Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 34 (2008) 1804–1812
Expert Systems with Applications
www.elsevier.com/locate/eswa

Visualization of patent analysis for emerging technology

Young Gil Kim, Jong Hwan Suh, Sang Chan Park *

Department of Industrial Engineering, Korea Advanced Institute of Science and Technology, Guseong-dong, Yuseong-gu, Daejeon 305-701, South Korea

Abstract

Many methods have been developed to recognize those progresses of technologies, and one of them is to analyze patent information. And visualization methods are considered to be proper for representing patent information and its analysis results. However, current visualization methods for patent analysis patent maps have some drawbacks. Therefore, we propose an alternative visualization method in this paper. With colleted keywords from patent documents of a target technology field, we cluster patent documents by the *k*-Means algorithm. With the clustering results, we form a semantic network of keywords without respect of filing dates. And then we build up a patent map by rearranging each keyword node of the semantic network according to its earliest filing date and frequency in patent documents. Our approach contributes to establishing a patent map which considers both structured and unstructured items of a patent document. Besides, differently from previous visualization methods for patent analysis, ours is based on forming a semantic network of keywords from patent documents. And thereby it visualizes a clear overview of patent information in a more comprehensible way. And as a result of those contributions, it enables us to understand advances of emerging technologies and forecast its trend in the future.

Keywords: Visualization; Patent analysis; *k*-Means clustering; Semantic network; Ubiquitous computing technology

1. Introduction

It has been a critical issue to understand technological trends not only to avoid unnecessary investment but also to gain the seeds for technological development. So, many methods have been developed to recognize those progresses of technologies, and one of them is to analyze patent information. However, it is hard for non-specialists to analyze patent information because patent information is enormous and rich in technical and legal terminology. Therefore, patent information needs to be transformed into something simpler and easier to understand.

And visualization methods are considered to be proper for representing patent information and its analysis results. Generally, visualization methods are known as one of the best data mining ways to understand because graphical display methods often offer superior result compared to other conventional techniques (Westphal & Blaxton, 1998). Especially to top managers who decide directions of technology investments, visualization methods are more useful than conventional ways such as textual, tabular, and list for quick and easy knowledge discovery (Ganapathy, Ranganathan, & Sankaranarayanan, 2004).

Those visualization methods for patent analysis are called a patent map at large. A patent map is the visualized expression of total patent analysis results to understand complex patent information easily and effectively. And it is produced by gathering related patent documents of a target technology field, processing, and analyzing them (WIPO, 2003). In general, a patent document contains dozens of items for analysis which are classified into structured and unstructured item groups. Structured items are uniform in semantics and in format across a patent document such as patent number, filing date, or investors. On the other hand, the unstructured items are free texts and quite different in length and content for a patent document like claims, abstracts, or descriptions of the invention. The visualized analysis results of the former items are called

* Corresponding author. Tel.: +82 42 8692920; fax: +82 42 8693110.
E-mail address: sangchanpark@kaist.ac.kr (S.C. Park).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.01.033

patent graphs, and those of the latter are called patent maps, although loosely a patent map may refer to both cases (Liu, 2003) Likewise, current visualization methods for patent analysis are based on mapping patent information and its analysis results.

However, current patent maps have some drawbacks. First, most of them are time based, ranking based or matrix maps which consider only one aspect between structured items or unstructured items of each patent document. More integrated and balanced visualization approach is required to provide the overall structure of patent information effectively. Second, they are complicated networks of patent documents though they use different methods. As a result, they make patent analysis results incomprehensible and unclear to analyzer. Consequently, those deficient patent maps fail to provide an intuitive insight into the concerned technology field. And this is the third drawback we'd like to make mention of. Especially for an emerging technology like ubiquitous computing and bio informatics, it is essential to recognize its advance and make an estimate of the hereafter.

Hence, we propose an alternative visualization method for patent analysis to overcome drawbacks of current patent maps. Our approach contributes to establishing a patent map which considers both structured and unstructured items of a patent document. We expect to keep the balance of analysis features by using filing dates as structured items and keywords and its frequency as unstructured items. Besides, differently from previous visualization methods for patent analysis, ours is based on forming a semantic network of keywords from patent documents. And thereby it visualizes a clear overview of patent information in a more comprehensible way. And as a result of those contributions, it enables us to understand advances of emerging technologies and forecast its trend in the future.

The rest of the paper is structured as follows. In Section 2, we introduce related works with visualization methods for patent information and its analysis results. In Section 3, we explain an overview of our approach. In Section 4, we apply our visualization method to develop a patent map of the ubiquitous computing technology as an emerging technology field. And in Section 5, finally we conclude the paper with a discussion of the proposed patent map's implications in the ubiquitous computing technology.

2. Literature review

Related to researches using patent information, there are two mainstreams. One of them is to study visualization methods for patent information and its analysis results. And in this paper, we are interested in the former research area. This research area has attracted attention of the persons concerned. That's because current technological development necessitates it to avoid unnecessary investment as well as gaining the seeds for technological development and the applicable fields. Also, the attention is increased to promote the efficient use of patent information by deepening related institutions' understanding of patent information.

On the basis of these awareness, the Japan Patent Office has been producing and providing more than 50 types of expressions and more than 200 maps for several technology fields since 1997 (JIII, 2000). In addition, many other countries such as Korea (Ryoo & Kim, 2005), Italy (Camus & Brancaleon, 2003; Fattori, Pedrazzi, & Turra, 2003) and the USA (Morris, DeYong, Wu, Salman, & Yemenu, 2002) also provide many kinds of patent maps.

Currently, most patent analyses use patent citations to represent the meaningful relationship in patent information. But it is known that patent citation analysis has some serious drawbacks. According to Yoon and Park (2004), there are four drawbacks of patent citation analysis described. First, it is difficult to grasp the overall relationship among patent documents. Second, related to the first problem, the scope of analysis and the richness of potential information are limited. Third, citation has no capability of considering internal relationship between patent documents. Finally, citation analysis is a time-consuming task because it needs only an exhaustive search. So, Yoon and Park (2004) proposed a networ*k*-based patent analysis as an alternative method. But the network patent analysis still has some limitations.

Researches on the intelligent methods for patent analysis have been made as well. The neural methods for mapping scientific and technical information (articles, patents) and for assisting a user in carrying out the complex process of analyzing large quantities of such information are concerned by Lamirel, Shehabi, Hoffmann, and Francois (2002). Based on text mining techniques, Tseng, Wang, Juang, and Lin (2005) created a real world patent map for an important technology domain: carbon nano-tube experimentally.

And the other mainstream is concerned about patent classification. By Black and Ciccolo (2004), machine learning technology is applied to text classification on United States' patent information to automatically differentiate between patents relating the biotech industry and those unrelated. Fall, Törcsvári, Fiévet, and Karetka (2004) reported the results of applying a variety of machine learning algorithms for training expert systems in Germa*n*-language patent classification tasks. And Trappey, Hsu, Trappey, and Lin (2006) developed a platform for patent document classification and search using a bac*k*-propagation network.

3. Methodology

Our visualization method steps are summed up as follows. With colleted keywords from patent documents of a target technology field, we cluster patent documents by the *k*-Means algorithm. With the clustering results, we form a semantic network of keywords without respect of