RESEARCH ARTICLE

User Friendly Data Searching

D.Sravani[1], K. Madhurima[2]
[1]M.Tech 2nd Year, Dept. of CSE, JNTU, HYDERABAD, INDIA
[2]Asst Professor, Dept. of CSE, JNTU, HYDERABAD, INDIA
[1] sravani1504@gmail.com; [2] madhu.kmadhurima@gmail.com

_____

Abstract — In an Intellectual property protection patents play an important role. Patent search attracted user's attention due to important operation for finding existing relevant patents and validating a new patent application. But many users have limited knowledge about the underlying patents, and they have to use a try-and-see approach to repeatedly issue different queries and check answers, which is a very difficult process. To overcome this problem our proposed system introduces a new user-friendly approach for patent search to find the relevant patents more easily and improve user search experience. Here proposed system uses three efficient techniques to improve the usability of patent search: error correction, Topic-based query suggestion and query expansion. We also study how to efficiently find relevant answers from a large collection of patents. For this, first partition patents into small partitions based to their topics and classes. Then given a query, we find highly relevant partitions and answer the query in each of such highly relevant partitions. Finally we combine the answers of each partition and generate top answers of the patent-search query.

Index Terms— Patent search, error correction, query suggestion, query expansion

INTRODUCTION

PATENTS play a very important role in intellectual property protection. As patent search can help the patent examiners to find previously published relevant patents and validate or invalidate new patent applications, it has become more and more popular, and recently attracts much attention from both industrial and academic communities. For example, there are many online systems to support patent search, such as Google patent search. As most patent-search users have limited knowledge about the underlying patents, they have to use a try-and-see approach to repeatedly issue queries and check answers, which is a very tedious process. To help users easily find relevant patents, the first step for the patent search is to capture users' search intention. In other words, suggesting search keywords for users is the most critical part of the search strategy. After selecting the precise search keywords, the next step is finding and ranking the relevant answers.

Most of existing methods focus on devising a complicated ranking model to rank patents and finding the most relevant answers. However, they do not pay enough attention to effectively capturing users' search intention,

which is at least as important as ranking patents. To address this problem, in this paper, we propose a new user-friendly patent search paradigm, which can help users find relevant patents more easily and improve user search experience. As users' query keywords may have typos, existing methods will return no answer as they cannot find patents matching query keywords. To alleviate 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 formulate high-quality queries, as users type in keywords, we suggest keywords that are topically relevant to the query keywords. In this way, users can interactively issue queries and modify their keywords if there is no relevant answer, which can provide users with gratifications. As users may not understand the underlying patents precisely, they may type in ambiguous keywords or inaccurate keywords.

To this end, we propose a query expansion-based technique to recommend users relevant keywords. We discuss two methods to efficiently suggest relevant keywords. To summarize, we use these three techniques to help users search patents more easily and improve the usability of patent search. In addition, existing methods only focus on the effectiveness of patent search and neglect the fact that the search efficiency is also very important. To address this problem, we propose a new method to improve search efficiency. We note that the patents are usually classified into different classes based on the topics. Classify the patents based on the classes and the topics of the patents using the topic model, and generate several patent partitions, such that patents in the same partition are very topically relevant and those in different partitions are not very relevant. Then, given a query, we find highly relevant partitions and use each partition to efficiently find relevant patents of the query. Finally, we combine the results from each partition and generate the top-k answers. Experimental results show that our method achieves high efficiency and result quality.

EXISTING SYSTEM

A search mechanism improves the performance of a queue system including a queue for storing a plurality of data items and search mechanism by maintaining a key cache data structure having an array of entries, each of which has a key field and a pointer field. The key and pointer fields respectively of each cache entry are used for storing a key value of a different one of the enqueued data items of the queue and a pointer to that enqueued item. The key of each Data item to be enqueued is used to generate an index value for accessing a location of the key Cache array to obtain immediate access to the corresponding enqueued data item thereby reducing the search time for determining the proper point within the queue for inserting the data item to be added.

Most of the existing methods focus on devising a complicated ranking model to rank patents and finding the most relevant answers. They do not pay enough attention to effectively capturing users search intention which is at least as important as ranking patents. As users query keywords may have typos existing methods will return no answer as they cannot find patent matching query keywords.

PROPOSED SYSTEM

a user-friendly patent search paradigm which help users to find relevant patents more easily and improve user search experience. Error correction technique is used to suggest similar terms for query keywords and return answers of similar terms when user keyword has any typos. We also suggest keywords that are topically relevant to the query keywords. On the other hand, the same 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 relevant keywords.

We can classify the patents based on the classes and the topics of the patents using the topic model and generate several patent partitions, such that patents in the same partition are very topically relevant and those in different partitions are not very relevant. Patent partitions are stored in different nodes in the cluster. The Indexing component builds inverted indexes on top of each partition. Then for each query, the Patent Partition Selection component selects top highly relevant data partitions and routes the query to such relevant partitions to find local answers. The Query Processing component computes answers in the local partitions. Finally the Query Aggregation

component combines the local results and the Ranking component ranks the answers to return the final top-k answers. Finally we combine the results from each partition and Generate the top-*k* answers.

SYSTEM ARCHITECTURE

Patent search is for finding the patents that are already there. There are several unique challenges in patent search, mainly due to the difficulty of understanding users query intent and efficiently matching the query keywords to patents. Our project proposes three effective techniques **error correction, topic based query suggestion, query expansion** to improve the usability of patent search. In this architecture user enters the patent search page using the login. If he is a new user he is registering with the page.

The user enters the patent search keyword then the key words will error corrected if is there any error. And keyword got expanded regarding with the search term. Next Patent will searched in the patent partitions that are stored in patent Database. The results will group and the top answers for the searched patent will be displayed to the user.
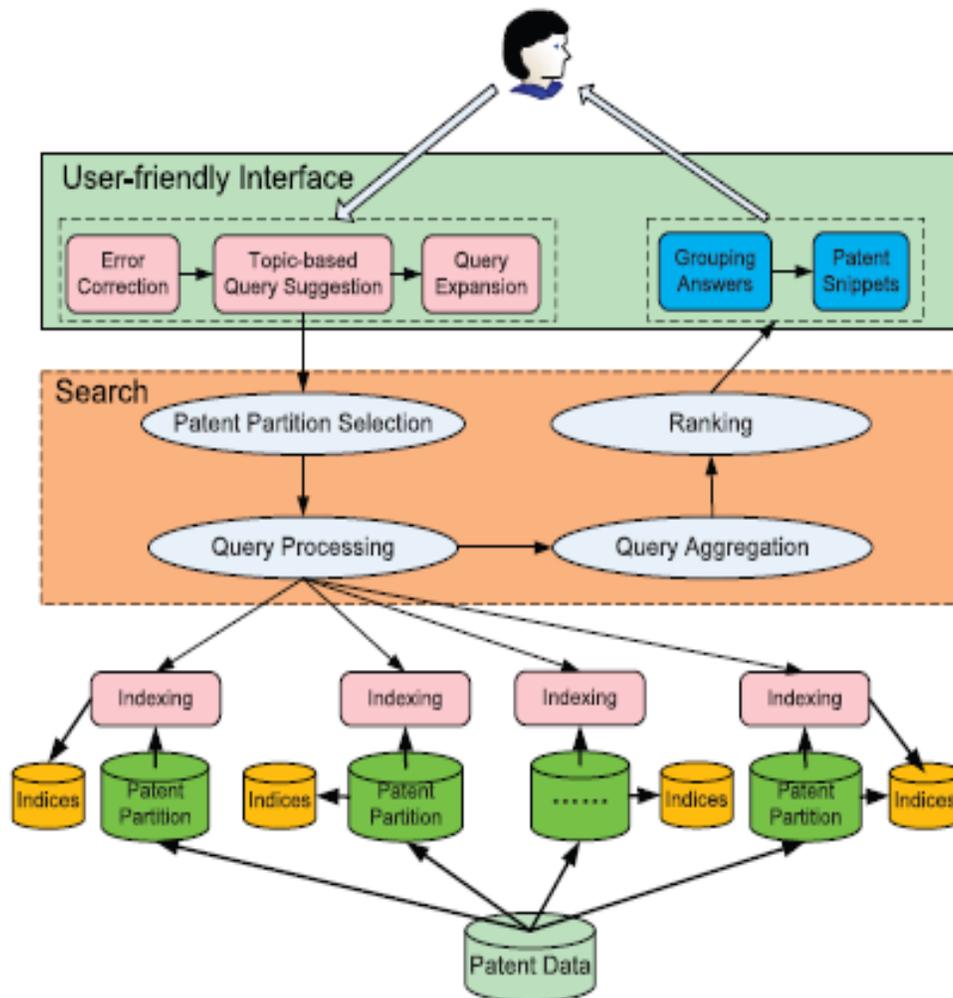
Fig 1: User-friendly patent search architecture.

User-Friendly Interface

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

Automatic Error Correction

In the automatic error correction we are using trie structure to do efficient keyword correction and completion. We are considering the prefix of the query word if it is not familiar with the trie node then we don't want to consider that keyword.
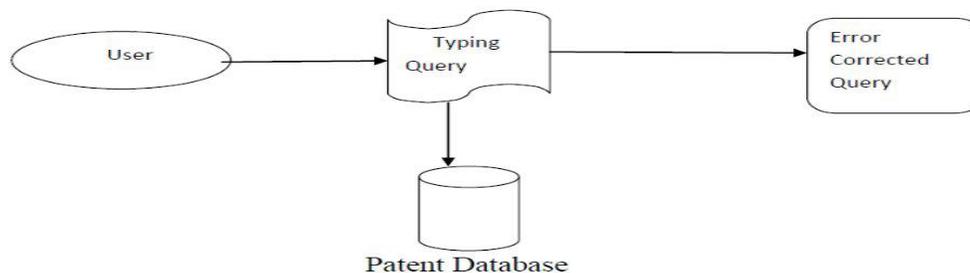
Fig 2 Automatic error correction module diagram

Topic-Based Query Suggestion

The topic based model is estimating the probability of the next query keyword. If a keyword in patents is more topically coherent with the previously typed query word it will be getting the higher score.
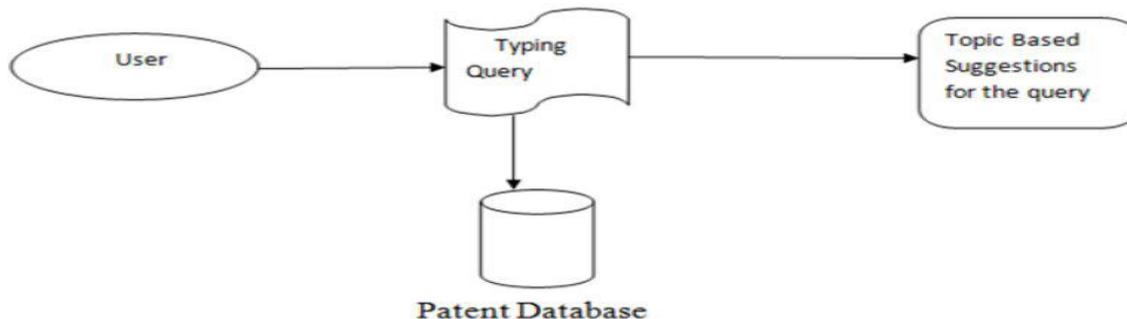
Fig 3 Topic based query suggestion module diagram

Query Expansion

In the query expansion we will be using the search engine for suggesting the relevant keyword. And we are using the relevant keywords from the query log for the expansion purpose.
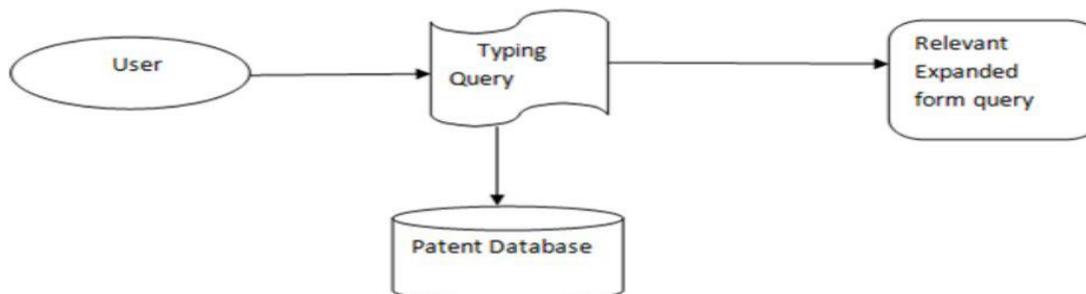
Fig 4 Query expansion module diagram

Ranking:

In this module we are ranking the answers that are obtained for our query search by the probability of most relevant patent. We are finding the most relevant patent regarding with the patent search.
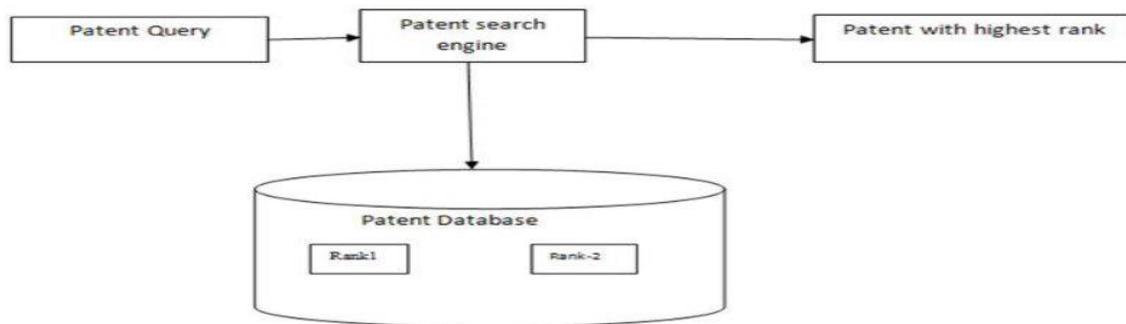

Fig 5 Ranking module diagram

Patent partition selection

In this module we are selecting the partition regarding with our patent search using two relevancies. That is topic relevancy and keyword relevancy. Using these two relevancies we are finding the top relevant partitions.
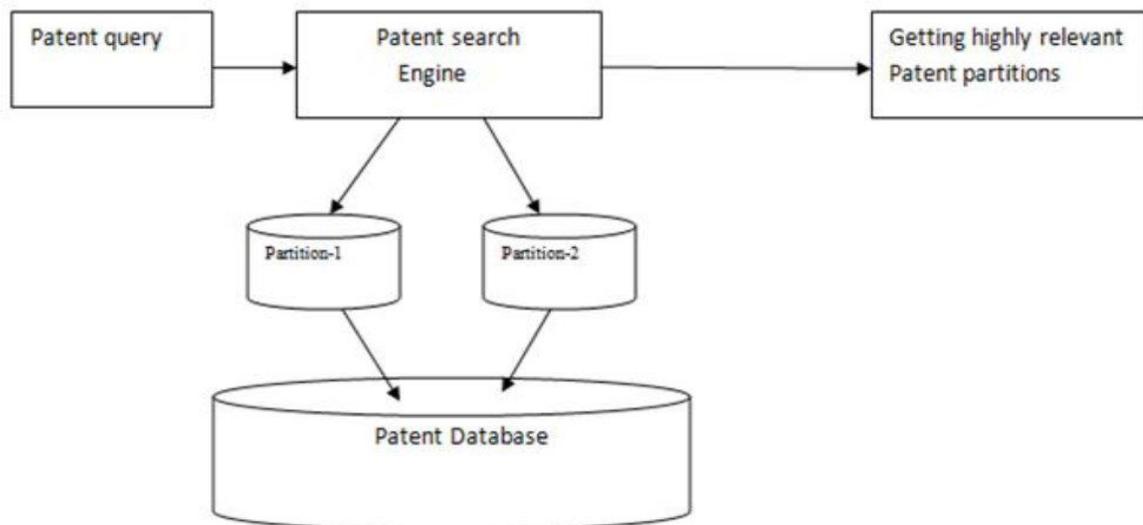

Fig 6 Patent partition selection module diagram

Query Processing

Query processing module is for find the top answers regarding with our search. In this Process we are combining all the ranking and selected partitions for finding the top answer.
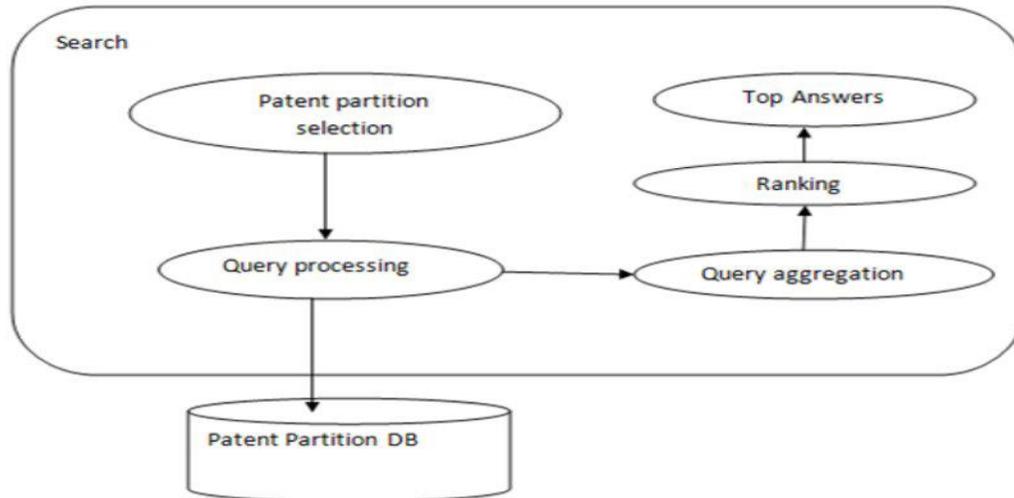

Fig 7 Query processing module diagram

CONCLUSION

In this paper, we proposed a new patent-search paradigm. We 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 correlation can provide users accurate keywords and correct the typing errors. Topic-based query suggestion can suggest topically coherent keywords as users type in query keywords. Query expansion can suggest synonyms and those relevant keywords of query keywords which are in the same concept with query keywords. We proposed a partition-based method to improve the search performance. Experimental results show that our method achieves high efficiency and quality.

REFERENCES

1. 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.

2. S. Bashir and A. Rauber, "Improving Retrievability of Patents in Prior-Art Search," Proc.   European Conf. Information Retrieval (ECIR),   pp. 457-470, 2010.

3. D.M Blei, A.Y Ng, and M.I Jordan, "Latent Dirichlet Allocation," J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.

4. 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.

5. 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.

6. 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.

7. L.S. Larkey, "A Patent Search and Classification System," Proc. Fourth ACM Conf. Digital Libraries, pp. 179-187, 1999.

8. C. Li, J. Lu, and Y. Lu, "Efficient Merging and Filtering Algorithms for Approximate String Searches," Proc. Int'l Conf. Data Eng. (ICDE), pp. 257- 266, 2008.