Creation of Software focusing on Patent Analysis

Kandare Dipak Prabhu, Bhosale Sayali Ravindra, Dhande Priyanka Umesh, Shaikh Nida Zahid

Kandare Dipak Prabhu BE Student, PUNE UNIVERSITY, GESCOE, Nashik, Maharashtra, India.
Bhosale Sayali Ravindra BE Student, PUNE UNIVERSITY, GESCOE , Nashik, Maharashtra, India.
Dhande Priyanka Umesh BE Student, PUNE UNIVERSITY, GESCOE, Nashik, Maharashtra, India.
Shaikh Nida Zahid BE Student, PUNE UNIVERSITY, GESCOE, Nashik, Maharashtra, India.
Prof. C.B.Patil Internal guide GESCOE, Nashik, Maharashtra India

---

Abstract - *Within the early phases of technology management processes, patents are often used as a source of inspiration for new ideas. Patents contain detailed technical information about a technical problem and the preferred technical solution. This information can be used for example to assess the state of the art or as a basis to identify possible gaps in a technology field. But often it is a very time consuming process to analyse the information provided by patents, because huge amounts of patents have to be considered. Therefore special text-mining and data mining concept are used to help extracting the desired information in short time. Classification is used to classify the problem and its solution. Our approach to make an effective Pre-Processing steps to save both space and time requirements by using improved Stemming Algorithm. Stemming algorithms are used to transform the words in texts into their grammatical root form.*

***Key Words*: Extraction, Stemming, StopWordRemoval**.

1.INTRODUCTION

Patent documents contain important research results that are valuable to the industry, business, law, and policy-making communities. If carefully analysed, they can show technological details and relations, reveal business trends, inspire novel industrial solutions, or help make investment policy (Campbell, 1983; Jung, 2003)[2]. In recent years, patent analysis had been recognized as an important task at the government level in some Asian countries.

1.1 A typical patent analysis scenario

1. Task identification: define the scope, concepts, and purposes for the analysis task
2. Searching: iteratively search, filter, and download related patents
3. Segmentation: segment, clean, and normalize structured and unstructured parts
4. Abstracting: analyse the patent content to summarize their claims, topics, functions, or technologies
5. Clustering: group or classify analysed patents based on some extracted attributes
6. Visualization: create technology-effect matrices or topic maps
7. Interpretation: predict technology or business trends and relations[4].

2. PROBLEM STATEMENT

To analyse and register patent through software by using stemming and classification algorithm which were earlier register after checking problem statement and solution manually.

2.1 A GENERAL METHODOLOGY

Patent analyses based on structured information such as filing dates, assignees, or citations have been the major approaches. These structured data can be analysed by bibliometric methods, data mining techniques, or well-established database management tools such as OLAP (On-Line Analytical Processing) modules[1].

Therefore, based on the patent analysis scenario introduced above, a text mining methodology specialized for full-text patent analysis is proposed.
This may involve a repeated process of devising a set of query terms (query formulation), searching a couple of patent databases (collection selection), filtering undesired patents (relevance judgment), and downloading patents for local analysis (data crawling). Depending on the analysis purpose, the step can be as easy as, for example, fetching all the patents under some IPC (International Patent Classification) categories[2] .

The general text mining methodology for patent analysis.

- Document Pre-processing
  - Collection Creation
  - Document Parsing and Segmentation
  - Text Summarization
  - Document Surrogate Selection
- Indexing
  - Keyword/Phrase Extraction
  - Morphological Analysis
  - Stopword Filtering
  - Term Association and Clustering
- Topic Clustering

- Term Selection
- Document Clustering/Categorization
- Cluster Title Generation
- Category Mapping

o Topic Mapping
- Trend Map
- Query Map
- Aggregation Map
- Zooming Map

3. TECNICH DETAILS

3.1 Extraction
This method is used to tokenize the file content into individual word.

3.2 Stemming
This method is used to find out the root/stem of a word. For example, the words user, users, used, using all can be stemmed to the word "**USE**". The purpose of this method is to remove various suffixes, to reduce number of words, to have exactly matching stems, to save memory space and time. The stemming process is done using various algorithms. Most popularly used algorithm is "M.F. Porters Algorithm[5].

3.3 Stop word removal
Most frequently used words in English are useless in Text mining. Such words are called Stop words. Stop words are language specific functional words which carry no information. It may be of the following types such as pronouns, prepositions, conjunctions. Our system uses the SMART stop word list[5].

4. PROPOSED SYSTEM

In this paper we aim to analyse and register patent through software by using stemming and classification algorithm which were earlier register after checking problem statement and solution manually. The patent databases world-wide grow continuously,
there is a growing need for software solutions assisting the user to handle the patent analyses, because the analysis of hundreds of patents is very complex and time consuming. To deploy methods, we have proposed a new architecture for identifying problem and solution of particular patent.
Our system involves following steps:

During our work with patents it occurred that there are many phrases like e.g. What is claimed is: or A method comprising . that are very frequently used in patents[1]. In addition, patent documents of various countries are generally structured in a similar way: they all provide an abstract, the claim section, a description of the invention as well as figures 1. This similar structure makes it easier to quickly identify the elements that are of interest for the various patent analysis reasons.

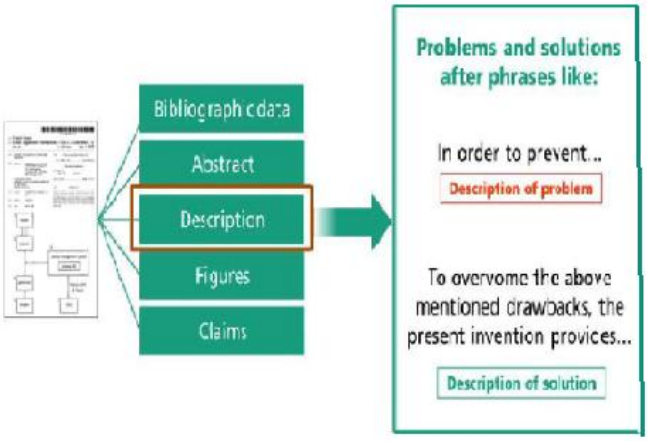

Fig. 1: Core element of Analysis: extraction and analysis of problems and solutions

To extract problems and solutions because in the majority of cases not only the solutions (the invention) are described in this part but also the problems (why the invention was made)[6]. Generally a patent provides more than one problem as well as more than one solution. But the description of problems is not always very detailed. In addition the relation between problems and solutions must not always be apparent, in some cases there is no close relationship or even no problem mentioned. This makes an extraction of any sort of relationships quite difficult.

Objective of the problem and solution extraction first of all is to retrieve the main claim of a
patent and then trying to identify at least one problem that refers to the first claim. Some patents actually provide a short summary of the main problem referring to the main claim, but most of them don't[7]. In those cases it could be possible that the extracted problem refers not to the first solution (main claim) but to further sub-solutions. Thus it is mandatory to check the problem retrieving results of afterwards.

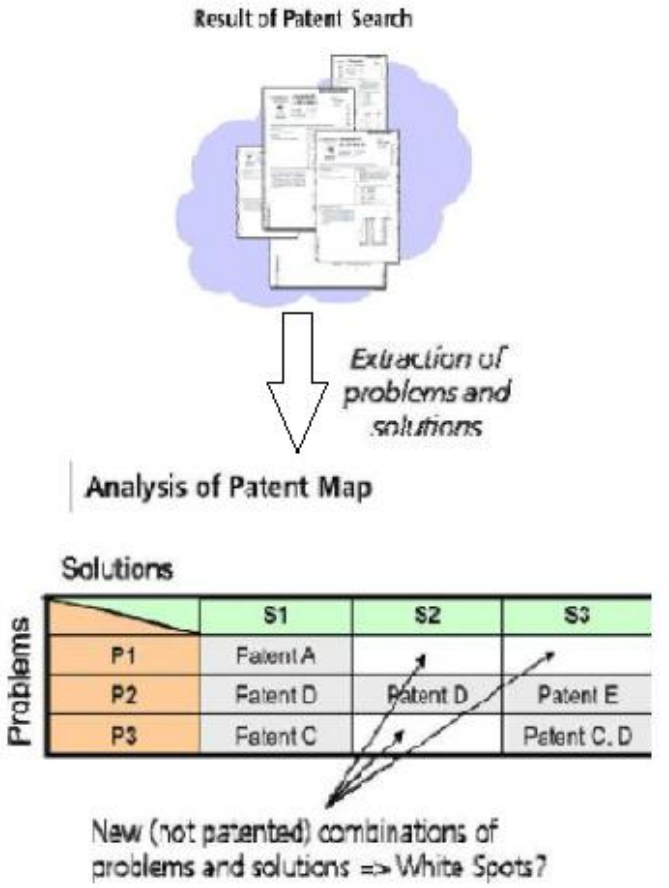

Fig. 2: Analysis of extraction and analysis of problems and solutions

In the following, problem or solution indicating phrases are parts of sentences that are surrounding a problem or solution, i.e. directly before, in between or after those phrases a problem or solution is described. For example within the sentence[1].

The Patent Skill Cartridge is able to extract problems and solutions by searching for those indicating phrases in the text and then displaying the sentence parts before or after the phrases in a pre-defined length. So within the Patent Skill Cartridge not only the problem and solution indicating phrases are defined, but also where the problem or solution text can be found, that means before, after or in between the indicating phrases.

For the development of the prototype of the Patent Skill Cartridge a first set of 57 patents from randomly chosen technology fields like electric vehicles and women hygiene articles was selected. Based on these patents problem and solution indicating key phrases were identified. The result was a list of over 100 phrases that were implemented in the Patent Skill Cartridge. A short section of the complete phrase list is shown in the following Table 1.

For the development of the prototype of the Patent Skill Cartridge a first set of 57 patents from randomly chosen technology fields like electric vehicles and women hygiene articles was selected. Based on these patents problem and solution indicating key phrases were identified. The result was a list of over 100 phrases that were implemented in the Patent Skill Cartridge. A short section of the complete phrase list is shown in the following Table 1. We found that some phrases, especially solution indicating phrases, occurred several times in more than one patent even
though the wording of some phrases was slightly different in the selected set of patents. For example the most frequent phrase indicating a solution was it is an object of the invention / general object of the (see Fig. 2). The challenge during the implementation of phrases in the Patent Skill Cartridge was to describe the phrases as universal as possible in order to also cover slightly different phrases. For example it was not sufficient to only implement the phrase A in Fig. 2 object of (the invention) . Because also common phrases like goal of the invention and aim of the invention appeared in the patent data set and had to be considered in order to retrieve the corresponding solutions[2]. In comparison to the solution key phrases the problem indicating phrases were more diverse and therefore complex to implement in the Patent Skill Cartridge. In only few cases very clear phrases like The present invention was made to solve were found. Instead a wide variety of phrases like a disadvantage of / a drawback associated with or therefore there is a need for or None of the prior attempts... were found and had to be implemented.

TABLE 1: EXCERPT OF LIST OF PROBLEM AND SOLUTION INDICATING PHRASES

Patents	Problems indicating phrases	Solutions indicating phrases
US2006042846	Background to the Invention Main challenges for	Summery for the Invention The present invention overcomes the aforentioned drawbacks for provindings
US2001568676	Discussion of the prior art While this idea is known in the prior art… do not utilize the full potential of	Field of invention The present inventions relates generally to… and specially
EP1067876	Background to the invention In general… is limited by there arises a problems that…	Summery of invention It is therefore an object of the present invention to provide
US2008316755	Background to the invention Thus it may be difficult to	Summery of the invention The present invention has been made in consideration of the foregoing.
US 2002725671	Background to the invention Common to all… therefore need … In order to maximize the efficiency therefore there is a need for	Summery of the invention In one exemeplary embodiment … is disclosed

Conclusion

Analysis of patents can be done using stemming algorithm and requirement of special patent analysis method like white spot analysis can be met. With the developed patent skill , It is possible to automatically identify text element. Like problems or solution in patents and retrieve them. There is a need for more problem and solution indicating phrases. As for the development of the cartridge only very few patents were considered (in comparison with the patent data that is already available world-wide) this result was expected. In addition the Patent Skill Cartidge should provide a systematic approach on how to differentiate between main problems, solutions, sub-problems, and sub-solutions. Concluding, the results of problems and solutions are currently not clustered or classified in a technology specific way. Especially if a lot of patents are analysed it is often advantageous to cluster the results e.g. by the use of a technology specific ontology and therefore minimize the patent map of problems and solutions in order to support the expert to work more efficiently.

REFERENCES

[1] Yvonne Siwczyk, Joachim Warschat, Dieter Spath Software-based Patent Analysis: How to Leverage a Text-mining Tool December 2014.

[2] Yuen-Hsien Tseng , Chi-Jen Lin , Yu-I Lin Text mining techniques for patent analysis 26 January 2007

[3] Nizar Ghoula, Khaled Khelif and Rose Dieng-Kuntz Supporting Patent Mining by using Ontology-based Semantic Annotations 2007.

[4] Khaled Khelif, Aroua Hedhili and Martine Collard Semantic Patent Clustering for Biomedical Communities 2008.

[5] C.Ramasubramanian1, R.Ramya Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm 2013.

[6] Peter Anick, Marc Verhagen and James Pustejovsky Extracting Aspects and Polarity from Patents 2014.

[7] Hei Chia Wang Patent Threat Analysis Search Engine 2015.