Scientometrics (2011) 88:213–228
DOI 10.1007/s11192-011-0383-0

# Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks

**Janghyeok Yoon · Kwangsoo Kim**

Received: 29 November 2010 / Published online: 9 April 2011


**Abstract** Patents constitute an up-to-date source of competitive intelligence in technological development; thus, patent analysis has been a vital tool for identifying technological trends. Patent citation analysis is easy to use, but fundamentally has two main limitations: (1) new patents tend to be less cited than old ones and may miss citations to contemporary patents; (2) citation-based analysis cannot be used for patents in databases which do not require citations. Naturally, citation-based analysis tends to underestimate the importance of new patents and may not work in rapidly-evolving industries in which technology life-cycles are shortening and new inventions are increasingly patented worldwide. As a remedy, this paper proposes a patent network based on semantic patent analysis using subject-action-object (SAO) structures. SAO structures represent the explicit relationships among components used in a patent, and are considered to represent key concepts of the patent or the expertise of the inventor. Based on the internal similarities between patents, the patent network provides the up-to-date status of a given technology. Furthermore, this paper suggests new indices to identify the technological importance of patents, the characteristics of patent clusters, and the technological capabilities of competitors. The proposed method is illustrated using patents related to synthesis of carbon nanotubes. We expect that the proposed procedure and analysis will be incorporated into technology planning processes to assist experts such as researchers and R&D policy makers in rapidly-evolving industries.

**Keywords** Patent mining · Semantic patent similarity · Subject-action-object (SAO) structures · Patent network · Research and development (R&D) trend · Natural language processing (NLP)

J. Yoon · K. Kim (✉)
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
e-mail: kskim@postech.ac.kr

J. Yoon
e-mail: janghyoon@postech.ac.kr

**JEL classification** C63 · C82

## Introduction

Patent documents constitute up-to-date and reliable source of knowledge for reflecting technological advance, so patent analysis has been widely used for identification of technological trends and formulation of technology strategies (Narin 1993; Reitzig 2004). Many studies have adopted patent analysis to assist researchers and practitioners to plan technology development (Liu and Shyu 1997), to detect emerging technologies (Daim et al. 2006; Ernst 1997; Trappey et al. 2011; Shibata et al. 2008), to assess innovation (Abraham and Moitra 2001) and to identify potential technological opportunities (Dewulf 2006; Porter and Detampel 1995; Yoon and Park 2005; Lee et al. 2009).

Among methods of patent analysis, citation-based patent analysis has been most widely adopted due to its simplicity and ease of use. The idea of citation-based analysis is that patents cited by many later patents have a strong possibility of containing important ideas upon which many later inventors are building; many studies have revealed a positive relationship between technological importance and the number of citations in later patents (Albert et al. 1991; Breitzman and Thomas 2002; Narin 1993).

To identify up-to-date trends in technological change, new patents should be included in the analysis. However, patents that are new or the under patenting process have less chance to be cited than older ones due to the time lag of the patenting process. This indicates that citation-based patent analysis may not work well for relatively new patents. Furthermore, some of the new patents share a common technological basis, but they may not cite patents that were applied for concurrently. Therefore, the importance of new patents tends to be underestimated in citation-based patent analysis. Another practical limitation of citation-based analysis is that the customary citation-based analysis can never successfully incorporate information from patents in databases which do not require citations, such as those of Japan and Korea. These problems impede efficient analysis of research and technology trends in rapidly-evolving industries, in which technology life-cycles are shortening and new inventions are increasingly patented world-wide.

As a remedy, this paper proposes a patent network based on semantic patent analysis. Because the proposed approach is based on the analysis of patent contents, it reveals the importance and an overall relationships among patents in a given technology area, while avoiding the limitations presented in the previous paragraph. In addition, even though patents are not granted, the proposed method can identify relationships among them if they are accessible online on the patent databases; citation-base approach can use only granted patents but not patents under the patent process. For semantic patent similarity measurement, the proposed method extracts subject-action-object (SAO) structures using natural language processing (NLP) of patent text. SAO structures are the key concept that shows a relationship between components used in the relevant patent (Cascini et al. 2004); the set of SAO structures is considered to represent a far more detailed picture of the inventor's expertise and capacities (Moehrle et al. 2005). Textual information of each patent contains many SAO structures that are the basis for measuring semantic similarities among patents.

First, this paper proposes a procedure for generating a patent network based on semantic patent similarity; this process consists of (1) collecting patent data sources, (2) extracting SAO structures from patent text, (3) computing patent similarities by measuring semantic sentence similarities between SAO structures, and (4) visualizing a patent network.

Although the automatically generated patent network gives an intuitive understanding of a given patent set, this paper proposes new indices for further analysis of the proposed patent network including the technological importance of patents, the characteristics of patent clusters and the technological capabilities of competitors. The procedure and analysis are illustrated using a patent set related to synthesis of carbon nanotubes (CNTs). The proposed analysis indices can be used as valuable information for decision-making of researchers and R&D policy makers. Therefore, we expect that the proposed method will be incorporated into technology planning processes to assist experts in rapidly-evolving industries.

The second section presents an overview of SAO-based patent analysis and network-based patent analysis, and the third section gives a description of the data source used in this paper. The fourth section proposes a procedure for generating semantic analysis-based patent networks, and the fifth section identifies technological implications from the patent network using the proposed analysis indices. Finally, the last section presents conclusions and future research.

## Related work

The proposed method visualizes the relationships between patents as a network using SAO structures extracted from patent text, and then identifies technological insight from the network. Therefore, we first review SAO-based patent analysis and network-based patent analysis.

### SAO-based patent analysis

The method proposed in this paper is based on semantic similarities between SAO structures in patents, which are the syntactically ordered structure of subject (noun phrase), action (verb phrase) and object (noun phrase). This structure clearly provides a relationship between components that appear in technological text. Given a simple sentence ‘soap cleans hands’, a subject is ‘soap’, an action is ‘cleans’ and an object ‘hands’. The action ‘cleans’ explicitly represents a structural relationship between the subject ‘soap’ and the object ‘hands’. SAO structures are fundamentally related to the concept of function, which is defined as “the action changing a feature of any object” (Savransky 2000). Therefore, in a technological document, subjects and objects may refer to components of a system, and actions may refer to functions performed by and on components (Cascini et al. 2004; Choi et al. 2010). SAO structures assist designers to decompose and analyze a complex system effectively (Cascini et al. 2004; Mann 2002). Especially, the SAO structures extracted from claims of a patent are considered to represent intensive knowledge related to the inventor’s expertise and the patent’s key findings (Moehrle et al. 2005).

NLP is a useful tool to extract SAO structures from patent text. The PAT-Analyzer, a customized NLP tool for design structure analysis of patents, identifies SAO structures of a patent and visualizes the design structure of the patent as an SAO-based network using subjects and objects as nodes, and actions as links (Cascini et al. 2004). This visual provides researchers with an intuitive understanding of a patent’s technological fundamentals. Measurements of the similarity between SAO networks can be used to identify similarity between patents (Cascini and Zini 2008). Using Knowledgist2.5™ (www.invention-machine.com), a commercial linguistic analyzer, several SAO-based studies have researched patent map-based merger and acquisition strategies (Moehrle and Geritz 2004), patent-based inventor profiles for human resource decisions (Moehrle et al.

2005), patent infringement risk evaluation (Bergmann et al. 2008) and product forecasting (Gerken et al. 2010).

Although various SAO-based approaches have been suggested, patent-level network analysis using SAO structures has not been proposed yet. Therefore, this paper aims at visualizing internal relationships among patents as a patent network and on acquiring technological insights from the network.

### Patent citation network analysis

Network analysis allows identification of characteristics of a small complex system which consists of actors (nodes) and interactions (links). In most cases, network-based patent analysis uses patents as the actors and cited-citing relationships as the interactions. Citations reveal meaningful implications of individual patents. Patents that are cited by many later patents have a strong possibility of being an original invention, and patents that cite numerous previous patents are likely to be among the advanced inventions (Narin 1994). If patents cite or are cited by various science articles, the technology related to the patents is being actively researched in both academies and industries (Karki 1997).

Generating patent citation networks provides general information regarding how inventions are related to each other in a given patent set (Hung and Wang 2008). By incorporating bibliographic information of patents, the patent citation network facilitates understanding of the knowledge transfer among technical fields, institutions and countries (Chang et al. 2009; Fontana et al. 2009; Huang et al. 2003a; Huang et al. 2003b; Hung and Wang 2010a; Li et al. 2007). For more in-depth analysis, statistical approaches can be used to extract technological implications from a given patent set. Such analysis applies a variety of network analysis techniques including centrality analysis of patents (Hung and Wang 2010b), and identification and density analysis of technology clusters (Chang et al. 2009; Li et al. 2007).

However, as stated in the introduction, citations are easily obtainable, but they have some limitations in rapidly-evolving industries. Therefore, this paper proposes a method that includes use of semantic patent analysis based on SAO structures to uncover the relationships among patents related to a given technology area.

## Data source

CNTs are allotropes of carbon that have a cylindrical nanostructure. These cylindrical molecules exhibit extraordinary strength and unique electrical properties, and are efficient thermal conductors, so they have many potential applications in nanotechnology, electronics, optics and material science. CNT technology is evolving rapidly, and over the past decade, many inventions in this field have been patented. To illustrate the proposed SAO-based method, this paper located 136 patents related only to CNT synthesis methods from patent databases of Europe (EU), Japan (JP), Korea (KR) and the United States (US) by exploiting CNT-related text and application date conditions (Table 1). Because the SAO-based analysis in this paper considers only English text, text in KR and JP databases was translated into English text using language translation services such as K2E-PAT (KIPO 2010).

The representation format of application numbers differs among patent databases, so the gathered patents were labeled from P1 (oldest patent JP1992-172242) to P136 (newest patent JP2009-223352) in application date order. Of these patents, 84.2% were issued or applied for after 2000, 20.9% were issued from 2005 to 2010 inclusive, and 50.7% are still

**Table 1** Patent retrieval query for EU and US patent databases; similar patent retrieval queries for JP and KR patent databases were represented in Japanese and Korean

| Patent retrieval query | Collected patents |
|---|---|
| (((carbon* adj nanotube) (carbon* adj nano* adj tube*) (carbonnanotube*) CNT* carbonnano* (carbon* adj nano*) (carbon* adj tube*) carbontube* MWCNT* SWCNT*) and (((DC* adj arc* adj discharge*) (arc-discharge*)) ((plasma* adj CVD*) PECVD* (plasma* adj (chemvapor* adj depos*)) (plasma* adj enhance* adj CVD*)) ((therm* adj chemvapor* adj despos*) TCVD* (therm* adj CVD*)))) AND @AD <=20100430 | Total 139 patents (after removing irrelevant patents) |

pending. The data source was obtained by searching not only the US patent database but also three other patent databases, so the customary citation-based analysis cannot be applied; most US patents of the data source were rarely cited and the citation is not included in the three other patent databases. Therefore, this paper identifies technological implications from the data source using the proposed SAO-based method as an alternative.

## Development procedure of patent networks

The procedure for developing patent networks based on semantic patent similarity is composed of four steps (Fig. 1): (1) patent data source is collected, (2) SAO structures are extracted by using syntactic analysis of patent text, (3) a patent similarity matrix is computed by measuring semantic sentence similarities between SAO sets of pairs of patents, and (4) the patent similarity matrix is visualized as a network. For in-depth analysis, this paper analyzes the patent network using the indices proposed in the next section.

### Collection of patent data source

The proposed patent network is generated using a specific patent set. Gathering of patents requires a patent retrieval query that is composed of textual information related to a target

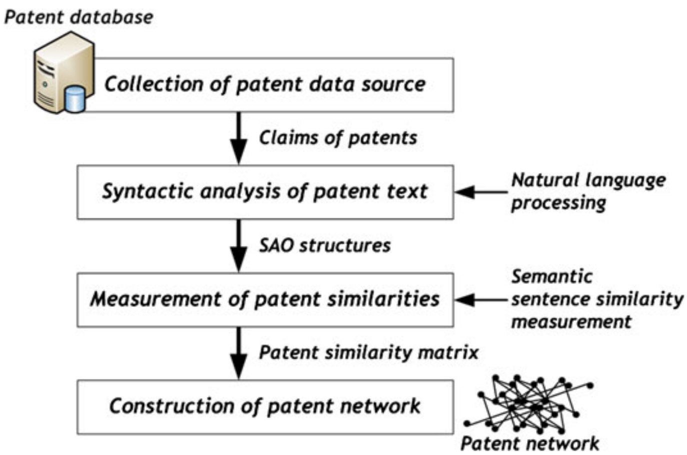


**Fig. 1** Overall procedure for construction of patent networks

technology and bibliographic information such as international patent code, applicants and application date. Next, a final patent set for analysis can be prepared by eliminating irrelevant patents.

Each patent has various sections: some contain numerical information including patent number and application date; some are specific pieces of text information including inventors, country and applicants; the others are narrative sections including title, abstract, background summary, detailed description and claims. Among the various sections, syntactic analysis of patents can use only narrative sections. Several studies have considered the claims of the patent most comprehensively during patent processing because they state clearly the innovative knowledge that requires legal protection (Fujii et al. 2007; Yoon et al. 2011). Therefore, the proposed method uses only claims among the narrative sections. The data source is converted to Microsoft Excel format to be used for automated extraction of SAO structures in the next step.

Syntactic analysis of patent text

In this step, SAO structures are output after syntactic analysis of patent text. Although many NLP tools can be used to extract SAO structures from patent text, this paper exploits Knowledgist2.5™, which is a textual language processing system, and which has been used in several recent content-based patent analyses (Bergmann et al. 2008; Moehrle et al. 2005). The system processes text to find, extract, and organize concepts from documents in a way that is similar to that by which humans infer meaning. By exploiting NLP, the system first extracts syntactically ordered SAO structures from patent text, then filters out duplicated or irrelevant SAO structures using a set of stopwords, followed by screening by a human expert. Finally, 1174 SAO structures (average 8.63 per patent) were extracted from the given patent set. As a complete sentence, each SAO structure explicitly reveals a relationship among components including tools and materials used in the patent (Table 2).

Measurement of semantic similarity between patents

In this step, a patent similarity matrix is generated by computing semantic similarities between pairs of patents (Fig. 2). Because the similarity between two patents is based on the similarity between SAO structures of the patents, semantic sentence similarity between two SAO structures must be measured.

**Table 2** An example of SAO extraction (Patent EP02749069)

| S (subject) | A (action) | O (object) |
|---|---|---|
| Carbon | Contain | Gas plasma |
| Chemical vapor deposition | Use | Carbon |
| Group | Comprise | Nitrogen, hydrogen argon and ammonia |
| Microwave energy | Generate | Plasma |
| Plasma chamber | Cool | Electrodes |
| Powder | Have | Particle size |
| CVD chamber | Inject | Catalyst |
| Vacuum chamber | Generate | Gaseous plasma |
| Vacuum chamber | Maintain | Gaseous plasma |

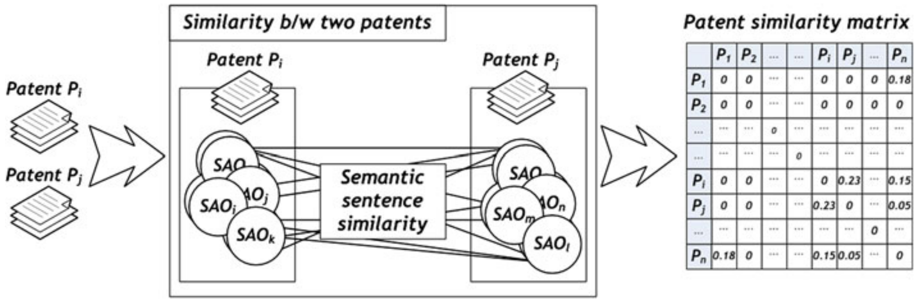


**Fig. 2** Concepts for generation of patent similarity matrices

In general, the process of measuring the semantic similarity of two sentences entails five steps: (1) tokenizing the sentences, (2) word stemming, (3) part-of-speech tagging, (4) determining the most likely meaning of each word in each sentence, and (5) computing the similarity of the sentence pair based on the similarities between pairs of corresponding words (Simpson and Dao 2005). First, a measure of similarity between two concepts (Resnik 1999) is defined as follows:

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

where *lcs* is the lowest common subsume of two concepts $c_1$ and $c_2$, and *depth* is the distance from a concept node $c_i$ to the root of a concept hierarchy, and $0 < sim(c_1, c_2) \leq 1$ where 1 means that the concepts are identical. Building on the measure of similarity between two concepts, the matching average (Simpson and Dao 2005) to compute similarity between sentences $X$ and $Y$ is:

$$MatAvg(X, Y) = \frac{2 \times Match(X, Y)}{|X| + |Y|}, \tag{2}$$

where $|\cdot|$ ($\cdot = X$ or $Y$) is the number of set tokens in the sentence, $Match(X, Y)$ is the sum of similarity of the matching word tokens between sentences $X$ and $Y$, and $0 < MatAvg(X, Y) \leq 1$, where 1 means that the sentences are identical. Using the WordNet semantic dictionary (Miller 1995) as the concept hierarchy, the measure for similarity measurement between two words (1), and the matching average for similarity measurement between two sentences (2), a.NET based semantic sentence similarity measurement has been implemented as a C# library (Simpson and Dao 2005). WordNet defines a concept hierarchy of most words; for example, the chemical symbol for element ‘Mg’ is the same as ‘magnesium’. However, it does not contain abbreviations that are very domain-specific; for example, in the field of CNT research ‘SWNT’ represents ‘single walled nanotube’ and ‘CVD’ means ‘chemical vapor deposition’. Therefore, after grouping synonyms of the data source, we modified the C# library source code to refer to the defined synonym set.

To determine whether two SAO structures are same or not, a threshold value $t$ is used. For example, when $t = 0.90$ and $MatAvg(SAO_i, SAO_j) = 0.92$, then the two SAO structures can be considered semantically the same. Determination of similarity between $SAO_i$ and $SAO_j$ is:

$$SAO_{ij} = \begin{pmatrix} 1, & if\ MatAvg(SAO_i, SAO_j) \geq t \\ 0, & otherwise \end{pmatrix}. \tag{3}$$

Finally, a measure of similarity between patents $A$ and $B$ can be simply defined based on how many SAO structures the two patents share:

$$SIM_{AB} = \frac{2 \times N_{SAO}(A, B)}{N_{SAO}(A) + N_{SAO}(B)}, \tag{4}$$

where $N_{SAO}(X)$ is the number of SAO structures in the patent $X$, and similarity $N_{SAO}(X, Y)$ is the number of the semantically identical SAO structures shared by patents $X$ and $Y$. The similarity score of two patents is $0 \leq SIM_{AB} \leq 1$, where 1 means that they are identical. By measuring the semantic similarities between all pairs of patents, a patent similarity matrix can be obtained that codifies the internal relationships among the patents. Given that the size of the patent set is $M$, an $M \times M$ (in this paper, $136 \times 136$) patent similarity matrix is generated. In the present paper, $t$ was set to 0.8.

Construction of patent networks

This step visualizes the patent similarity matrix as a patent network using network formulation techniques and NetMiner3.0 (www.cyram.com). Generating a well-displayed network often requires a sensitivity analysis, and good visuals provide an easily-understandable illustration of the relationships among patents in the given set. To construct a patent network, only valid relationships between patents are considered using a cut-off value $P$. Two patents are strongly related if $SIM_{AB} \geq p$, and weakly or unrelated otherwise. Experts must set a reasonable cut-off value to obtain a well-displayed patent network. By trial-and-error, we found that $P = 0.53$ provided a network that clearly displays relationships among patents in the data source (Fig. 3).

## Analysis of patent networks

Although the automatically generated patent network gives an overall understanding, the network becomes too complex to be understood intuitively as the number of patents increases. Furthermore, identifying further implications from the network requires statistical analysis. Therefore, using newly proposed indices and methods, this section suggests ways to identify the technological importance of patents, the characteristics of patent clusters and the technological capabilities of competitors.

Technological importance of patents

Within network analysis, the *degree* of a node is defined as the number of links or the sum of values of links incident to the node (Diestel 2005), and the *centrality* of a node is defined as its proximity to the center of the network (Freeman 1979). These two concepts are often used to determine the relative importance of a node in a network. For example, the most popular student in a friendship network has many intimate friends or is located nearest the center in the network. Likewise, this paper adopts degree and centrality to identify the relative importance of patents. Instead of using citations, the two proposed indices are based on the semantic relatedness, i.e., the degree to which a patent shares the same

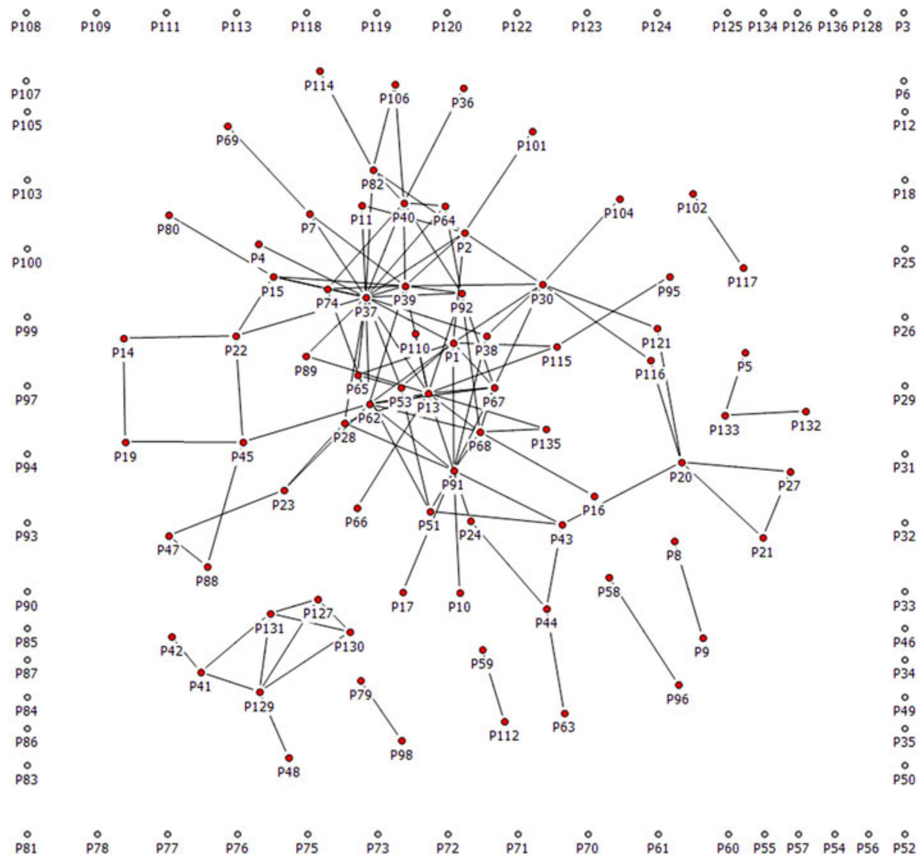


**Fig. 3** Visualization of the patent network ($P = 0.53$)

technological basis with others. Using the definition of degree, the degree sum index (*DSI*) of a patent $i$ in a patent network is defined as:

$$DSI_i = \sum_{j=1}^{n} link_{ij}, \tag{5}$$

where $link_{ij}$ is the semantic similarity score between patents $i$ and $j$. Some network-based research indicated a positive relationship between the degree of technological concepts and the technological importance (Yoon et al. 2011). Because a patent with high *DSI* is directly related to various similar patents, it has a strong possibility of being a dominant invention in the technology area. Next, the global centrality index (*GCI*) of a patent $i$ in a patent network is defined using closeness centrality:

$$GCI_i = \left(\sum_{j=1}^{n} distance_{ij}\right)^{-1}, \tag{6}$$

where $distance_{ij}$ is the shortest path between patents $i$ and $j$. Because a patent with high *GCI* is strongly related to various patents in an overall sense, it has a possibility of being a widely-used or leading invention. From a technological perspective, an old patent with high *DSI* and *GCI* is likely to be a basis for other later patents related to the patent, whereas a new patent with high *DSI* and *GCI* has a strong possibility of being a more-advanced technology that is based on earlier patents.

Patents with high *DSI* and *GCI* were identified from the patent network (Table 3). Overall, patents applied for or granted from 1999 to 2004 were found to be important, and actually during this period many researchers and companies applied for patents for inventions that use arc-discharge and laser vaporization (RPCCT 2010). Many of the applicants of patents with high *DSI* and *GCI* are among the major companies involved in synthesis of CNTs including NEC, SONY and SAMSUNG SDI. For example, P2 has relatively high *DSI* and *GCI*; it is an original patent of NEC, which is a leading company with original technology in the arc-discharging synthesis of CNTs. Several new patents applied for from 2005 to 2008 also had high *DSI* and *GCI*. New patents with high *GCI* such P110 and P135 imply that chemical vapor deposition is replacing arc-discharge and laser vaporization as a CNT-synthesis method. Although P110 patented by SAMSUNG SDI is under the patenting process, the patent is considered to be important for large-scale production because it proposes a new apparatus to remove production delays caused by frequent replacement of electrodes. Therefore, the proposed indices help experts to identify patents that have a strong possibility of being key patents, based on semantic relatedness between patents. We expect that the proposed indices will be a valuable input for evaluating the technological value and potential of new patents in rapidly-evolving industries.

## Characteristics of patent clusters

*Density* is a measure of the characteristics of networks; it is defined as the proportion of ties in a network relative to the total number possible. Because the density indicates how closely all nodes in a network are related each other, it measures the completeness of networks. Using methods such as Bi-Component (Hopcroft and Tarjan 1973) and K-Core (Bollobas 1983), the patent network can be clustered into one or more sub-networks. Applying the concept of density to the patent network, implications from a technological perspective can be formulated (Table 4). If a patent cluster has large size and high density, many inventions have been already developed and many similar patents exist in the cluster. Therefore, the cluster has a strong possibility of being a technologically verified or

**Table 3** Patents with high *DSI* and *GCI*

| Patent label$_{\text{(real application number)}}$ | DSI | Patent label$_{\text{(real application number)}}$ | GCI |
|---|---|---|---|
| P37$_{\text{(JP2001-031190)}}$ | 12.8 | P37$_{\text{(JP2001-031190)}}$ | 0.200 |
| P13$_{\text{(KR1999-0014307)}}$ | 7.78 | P62$_{\text{(JP2002-314127)}}$ | 0.188 |
| P91$_{\text{(JP2004-060928)}}$ | 6.24 | P39$_{\text{(JP2001-031238)}}$ | 0.188 |
| P39$_{\text{(JP2001-031238)}}$ | 6.17 | P13$_{\text{(KR1999-0014307)}}$ | 0.182 |
| P62$_{\text{(JP2002-314127)}}$ | 5.95 | P67$_{\text{(JP2002-378840)}}$ | 0.182 |
| P30$_{\text{(JP2000-163489)}}$ | 5.5 | P2$_{\text{(JP1993-337937)}}$ | 0.170 |
| P129$_{\text{(KR2008-0064667)}}$ | 4.93 | P91$_{\text{(JP2004-060928)}}$ | 0.166 |
| P67$_{\text{(JP2002-378840)}}$ | 4.78 | P115$_{\text{(JP2006-088419)}}$ | 0.166 |
| P2$_{\text{(JP1993-337937)}}$ | 4.65 | P135$_{\text{(JP2009-059208)}}$ | 0.164 |
| P20$_{\text{(JP2000-000299)}}$ | 3.26 | P110$_{\text{(KR2005-0069649)}}$ | 0.143 |

**Table 4** Density analysis of patent clusters

| Type of patent cluster | Characteristics of patent cluster |
|---|---|
| Large size; high density | Sub-domain of technologically verified or dominant inventions; many inventions already developed; many similar patents |
| Large size; low density | Sub-domain of many promising inventions; inventions not yet dominant and technologically verified |
| Small size; high density | Sub-domain of novel and fresh inventions; weak signals about new types of technology |

dominant patent cluster in a given technology area. If a patent cluster has large size and low density, it may not be dominant or technologically verified, but it is likely to be a promising technology area because no dominant designs exist and many designs are being invented. If a patent cluster has small size and high density, it has a possibility of containing novel and fresh inventions with respect to a specific technology. Therefore, the cluster may signal development of new technology.

From the CNT patent network, eight patent clusters were identified using the Bi-Component method (Fig. 4), and their members, size and density were obtained (Table 5). Technological implications could be identified from patent clusters G1 and G2. Cluster G1 has a large number of patents including P135, P121, P116 and P115, and its density is low. Therefore, the possibility is high that innovative methods of synthesizing CNTs including arc-discharging, laser vaporization, plasma enhanced chemical vapor deposition and thermal chemical vapor deposition, are still being invented frequently; the fact that 50.7% of the patents are still pending emphasizes the leading-edge status of this branch of CNT research. Patent cluster G2 is composed of P131, P130, P129, P127, P48, P42 and P41; it is small and very dense, so its members are strongly related each other and it may represent a novel method. Bibliographic information of the members of G2 reveals that one individual applicant (4 patents) and one company (2 patents) applied for patents for similar methods related to arc-discharge synthesis of CNTs. From patents of the individual applicant in G2, we could identify development of a novel invention. Although these patents are slightly different from each other in that they are based on different methods including pulse-based arc-discharging and magnetic field-based arc-discharging, they are characterized in common by the consecutive injection of carbon rods and the selective collection of CNTs by degree of purity. The injection module stores several consumable carbon rods and injects them consecutively into reactor chambers without production delays. Furthermore, the CNT collection module scrapes CNTs coagulated on the inner wall of the reaction chambers and stores them selectively by degree of purity. Therefore this invention can dramatically improve productivity of synthesis process and effectively collect high purity CNTs. In fact, one of the recent tasks in arc-discharging synthesis is large-scale and cost-effective production of CNTs because arc-discharging synthesis has been technologically verified.

## Technological capabilities of competitors

As a customary technological capability index, the current impact index (*CII*) is a normalized indicator of the importance of an applicant's patents, using how often they have been cited in other patents (Narin 1993). If *CII* of a company the technology area is higher than those of its competitors, its patents are better in quality in that area. However, as stated in the introduction, *CII* may not address the up-to-date status of a patent set composed of

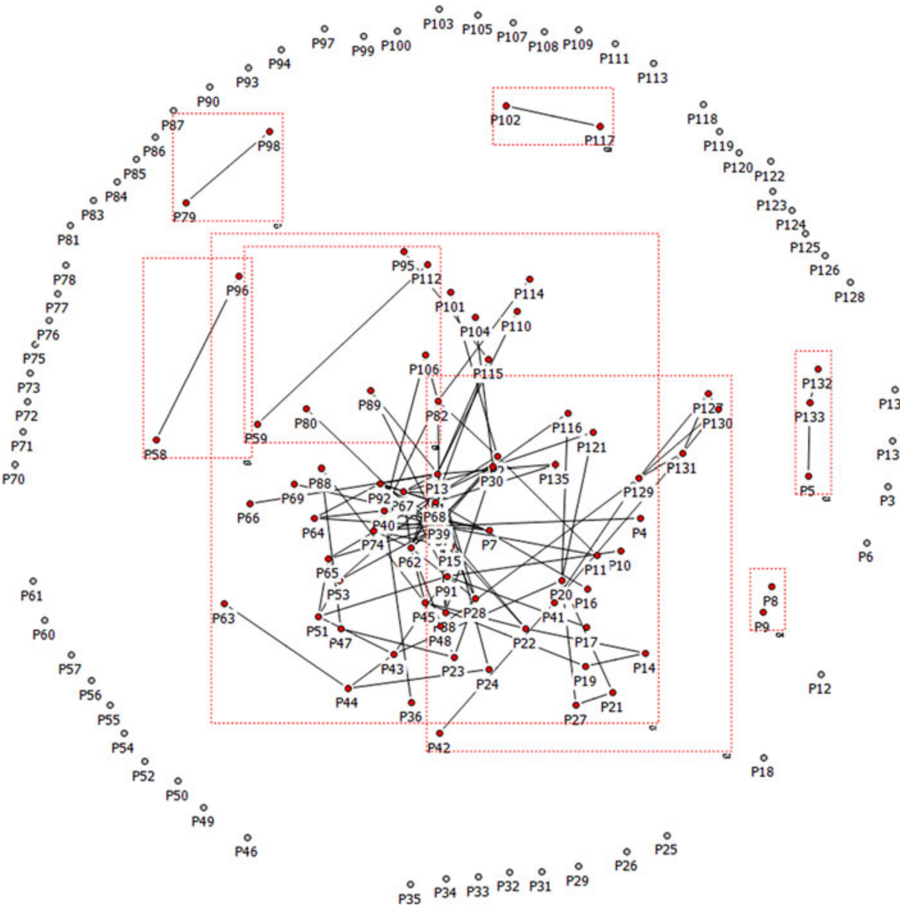


**Fig. 4** Patent clusters of the patent network

new patents and cannot successfully incorporate information from patent databases (e.g. Japan and Korea patent databases) which do not include citations. Therefore, using *GCI* (6), this paper proposes the technological impact index (*TII*) of an applicant *X*:

$$TII_X = \frac{avg(\sum_{i \in P(X)} GCI_i)}{avg(\sum_{i \in P(ALL)} GCI_i)}, \tag{7}$$

where *P(X)* is the set of patents of the applicant X, and *avg* is the average value of *GCI*s. This index represents the average qualitative capability of inventions that an applicant possesses. The technology strength index (*TSI*) is a customary index of the quantitative and qualitative technological capability of an applicant. *TSI* is the number of patents an applicant has obtained multiplied by *CII* (Kayal and Waters 1999). If a company has a higher *TSI* than its competitors, the company is more powerful in that technology area. Because *TSI* is computed using *CII*, *TSI* also has the limitations of *CII*. Therefore, using *TII*, this paper proposes the technological capability index (*TCI*) of an applicant *X*:

$$TCI_X = TII_X \times N_p(X), \tag{8}$$

**Table 5** Patent clusters identified from the patent network

| Patent cluster | Members [Patent label$_{\text{(real application number)}}$] | Size | Density |
|---|---|---|---|
| G1 | P135$_{\text{(JP2009-059208)}}$, P121$_{\text{(KR2007-001257)}}$, P116$_{\text{(JP2006-210551)}}$, P115$_{\text{(JP2006-088419)}}$, P114$_{\text{(JP2006-003909)}}$, P110$_{\text{(KR2005-0069649)}}$, P106$_{\text{(JP2005-082385)}}$, P104$_{\text{(JP2004-280179)}}$, P101$_{\text{(JP2004-217677)}}$, P95$_{\text{(JP2004-091701)}}$, P92$_{\text{(JP2004-106991)}}$, P91$_{\text{(JP2004-060928)}}$, P89$_{\text{(JP2004-025540)}}$, P88$_{\text{(KR2004-0006775)}}$, P82$_{\text{(JP2003-371351)}}$, P80$_{\text{(JP2003-197339)}}$, P74$_{\text{(JP2003-101183)}}$, P69$_{\text{(US2003-335691)}}$, P68$_{\text{(KR2002-0086799)}}$, P67$_{\text{(JP2002-378840)}}$, P66$_{\text{(JP2002-359005)}}$, P65$_{\text{(KR2002-0078168)}}$, P64$_{\text{(JP2002-331816)}}$, P63$_{\text{(JP2002-331815)}}$, P62$_{\text{(JP2002-314127)}}$, P53$_{\text{(KR2002-0031444)}}$, P51$_{\text{(JP2002-086858)}}$, P47$_{\text{(KR2001-0060349)}}$, P45$_{\text{(JP2001-219202)}}$, P44$_{\text{(KR2001-0040444)}}$, P43$_{\text{(KR2001-0037496)}}$, P40$_{\text{(JP2001-041628)}}$, P39$_{\text{(JP2001-031238)}}$, P38$_{\text{(JP2001-031217)}}$, P37$_{\text{(JP2001-031190)}}$, P36$_{\text{(JP2001-028987)}}$, P30$_{\text{(JP2000-163489)}}$, P28$_{\text{(KR2000-0019559)}}$, P27$_{\text{(JP2000-108319)}}$, P24$_{\text{(KR2000-0013039)}}$, P23$_{\text{(KR2000-0012099)}}$, P22$_{\text{(JP2000-002091)}}$, P21$_{\text{(JP2000-000300)}}$, P20$_{\text{(JP2000-000299)}}$, P19$_{\text{(JP1999-229091)}}$, P17$_{\text{(KR1999-0030700)}}$, P16$_{\text{(KR1999-0030697)}}$, P15$_{\text{(JP1999-205447)}}$, P14$_{\text{(JP1999-159180)}}$, P13$_{\text{(KR1999-0014307)}}$, P11$_{\text{(JP1999-008392)}}$, P10$_{\text{(JP1998-301259)}}$, P7$_{\text{(JP1998-082409)}}$, P4$_{\text{(JP1994-278287)}}$, P2$_{\text{(JP1993-337937)}}$, P1$_{\text{(JP1992-172242)}}$ | 56 | 0.072 |
| G2 | P131$_{\text{(KR2008-0064686)}}$, P130$_{\text{(KR2008-0064683)}}$, P129$_{\text{(KR2008-0064667)}}$, P127$_{\text{(KR2007-0121569)}}$, P48$_{\text{(JP2001-353793)}}$, P42$_{\text{(JP2001-191424)}}$, P41$_{\text{(JP2001-056330)}}$ | 7 | 0.476 |
| G3 | P133$_{\text{(KR2008-0087841)}}$, P132$_{\text{(KR2008-0087840)}}$,P5$_{\text{(JP1996-003636)}}$ | 3 | 0.667 |
| G4 | P9$_{\text{(JP1998-159520)}}$, P8$_{\text{(JP1998-121022)}}$ | 2 | 1 |
| G5 | P96$_{\text{(US2004-827915)}}$, P58$_{\text{(US2002-237695)}}$ | 2 | 1 |
| G6 | P112$_{\text{(US2005-246063)}}$, P59$_{\text{(US2002-237729)}}$ | 2 | 1 |
| G7 | P98$_{\text{(JP2004-170379)}}$, P79$_{\text{(JP2003-132770)}}$ | 2 | 1 |
| G8 | P117$_{\text{(JP2006-224896)}}$, P102$_{\text{(JP2004-243755)}}$ | 2 | 1 |

where $N_p$ is the number of patents of the applicant *X*. Similarly to *TSI*, *TCI* represents the quantitative and qualitative capability of inventions which an applicant is making an application for.

From the patent network, applicants with high *TII* and *TCI* were identified (Table 6). Overall, well-known companies such as SAMSUNG, LG, SONY and NEC had strong technological capability in the synthesis of CNTs. Interestingly, ULVAC had relatively low *TII* in the list but was ranked second in *TCI*. In fact, it is one of the leading companies in vacuum technology related to the display and semiconductor industries. For the last decade, the company has patented arc-discharging methods, and lately it is applying for patents for plasma enhanced chemical vapor deposition methods using its own vacuum technologies such as sputtering and vacuum coating. Therefore, this suggests that the company has a strong possibility of being an emerging competitor. In this way, by analyzing *TII* and *TCI*, R&D policy makers can understand the current status of applicants' technological capability. We expect that the proposed indices will assist experts to analyze the technological capability of potential partners or emerging competitors.

## Discussions and future research

As a supporting tool for technology planning, patent analysis has been actively researched. Although citation-based patent analysis has been widely adopted due to its simplicity and ease of use, it has two main limitations: it tends to underestimate the importance of new patents, and it cannot successfully incorporate information from patents in databases which

**Table 6** Applicants with high *TII* and *TCI*

| Applicant | TII | Applicant | TCI |
|---|---|---|---|
| SAMSUNG SDI | 1.53 | SONY CORP | 12.13 |
| LG ELECTRONICS | 1.36 | ULVAC CORP | 7.74 |
| SONY CORP | 1.35 | NKK CORP | 5.98 |
| JANG (PERSONAL APPLICANT) | 1.32 | JANG (PERSONAL APPLICANT) | 5.28 |
| NEC CORP | 1.30 | SAMSUNG SDI | 4.60 |
| LEE (PERSONAL APPLICANT) | 1.28 | ISE ELECTRONICS CORP | 4.32 |
| SHIZUOKA UNIV | 1.22 | NEC CORP | 3.89 |
| HANYANG UNIV | 1.14 | LG ELECTRONICS | 2.72 |
| ISE ELECTRONICS CORP | 1.08 | LEE (PERSONAL APPLICANT) | 2.57 |
| ULVAC | 0.65 | SHIZUOKA UNIV | 2.45 |

Applicants with only one or two patents were ignored to avoid skewing the technological capabilities

do not include citations. To solve the limitations, this paper proposed a patent network based on SAO-based semantic patent analysis using NLP. SAO structures, which are syntactically ordered structures of subject, action and object, explicitly reveal both the key findings of an invention and intensive knowledge related to the inventor's expertise. Using SAO-based semantic patent similarities instead of citations, this paper proposed a procedure for generation of the patent network that is composed of (1) collecting patent data sources, (2) extracting SAO structures from patent text, (3) computing patent similarities based on semantic sentence similarities between SAO structures, and (4) visualizing a patent network.

This network based on internal relationships of patents gives an overall understanding of a given patent set. Additionally, this paper proposed new indices and methods for further analysis of the patent network, i.e., the technological importance of patents (*DSI* and *GCI*), the characteristics of patent clusters (density analysis), and the technological capabilities of competitors (*TII* and *TCI*). Specifically, analyses of semantic patent networks will be useful for identifying technological trends in rapidly-evolving industries in which technology life-cycles are shortening and new inventions are frequently patented world-wide. Insights from the analysis can be a valuable source for R&D trend analysis, so we expect that the proposed method can be incorporated into technology planning processes to assist researchers and R&D policy makers in fast-moving industries.

Despite those advantages, the proposed method has some challenges. First, the method exploited several existing tools including commercial software for extraction of SAO structures, measurement of semantic sentence similarities and visualization of networks. However, these tools need to be incorporated and developed on an integrated platform for further practicality of this research. Second, this research defined only a synonym set to measure semantic similarities of SAO structures because we analyzed only a specific technological field. Further work should define domain-specific concept hierarchies for more accurate measurement of structural and semantic similarities among patents. Third, in this research, analysis indicators were based only on degree, closeness centrality and density. However, further research will examine technological implications of the proposed patent networks using other network indicators such as betweenness centrality and structural holes. Finally, this paper only considered inventions related to synthesis of CNTs. Therefore, a future topic will investigate identifying technological trends from several rapidly-evolving fields with technological overlaps. In this case, the proposed method will assist experts to identify technological missing links, i.e., convergence technologies, among different fields.

**Acknowledgments** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2009-0088379).

## References

Abraham, B., & Moitra, S. (2001). Innovation assessment through patent analysis. *Technovation, 21*(4), 245–252.

Albert, M., Avery, D., Narin, F., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research Policy, 20*(3), 251–259.

Bergmann, I., Butzke, D., Walter, L., Fuerste, J., Moehrle, M., & Erdmann, V. (2008). Evaluating the risk of patent infringement by means of semantic patent analysis: The case of DNA chips. *R&D Management, 38*(5), 550–562.

Bollobas, B. (1983). *The evolution of sparse graphs (Graph theory and combinatorics)*. London: Academic Press.

Breitzman, A., & Thomas, P. (2002). Using patent citation analysis to target/value ma candidates. *Research-Technology Management, 45*(5), 28–36.

Cascini, G., Fantechi, A., & Spinicci, E. (2004). Natural language processing of patents and technical documentation. *Lecture Notes in Computer Science*, *3163*, 508–520.

Cascini, G., & Zini, M. (2008). Measuring patent similarity by comparing inventions functional trees. *Computer-Aided Innovation (CAI), 277*, 31–42.

Chang, S., Lai, K., & Chang, S. (2009). Exploring technology diffusion and classification of business methods: Using the patent citation network. *Technological Forecasting and Social Change, 76*(1), 107–117.

Choi, S., Lim, J., Yoon, J., & Kim, K. (2010). *Patent function network analysis: A function based approach for analyzing patent information*. Cairo, Egypt: IAMOT 2010.

Daim, T., Rueda, G., Martin, H., & Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change, 73*(8), 981–1012.

Dewulf, S. (2006). *Directed variation: Variation of properties for new or improved function product DNA, a base for 'connect and develop'*. Kortrijk, Belgium: ETRIA TRIZ Futures.

Diestel, R. (2005). *Graph theory (graduate texts in mathematics)*. Heidelberg: Springer.

Ernst, H. (1997). The use of patent data for technological forecasting: The diffusion of cnc-technology in the machine tool industry. *Small Business Economics, 9*(4), 361–381.

Fontana, R., Nuvolari, A., & Verspagen, B. (2009). Mapping technological trajectories as patent citation networks. An application to data communication standards. *Economics of Innovation and New Technology, 18*(4), 311–336.

Freeman, L. (1979). Centrality in social networks conceptual clarification. *Social networks, 1*(3), 215–239.

Fujii, A., Iwayama, M., & Kando, N. (2007). Introduction to the special issue on patent processing. *Information Processing and Management, 43*(5), 1149–1153.

Gerken, J., Moehrle, M., & Walter, L. (2010). Patents as an information source for product forecasting: Insights from a longitudinal study in the automotive industry. In *The R&D Management Conference 2010*, Manchester, England: Managing science and technology.

Hopcroft, J., & Tarjan, R. (1973). Algorithm 447: Efficient algorithms for graph manipulation. *Communications of the ACM, 16*(6), 372–378.

Huang, Z., Chen, H., Yip, A., Ng, G., Guo, F., Chen, Z., et al. (2003a). Longitudinal patent analysis for nanoscale science and engineering: Country, institution and technology field. *Journal of Nanoparticle Research, 5*(3), 333–363.

Huang, M., Chiang, L., & Chen, D. (2003b). Constructing a patent citation map using bibliographic coupling: A study of taiwan's high-tech companies. *Scientometrics, 58*(3), 489–506.

Hung, S., & Wang, A. (2008). A small world in the patent citation network. In *International Conference on IEEM 2008* (pp. 1–5). Singapore: IEEM.

Hung, S., & Wang, A. (2010a). Examining the small world phenomenon in the patent citation network: A case study of the radio frequency identification (rfid) network. *Scientometrics, 82*(1), 121–134.

Hung, S., & Wang, A. (2010b) *Small world phenomenon in the radio frequency identification (rfid) patent citation network* (pp. 23–26). Kyoto, Japan: International Conference on Mechanical and Electronics Engineering (ICMEE).

Karki, M. (1997). Patent citation analysis: A policy analysis tool. *World Patent Information, 19*(4), 269–272.

Kayal, A., & Waters, R. (1999). An empirical evaluation of the technology cycle time indicator as a measure of the pace of technological progress in superconductor technology. *IEEE Transactions on Engineering Management, 46*(2), 127.

KIPO. (2010). *K2e automatic translation*. Retrieved from http://eng.kipris.or.kr/eng/other_service/k2e_automatic_translation.jsp. Accessed 1 Oct 2010.

Lee, S., Yoon, B., & Park, Y. (2009). An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation, 29*(6–7), 481–497.

Li, X., Chen, H., Huang, Z., & Roco, M. (2007). Patent citation network in nanotechnology (1976–2004). *Journal of Nanoparticle Research, 9*(3), 337–352.

Liu, S., & Shyu, J. (1997). Strategic planning for technology development with patent analysis. *International Journal of Technology Management, 13*(5), 661–680.

Mann, D. (2002). *Hands-on systematic innovation*. Belgium: Creax press.

Miller, G. (1995). Wordnet: A lexical database for english. *Communications of the ACM, 38*(11), 41.

Moehrle, M., & Geritz, A. (2004). Developing acquisition strategies based on patent maps. In *Proceedings of the 13th International Conference on Management of Technology,* (pp. 1–9). Washington, USA: R&D Management.

Moehrle, M., Walter, L., Geritz, A., & Muller, S. (2005). Patent-based inventor profiles as a basis for human resource decisions in research and development. *R&D Management, 35*(5), 513–524.

Narin, F. (1993). Technology indicators and corporate strategy. *Review of Business, 14*(3), 19–23.

Narin, F. (1994). Patent bibliometrics. *Scientometrics, 30*(1), 147–155.

Porter, A., & Detampel, M. (1995). Technology opportunities analysis. *Technological Forecasting and Social Change, 49*(3), 237–255.

Reitzig, M. (2004). Improving patent valuations for management purposes—validating new indicators by analyzing application rationales. *Research Policy, 33*(6–7), 939–957.

Resnik, P. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of artificial intelligence research, 11*, 95–130.

RPCCT. (2010). *Carbon nanotue (cnt): Recent trends*. Daejun, Korea: Research and Policy Center for Chemical Technology.

Savransky, S. (2000). *Engineering of creativity: Introduction to triz methodology of inventive problem solving.* Boca Raton, USA: CRC.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation, 28*(11), 758–775.

Simpson, T., & Dao, T. (2005). Wordnet-based semantic similarity measurement. Retrieved from http://www.codeproject.com/KB/string/semanticsimilaritywordnet.aspx. Accessed 1 Oct 2010.

Trappey, C., Wu, H., Taghaboni-Dutta, F., & Trappey, A. (2011). Using patent data for technology forecasting: China rfid patent analysis. *Advanced Engineering Informatics, 25*(1), 53–64.

Yoon, J., Choi, S., & Kim, K. (2011). Invention property-function network analysis of patents: a case of silicon-based thin film solar cells. *Scientometrics, 86*(3), 687–703.

Yoon, B., & Park, Y. (2005). A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technological Forecasting and Social Change, 72*(2), 145–160.