Journal of Informetrics 12 (2018) 217–236

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Deriving technology intelligence from patents: Preposition-based semantic analysis☆

Jaehyeong An[a], Kyuwoong Kim[b], Letizia Mortara[c], Sungjoo Lee[b,*]

[a] Technology Intelligence Team, Hyundai NGV, Gwanak-ro 1, Gwanak-gu, Seoul 08826, Republic of Korea
[b] Department of Industrial Engineering, Ajou University, Worldcup-ro 206, Yeongtong-gu, Suwon 16499, Republic of Korea
[c] Centre for Technology Management, Institute for Manufacturing, University of Cambridge, Department of Engineering, Alan Reece Building, 17 Charles Babbage Road, Cambridge CB3 0FS, UK

ARTICLE INFO

Article history:
Received 21 December 2016
Received in revised form 3 January 2018
Accepted 3 January 2018
Available online 22 February 2018

Keywords:
Technology intelligence
Technology search
Technology trends
Patent analysis
Semantic
Preposition
Text mining
Key-words
Text mining

ABSTRACT

Patents are one of the most reliable sources of technology intelligence, and the true value of patent analysis stems from its capability of describing the content of technology based on the relationships between keywords. To date a number of techniques for analyzing the information contained in patent documents that focus on the relationships between keywords have been suggested. However, a drawback of the existing keyword approaches is that they cannot yet determine the types of relationships between the keywords. This study proposes a novel approach based on preposition semantic analysis network which overcomes the limitations of the existing keywords-based network analysis and demonstrates its potential through an application. A preposition is a word that defines the relationship between two neighboring words, and, in the case of patents, prepositions aid in revealing the relationships between keywords related to technologies. To demonstrate the approach, patents regarding an electric vehicle were employed. 13 prepositions were identified which could be used to define 5 relationships between neighboring technological terms: "inclusion (utilization)," "objective (purpose)," "effect," "process," and "likeness." The proposed approach is expected to improve the usability of keyword-based patent analyses and support more elaborate studies on patent documents.

1. Introduction

Technology intelligence is defined as "the capture and delivery of technological information as part of the process whereby an organization develops an awareness of technological threats and opportunities" (Kerr, Mortara, Phaal, & Probert, 2006, p.75). Understanding the technological landscape and identifying the changes in the landscape could be one of the core technology intelligence activities considering that companies that fail to adapt to the changing technological environment are more likely not to survive. Patent documents largely consist of technical terms, using which the characteristics of the inventions are clearly explained; if the relationships between the keywords are well identified, the keyword analysis can be a useful tool for revealing the overall technological structure of the invention and, furthermore, the technological landscape.

☆ This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIP) (NRF-2013R1A2A2A03016904) and also by the Ajou University research fund (grant number: S-2015-G0001-00333).

* Corresponding author.
E-mail addresses: jaehyeongan@hyundai-ngv.com (J. An), dongul@ajou.ac.kr (K. Kim), lm367@cam.ac.uk (L. Mortara), sungjoo@ajou.ac.kr (S. Lee).

https://doi.org/10.1016/j.joi.2018.01.001
1751-1577/

Patents are one of the most reliable sources of information on technologies, and hence, they have been used as an essential tool for analyzing technological developments (Pilkington, Lee, Chan, & Ramakrishna, 2009; Zhang, 2011). Accordingly, patent analysis has been frequently employed in innovation studies (Park, Yoon, & Lee, 2005), for example for the identification of industry, technology, and competitor trends; investigation of emerging technologies; and the search for potential collaborators (Ding, 2011; Mogee, 1991; Narin, 1993; Reitzig, 2004). In particular, the recent advancement in data analysis techniques through text mining has enabled the use of the descriptive parts of patent documents for analysis, extending the scope of patent analysis significantly, receiving considerable attention (Choi, Kim, Yoon, Kim, & Lee, 2013; Yoon, Park, & Kim, 2013; Yoon et al., 2015).

To extract meaningful implications from text-mining-based patent analysis, a comprehensive selection of keywords is needed. The keywords chosen for the analysis should represent the content of all the concerned patent documents, and they should be sufficiently specific to distinguish one document from another. The general approach is to use a specific algorithm, such as term-frequency–inverse-document-frequency (TF–IDF) analysis or latent semantic analysis (Li, Wang, & Hong, 2009; Tonta & Darvish, 2010), or to use an expert panel with domain knowledge to choose only significant keywords (Jeon, Lee, & Park, 2011; Lee, Kang, & Shin, 2015). Once a keyword set is properly defined, further analysis can be conducted based on the set.

Among various types of analyses, this study focuses on the analysis of the relationships between the keywords; the true value of text-mining-based patent analysis lies in its ability to describe the technological content of a patent based on the relationships between the keywords. The relationship between the keywords may change the meaning of the patent documents. After recognizing the significance of keyword relationships; a subject-action-object (SAO) analysis that considers sentence structure for analyzing keywords was proposed (Moehrle, Walter, Geritz, & Müller, 2005) and applied to patent analysis (Choi, Park, Kang, Lee, & Kim, 2012; Guo, Wang, Li, & Zhu, 2016; Park, Kim, Choi, & Yoon, 2013; Park, Ree, & Kim, 2013; Wang et al., 2015). The SAO analysis defines the relationship between two keywords (subject and object) using a verb (action). However; despite its contribution and potential; because of the variety of verbs often used in patents; SAO inevitably increases the complexity of the analysis and cannot be used to easily investigate the technological content of a patent document. In addition; the SAO analysis focuses primarily on functional relationships between the components of a technical system (Cascini, Fantechi, & Spinicci, 2004; Park, Kim, et al., 2013; Park, Ree, et al., 2013); while other non-functional relationships can also be investigated to offer valuable technology intelligence from patents.

To overcome the limitations of the existing keywords-based network analysis such as SAO, this study proposes a novel approach called preposition-based semantic analysis and applies it to patent analysis for understanding the relationship between technologies, developing a technology-relation-technology (TRT) network. This method is based on the observation that number of prepositions, i.e. words that define the relationship between two neighboring words, are very limited in the English grammar. Their employment would make it easier to define the relationship between technological terms than using verbs. A preposition can also be used to describe a non-functional relationship—referred to as structural relationship in this study—which can be used to explain the relationships "between the whole and a part" or "between parts of the whole."

Thus, the research question posed in this study is: how can semi-automatic methods for deriving technical insight from large number of patents be improved by using semantic analysis based on the evaluation of how technology relevant terms are linked by prepositions? In response we develop and then test a new method (TRT), comparing the insight it obtains on a particular case-study with that obtained with other state-of-the-art keyword-based methods (e.g., SAO).

The rest of this paper is organized as follows. Section 2 reviews patent-based network analyses and existing approaches for developing a network. In Section 3, based on the background knowledge, a new approach (TRT) proposed in this study is described by employing the case of electric vehicles. Section 4 introduces the method for applying TRT. Finally, Section 5 discusses the limitations of this study and future research directions.

2. Background

2.1. Patent network analysis

Networks have been widely used as a way to systematically structure an enormous amount of data, visualize the relationships between analysis units, and even quantify such relationships and the effects of one unit on the other (Borgatti & Cross 2003, Borgatti, Mehra, Brass, & Labianca, 2009; Kho et al., 2013; Otte & Rousseau, 2002). A network consists of *nodes* and *links*: nodes are a unit of analysis that can be defined at various levels for patent documents, including keywords, papers, and authors (Noyons, 2001). Link represent the relationships between the nodes and are defined using several analysis indicators such as co-citation, co-authorships, and co-words (Chen, Huang, Hsieh, & Lin, 2011; Lee & Jeong, 2008; White & McCain, 1997).

Existing studies have stressed on the use of networks to investigate the technological structure, technological changes, and emerging technologies, as a patent-based network helps to extract valuable technological insights from patents. For the circumstances wherein new technologies emerge by combining the knowledge of different fields (Curran & Leker, 2011), a network has been used as a critical tool for facilitating technology intelligence activities. Although various resources, including patents, publications, and project documents, are available for network analysis, patents are one of the most frequently adopted resources as they provide fruitful information regarding technology innovation, being outcomes of research and development (R&D) activities (Beneito, 2006; Lee, Kim, & Cho, 2010; Sun, Lu, Wang, Ma, & He, 2008).