World Patent Information 37 (2014) 3–13

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

Review

A literature review on the state-of-the-art in patent analysis

Assad Abbas*, Limin Zhang, Samee U. Khan

North Dakota State University, USA

Keywords:
Patent analysis
Text mining
Visualization techniques
Natural language processing

A B S T R A C T

The rapid growth of patent documents has called for the development of sophisticated patent analysis tools. Currently, there are various tools that are being utilized by organizations for analyzing patents. These tools are capable of performing wide range of tasks, such as analyzing and forecasting future technological trends, conducting strategic technology planning, detecting patent infringement, determining patents quality and the most promising patents, and identifying technological hotspots and patent vacuums. This literature review presents the state-of-the-art in patent analysis and also presents taxonomy of patent analysis techniques. Moreover, the key features and weaknesses of the discussed tools and techniques are presented and several directions for future research are highlighted. The literature review will be helpful for the researchers in finding the latest research efforts pertaining to the patent analysis in a unified form.

1. Introduction

Contemporary advances in the technological arena have elevated the need for managing organizational knowledge scattered across diverse sources of information. A key challenge for knowledge management systems is the effective discovery and utilization of the contents stored in knowledge bases. Applying the same analogy to patents, it can be inferred that analyzing patents is essentially worthwhile to manage the complexities of searching and inter-relating patent information [1]. A patent represents an invention in a particular field of technology and also previous studies portray that a considerable part of the information presented in patents is relatively new [1,2].

With the ever-increasing volumes of patent information, the tasks of patent search and analysis have become vital from both legal and managerial perspectives [3]. Consequently, the patent data is analyzed in a variety of ways to fulfill different purposes. For instance, organizations are interested in analyzing patents for: (a) determining novelty in patents, (b) analyzing patent trends, (c) forecasting technological developments in a particular domain, (d) strategic technology planning, (e) extracting the information from patents for identifying the infringements, (f) determining patents quality analysis for R&D tasks, (g) identifying the promising patents, (h) technological road mapping, (i) identification of technological vacuums and hotspots, and (j) identifying technological competitors.

Various tools and techniques have been developed to assist patent analysis experts, business managers, and technology offices to fulfill diverse requirements. The task of analyzing the patent data using the automated tools to discover the patent intelligence through visualization, citation analysis, and other techniques, such as text mining is termed as patent informatics [54,55]. The techniques can be mainly classified into text mining techniques and visualization techniques. The text mining techniques further utilize Natural Language Processing (NLP) based approaches, semantic analysis based approaches, rules based approaches, property–function based approaches, and neural networks based approaches. On the other hand visualization techniques for patent analysis also use certain text mining approaches to present the results of patent analysis in visual form. The visual output of the task of patent analysis is in the form of patent networks, patent maps, and data clusters that emerge as a consequence of applying a particular algorithm. The patent map is a tool that is used to visualize the relationships among the patents by constructing the maps through the keywords and the key phrases [18]. The notion of patent networks is analogous to the widely used concept of networks. However, in patent networks the nodes represent the patents whereas the links in the network represent the relationship among the nodes or patents [8]. Clustering is a data mining concept that is used to group data items into clusters or groups based on their categories. The clustering techniques use unsupervised classification of data [59] and also have been used in patent analysis for clustering patent data according to the relevance.

* Corresponding author.
E-mail addresses: assad.abbas@my.ndsu.edu (A. Abbas), limin.zhang@ndsu.edu (L. Zhang), samee.khan@ndsu.edu (S.U. Khan).

0172-2190/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wpi.2013.12.006

We present a literature review on research efforts carried out for analyzing patents belonging to different technological areas. We also present taxonomy of techniques used to extract and analyze patent data. A few prior studies have surveyed the literature related to patent analysis. For example, the survey presented in by Bonino et al. [1] offer an insight to the contemporary patent informatics works. However, the study mainly focuses on semantic based techniques used for patent analysis. Another piece of research conducted by Saad and Nürnberger [4] studied prior-art cross lingual information retrieval approaches from patents. Liu [60] encompassed the visualization analysis of the patents and papers pertaining to the terahertz technology. Hanbury et al. [61] presented a survey that focuses on analysis and retrieval of the images, drawings, diagrams, charts, and plots from patent documents. In this literature review, we present a detailed review of major patent analysis techniques developed in recent years and classify them into text mining and visualization techniques. We also present taxonomy of the patent analysis techniques. Moreover, the strengths and weakness of the patent analysis techniques are presented and some possible future research directions are also highlighted.

The literature review is organized as follows. Section 2 presents the methodology adopted to conduct the literature review. Section 3 discusses the background and significance of patent analysis. Section 4 discusses the patent analysis techniques, and Section 5 concludes the literature review and highlights some future research directions.

2. Methodology

To carry out the literature review on patent analysis, the articles on patent analysis were searched through web searches. To find out the most relevant research work on patent analysis, the search was applied to databases, such as ScienceDirect, ACM digital library, IEEE digital library, and CiteSeerX. The search was narrowed using various terms, such as “patent analysis tools and techniques,” “visualization approaches in patent analysis,” “text mining and patent analysis.” The articles published in more recent years were selected for presenting a detailed discussion on tools and techniques used for patent analysis. The reference lists of the published research articles were also scanned. Out of the retrieved articles from the databases and based on the specified search criteria, a total of 22 articles were selected for detailed discussion in the literature review. Research articles published in last five years with a focus on the development of tools, techniques, and algorithms for analyzing patents using text mining and visualization techniques were selected. The purpose of presenting the research articles in detail is to provide the readers with the latest research on patent analysis in a unified form.

3. Background and significance of patent analysis

The existing patent search corpus comprises millions of patents scattered across different databases integrated through up-to-date web sources. For example, the most popular patent repositories for the patent documents are the United States Patent and Trademark Office (USPTO) [5], the European Patent Office (EPO) [6], and the Japan Patent Office (JPO) [7]. However, the ever increasing volumes of technical data pertaining to the inventions in certain fields of technology are difficult to analyze for the various purposes stated in the previous section. Therefore, one can no longer completely rely on experts’ knowledge and skills to analyze the patents [9] and this has consequently necessitated the use of computer aided tools for analyzing the patents [8]. Utilizing automated tools for patent analysis not only relieves the patent analysis experts of the laborious tasks of manually analyzing the patents but also speeds up the analysis process. Patent analysis involves a series of steps, including extracting patents from patent databases, extracting the information from the patents, and analyzing the extracted information to infer the logical conclusions. Fig. 1 shows the generic workflow of patent analysis. The patents contain various types of content, such as structured and unstructured data. The unstructured patent data comprises narrative text including the patent title, abstract, claims, and description. The structured patent data contains information, such as the inventor of the patent, assignee of the patent, and citation information [3,10].

However, patent analysis also entails some challenges related to the capabilities of the patent experts; whereas, the others pertain to the information presented in patents and the patent databases. Patent analysts with different levels of expertise require patent analysis tools with versatile capabilities [1]. On the other hand, the issues related to patent information are more complex and critical from the perspectives of searching the patent databases and retrieving information. The task of searching the patent databases to find relevant patents is supported by various data and text mining tools. Text mining tools with capabilities of mining text from structured and unstructured data have been developed. Mining the structured information from patent documents is relatively easier as compared to unstructured data because of its textual nature. Therefore, the task of parsing the unstructured data requires extraction tools having capabilities of segmentation of textual data into meaningful structures [10]. The visual output of the structured patent data is represented in the form of graphs and networks whereas the results from the unstructured patent data are represented as patent maps.

Patents are representatives of the technological innovations of a country or an organization and are indeed an agreement between the inventor of the patent and government or any agency designated by the government [11,12]. Therefore, analyzing the patents within a particular domain supports organizations in various aspects. For example, organizations or individuals interested in filing patents are often concerned whether a certain invention is indeed novel [1]. Patent analysis is beneficial for organizations in determining the novelty of their inventions, as well as identifying the Intellectual Property (IP) and technological competitiveness (strengths and weaknesses) of the competitors [13]. Besides the technological competitiveness, using IP information also helps in estimating the developments of a particular firm in a specific time interval [11]. Moreover, research in patent analysis has been conducted to determine the relationship between the technological advances and economic development. Furthermore, patent analysis is also useful in identifying the future technological trends in a specific field of technology.

In a patent that belongs to particular field of technology a substantial part of the information is relatively new that witnesses the novelty of the patent claims. On the other hand, organizations should investigate the potential infringement risks before investing in new products because patent litigation may possibly result in huge financial bearings. However, analyzing patents manually by domain experts to identify the infringements is labor intensive and time consuming for huge volumes of textual data [14]. Therefore, the process requires well defined automated procedures and tools to determine the infringements. There are two types of approaches for the patent infringement analysis. One approach works with patent citations, thereby looking at the patents cited by a target patent [15]. The other approach involves identifying similarities among the patent documents and converting the unstructured patent text into structured patent text using particular text mining techniques [14].

Several tools have been developed to assist business managers in identifying technological trends and formulating the strategies for