Expert Systems with Applications 39 (2012) 11443–11455

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

An SAO-based text mining approach to building a technology tree for technology planning

Sungchul Choi [a], Hyunseok Park [b], Dongwoo Kang [c], Jae Yeol Lee [d], Kwangsoo Kim [c,*]

[a] CTO Office, Samsung Advanced Institute of Technology Mt. 14-1, Nongseo-dong, Giheung-gu, Yongin-si Gyeonggi-do, 446-712, Republic of Korea
[b] Department of Technology and Innovation Management, Pohang University of Science and Technology, San 31, Hyoja-dong, Nam-gu, Pohang, 790-784, Republic of Korea
[c] Department of Industrial and Management Engineering, Pohang University of Science and Technology, San 31, Hyoja-dong, Nam-gu, Pohang, Kyungbuk 790-784, Republic of Korea
[d] Department of Industrial Engineering, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju 500-757, Republic of Korea

ARTICLE INFO

Keywords:
Technology tree
Subject–Action–Object (SAO)
Function
Patent mining
Patent analysis
Technology trends analysis

ABSTRACT

A technology tree (TechTree) is a branching diagram that expresses relationships among product components, technologies, or functions of a technology in a specific technology area. A TechTree identifies strategic core technologies and is a useful tool to support decision making in a given market environment for organizations with specified capabilities. However, existing TechTrees generally overemphasize qualitative and expert-dependent knowledge rather than incorporating quantitative and objective information. In addition, the traditional process of developing a TechTree requires vast amounts of information, which costs considerably in terms of time, and cannot provide integrated information from a variety of technological perspectives simultaneously. To remedy these problems, this research presents a text mining approach based on Subject–Action–Object (SAO) structures; this approach develops a TechTree by extracting and analyzing SAO structures from patent documents. The extracted SAO structures are categorized by similarities, and are identified by the type of technological implications. To demonstrate the feasibility of the proposed approach, we developed a TechTree regarding Proton Exchange Fuel Cell technology.

1. Introduction

A technology tree (TechTree) is a branching diagram (tree structure) that represents relationships among technologies. Generally, a TechTree is used to represent the relationships among product components, technologies, or functions of a technology in a specific technology area. A TechTree is a useful tool for supporting a decision making process to identify strategic core technologies in a given market environment and organization capabilities. Furthermore, it helps to define element technologies based on identified core technologies. Due to these advantages, TechTrees are important decision-making tools for technology planning (Cheong, 2006), and are utilized in technology planning (Aude & Kahn, 1986; Choudhury & Fallah, 2009; Durand, 1992; Guglielmi, Williams, Groepper, & Lascar, 2010; Visentin, 2008; Yoon, Phaal, & Probert, 2008).

The current process of developing a TechTree has the following limitations. The first is exclusive dependence on experts during the development process. Currently, most TechTree diagrams are created by reflecting the opinions of technology domain experts. However, developing a TechTree by this process has considerable costs in terms of time and human resources; and obtaining the participation of an expert can be difficult (Heiss & Jankowsky, 2001). Recently, new technologies have continuously appeared and replaced old technology. To establish strategic technology planning that considers the rapid technological changes, a TechTree should be constantly updated to reflect the changes. However, because of the current process, updating and managing TechTrees are difficult. The second limitation is that current TechTrees only represent limited technology information. A TechTree can represent various technology perspectives such as product taxonomy (Aude & Kahn, 1986; Cascini & Zini, 2008; Durand, 1992; Yoon et al., 2008), technology taxonomy (Choudhury & Fallah, 2009; Durand, 1992; Guglielmi et al., 2010; Visentin, 2008), and function taxonomy of technology (Cascini & Zini, 2008; Cheong, 2006). However, the TechTree represents technology information from only the technology perspective for which the tree was developed. Another limitation is that existing TechTrees cannot represent subsidiary information related to technologies. In technology analysis, the subsidiary information of technology is as important as the technology itself. We called this 'technology meta-information'. This information includes the inventor of the technology,

* Corresponding author. Tel.: +82 54 279 2195; fax: +82 54 279 5998.
E-mail addresses: blissray@postech.ac.kr (S. Choi), howgood@postech.ac.kr (H. Park), hyunil@postech.ac.kr (D. Kang), jaeyeol@chonnam.ac.kr (J.Y. Lee), kskim@postech.ac.kr (K. Kim).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.04.014

the time at which the technology was developed, or where it was developed. By integrating these various perspectives of technology and technology meta-information, we can generate a TechTree that helps decision-making in technology planning more effectively than do existing TechTrees.

To address these problems, we suggest using Subject–Action–Object (SAO)-based text mining techniques to develop TechTrees. This approach analyzes patent documents and exploits the analysis results to create a TechTree that represents technology perspectives. Patent information includes valuable up-to-date technological information and is continuously updated. Patents also include bibliometric information and this information can be utilized as technology meta-information.

The proposed approach consists of a procedure for constructing source data from patents and a method of developing a TechTree using the data. The procedure uses Natural Language Processing (NLP) technology to extract SAO structures from patents, and text-mining techniques to analyze the SAO structures. The resulting information is used as the source data for development of a technology tree. The method develops a TechTree that represents various technology perspectives such as product taxonomy, technology taxonomy, and function taxonomy of technology. The suggested TechTree is represented with technology meta-information.

The rest of paper is organized as follows. In Section 2 we describe work related to the proposed approach. In Section 3 we provide a detailed description of our approach. In Section 4 we illustrate the proposed framework by describing a case study of developing a TechTree that represents the technology used in Proton Exchange Membrane Fuel Cells (PEMFCs). In Section 5 we provide concluding remarks and directions for further study.

2. Related work

2.1. Technology tree

Although a TechTree includes the word "technology", it does not mean the only tree structure for representing technological information. TechTrees can be classified into three categories based on the purpose for which they are developed: product taxonomy, technology taxonomy, and function taxonomy of technology.

A 'product taxonomy' TechTree depicts components used in a product and their interrelatedness. Aude and Kahn (1986) used a TechTree to describe structures of a silicon product, and Yoon et al. (2008) used a TechTree to represent the components of a mobile phone. Other researchers used TechTrees to express product taxonomy information (Buderi, 2000; Cascini & Zini, 2008; Durand, 1992).

A 'technology taxonomy' TechTree depicts technologies applied to a product and their interrelatedness (Lee, Kang, Park, & Park, 2008). Lee et al. (2008) used a TechTree to represent the technology taxonomy of Broadband Network a preliminary step to creating a technology roadmap. The European Space Agency (ESA) used a TechTree to describe technology taxonomy. ESA reviewed the European and worldwide trends in the Space Automation and Robotics field to identify long term development trends, and used a TechTree represent the outcome (Guglielmi et al., 2010; Visentin, 2008). Choudhury and Fallah (2009) used the VND (Veterans, Newbies and Downgraded) model and a "green TechTree" to show the general trajectory followed by the green energy sector.

A 'function taxonomy of technology' TechTree depicts the functions of technology applied to products and technologies and their interrelatedness. The function of technology presents the information regarding the purposes of technologies, the effects of technologies, or the interrelationships of products. Cascini and Zini (2008) used a computer-aided patent analyzer system (PatAnalyzer) to develop functional TechTrees of products. Unlike previous studies, this research examined ways to use computing resources to generate a TechTree. Cheong (2006) suggested a functional TechTree used at the Samsung Advanced Institute of Technology (SAIT). SAIT used the functional TechTree to identify the core functions of technologies by representing the interrelationships among core functions and secondary functions that support the core functions.

2.2. SAO structures for technology analysis

In this paper, an SAO structure is used to construct a TechTree. SAO structures are composed of Subject (noun phrase), Action (verb phrase) and Object (noun phrase). SAO structures are commonly used to represent functions of technology. This simple structure explicitly describes a relationship between components that appear in a patent text. For example, an SAO structure can simply represent the function of a battery as "Battery Energizes Bulb". In this example, "Battery" is the subject, "Energizes" is the action, and "Bulb" is the object. The technology purpose of "Battery" is to energize "Bulb" and the function of "Battery" is "To energize Bulb". Recently, many researchers have attempted to determine how to use SAO structures for technology analysis (Moehrle, 2010; Sternitzke & Bergmann, 2009; Yoon & Kim, 2012; Yoon, Lim, Choi, Kim, & Kim, 2011).

SAO structures can be used to represent a variety of technology information. Subjects and objects may refer to components of a system; actions may refer to functions performed by and on components (Cascini, Fantechi, & Spinicci, 2004a). Moreover, an SAO structure can be organized in a problem–solution format if the action–object (AO) states the problem and the subject (S) forms the solution (Moehrle, Walter, Geritz, & Müller, 2005). SAO structure also states partative relationships among products or technologies (Cascini, Lueehesi, & Rissone, 2001). If the action word is partative verb such as such as "have", "composed of", and "be made of", the component of a subject may include the object as a component.

To computationally extract SAO structure from patent textual descriptions, use of natural language processing (NLP) is necessary (Cascini, Fantechi, & Spinicci, 2004b) in the Part-Of-Speech tagging process. Because previous studies regarding NLP have been already conducted to extract SAO structures from a sentence (Charniak et al., 1996; Choi, Lim, Yoon, & Kim, 2010; Liu & Singh, 2004; Tsourikov, Batchilo, & Sovpel, 2000), we use methods described in those studies to extract SAO structures.

2.3. WordNet based sentence similarity

With SAO structures, we use WordNet-based Sentence Similarity (Simpson & Dao, 2005) to determine extracted SAO structures. To utilize large amounts of SAO structures from patent text, a clustering method is needed that can categorize the SAO structures by using similarities. Because an SAO structure is a simple sentence, the WordNet-based Sentence Similarity method is a useful tool to calculate similarities among SAO structures.

In general, the measurement process of assessing semantic similarity between two sentences is composed of (1) tokenizing the sentences, (2) stemming words, (3) tagging parts of speech, (4) determining the most likely meaning of each word in each sentence, and (5) computing the sentence similarity based on the similarity between pairs of corresponding words (Simpson & Dao, 2005). A measure of similarity (*sim*) between two concepts (c_1, c_2) is defined as follows (Resnik, 1999):

$$sim(c_1, c_2) = \frac{2 \times depth\ (lcs(c_1, c_2))}{depth\ (c_1) + depth\ (c_2)}, \tag{1}$$

where *lcs* is the lowest common subsume of c_1 and c_2, and *depth* is the distance from a concept node c_i to the root of a concept hierarchy, and $0 < sim(c_1, c_2) \leqslant 1$ where 1 means that the concepts are