www.ijmserh.com
IJMSERH
Volume 4, Issue 3, July 2016
ISSN: 2349-6037
International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

Efficient Approach of Patent Search Paradigm

Haritha.V[1], Dr. A.Senthil Kumar[2], Dijith.M.S[3]

P.G. Scholar, Department of CSE, RVS College of Engineering and Technology, Coimbatore, India[1]

Professor, Department of CSE, RVS College of Engineering and Technology, Coimbatore, India[2]

P.G. Scholar, Department of CSE, R.V.S. College of Engineering and Technology, Coimbatore, India[3]

ABSTRACT: As an important action of finding existing relevant patents and validating or invalidating new patent application, patent search attracts much attention from both industrial and academiccommunities. But most of the users have limited knowledge about the underlying patents and they have to repeatedly issue different queries and check answers, which is a very time consuming and boring process. To overcome this problem, we propose an efficient approach of user friendly patent search paradigm, which improves the user search experience by helping the user in finding the relevant patents more easily. Automatic error correction, topicbased query suggestion, and query expansion, are the three effective techniques that we have proposed to improve the usability of patent search. We also focus on how to efficiently find relevant patents from a large collection of patents. First, partition the patents into small partitions based on their topic and classes. Then for a given query, we find the highly relevant partitions and answer the query in each of such pertinent partitions. Finally, we combine the answers of each partition and generate top-k answers of the patent-search query. Thus it produces more pertinent answers through sophisticated graphical user interaction.

KEYWORDS: patent search, error correction, query suggestion, query expansion

I. INTRODUCTION

A patent is a set of exclusive rights granted by a sovereign state to inventor or assignee for a limited period of time in exchange for detailed public disclosure of an invention. Patents play a very important role in intellectual property protection.Patent search can help the patent examiners to find previously published relevant patents and validate or invalidate new patent applications. It has turn out to be more and more popular, and recently gained much recognition from both industrial and academic communities.For example, there are many online systems such as Google patent search, Derwent Innovations Index (DII), and USPTO to support the patent search. As most patent-search users have limited knowledge about the underlying patents, they have to employ a try-and-see approach to repeatedly issue queries and check answers, which is a very monotonous process. To help users easily find pertinent patents, the first step for the patent search is to capture users' search purpose. In other words, suggesting search keywords for users is the most critical part of the search strategy. After selecting the correct search keywords, the succeeding step is finding and ranking the relevant answers.

Providing patent test collections and a spate of workshops and symposiums on patent retrieval, there has been renewed interest in researching and developing Information Retrieval (IR) tools, techniques and theory for patent search. Patent analysts perform a number of difficult and challenging search tasks and rely upon sophisticated search functionality, tools, and specialized products. These search tasks are often performed under stringent conditions and they also require different search strategies to achieve the end goal. Whilst there has been substantial research on patent search and the tasks and tools involved, little work has been performed investigating the requirements of patent searchers, and what they want. It is vital that users are consulted and their needs understood.

Most of existing methods focus on devising a complicated ranking model to rank patents and finding the most pertinent answers .However, they do not pay enough attention to effectively capturing users' search purpose, which is at least as important as ranking patents. To address this problem, this paper propose a new user-friendly patent search paradigm, which can help users find pertinent patents more easily and improve user search experience. As users' query keywords may have typing errors, existing methods will return no answer as they cannot find patents matching query keywords. To palliate this problem, we propose an error-correction technique to suggest similar terms for the query keywords and return answers of the similar terms. In addition, to help users produce high-quality queries, as users type in keywords, we recommend keywords that are topically relevant to the query keywords. In this way, users can

interactively supply queries and modify their keywords if there is no pertinent answer, which can provide users with satisfaction. As users may not understand the underlying patents precisely, they may type in ambiguous keywords or incorrect keywords. On the other hand, the same concept or entity may have different representations. For example, "car" and "sedan" are relevant to "automobile." Thus, if users type in a keyword "car," we may need to expand the keyword to "automobile." To this end, we propose a query expansion-based technique to recommend users pertinent keywords. Two methods are discussed here whichefficiently suggest pertinent keywords. To summarize, we use these three techniques to help users search patents more easily and improve the usability of patent search.

II. RELATED WORK

Guo and Gomes [10] proposed SVM patent ranking model to improve the search quality. Larkey [5] studied the patent classification problem; however,the paper neglected the prior-art search (novelty search). Xue and Croft [8] studied how to automatically transform a query patent into a search query and find the answers using the search query. They focused on how to extract query words from patents, how to weight them and whether to use nounphrases. Here the problem is different from that as we focus on improving efficiency and quality to answer a keyword query. Azzopardi et al. [4] surveyed eighty patent analysts in order to obtain a better picture of their search habits, types of functionality, preferences, and gave some findings from this survey. Magdy et al. [8] discussed two approaches for the patent prior-art search. The first one is a simple method with the demand of low-resources, and the second one is a sophisticated method, using an advanced level of content examination. Bashir and Rauber [6] evaluated the coverage of prior-art queries extracted from query patents using retrievability measurement. Different from prevailing studies, we propose an efficient user-friendly patent search paradigm.

III. OVERVIEW

In this paper, we propose an easy patent search techniquewhich mightfacilitate users simplynotice pertinent patents and improve user search expertise. Fig.1 illustrates the architecture of our patent-search paradigm.

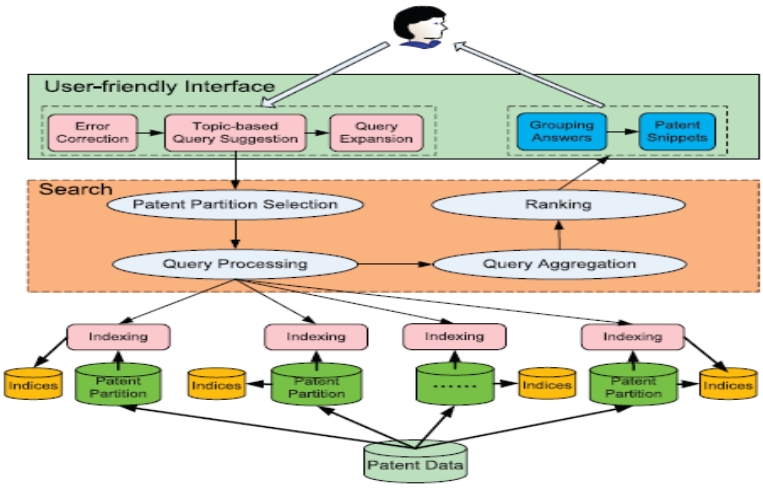

Fig.1 User-Friendly Patent Search Architecture

The User-friendly Interface component is used to capture users' search intention and refine query keywords so as to find pertinent answers. It consists of three subparts, error correlation, topic-based query suggestion, and query expansion. Also, it groups the answers based on their topics to help users navigate answers. It additionally provides users with the patent snippets of the answers to assist users quickly check whether or not the returned answers are pertinent. Thus, users can interactively issue queries, browse the results and get the ultimate answers, which can help them find relevant answers more easily.

To improve the efficiency, patents are partitioned into different data partitions based on their topics. The Indexing component builds inverted indexes on top of each partition. Then, the Patent Partition Selection component selects top-

'highly pertinent data partitions for each query and routes the query to such pertinent partitions to find local answers. The Query Processing component determines answers in the local partitions. ultimately, the Query Aggregation component combines the local results and the Ranking component ranks the answers to return the final top-k answers.

IV. USER-FRIENDLY PATENT SEARCH PARADIGM

There are several unique challenges in patent search, mainly because of the difficulty of understanding users's query intent and efficiently matching the query keywords to patents. Here, we present several effective techniques to address these challenges.

1.Patent partition

Patents are partitioned into different data partitions based on the following reasons. First, patents have different classes. Second, the number of patents is usually very large. Moreover, the number of patents is increasing quickly. Third, for a patent search query, only some of the classes/subclasses of patents could be pertinent to the patent query. Based on these, we partition the patents based on their classes and topics using the topic model as follows: We first extract the topic of each patent. Then, patents are partitioned with the same topic into the same data partition, and each and every topic corresponds to a data partition. Patents in the same partition are highly pertinent and those in different partitions are inappropriate.

2. Effective indexing

For every partition, we construct a well-known inverted index structure. For every query keyword, we make use of the index structure to locate patents containing the keyword. Then, we intersect the patents corresponding to different keywords to generate the most pertinent patents. In each partition, we can use any effective ranking function to rank the patents in the partition. As patents in each partition are very pertinent, we can do more deep ranking by taking into account the correlation between different patents.webuilda triestructure ontop of keywords in the patent partition, to assist the query suggestion. Every node on the path has a label of a character in the keyword. For each and everyleaf node, we store an inverted list of Idsof records that contain the corresponding keyword.

3. User-friendly interface

To capture users' query intention, we bring in several effective techniques to make patent search user friendly and help users easily find pertinent patents.

3.1 Automatic error correction

As query keywords that users have typed in may have typing errors, traditional methods will return no answer as they cannot locate answers that contain the query keywords. Clearly, this method is not user friendly. Instead, it is better to correct the typing errors, suggestusers similar keywords, and return the answers of the related keywords. To measure the similarity between keywords, existing methods usually espouse edit distance. The edit distance between two keywords is the minimum number of edit operations (i.e., insertion, deletion, and substitution) of single characters needed to convert the first one to the second. For example, the edit distance of "patent" and "paitant" is 2. Two keywords are said to be akin if their edit distance is within a given threshold. In this section, the method first uses the filter step to find a subset of keywords which may be potentially akin to the query keyword. Then, it uses a verification step to eliminate those false positives and get the final akin keywords.

3.2 Topic based query suggestion

We develop a completely unique model for effectively suggesting keywords as user's type in queries letter by letter. The basic notion of this method is to make use of the topic model to estimate the probability of the next query keyword. Intuitively, if a keyword in patent is more topically coherent with the previously typed query keywords, it would get a higher score. Specifically, the emphasis is on estimating two important probabilities: the probability of a keyword conditioned on topics, and the probability of sampling a keyword from a patent. Each of the two probabilities are used to evaluate the score of each keyword. An LDA model can be used to learn the keyword distribution over each topic from the underlying patents. LDA can be stated as a soft-clustering technique that permits a keyword to appear in multiple topics and takes into account the degree of a keyword belonging to each and every topic. The keyword distribution over a set of patents is learned by using a language model. The language model approaches captures the property of the patents and calculate the likelihood of sampling an exact keyword. Thus, combine the two probabilities and use the topic-based method to suggest pertinent keywords.

3.3 Query expansion

WordNet can be used to expand a keyword. If the query keyword is indexed by WordNet, it is easy to get the pertinent keywords of the query keyword by using an inverted list structure. WordNet is artificially generated for common words. If the query keywords are not present in WordNet, it is not possible to suggest pertinent keywords.

To address this problem, the first way is to utilize search engines, since most search engines will recommend pertinent keywords as user's type in queries. The patent query is issued to search engines and get the pertinent keywords from the search engines(such as Google). The second way is to mine the pertinent keywords from the query logs. Use the click-through data to mine the related queries as follows: For two queries, if users click the same returned result (patent), they are potentially relevant. This property is utilized to mine relevant queries. For two queries, use the number of times user clicked on the same patent to show their relevance. If a keyword pair with their co-occurrence is larger than a given threshold, the two keywords are pertinent and use them to perform query expansion.

3.4 Ranking and patent partition selection

To improve the efficiency, the query is not issued to each and every patent partition. Instead, select the top-l pertinent patent partitions and use them to answer the query. The relevancy of a query to a patent partition is need to be evaluated. There are three factors that are need to be considered to rank a patent partition. The first is the topic relevancy. That is, whether the patent partition is topically rpertinentt to the query keywords. The second is keyword relevancy. That is, whether the patent partition contains query keywords. tf-idf model is used to evaluate the relevancy. The third is prior-art relevancy. That is, whether the patent partition is novel enough to the query.

Here focus is on effective ranking models to improve the result quality by evaluating the relevancy between a query and a patent. Important factors considered for this calculation are:

1. The importance of a patent p, represented by Wp. The more important a patent, the higher probability pertinent to a query. Patents can be designed as a graph where nodes are patents and edges are citations between patents. Thus, graph can be employed to compute the weight of a patent.
2. The keyword relevancy of a patent p to a query Q, represented by R(p;Q). The well-known IR method (e.g., tf-idf) is used to compute the relevancy.
3. The topic relevancy of patent p to query Q, represented by T(p;Q). above topic-based method can be used to compute the value.
4. The prior-art relevancy of a patent Pp

Combine the above factors to rank a patent p given with respect to a query Q, represented by S(p|Q), as follows :

$$S(p|Q) = \alpha * W_p + \beta * R(p,Q) + \gamma * T(p,Q) + (1 - \alpha - \beta - \gamma) * P_p \quad \rightarrow \text{Eqn 5.1}$$

The above function can be used to compute the relevancy between patent p and query Q and return the top-k most relevant patents.

Given a query, to find its top-k answers, first select top- l relevant patent partitions, and issue the query to such pertinent patent partitions. Use the above ranking functions to compute the scores of different patent partitions. For each partition, efficiently find top-k answers using our indexing structures and ranking model. Then, combine the answers from each selected partition and generate the final top-k answers based on our ranking model. This method can prune many irrelevant patent partitions and can improve the efficiency significantly.

V. RESULTS AND DISCUSSION

In the present system, minimum number of queries takes less time as shown in Fig.5a. It is the graphdrawn betweenNumber of query given and the time taken for that query processing. If the number of queries increases, timing performance also increases. It is the timing of query heading selection process. We give two or three query means the resulting time is minimum, otherwise it takes huge time. In that, present system timing performance is based on the Query. It improves user search experience.

The overall performance of the system is calculated by comparing timing of patent result and top k values as shown in Fig.5b. It is the graph drawn between Top K result values and the time taken. If number of top k value increases, timing performance of the system also increases. User-Friendly Patent Search Paradigm method improves

the search performance. Patent Search Paradigm method improves the search performance. The performance graph represents how fast the patent search result is retrieved.

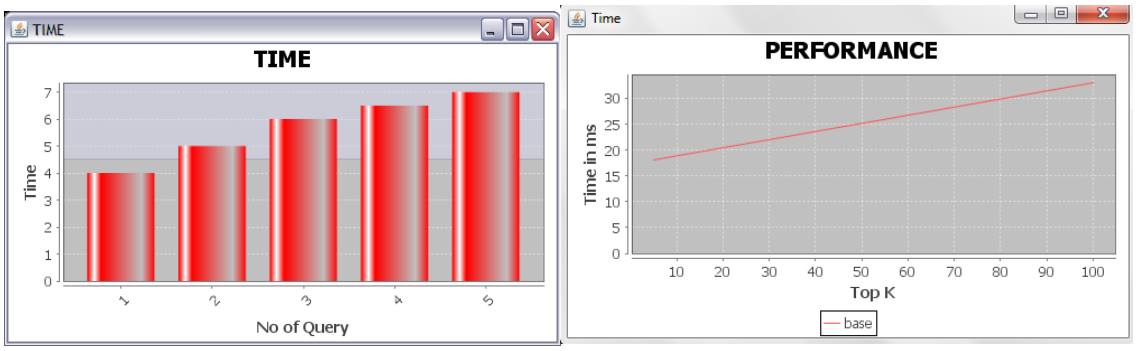

Fig.5a. Time graphFig.5b. Performance Graph

VI. CONCLUSION AND FUTURE WORK

New patent-search paradigm developed three effective techniques, error correction, topic-based query suggestion, and query expansion, to make patent search more user friendly and improve user search experience. Error correction technique can provide users precise keywords and correct the typing errors. Topic-based query suggestion can recommend topically coherent keywords as users type in query keywords. Query expansion can recommend synonyms and those pertinent keywords of query keywords which are in the same concept with query keywords. A partition-based method is implemented to improve the search performance. Experimental results show that this method achieves high efficiency and quality. As future work,more user interaction can be made possible by including user suggestion session to be analysed to infer user search goals. We try to reduce the computation time required to partition the data set. It will reduce the original data set in to simplified data set and find pertinent documents based on user suggestion. Also it reduces time complexity for displaying the top-k answers for the patent search query.

REFERENCES

[1] D.M Blei, A.Y Ng, and M.I Jordan, "Latent Dirichlet Allocation," J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.
[2] G. Li, S. Ji, C. Li, and J. Feng, "Efficient Fuzzy Full-Text Type-Ahead Search," VLDB J., vol. 20, no. 4, pp. 617-640, 2011.
[3] J. Fan, H. Wu, G. Li, and L. Zhou, "Suggesting Topic-Based Query Terms as You Type," Proc. Int'l Asia Pacific Web Conf. (APWEB), pp. 61-67, 2010.
[4] L. Azzopardi, W. Vanderbauwhede, and H. Joho, "Search System Requirements of Patent Analysts," Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 775-776, 2010.
[5] L.S. Larkey, "A Patent Search and Classification System," Proc. Fourth ACM Conf. Digital Libraries, pp. 179-187, 1999.
[6] S. Bashir and A. Rauber, "Improving Retrievability of Patents in Prior-Art Search," Proc. European Conf. Information Retrieval (ECIR), pp. 457-470, 2010.
[7] S. Ji, G. Li, C. Li, and J. Feng, "Efficient Interactive Fuzzy Keyword Search," Proc. Int'l Conf. World Wide Web (WWW), pp. 371-380, 2009.
[8] W. Magdy, P. Lopez, and G.J.F Jones, "Simple vs. Sophisticated Approaches for Patent Prior-Art Search," Proc. European Conf. Advances in Information Retrieval, pp. 725-728, 2011.
[9] X. Xue and W.B Croft, "Automatic Query Generation for Patent Search," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 2037- 2040, 2009.
[10] Y. Guo and C.P. Gomes, "Ranking Structured Documents: A Large Margin Based Approach for Patent Prior Art Search," Proc. Int'l Joint Conf. Artificial Intelligence (IJCAI), pp. 1058-1064, 2009.