Patent Analysis using Text Mining
Sanuja Amritkar1 Gayatri Maharnor2 Prajakta Shirsath3 Aishwarya Thube4
1,2,3,4NDMVP's K.B.T.College of Engineering, Nashik, SavitribaiPhule Pune University

Abstract— Patent documents contain important research results. A patent for an invention is the grant of a property right to the inventor. Patents are granted for new, useful and non-obvious inventions for a period of 20 years from the filing date of a patent application, and provide the right to exclude others from exploiting the invention during that period. The manual process of filing a patent is very time consuming. It takes one to two years for getting a patent. Therefore we introducing a web application which takes a patent file as an input and analyze that patent file and gives the output as patent is granted if it is unique otherwise it shows that patent is not granted. For this application special text-mining techniques are used to help extracting the desired information in short time. Text mining also referred to as text data mining, it is a process of deriving high quality information from the text. Text analysis involves information retrieval, lexical analysis to study pattern recognition and information extraction. This application will give the result in few minutes; this is the advantage of this application. also patent map will be generated and The White Spot Analysis is based on a this patent map: building a problem solution matrix of patent data within a defined technology field, gaps so called White Spots can be identified. The approach presented in this paper utilizes a text mining based method in order to support the extraction of problems and solutions from patent file.

Key words: Text mining, tokenization, White Spot Analysis, Patent map generation

I. INTRODUCTION

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 technical problem and the preferred technical solution. This information can be used for example to assess the state of the ardor as a basis to identify possible gaps in a technology field. But often it is a very time consuming process to analyze the information provided by patents, because huge amounts of patents have to be considered. Therefore special text-mining tools are used to help extracting the desired information in short time. The main advantage of this approach is that the user need not read whole patent documents but is able to retrieve the relevant parts of the text in short time for further analysis steps.

It is well known that patents are used to protect the innovative ideas during research and development. And it is well known that patents are not only used by companies for reasons of protection, they are also a very powerful tool to generate new ideas and solve technical problems. So the analysis of patent information is also very important for companies and there exist several methods and tools for various patent analysis reasons.

As 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. Patent documents contain important research results. However, they are lengthy and rich in technical terminology such that it takes a lot of human efforts for analyses. Automatic tools for assisting patent engineers or decision makers in patent analysis are in great demand.

II. LITERATURE REVIEW

A. Existing Approaches:

1) System:
In the current existing system the Patent application should be filed before the appropriate patent office based on patent applicant’s or company’s location of registered office. The Indian Patent offices (IPOs) are located at Delhi, Kolkata, Mumbai and Chennai in India.

2) Process of Filling Application Form:
Required Details and patent Forms for filling of patent application in India:
1) Complete Patent specification including patent claims (patent forms)
2) Patent Drawings (if any)
3) Patent Priority Documents
4) Patent Abstract
5) International Search Report
6) Power of Attorney

Fig. 1: Traditional Patent Analysis Process

3) LUXID:
A text mining tool: this tool is used to extract the problem statement and solution statement and verifies that these statements are matched with previous statements stored in database, but the disadvantage of this tool is user have to install this tool first and then he can use it. And this tool is not easily available.
III. PROPOSED METHOD

In proposed system, Instead of filling the application form user needs to upload the patent file and it will mapped with the Available domains and the inventor will get the result that weather patent is already exist or not in short span of time. Fig (a) shows architecture diagram for system.

![Architecture Diagram](image)

**A. Text Mining:**
Text mining also referred to as text data mining, roughly equivalent to text analytics, refers to the process of deriving high quality information from the text. Text analysis involves information retrieval, lexical analysis to study word frequency distribution, pattern recognition and information extraction.

**B. Tokenization:**
Tokenization process is an integral part of IR systems, involve pre-processing of given documents and generates respective tokens.

**C. Syntax and Semantic Analysis:**
 Parsing or syntactic analysis is the process of analyzing a string of symbols, either in natural language or in computer languages, conforming to the rules of a formal grammar. The term parsing comes from Latin pars, meaning part. Semantic analysis, also context sensitive analysis, is a process in compiler construction, usually after parsing, to gather necessary semantic information from the source code. It usually includes type checking, or makes sure a variable is declared before use which is impossible to detect in parsing.

**D. White Spot Analysis:**
For the White Spot Analysis especially the description part of patents is used 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). 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.

**E. Patent Map:**
A patent map is a graphical model of patent visualization. Patent map contains the problem statement and solution statement in a tabular form. Patent mapping is essentially the visualization of the results of statistical analyses and text mining processes applied to patent documents. Patent mapping allows you to create a visual representation of information from and about patent documents in a way that is easy to understand. i.e. from patent map user will understand that in which domain more patents are filed.

![Block diagram of the system](image)

**IV. CONCLUSION**

In this way, we use a Text mining technique and White spot analysis for analyzing the patent database and generating a patent map and giving the result to the user that patent will be granted to him or not. It is possible to use a text mining for patent analysis purposes and meet the requirements of special patent analysis methods like the White Spot Analysis. 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.

![Data flow diagram of the system](image)
V. FUTURE SCOPE

The proposed systems are worked on some text mining techniques like stop word, stemming and using white spot analysis but we can also use the Natural language Processing for keyword mapping purpose.

REFERENCES