depths of examiner search (il)-logic: A patent searching perspective*. Presentation given at PIUG 2011 Northeast Conference, New Brunswick (New Jersey), USA. <http://www.demarcoip.com/wordpress1/wp-content/uploads/2013/07/ExaminerSearch.pdf>.
- Fujita, S. (2007). Technology survey and invalidity search: An comparative study of different tasks for Japanese patent document retrieval. *Information Processing and Management, An International Journal*, 42(5), 1154–1172.
- Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus annotation: Linguistic information from computer text corpora* (pp. 102–121). London: Longman.
- Hang, C., Ji-Rong, W., Jian-Yun, N., & Wei-Ying, M. (2002). Probabilistic query expansion using query logs. In *Proceedings of the 11th international conference on world wide web (WWW 2002)*. Hawaii, USA, pp. 325–332.
- Herbert, B., Szarvas, G., & Gurevych, I. (2009). Prior art search using international patent classification codes and all-claims-queries. In *Proceedings of the 10th cross-language evaluation forum conference on multilingual information access evaluation (CLEF 2009)*. Corfu, Greece, pp. 452–459.
- Hunt, D., Nyugen, L., & Rodgers, M. (2007). *Patent searching: Tools & techniques*. Hoboken: Wiley.
- Jochim, C., Lioma, C., Schütze, H. (2011). Expanding queries with term and phrase translations in patent retrieval. In *Proceedings of the second international conference on multidisciplinary information retrieval facility (IRFC 2011)*. Vienna, Austria, pp. 16–29.
- Jochim, C., Lioma, C., Schütze, H., Koch, S., & Ertl, T. (2010). Preliminary study into query translation for patent retrieval. In *Proceedings of the patent information retrieval workshop (PAIR 2011)*. Toronto, Canada, pp. 57–66.
- Jürgens, J., Hansen, P., & Womser-Hacker, C. (2012). Going beyond CLEF-IP: The 'reality' for patent searchers? In *Proceedings of the third international conference of the CLEF initiative (CLEF 2012)*. Rome, Italy, pp. 30–35.
- Kato, M., Sakai, T., & Tanaka, K. (2013). When do people use query suggestion? A query suggestion log analysis. *Information Retrieval*, 16(6), 1–22.
- Konishi, K. (2005). Query terms extraction from patent documents for invalidity search. In *Proceedings of NTCIR 2005: NTCIR-5 workshop meeting*. Tokyo, Japan.
- Kunpeng, Z., Xiaolong, W., & Yuanchao, L. (2009). A new query expansion method based on query logs mining. *International Journal on Asian Language Processing*, 19, 1–12.
- Magdy, W., & Jones, G. J. F. (2011). A study of query expansion methods for patent retrieval. In *Proceedings of PAIR 2011*. Glasgow, Scotland, pp. 19–24.
- Mahdabi, P., & Crestani, F. (2011). Learning-based pseudo-relevance feedback for patent retrieval. In *Proceedings of the second international conference on multidisciplinary information retrieval facility (IRFC 2011)*. Vienna, Austria, pp. 1–11.
- Mahdabi, P., Keikha, M., Gerani, S., Landoni, M., & Crestani, F. (2011). Building queries for prior-art search. In *Proceedings of the second international conference on multidisciplinary information retrieval facility (IRFC 2011)*. Vienna, Austria, pp. 3–15.
- Miller, G. (1995). WordNet: A lexical database for english. *Communications of the ACM*, 38(11), 39–41.
- Russo, D. 2011. Knowledge extraction from patent: Achievements and open problems. A multidisciplinary approach to find functions. In *Proceedings of the 20th CIRP design conference (CIRP Design 2012)*. Nantes, France, pp. 567–576.
- Sekine, S., & Suzuki, H. (2007). Acquiring ontological knowledge from query logs. In *Proceedings of the 16th international conference on world wide web (WWW 2007)*. Banff, Canada, pp. 1223–1224.
- Silvestri, F. (2010). Mining query logs: Turning search usage data into knowledge. *Foundations and Trends in Information Retrieval*, 4(1–2), 1–174.
- Tannebaum, W., & Rauber, A. (2012). Acquiring lexical knowledge from query logs for query expansion in patent searching. In *The IEEE sixth international conference on semantic computing (IEEE ICSC 2012)*, Italy, Palermo, pp. 336–338.

- Tannebaum, W., & Rauber, A. (2012). Analyzing query logs of USPTO examiners to identify useful query terms in patent documents: A preliminary study. In *Proceedings of the information retrieval facility conference (IRFC 2012)*. Vienna, Austria, pp. 127–136.
- Xue, X., & Croft, W. (2009). Automatic query generation for patent search. In *Proceedings of CIKM 2009*. Hong Kong, China, pp. 2037–2040.
- Xue, X., Croft, W. (2009). Transforming patents into prior-art queries. In *Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval*. Boston, USA, pp 808–880.
- Zhang, J., Xiong, M., & Yu, Y. (2006). Mining query log to assist ontology learning from relational database. In *Proceedings of the 8th Asia Pacific web conference (APWeb 2006)*. Harbin, China, pp. 437–448.