Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of High Technology Management Research 15 (2004) 37–50
THE JOURNAL OF HIGH TECHNOLOGY MANAGEMENT RESEARCH

A text-mining-based patent network: Analytical tool for high-technology trend

Byungun Yoon, Yongtae Park*

Department of Industrial Engineering, School of Engineering, Seoul National University, San 56-1, Shillim-Dong, Gwanak-Gu, Seoul 151-742, South Korea

Accepted 22 September 2003

Abstract

Patent documents are an ample source of technical and commercial knowledge and, thus, patent analysis has long been considered a useful vehicle for R&D management and technoeconomic analysis. In terms of techniques for patent analysis, citation analysis has been the most frequently adopted tool. In this research, we note that citation analysis is subject to some crucial drawbacks and propose a network-based analysis, an alternative method for citation analysis. By using an illustrative data set, the overall process of developing patent network is described. Furthermore, such new indexes as technology centrality index, technology cycle index, and technology keyword clusters are suggested for in-depth quantitative analysis. Although network analysis shares some commonality with conventional citation analysis, its relative advantage is substantial. It shows the overall relationship among patents as a visual network. In addition, the proposed method provides richer information and thus enables deeper analysis since it takes more diverse keywords into account and produces more meaningful indexes. These visuals and indexes can be used in analyzing up-to-date trends of high technologies and identifying promising avenues for new product development.

Keywords: Patent network; Technoeconomic analysis; Citation analysis

1. Introduction

Patent documents are an ample source of technical and commercial knowledge in terms of technical progress, market trend, and proprietary ownership and, thus, patent analysis has long been considered as a useful vehicle for R&D management in corporate setting and technoeconomic analysis in macro

* Corresponding author. Tel.: +82-2-880-8358.
E-mail address: parkyt@cybernet.snu.ac.kr (Y. Park).

1047-8310/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.hitech.2003.09.003

context. Furthermore, patents facilitate analytical work due to their relative advantages, vis-à-vis other indexes, with respect to availability of database, scope of coverage, and richness of information (Kuznets, 1962; OECD, 1994). Recently, the strategic importance of patent analysis is more highlighted in high-technology management as the process of innovation becomes more complex, the cycle of innovation becomes shorter, and the market demand becomes more volatile.

The research spectrum and practical applicability of patent analysis is wide among a variety of technology agents—R&D managers, academicians, and policy makers. In a macro sense, patent analysis has often been employed to generate economic indicators that gauge the linkage between technology development and economic growth (Grandstrand, 1999; Grilliches, 1990; Holl, Jaffe, & Trajtenberg, 2000), estimate technological knowledge flows and their impact on productivity (Evenson & Puttnam, 1988; Scherer, 1982), or compare innovative performance in international context (Paci, Sassu, & Usai, 1997). In a micro level, patent analysis has been used to evaluate the competitiveness of firms (Narin & Noma, 1987), develop technology plans (Mogee, 1991), prioritize R&D investment (Hirschey & Richardson, 2001), or monitor technological change in firms (Archibugi & Pianta, 1996; Basberg, 1987).

In general, patent analysis utilizes bibliometric data that include such information as patent number, type of document, title, inventor, international patent classification, date of application, and so forth (Gupta & Pangannaya, 2000). Then, a number of techniques may be used to manipulate and analyze bibliometric data. Amongst others, the most frequently adopted tool is patent citation analysis (Narin, 1994). Patent citations are defined as the count of citations of a patent in subsequent patents, and citations per patent represent the relative importance of the patent. Based on this idea, patent citation analysis executes a bibliometric analysis on patent documents. In essence, the methodology is a citation-based technique in that it attempts to link patents in a patent database in the same way as science citation analysis links references in a scientific paper database (Karki, 1997). Ultimately, patent citation analysis produces such technological indexes as citations per patent, highly cited patents, nonpatent link, technical impact index, current impact index, technology cycle time, and so forth. These indexes then have been used as measures of quality of technical assets (Hirschey & Richardson, 2001), negotiation power between firms (Mowery, Oxley, & Silverman, 1998), economic value of innovative outputs in market value equation (Holl et al., 2000), or domestic or cross-border technology linkages and knowledge flows (Tijssen, 2001).

Patent citation analysis, albeit easy to understand and simple to use, is subject to some serious drawbacks. First, it is very difficult to grasp the overall relationship among all the patents because citation analysis merely indicates individual links between two particular patents. Second, related to the first problem, the scope of analysis and the richness of potential information are limited because citation analysis takes only citing–cited information into account. Third, citation analysis has no capability of considering internal relationship between patents. It takes only existence or frequency of citations into account and hence may produce superficial or even misleading indexes. Finally, citation analysis is a time-consuming task because it needs an exhaustive search.

Recognizing the shortcomings of citation analysis, the main objective of current research is to propose a network-based patent analysis, an alternative method for patent citation analysis. Although network analysis shares some commonality with citation analysis, its relative advantage is substantial. First, network analysis shows the relationship among patents as a visual network and therefore assists the analyzer in intuitively comprehending the overall structure of a patent database. Second, network analysis enriches the potential utility of patent analysis because it takes more diverse keywords into account and produces more meaningful indicators. Third, the proposed method is more economical, in