Scientometrics (2013) 94:313–331
DOI 10.1007/s11192-012-0830-6

Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis

Janghyeok Yoon · Hyunseok Park · Kwangsoo Kim

Received: 26 January 2012 / Published online: 1 September 2012

Abstract Patent maps showing competition trends in technological development can provide valuable input for decision support on research and development (R&D) strategies. By introducing semantic patent analysis with advantages in representing technological objectives and structures, this paper constructs dynamic patent maps to show technological competition trends and describes the strategic functions of the dynamic maps. The proposed maps are based on subject-action-object (SAO) structures that are syntactically ordered sentences extracted using the natural language processing of the patent text; the structures of a patent encode the key findings of the invention and expertise of its inventors. Therefore, this paper introduces a method of constructing dynamic patent maps using SAO-based content analysis of patents and presents several types of dynamic patent maps by combining patent bibliographic information and patent mapping and clustering techniques. Building on the maps, this paper provides further analyses to identify technological areas in which patents have not been granted ("patent vacuums"), areas in which many patents have actively appeared ("technological hot spots"), R&D overlap of technological competitors, and characteristics of patent clusters. The proposed analyses of dynamic patent maps are illustrated using patents related to the synthesis of carbon nanotubes. We expect that the proposed method will aid experts in understanding technological competition trends in the process of formulating R&D strategies.

J. Yoon
Department of Industrial Engineering, Konkuk University,
120 Neungdong-ro, Gwangjin-gu, Seoul 143-701, Korea
e-mail: janghyoon@konkuk.ac.kr

H. Park
Department of Technology and Innovation Management, Pohang University of Science and Technology, 77 Cheongam-ro, Nam-gu, Pohang 790-784, Korea
e-mail: howgood@postech.ac.kr

K. Kim (✉)
Department of Industrial and Management Engineering, Pohang University of Science and Technology, 77 Cheongam-ro, Nam-gu, Pohang 790-784, Korea
e-mail: kskim@postech.ac.kr

Keywords Dynamic patent map · Subject-action-object (SAO) structure · Semantic patent similarity · Natural language processing (NLP) · Research and development (R&D) strategy

JEL Code C63 · C82

Introduction

Patent documents are up-to-date and reliable information sources that reflect technological change (Yoon et al. 2011). Patent analysis has a narrow functional focus related only to technological knowledge of inventions, but it nevertheless has wide applicability to many different businesses (Feldman and Sanger 2007) and has been considered a vital tool for the analysis of technological trends (Yoon and Park 2004). Recently, the number of patents granted increased rapidly, and the need to analyze large quantities of patents has made relying only on experts' knowledge to analyze technological trends almost impossible (Kostoff 1998). In response, tools for patent analysis have been developed to assist researchers and research and development (R&D) policy makers to better concentrate on their own knowledge services, including product development planning and technology strategy formulation.

One tool for patent analysis is a patent map that visualizes overall relationships among patents in a given technology. A widely adopted method for patent mapping is the keyword-based approach that constructs patent maps using the keywords and key phrase patterns that occur in a given patent set (Chang et al. 2010; Lee et al. 2009; Yoon and Park 2004; Yoon et al. 2002). By exploiting the vector space model (Salton et al. 1975) that was originally developed for information retrieval (IR), the approach predefines a set of patterns, represents each patent according to the vector codifying occurrences of the patterns, and then maps the relationships among the vectors onto a lower-dimensional space. Although the keyword-based approach has been widely used because of its simplicity and applicability, it has some substantial limitations: (1) predefining keywords and key phrases relies heavily on the knowledge of experts who may be expensive or unavailable (Choi et al. 2011; Yoon et al. 2011; Yoon and Kim 2012), (2) frequencies and co-occurrences of the patterns may not reflect substantial characteristics of inventions (Wanner et al. 2008), and (3) the patent maps may differ according to which expert was employed because experts select the keywords and key phrases (Yoon and Kim 2012). In the strictest sense, the uniqueness of an invention can be determined not by the occurrences of keywords and key phrases but by inventive key findings or fresh expertise.

For these reasons, by introducing the subject-action-object (SAO)-based approach, this paper constructs dynamic patent maps to identify technological competition trends effectively in a given technology. SAO structures are syntactically ordered sentences extractable using the natural language processing of the patent text; they are the key concepts that explicitly include technological objectives and structures of a relevant patent (Cascini et al. 2004; Mann 2002), and the set of SAO structures is considered a detailed picture of the inventor's expertise (Moehrle et al. 2005; Park et al. 2012). In this study, we produce dynamic patent maps by combining textual and bibliographic analysis of patents and present in-depth strategic analyses of the maps to identify useful insights into technology areas in which patents have not been granted ("patent vacuums"), areas in which many patents have actively appeared ("technological hot spots"), R&D trends of competitors, and characteristics of patent clusters. The proposed SAO-based dynamic patent maps and their strategic analyses are illustrated using patents related to the synthesis of carbon nanotubes (CNTs). We expect that the proposed SAO-based dynamic patent maps and their strategic analyses can be incorporated into R&D planning processes to assist technology strategy decision makers and R&D policy makers in identifying technological competition trends.

Section "Theoretical background" presents an overview of the theoretical background. Based on the groundwork, section "SAO-based dynamic patent maps" briefly introduces a procedure for SAO-based dynamic patent maps, and section "Illustration and strategic analyses of dynamic patent maps" presents the practical analyses of the dynamic maps. Section "Conclusion and future research" concludes the paper and discusses future research topics.

Theoretical background

Constructing SAO-based dynamic patent maps requires SAO-based patent textual analysis, measurement of similarities between SAO structures, and patent mapping. Therefore, this section reviews the theoretical literature on SAO-based patent analysis, semantic sentence similarity measurement, and multidimensional scaling (MDS).

SAO-based patent analysis

The method proposed in this paper requires extracting SAO structures from patent text; these are composed of a subject (noun phrase), action (verb phrase), and object (noun phrase). This syntactically ordered sentence explicitly describes a relationship between components that appear in the relevant patent. Given a simple sentence, "soap cleans hands," the subject is "soap," the action is "cleans," and the object is "hands." The action "cleans" clearly defines a relationship between the subject "soap" and the object "hands." SAO structures are substantially related to the concept of function, which is defined as "the action changing a feature of any object" (Savransky 2000), so 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). Furthermore, SAO structures from patent claims are considered to represent intensive knowledge related to the inventor's expertise and the patent's key concepts (Moehrle et al. 2005; Cascini et al. 2004).

Recently, several SAO-based patent design structure analyses have been used to visualize the design structures of patents as SAO-based networks (Cascini et al. 2004) and to identify patent similarity by measuring similarity among the SAO networks (Cascini and Zini 2008). Other SAO-based patent analyses have operated on the basis of patent-based merger and acquisition strategies, patent-based human resource decisions (Moehrle et al. 2005), patent infringement risk evaluation (Bergmann et al. 2008; Park et al. 2012), product forecasting (Gerken et al. 2010), and the identification of rapidly evolving technological trends (Yoon and Kim 2012).

Although the patent maps of this paper introduce an SAO-based patent map approach similar to that introduced in our previous works (Park et al. 2012; Yoon and Kim 2011; Yoon and Kim 2012), this paper has a differentiation in that it constructs SAO-based patent maps from a "dynamic" perspective and presents practical analyses of the dynamic maps to address several dimensions, including patent vacuums, technological hot spots, R&D overlaps of competitors, and characteristics of patent clusters.

WordNet-based semantic sentence similarity measurement

The semantic patent similarity between two patents is computed using semantic sentence similarities between pairs of SAO structures of the patents. In general, the measurement process of semantic similarity between two sentences is composed of (1) tokenizing the sentences, (2) stemming works, (3) tagging parts of speech, (4) determining the most likely meaning of each word in each sentence, and (5) computing the sentence similarity on the basis of the similarity between pairs of corresponding words (Simpson and Dao 2005; Yoon and Kim 2012). A measure of similarity between two concepts (Resnik 1999) is defined as follows:

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

where lcs is the lowest common subsumer 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. The similarity score of the two concepts is $0 < \mathrm{sim}(c_1, c_2) \leq 1$, 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 $|X|$ and $|Y|$ denote the number of set tokens in sentences X and Y, respectively, and $\mathrm{Match}(X, Y)$ is the sum of similarity of the matching word tokens between the sentences. The matching average score between the two sentences is $0 < \mathrm{MatAvg}(X, Y) \leq 1$, where 1 means that the sentences are identical. Using the WordNet semantic dictionary (Miller 1995) as a concept hierarchy, the measure for similarity measurement between two words, and the measure for similarity measurement between two sentences, a .NET-based semantic sentence similarity measurement has been implemented as an application programming interface library (Simpson and Dao 2005). Because SAO structures are complete sentences, the procedure proposed in this paper uses the C# library as a basis for the automatic measurement of semantic similarity between SAO structures.

Although an advanced semantic similarity measurement method was developed to support the exchange of information among enterprise applications (Jeong et al. 2005), it requires specific typed data, including a tree structure. However, SAO structures extracted from patents do not construct such tree structures, so we adopt the semantic sentence similarity measurement to measure similarities among patents.

Multidimensional scaling

MDS is a set of related statistical techniques often used in information visualization for exploring similarities or dissimilarities in data. Using dissimilarity between objects, or attributes of objects, MDS maps their relationships onto a lower-dimensional space. The quality of an MDS can be identified using the stress value V, ranging from 0 (no stress) to 1 (maximum stress). In general, MDS with $V < 0.2$ is accepted to avoid degeneration (Kruskal 1964). The strength of MDS is that it provides analyzers with a visual understanding of data. MDS has been applied to chemical engineering science maps (Peters and Van Raan 1993a, b), the evaluation of technological strategies (Schmoch 2009), and analysis of technology opportunities (Lee et al. 2009). These studies were based on co-occurrences of predefined keywords or key phrase patterns; this paper is different from previous studies in that it exploits SAO-based semantic patent similarity.

SAO-based dynamic patent maps

The procedure for SAO-based dynamic patent maps is composed of four steps (Fig. 1): (1) collecting patent data, (2) extracting SAO structures from patent text using NLP, (3) generating a patent dissimilarity matrix by computing semantic patent dissimilarities based on semantic sentence similarities between the SAO structures of patents, and (4) visualizing the patent dissimilarity matrix as dynamic patent maps by exploiting MDS and patent bibliographic information. To provide this procedure in an automated way, we developed a .NET-based prototype system that interfaces external systems including information mapping software. The prototype system provides several functions for strategic analyses of the dynamic patent maps in section "Illustration and strategic analyses of dynamic patent maps".

Collection of patent data

The proposed patent map is generated using a specific patent set. In general, gathering patents requires a retrieval query composed of textual information related to a target technology and bibliographic information such as international patent codes, applicants, and application date. A final patent set for analysis is prepared by eliminating irrelevant patents.

Although patent documents have various sections, syntactic analysis of patents can use only narrative sections, including title, abstract, background summary, detailed description, and claims. Among various narrative sections, the proposed procedure uses only claims because they clearly state the innovative knowledge that requires legal protection (Fujii

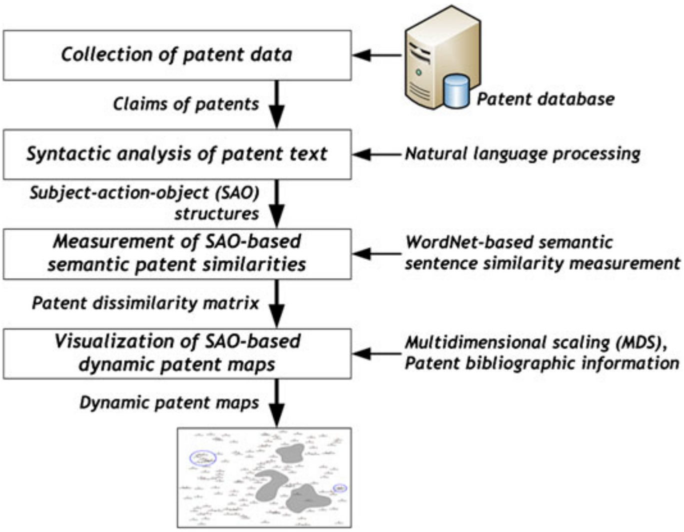

Fig. 1 Overall procedure for SAO-based dynamic patent maps

et al. 2007; Yoon et al. 2011). The data source is stored in an electronic format, such as Microsoft Excel, for syntactic analysis in the next step.

Syntactic analysis of patent text

In this step, SAO structures are output after the syntactic analysis of patent text. NLP parsers such as the Stanford parser (Stanford 2010) and MiniPar (Lin 2010) assist people without linguistic expertise to automatically extract the specific syntactic structures from the text. Commercial linguistic analyzers such as Knowledgist2.5™ (www.invention-machine.com) are also available. Because some of the SAO structures in patents may be duplicated or irrelevant, they are filtered out using a set of stopwords (STOPWORDS 2010); this is followed by human expert screening. Finally, SAO structures for the semantic analysis of the next step can be prepared. As a complete sentence, each SAO explicitly describes a relationship between components, that is, subjects and objects, including the tools and materials used in the relevant patent (Table 1).

Measurement of SAO-based semantic patent similarities

In this step, a patent dissimilarity matrix is generated by first computing the semantic similarities between pairs of patents (Fig. 2). The semantic patent similarity between two patents is computed by measuring the semantic similarity between the SAO structures of the patents.

Using the C# library for the WordNet-based semantic sentence similarity measurement (Simpson and Dao 2005), semantic similarity between two SAO structures can be computed. WordNet defines the concept hierarchies of most words, including synonyms, hypernyms, and hyponyms, by word meaning: for example, the chemical symbol for the element "Mg" is the same as "magnesium," and the noun "automobile" is a variant of the noun "vehicle". However, WordNet does not contain abbreviations that are strongly domain-specific: for example, CNT researchers use specialized acronyms, such as

Table 1 An example of SAO extraction (Patent US2008-232042)

S (subject)	A (action)	O (object)
Catalyst deposition step	Deposit	Metal catalyst particles
Desired growth density of carbon nanotubes	Control	Compounding ratio of metal catalyst particles and barrier particles
Electron emitting element	Use	Carton nanotube assembly
Field emission phenomenon	Provide	Electron source
Film-formed substrate	Include	Catalyst particle dispersed film
Metal catalyst particles	Have	Predetermined particle diameter
Photocatalyst	Exhibit	Photocatalytic ability
Reduction step	Perform	Reduction treatment
Simultaneous sputtering method	Form	Catalyst particle dispersed film
Thermal chemical vapor deposition method	Use	Catalyst particle dispersed film
Thermal chemical vapor deposition step	Grow	Carbon
Visible light	Have	Wavelength of 550 nm

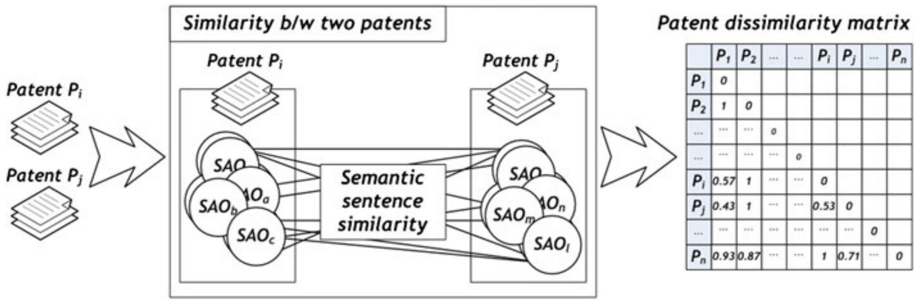

Fig. 2 Concepts for computation of the patent dissimilarity matrix

"MWNT" for "multiple-walled nanotube" and "CVD" for "chemical vapor deposition". Therefore, the C# library source code was modified to refer to a synonym set that can be grouped by investigating the SAO structures of the data source.

To determine whether two SAO structures are same, a threshold value p is used. For example, when $p = 0.90$ and MatAvg(SAO_i, SAO_j) = 0.92, then the two SAO structures can be considered semantically the same. The determination of similarity between SAO_i and SAO_j is

$$\text{SAO}_{ij} = \begin{cases} 1, & \text{if MatAvg}(\text{SAO}_i, \text{SAO}_j) \geq p \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Therefore, a measure for semantic similarity between patents X and Y can be defined on the basis of how many SAO structures the two patents share:

$$\text{SIM}(X, Y) = \frac{2 \times N_{\text{SAO}}(X, Y)}{N_{\text{SAO}}(Y) + N_{\text{SAO}}(Y)} \tag{4}$$

where $N_{\text{SAO}}(X)$ is the number of SAO structures in patent X, and $N_{\text{SAO}}(X,Y)$ is the number of the semantically identical SAO structures shared by patents X and Y. The dissimilarity between patents A and B is

$$\text{DSIM}(A, B) = 1 - \text{SIM}(A, B) \tag{5}$$

The semantic dissimilarity score of two patents is $0 \leq \text{DSIM}(A, B) \leq 1$, where 0 means that they are identical. Using the dissimilarities between pairs of patents, a patent dissimilarity matrix can be obtained that codifies the semantic distances between patents. Given that the number of patents is M, a lower-triangular $M \times M$ patent dissimilarity matrix can be prepared.

Visualization of SAO-based dynamic patent maps

This step visualizes the relative positions of patents on the lower-dimensional spaces by considering the semantic distances between patents. For this purpose, this paper uses MDS. Various MDS algorithms such as PREFSCAL, PROXCAL, and ALSCAL are available, and SPSS 17.0, a widely used commercial statistics software package, currently provides these algorithms. Using the $M \times M$ patent dissimilarity matrix, the software outputs the relative positions of patents on a two-dimensional space. After the positions of patents are combined with the patent bibliographic information, the patent map can provide the dynamic features of a given technology with respect to application dates, applicants, and patent clusters.

Illustration and strategic analyses of dynamic patent maps

To illustrate the proposed method, this section generates dynamic patent maps using patents related to the synthesis of CNTs, which are allotropes of carbon with a cylindrical nanostructure. These cylindrical molecules exhibit extraordinary strength and unique electrical properties and are efficient thermal conductors, so they are potentially applicable to nanotechnology, electronics, optics, and material science. Over the past decade, many CNTs-related patents have been applied for. This paper considers 136 patents collected from patent databases for Europe (EU), Japan (JP), Korea (KR), and the United States (US), the top four countries in the Patent Cooperation Treaty application (WIPO 2010). This paper used a patent set identical to that in our previous study (Yoon and Kim 2011); readers can refer to the retrieval query for the collected patents from the study. These patents were related only to methods of synthesizing CNTs, so most patents were classified into nano-materials and manufacture related international patent classifications (IPC) including C01B (non-metallic elements; compounds thereof), B82B (nano-structures; manufacture or treatment thereof) and C23C (coating metallic material; coating by vacuum evaporation etc.) (Table 2).

Because the SAO-based analysis in this paper considers only English text, patent text in the KR and JP databases were translated into English text using language translation services such as K2E-PAT (KIPO 2010); because the grammatical structure of Korean is identical to that of Japanese, the text of JP patents can be perfectly translated into Korean text using the Google translating program. 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.

Section "Dynamic patent maps related to the synthesis of carbon nanotubes" provides the generated SAO-based semantic patent map and section "Strategic analyses of patent maps for R&D planning" presents ways to identify technological insights for decision supports on R&D strategies.

Dynamic patent maps related to the synthesis of carbon nanotubes

The claims of each patent were stored in a Microsoft Excel file, and Knowledgist2.5^{TM} was then used to extract their SAO structures. After eliminating duplicate and irrelevant SAO structures, 1174 SAO structures (an average of 8.63 per patent) were extracted. Then, by

Table 2 4 digit IPC groups of patents

IPC	B05D	B29C	B32B	B82B	C01B	C23C	C30B	D01F	G03F	H01J	H01L	Sum
# of patents	2	1	1	19	87	8	1	8	1	4	4	136
% of patents	1.5 %	0.7 %	0.7 %	14.0 %	64.0 %	5.9 %	0.7 %	5.9 %	0.7 %	2.9 %	2.9 %	100 %

examining the extracted SAO structures, we defined a synonym set that includes abbreviations and phrases with the same meaning.

The semantic patent similarities were computed on the basis of the WordNet-based semantic sentence similarities between their SAO structures. To determine whether two SAO structures are same, the threshold value p was set to 0.80 after discussion with a nanotechnology domain expert. Then, dissimilarities between pairs of patents were calculated and a 136 × 136 patent dissimilarity matrix was obtained (Fig. 3), and PROXCAL of SPSS 17.0 was used to convert the matrix into a two-dimensional patent map with an MDS quality stress value of $V = 0.15$ (Fig. 4). In the map, patents that are near each other have a strong possibility of being technologically similar or complementary, whereas they are unrelated if they are far from each other. Although the patent map gives an intuitive understanding about a given patent set, this understanding becomes more difficult as the size of the patent set increases. Therefore, further analysis to effectively identify technological insights from the patent map is introduced in the next sub-section.

Strategic analyses of patent maps for R&D planning

By incorporating patent bibliographic information, the dynamic features of the SAO-based semantic patent map assist experts in effectively identifying technological insights. Specifically, this section identifies patent vacuums, technological hot spots, R&D trends of competitors, and characteristics of patent clusters. To analyze the patent map, a .NET-based application was developed; it shows the map in a dynamic sense by incorporating patent bibliographic information including application dates, applicants, and patent clusters (Fig. 4).

Patent vacuums and technological hot spots

Temporal analysis of the SAO-based patent map identifies patent vacuums and reveals how they change over time (the blue oval in Fig. 5); patent vacuums may represent opportunities for the creation of new technology. However, not all patent vacuums are promising, so several processes, including significance analysis, trend analysis, and feature analysis,

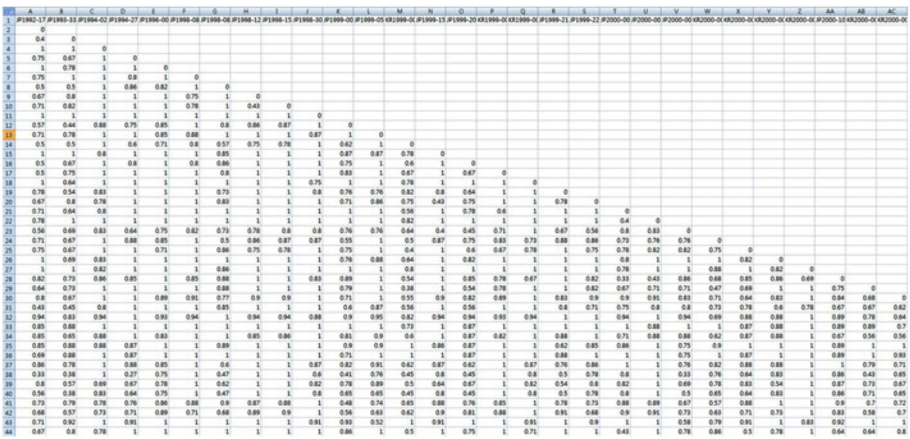

Fig. 3 A portion of the patent dissimilarity matrix

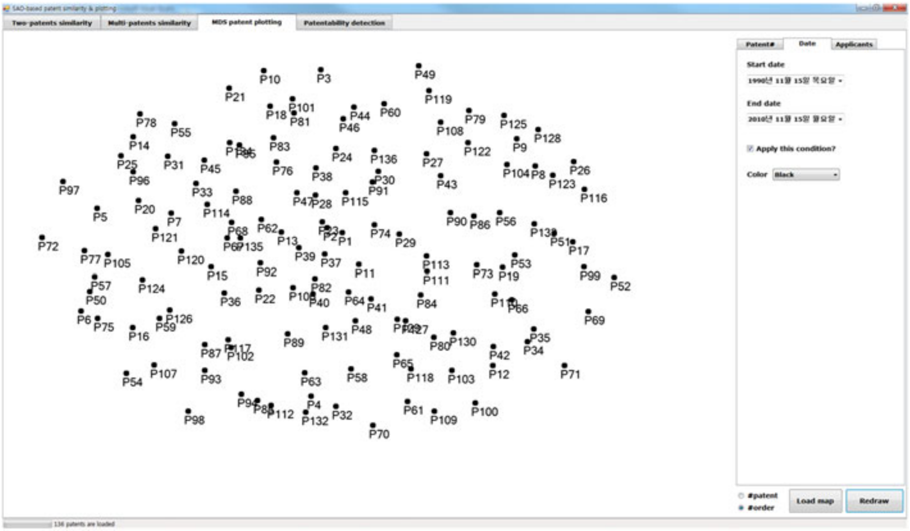

Fig. 4 Visualization of the SAO-based patent map

are required to determine whether they actually represent technology opportunities (Lee et al. 2009). In this paper, the SAO structures of patents located around or in the vacuum areas were used to identify the characteristics of patent vacuums. In this respect, the SAO-based approach is substantially different from the keyword-based approach, which shows the occurrence frequency of keywords but does not show the technological structures of inventions or the relationships among components used in the inventions. However, the SAO-based approach can explicitly describe them using SAO structures (Appendix 1). This advantage can help experts infer the characteristics of patent vacuums. One such vacuum (Fig. 5) can be characterized by three technologies, including single-walled CNTs (P5[JP1996-003636], P7[JP1998-082409]), growth of CNTs at a low temperature (P57[JP2002-253317], P59[US2002-237729]), and microwave-based synthesis (P20[JP2000-000299]). Even though the characteristics of these technologies seem to be different from one another, combinations may provide technological opportunities; for example, methods of growing single-walled CNTs (P5, P7) can be made more cost-efficient by using low-temperature CNT synthesis (P57, P59) or by avoiding heating of the substrate (P20). Actually, this vacuum was later occupied by patent P105[KR2004-0112852], which was for an apparatus and method for growing CNTs at a low temperature that uses plasma etch gas and a catalyst; it thus demonstrated that this vacuum did represent a technically feasible combination of methods. Although this finding used the information of P105 to predict the appearance of P105, its intention is to show that SAO structures can provide significant insight with regard to the technological objectives and structures of the patents that may appear in the relevant patent vacuums. As another example, from the characteristics of patents around another patent vacuum (the green oval in Fig. 5), methods of multi-walled CNTs can be made to provide telescopic motion ability for inner tubes by combining the precursor materials of CNTs and multi-walled structures. A striking telescoping property of this multiple concentric CNTs can be used to create useful machines, including the world's smallest rotational motor and a gigahertz mechanical oscillator.

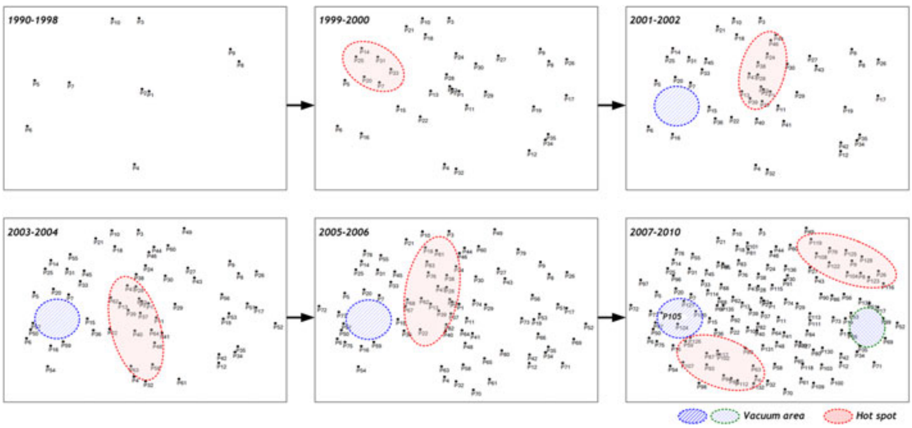

Fig. 5 Temporal analysis of patent maps (Vacuums: *blue oval*, *green oval*; Hot spots: *red oval*) (color figure online)

Temporal analysis of the patent map can also be used to detect technological hot spots where new patents are being actively applied for. Therefore, the hot spots directly show areas in which technological development is most competitive during a specific period of time (Fig. 5). The characteristics of the hot spots can be directly identified by examining the SAO structures of the patents that each hot spot contains (Appendix 1).

In the patent database examined, the characteristics of hot spots changed over time. During 1999 and 2000, they consisted mainly of CNT synthesis methods that use conductive materials and arc-discharging (P14[JP1999-159180], P25[KR2000-0014246]); during 2003 and 2004 they consisted mainly of heat treatment synthesis methods using metal catalysts and CVD (P58[US2002-237695], P62[JP2002-314127], P64[JP2002-331816]); during 2005 and 2006 they consisted mainly of arc-discharging methods using purification control (P67[JP2002-378840], P68[KR2002-0086799], P81[JP2003-205629], P82[JP2003-371351]); and during 2007 and 2010 they consisted mainly of methods related to plasma-enhanced CVD (P104[JP2004-280179], P108[JP2005-132893], P122[JP2007-072270]) and methods related to low-temperature thermal CVD (P112[US2005-246063], P117[JP2006-224896]).

The identification of patent vacuums and hotspots provides valuable information. An early understanding of patent vacuums may provide opportunities for the creation of new technologies that have rarely been explored by competitors. In addition, identifying technological hot spots provides an understanding of up-to-date technological competition trends, so it assists technology strategy decision makers in responding promptly to technological changes. We expect that this temporal analysis can help such decision makers perform technology creativity activities and that it can be exploited to aid decision support in product development and R&D directions.

R&D trends of competitors

Identifying trends followed by patent applicants provides a better understanding of the technological competence of competitors (Moehrle and Geritz 2004). For example, if a company's patents are located in a small area of a patent map, it means that the company is focusing on a specific technology. Conversely, if patents of a company are widely distributed on a patent map, this indicates that the company is exploring or developing new technologies. An R&D trend map of competitors including four companies and two individuals provides an intuitive understanding of the R&D activities by competitors (Fig. 6). The map shows that patents of ULVAC are far from other five competitors and that the company is developing its own technological area (Fig. 6). ULVAC is a leading company in vacuum technology, including vacuum coating and sputtering. In the last decade, this company has applied for patents (P27[JP2000-108319], P108[JP2005-132893], P119[JP2006-255299], P122[JP2007-072270]) related to plasma-enhanced chemical vapor deposition (PECVD), building on its vacuum technology (Appendix 2). SONY's patent pattern overlaps with those of other applicants, including NKK, CHO, and JANG. Some SONY patents (P42[JP2001-191424], P80[JP2003-197339]) are for arc-discharging, synthesis-using heat traps, and some patents of CHO (P127[KR2007-0121569], P130[KR2008-0064683]) are for arc-discharging and synthesis-using rotating cathodes (Fig. 6). Although these patents have a similar technological basis, they use slightly different methods to attain high yields of CNTs (Appendix 2). The SAO-based approach provides plentiful information—that is, the gathered SAO structures—to assist experts in understanding competitors' R&D trends. The occurrence frequency of keywords in the keyword-based approach cannot show the relationships among tools and materials used in the inventions, but SAO structures in the approach of this paper explicitly describe the technological structures and findings of inventions.

R&D trend maps of competitors visualize R&D trends of and technological overlaps among competitors. Therefore, they provide technology strategy decision makers with an intuitive understanding of the competitive outlook in a given technology area. We expect that the analysis of R&D trend maps can be used as a valuable source for technological decision making, including patent purchasing, M&A, and business partnerships.

Competing on different technologies

Using the relative position of each patent on the map, patent clusters can be identified. Managing patent clusters allows similar patents to be grouped as a technology package,

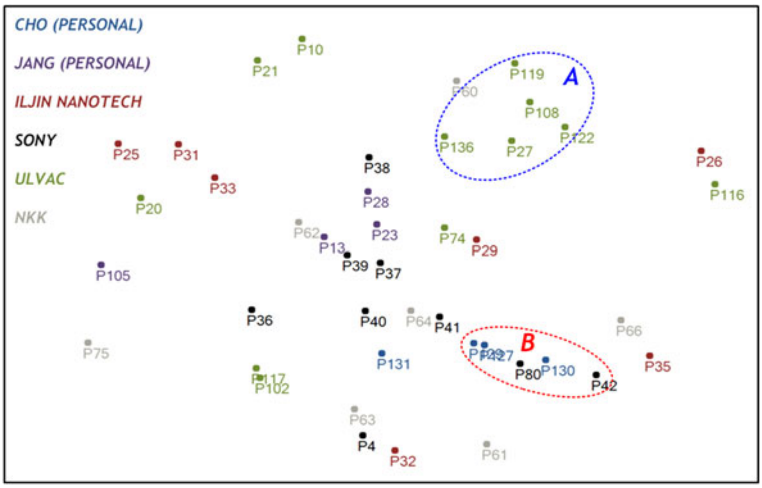

Fig. 6 R&D trend map of six competitors. **a** ULVAC patents; **b** Overlap between SONY patents and CHO patents

which enables experts to identify the idiosyncratic characteristics of patent clusters (Yoon et al. 2002). Although several clustering algorithms can be used to identify patent clusters, this paper has adopted the *k*-means clustering algorithm. *k*-means clustering is a method of cluster analysis that aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean (MacQueen 1967). In *k*-means clustering, the best number *k* of clusters tends to show the smallest average intra-cluster distance and the largest average inter-cluster distance (Yoon and Park 2004). The average intra-cluster distance is the average distance between individual elements and the cluster center, and the average inter-cluster distance is the average distance between cluster centers. However, a particular cluster may not be completely separable from the others, which means that the cluster is different from the others in an overall sense. Applying the analysis of Yoon et al. (2011), the technological implications of patent clusters can be formulated. Clusters with many elements, low average intra-cluster distance and high average inter-cluster distance have a strong possibility of being technologically verified, cost-efficient, or dominant technologies. In contrast, patent clusters with a small number of elements, low average intra-cluster distance, and high average inter-cluster distance may not be a dominant or technologically verified technology but may deliver weak signals for the development of new technology.

By trial and error, we found that the patent cluster map of the CNTs synthesis showed the best clustering result when $k = 12$ (Fig. 7; Table 3). Using the SAO structures (Appendix 3), decision makers can easily understand that patent cluster 2 is composed of many new patents that are mainly related to PECVD technology. In addition, patent cluster 2 is small, with a low average intra-cluster distance and a high inter-cluster distance, which indicates that the cluster has a possibility of consisting of patents related to a novel method. Many patents in this cluster (P27[JP2000-108319], P108[JP2005-132893], P119[JP2006-255299], P122[JP2007-072270]) were applied for by one company and all are related to PECVD technology, which is currently part of the most promising next-generation CNT synthesis methods. Interestingly, patent cluster 11 shows a result similar to our previous study related to semantic patent network analysis; the previous study identified an emerging patent cluster composed of seven patents, including P131, P130, P129, P127, P48, P42, and P41. Patent cluster 11 of the patent cluster map includes all patents except P42. The previous work uses a cut-off value to construct patent networks and identify patent clusters using the bi-component method (Bollobas 1983), while the proposed analysis of this paper uses *k*-means clustering to identify technology packages that are similar in an overall sense. The previous work and this research may show slightly different analysis results because they adopt different techniques. This means that the two approaches are complementary in some sense; each approach can analyze the dimensions that the other approach does not identify.

Analyzing the characteristics of patent clusters is useful for identifying patent packages that are composed of similar patents and for perceiving technological insights provided by the patent packages. Therefore, we expect that the analysis of patent cluster maps can be incorporated into the patent portfolio management process.

Conclusions and future research

Patents are considered a proxy reflecting technological advancement, and thus, patent analysis has been a vital activity to identify technological trends and formulate strategic decisions in the R&D planning process. One tool for effective patent analysis is a patent

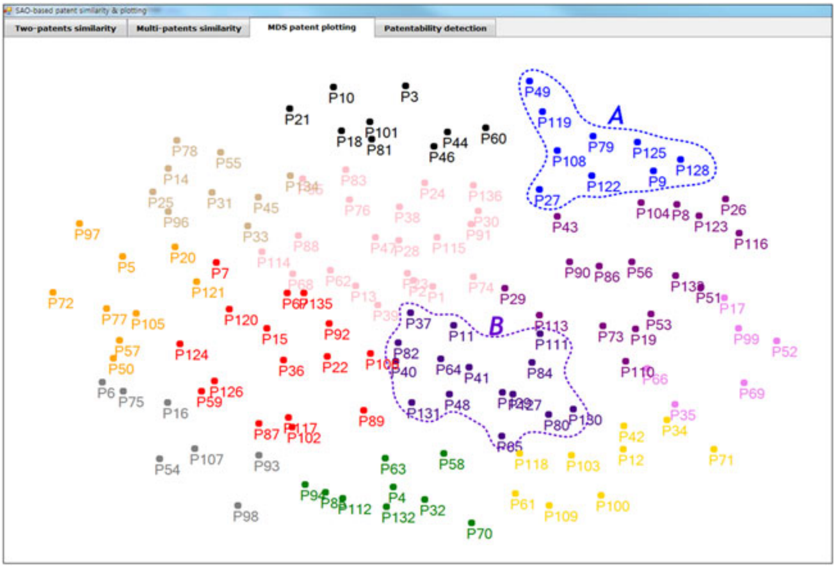

Fig. 7 Patent cluster map by k-mean clustering ($k = 12$). *Colors* represent clusters (Table 3) (color figure online)

Table 3 Patent clusters and their properties

Cluster number (Color)*	Cluster size	Average intra-cluster distance (1)	Average inter-cluster distance (2)	(1)/(2) ratio
1 (Red)	16	0.229	0.773	0.296
2 (Blue)	9	0.171	1.036	0.165
3 (Black)	9	0.210	1.005	0.209
4 (Green)	9	0.154	0.976	0.158
5 (Orange)	9	0.245	0.991	0.247
6 (Gray)	7	0.213	1.020	0.209
7 (Pink)	21	0.215	0.739	0.291
8 (Purple)	17	0.250	0.884	0.283
9 (Tan)	9	0.223	0.973	0.229
10 (Violet)	6	0.207	1.031	0.201
11 (Indigo)	15	0.188	0.741	0.254
12 (Gold)	9	0.224	1.008	0.222

* Each patent cluster is characterized by its corresponding color in Fig. 7

map; the dynamic features of patent maps can provide significant technological implications that cannot be easily obtained by human experts' qualitative analysis. In fact, many studies have provided techniques to identify technological trends using patent maps. To this end, the keyword-based approach has been widely used because of its simplicity and practicality in constructing patent maps. However, the keyword-based approach has some limitations in that it cannot effectively represent key technological concepts, including the structures and objectives of patents. For this reason, by introducing an SAO-based approach over the keyword-based approach in specific dimensions, this paper presents practical functions and capabilities of patent maps based on SAO structures of patents from a "dynamic" perspective. Proposed strategic analyses of the dynamic patent maps include several functions that analyze technological vacuums, technological hot spots, R&D overlaps of competitors, and competition on different technologies on patent maps. The strategic analyses of SAO-based dynamic patent maps are illustrated using patents related to the CNTs synthesis method. We expect that the functions of SAO-based dynamic patent maps will aid experts in identifying technological competition trends in the process of formulating R&D strategies.

Despite this contribution, this study has its limitations. First, this paper used the MDS to visualize the patent maps for strategic analyses and adopted the *k*-means clustering algorithm, a trial-and-error approach to group patents, for patent clustering. However, the MDS can cause information loss on distances between patents, and *k*-means clustering is considered a simple technique that is often not effective. Although the MDS and *k*-means clustering allowed us to identify reasonable technological implications, future research will adopt advanced techniques, such as *k*-medoids clustering (Park and Jun 2009) without dimension reduction and hierarchical clustering (Hair et al. 2010), to obtain more valid and effective analysis results. Second, although this research applied the strategic analyses to the synthesis of CNTs, further research will extend the analyses to strategic decision making for other fiercely competitive technology fields, including solar cells and flexible display technology. Finally, this paper only applied SAO-based semantic patent similarity to the development and analysis of patent maps. However, in our future research, we will apply the method to other applications, such as patentability analysis, strategic partner selection, and patent licensing opportunities.

Appendices

Because a full list of SAO structures of all patents requires too much space, only selected SAO structures of patents are listed in the following tables.

Appendix 1: A portion of SAO structures for analysis of vacuum areas and hot spots

$\text{Patent}_{[\text{real application number}]}$	S (subject)	A (action)	O (object)
$P5_{[JP1996\text{-}003636]}$, $P7_{[JP1998\text{-}082409]}$	Carbon	Use	Electrode metal additive
	Collecting soot	Characterize	Single layer
	Soot	Contain	Single walled carbon nanotubes
	Plasma	Supply	Powdered metal catalysts
$P57_{[JP2002\text{-}253317]}$, $P59_{[US2002\text{-}237729]}$	Arc discharge	Synthesize	Carbon nanotubes
	Low-temperature thermal CVD	Comprise	Deposition precipitation
	Carbon source gas	Comprise	Hydrocarbon or carbon monoxide

Table a continued

Patent$_{[\text{real application number}]}$	S (subject)	A (action)	O (object)
P20$_{[\text{JP2000-000299}]}$	ECR microwave generation system	Generate	Plasma deposition chamber
	Formation	Characterize	Thin film forming method
	Gas supply system	Supply	Hydrogen gas
P105$_{[\text{KR2004-0112852}]}$	Deposition gas	Form	Carbon-containing gas and hydrocarbon gas plasma etching gas mixture
	Plasma	Etch	Gas particulate catalyst
	Method of carbon nanotubes	Generate	Multiple electrodes of multi-electrode
P14$_{[\text{JP1999-159180}]]}$, P25$_{[\text{KR2000-0014246}]}$	Cathode	Comprise	Conductive material
	Cathode surface	Have	Sediment generated surface cathode
	Mixture of carbon and reactive gas or inert gas	Characterize	Atmosphere
	Growth of carbon nanotubes	Use	Carbon nanotube growth substrate
P58$_{[\text{US2002-237695}]}$, P62$_{[\text{JP2002-314127}]}$, P64$_{[\text{JP2002-331816}]}$	Coating	Form	Bonding metal layer
	Carbon source gas	Comprise	Hydrocarbon or carbon monoxide
	Electroplating	Form	Metal catalyst layer
	Metal catalyst layer	Comprise	Fe, Co, Ni and alloy
	Vacuum sputtering	Form	Metal catalyst layer and metal support layer
	Heat treatment temperature	Characterize	Carbon nanotubes
	Spraying inert gas from interior of tubular cathode method	Produce	Carbon
P67$_{[\text{JP2002-378840}]}$, P68$_{[\text{KR2002-0086799}]}$, P81$_{[\text{JP2003-205629}]}$, P82$_{[\text{JP2003-371351}]}$	Arc discharge of carbon electrode	Generate	Carbon soot
	Purity of carbon	Characterize	Carbon nanotubes
	Evaporation and heat	Relax	Mixed carbon electrode catalyst metal
	Cooling thermal relaxation	Produce	Acid-treated carbonaceous product
	Pulsed arc discharge	Trap	Carbon nanotubes
P104$_{[\text{JP2004-280179}]}$, P108$_{[\text{JP2005-132893}]}$, P122$_{[\text{JP2007-072270}]}$	Substrate holder disposed in reaction chamber	Apply	Bias voltage
	Plasma heating means	Characterize	Mesh
	Remote plasma shielding member	Characterize	Substrate, Mo, Ti, W and WC
	Sponge growth	Characterize	Plasma fabrication of carbon nanotube

Table a continued

Patent$_{[real\ application\ number]}$	S (subject)	A (action)	O (object)
P112$_{[US2005-246063]}$, P117$_{[JP2006-224896]}$	Low temperature	Comprise	Thermal chemical vapor deposition process
	Circulating reaction tube	Cool	Refrigerant
	Infrared radiation condenser tube	Heat	Lamp
	Reaction tube method	Produce	Carbon nanotubes

Appendix 2: A portion of SAO structures for R&D trend analysis of competitors

Patent$_{[real\ application\ number]}$	S (subject)	A (action)	O (object)
P42$_{[JP2001-191424]}$, P80$_{[JP2003-197339]}$	Cold cathode	Characterize	Electron emitter field emission device
	Cold cathode field emission device	Have	Tube shape
	Heating temperature	Form	Electron emission
	Inner surface of electrode	Have	Hollow part
	Guide roller	Constitute	Nanotube manufacturing method
P127$_{[KR2007-0121569]}$, P130$_{[KR2008-0064683]}$	Cathode rod movement guidance	Engage	Anode rod
	Consumable carbon rods gripper	Store	Cases
	Interior of non-consumable carbon plate	Move	Negative currents
	Mobile guide rod thread	Engage	Connecting member
	Scraper	Rotate	Magnetic field
P108$_{[JP2005-132893]}$, P119$_{[JP2006-255299]}$	Plasma surface treatment method	Generate	Film
	Film	Include	Diamond particles
	Temperature measuring unit	Measure	Temperature of chemical deposition system
	Plasma	Not touch	Substrate

Appendix 3: A portion of SAO structures for characteristic analysis of patent clusters

Patent$_{\text{[real application number]}}$	S (subject)	A (action)	O (object)
P27$_{\text{[JP2000-108319]}}$, P108$_{\text{[JP2005-132893]}}$, P119$_{\text{[JP2006-255299]}}$, P122$_{\text{[JP2007-072270]}}$	Slot antenna	Have	Microwave ECR plasma chemical vapor deposition
	Thin film	Use	Slot antenna
	Plasma heating means	Characterize	Mesh
	Contacting narrowing of anode	Initiate	Arc discharge
	Inserting anode	Provide	Cooling
	Main body of anode	Consist of	Cylinder
	Passing current through electrodes	Vaporize	Carbon
	Remote plasma shielding member	Characterize	Substrate, Mo, Ti, W and WC
	Sponge growth	Characterize	Plasma fabrication of carbon nanotube

References

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.

Cascini, G., Fantechi, A., & Spinicci, E. (2004). Natural language processing of patents and technical documentation. *Lect Notes Comput Sci, 3163*, 89–92.

Cascini, G., & Zini, M. (2008). Measuring patent similarity by comparing inventions functional trees. *IFIP International Federation for Information Processing, 277*, 31–42.

Chang, P., Wu, C., & Leu, H. (2010). Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. *Scientometrics, 82*(1), 5–19.

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

Choi, S., Yoon, J., Kim, K., Lee, J. Y., & Kim, C. (2011). SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells. *Scientometrics, 88*(3), 863–883.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. New York: Cambridge University Press.

Fujii, A., Iwayama, M., & Kando, N. (2007). Introduction to the special issue on patent processing. *Inf Process Manage, 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.

Hair, F., Black, C., Babin, J., & Anderson, E. (2010). *Multivariate data analysis* (7th ed.). New Jersey: Pearson Education Inc.
Jeong, B., Lee, D., Cho, H., & Lee, J. (2005). A novel method for measuring semantic similarity for XML schema matching. *Expert Syst Appl, 34*(3), 1651–1658.
KIPO. (2010). K2E automatic translation. http://eng.kipris.or.kr/eng/other_service/k2e_automatic_translation.jsp.
Kostoff, R. (1998). The use and misuse of citation analysis in research evaluation. *Scientometrics, 43*(1), 27–43.
Kruskal, J. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika, 29*(2), 115–129.
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.
Lin, D. (2010). MINIPAR. http://webdocs.cs.ualberta.ca/~lindek/minipar.htm.
MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations (vol. 1, pp. 14). California: University of California Press.
Mann, D. (2002). *Hands-on systematic innovation*. Belgium: Creax press.
Miller, G. (1995). WordNet: a lexical database for English. *Commun 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.
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.
Park, H., & Jun, C. (2009). A simple and fast algorithm for K-medios clustering. *Expert Syst Appl, 36*(2), 3336–3341.
Park, H., Yoon, J., & Kim, K. (2012). Identifying patent infringement using SAO-based semantic technological similarities. *Scientometrics, 90*(2), 515–529.
Peters, H., & Van Raan, A. (1993a). Co-word-based science maps of chemical engineering. Part I: representations by direct multidimensional scaling. *Res Policy, 22*(1), 23–45.
Peters, H., & Van Raan, A. (1993b). Co-word-based science maps of chemical engineering. Part II: representations by combined clustering and multidimensional scaling. *Res Policy, 22*(1), 47–71.
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.
Salton, G., Wong, A., & Yang, C. (1975). A vector space model for automatic indexing. *Commun ACM, 18*(11), 613–620.
Savransky, S. (2000). *Engineering of creativity: introduction to TRIZ methodology of inventive problem solving*. Boca Raton: CRC Press.
Schmoch, U. (2009). Evaluation of technological strategies of companies by means of MDS maps. *International Journal of Technology Management, 10, 4*(5), 426–440.
Simpson, T., & Dao, T. (2005). WordNet-based semantic similarity measurement. http://www.codeproject.com/KB/string/semanticsimilaritywordnet.aspx.
Stanford (2010). The Stanford Parser: A statistical parser. http://nlp.stanford.edu/software/lex-parser.shtml.
STOPWORDS. (2010). English stopwords. http://www.ranks.nl/resources/stopwords.html.
Wanner, L., Baeza-Yates, R., Brugmann, S., Codina, J., Diallo, B., Escorsa, E., et al. (2008). Towards content-oriented patent document processing. *World Patent Inf, 30*(1), 21–33.
WIPO (2010). PCT Yearly Review 2009. http://www.wipo.int/export/sites/www/ipstats/en/statistics/pct/pdf/901e_2009.pdf2011.
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, J., & Kim, K. (2011). Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics*, 88(1), 213–228.
Yoon, J., & Kim, K. (2012). Detecting signals of new technological opportunities using SAO-based semantic patent analysis and outlier detection. *Scientometrics, 90*(2), 445–461.
Yoon, B., & Park, Y. (2004). A text-mining-based patent network: analytical tool for high-technology trend. *The Journal of High Technology Management Research, 15*(1), 37–50.
Yoon, B., Yoon, C., & Park, Y. (2002). On the development and application of a self-organizing feature map-based patent map. *R&D Management, 32*(4), 291–300.