World Patent Information 46 (2016) 32–48

Contents lists available at ScienceDirect

World Patent Information

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

The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis

Farshad Madani*, Charles Weber

Portland State University, USA

ARTICLE INFO

Article history:
Received 16 June 2015
Received in revised form
12 May 2016
Accepted 16 May 2016
Available online 31 May 2016

Keywords:
Technology mining
Patent mining
Bibliometrics analysis
Keyword network analysis
Cluster analysis
CiteSpace

ABSTRACT

Text mining methods allow researchers to investigate technical documents (tech mining) and specifically explore patents for valuable information (patent mining. To the review literature and analyze the evolution of patent analysis and patent mining methods, bibliometrics analysis and keyword-based network analysis is applied on 143 papers extracted from the 'Web of science' database. Bibliometrics analysis was applied to determine top players researching in patent mining. Applying cluster analysis on the keyword network shows three main stages of patent analysis evolution. Also, it is discussed how patent mining is evolutionized in terms of information retrieval, pattern recognition and pattern analysis.

1. Introduction

Patent databases are an important source of information for innovators [1–3], R&D engineers [4,5], corporate executives [6–9], and policy makers in technology latecomer countries [10–12]. Innovators need information on prior art, in order to assess whether their inventions are commercially viable [13]. R&D engineers, who are trying to solve a particular technical problem, want to identify patents that may contribute to the solution of their problem [4,5]. Corporate executives, who are looking for a technology that fits their product strategy, will make use of patent searches to identify how and where they can gain access to the desired technology [6–9]. Policy makers in technology latecomer counties tend to conduct patent analyses, in order to identify particular gaps in the capabilities of their national innovation systems [10–12]. In all the above instances, patent databases serve as a critical source of information upon which policy decisions are based.

For patent databases to be helpful in decision making, the information that they provide must be accurate, presented in a comprehensible format and delivered in a timely manner. This can only be done if the users of patent databases have access to capabilities in keyword extraction, pattern recognition and pattern analysis. These crucial aspects of modern text mining have thus become an integral component of decision making, both at the strategic and tactical levels.

Patent citation analysis and even established statistical techniques like Term Frequency-Inverse Document Frequency (TF-IDF) for patent keyword analysis do not provide the user of patent databases with an understanding of the content and the context of the patent. The user cannot determine whether a patent contains relevant prior art, unless he/she actually reads the patent. This process is highly inefficient. Tens of millions of patents reside in the databases of the world's major patent offices. The innovator may take years to identify all patents that are relevant. In order for patent databases to be useful for the abovementioned stakeholders, the processes for extracting and analyzing relevant information must be highly efficient. Researchers in academia have made significant progress in the area of applying text mining for keyword extraction [14] and pattern recognition [15–18]. However, the field of pattern analysis is still in its infancy by comparison.

Due to advances in natural language processing, text mining methods and tools have become increasingly available in many different research areas including technology management where scholars try to extract useful information and textual patterns from technical documents, particularly patents. Applying text mining methods in technical documents is named 'tech mining' or

* Corresponding author.
E-mail addresses: fmadani@pdx.edu, farshad.madani@gmail.com (F. Madani).

http://dx.doi.org/10.1016/j.wpi.2016.05.008
0172-2190/

'technology mining', and for patent analysis purposes, it is named 'patent mining'. Porter, as one of the pioneers in technology mining, has defined 'tech mining' in his book [19, p. 19] as: *'the application of text mining tools to science and technology information, informed by understanding of technological innovation processes.'* Therefore, tech mining has two significant characteristics: 1) using 'text mining tools' and 2) applying these tools to 'technology management'. As shown in Fig. 1, the number of published papers and the number of citations in the tech mining area illustrates a hyperbolical progression; there is a jump in the number of publications after 2005 and a huge rise in the number of citations after 2012.

Given the rapid evolution of patent mining, it is not clear how patent mining has been developed and how the scholars are trying to apply different methodologies to expand this research area. Few papers that shed light on this area and find answers to the above-mentioned questions have been published to date. Abbas et al. [20] have reviewed 22 articles published in the field of patent analysis, and they have provided a general taxonomy of techniques for patent analysis. Also, in an editorial note [21], Porter and Chiavetta investigated six papers published in The First Global Tech Mining (GTM) Conference. They report four main analytics tools which are bibliometrics, data mining, network analysis and cluster analysis. In addition, they reveal eight application areas including emerging technologies and technology dynamics (trend analyses), technology forecasting, roadmapping and foresight, R&D management, engineering industries, science and technology (S&T) indicators, evolutionary economics, technology assessment and impact analysis, as well as science, technology and innovation policy studies.

To conduct a literature review that is as comprehensive as possible, we deploy a methodology that investigates all published patent mining papers in the Web of Science database. We applied bibliometrics analysis to recognize the main papers, authors, universities, and journals. More importantly, we applied cluster analysis on a keyword network, which was extracted from the abstracts of the papers. CiteSpace [22], a free Java application for visualizing and analyzing citations and contents in scientific literature, is applied as the main analysis tool to figure out, detect and visualize emerging trends. CiteSpace is developed by Chaomei Chen whose research is 'information visualization' [23–25]. By applying co-citation network analysis, CiteSpace enables us to identify co-citation clusters and trace how the trend of researches has been developed [24]. The main techniques implemented in the software are spectral clustering and feature selection algorithms [24]. Visualization of the results is the main characteristic of CiteSpace, which assists more analysts to make sense of the trends and evolution [23]. Information visualization in this software is beyond just visualizing graphical displays. This method deploys cognitive, social and collaborative activities [25].

There are some papers in which the authors used CiteSpace as the main tool for bibliometric analysis. Tonta and Darvish [26] used CiteSpace in their research to do social network analysis (cluster methods and centrality measures), and co-occurrence analysis on authors and journals to reveal the social structure of a discipline in terms of collaboration among scientists. In another study, Dhami and Olsson [27] applied CiteSpace to analyze the clusters of co-citations network to figure out the evolution of personal conflict theories. Furthermore, Citespace is deployed to study co-citation patterns from 1987 to 2006 to disclose the overall evolution of S&T Policy [28].

2. Data collection

To extract the right papers from Web of Science, it is important to apply keywords that refer to the concept of tech mining. To do so, we consider Porter's definition [19, p. 19] to build the concept of patent mining based on two pillars: 1) 'purposes' and 2) 'applications' of tech mining. This approach helps cover all papers related directly and indirectly to patent mining papers. Since there are many other terminologies used in patent mining that have not used known patent mining terminologies but applied 'text mining' tools for 'technology management purposes'.

To build a reliable keyword list, several tools and techniques were applied. As mentioned above, possible keywords that refer to the concept of patent mining are listed. To figure out the keywords, reviewing publications of renowned authors is a quick trick. For example, the publications of Alan Porter as one of the pioneers in 'tech mining' are beneficial to make a preliminary list of keywords. But there are many authors who use their own keywords, so 'Keyword planner', an option of 'Google Adwords', is utilized to figure out what keywords people have been looking for in Google while they search for 'patent mining'. For example, 'tech mining', 'text mining', and 'text classification' are the most relevant and applied keywords. To test the initial list of keywords, it is necessary to examine two aspects: 'relevance', and 'applicability'. In respect to relevance, this aspect is questioned if the keyword is discussed in both 'patent mining' papers and other research papers.

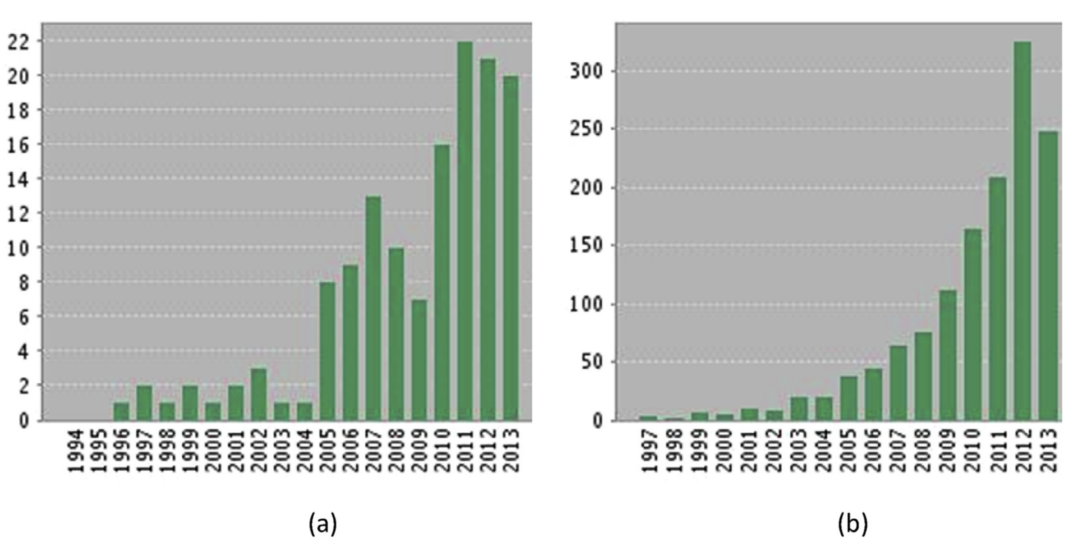

Fig. 1. a) published papers, b) Citations.

'Applicability' observes whether the keyword introduces a method, tool, etc. that is potentially applicable in 'patent mining'. For example, 'citation analysis', 'patent citation', and 'patent analysis' are all applicable methods in 'patent mining'. Three main techniques are used to test 'relevance' and 'applicability'. In the first technique, we check to see if there any papers with topics (title, abstract, and keywords) that contain the keyword. If there are no papers, or the papers contain other keywords, the searched keyword can be snubbed. Second, looking for the definition of a keyword in dictionaries, or synonyms in thesauruses, or checking encyclopedias like Wikipedia can help examine the relevancy of a keyword. Also, looking at some sample papers containing the keyword can help recognize its relevance and applicability.

Given the keywords list, we can design a query by applying Boolean operators and field tags based on Web of science standards. 'Search query design' is an iterative activity. Each search strategy must be tested several times to make sure that it addresses the right papers. The final query applied in this research is given below:

ts = ((“technology mining” or “tech mining” or “patent mining”) or ((“Technology Monitoring” or “Competitive technical intelligence” or “Technology Intelligence” or “technology foresight” or “Technology Forecasting” or “Technology analysis” or “Innovation Forecasting” or “Emerging Technologies” or “R & D” or “technology management”) and (“text mining” or “text classification” or “Document mining” or “web mining” or “document mapping” or “*Citation network” or “*Citation Analysis” or “Patent citation” or “patent mapping”))).

3. Methodology

The papers consist of three main parts: 1) Meta data including authors' names, title, keywords (offered by authors), authors' institution/university, publication year; 2) the main body of paper; and 3) references cited by authors. Each these parts has to be analyzed, respectively, by the following approaches: 1) bibliometric analysis, 2) keywords analysis wherein keywords represent main concepts mentioned in the main body of papers, and 3) citation analysis. To provide a comprehensive literature review—the main purpose of this research—researchers can pick one or a combination of these strategies. In our research, we picked bibliometrics analysis and keyword network analysis, and did not apply citation analysis. Citation analysis inherently relies completely on citations, and if author have not identified and cited related papers well, the literature review will not be comprehensive. This difficulty is more likely encountered in emerging areas like patent mining. Rather, keyword network analysis is much more powerful for emerging research areas since this method relies on the content of papers and no paper will be missed providing a well-defined query is applied to search into academic databases like Web of Science (WoS). It is worthy to note that the scope of this research is limited to WoS, since CiteSpace accepts only WoS as the primary source of input. This implies, for example, that our analysis does not include papers such as [29–36], which are contained in SCOPUS but not in the Web of Science.

Bibliometric analysis is applied to find a general view about top authors, journals, universities, and countries in patent mining. To find the most effective papers, the 'eigenvector centrality' measure is applied. Eigenvector acknowledges connections to a highly connected node. Eigenvector centrality assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes [37].

To obtain an in depth recognition of the evolution of patent mining, keywords are deemed as representatives of main streams in patent mining. So, a network is created based on the keywords that are extracted from the abstracts. This keyword-based network is analyzed by cluster analysis to find groups of keywords. The mean value of publishing time of the clustered papers lets us have a sense of evolution over time. Nonetheless, the keywords that received the top scores in the cluster analysis cannot lead us to specific conclusions. Therefore, to understand the evolution through the clusters, it is required to review the abstracts of the clustered papers. Three main concepts including methodology, application, and domain are considered to elicit the main aspects of clusters from the abstracts of clustered papers. If some clustered papers refer to a specific methodology, those papers are analyzed more deeply by reviewing their content more thoroughly. It helps to organize the main points of the papers, in order to understand how patent mining is evolving.

4. Results

In addition to the query, the search results are refined by choosing 'document type' as 'article' and 'proceeding papers'. As a result, the search came up with 143 papers which are listed in Appendix 1. In the analysis section, first, a bibliometric analysis on bibliometric data (including authors, papers, journals, universities, and countries) is applied, and then the keywords network is scrutinized by cluster analysis.

4.1. Bibliometric analysis

To figure out the most effective players in patent mining, four aspects of bibliometrics including authors, papers, universities, and countries are analyzed.

Of the top ten authors in 'patent mining' shown in Fig. 2, five are from Pohang University Science & Technology. This South Korean group of authors has 11 papers published in 'patent mining' that are concentrated in patent analyses. Also, they have developed a patent intelligent tool based on the Subject-Action-Object (SAO) method and applied it for various purposes such as R&D planning [38,39], roadmapping [15], technology trend identification [16], identifying patent infringement [40], and technology planning [41].

The most renowned author in 'patent mining' is Alan Porter who teaches at Georgia Tech University. He has published a book on 'tech mining' [19], and designed a five-step framework to incorporate external R&D information in decision making in management of technology [42]. He has also developed the Quick Technology Intelligence Processes (QTIP) framework to search, compose and quickly analyze information for technology analyses [43]. In addition, he has done notable research on applying 'patent mining' to nanotechnology [44–46] and on 'technology forecasting' [47,48].

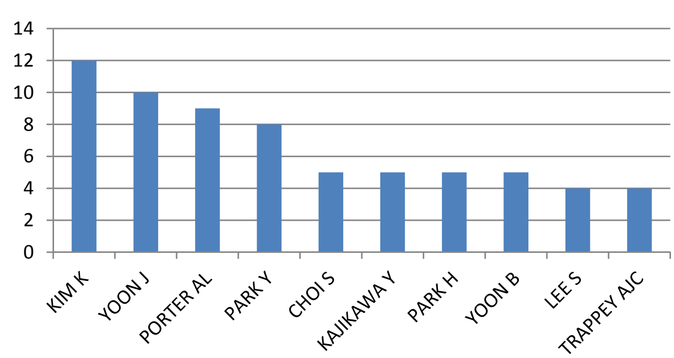

Fig. 2. Top 10 authors in 'patent mining'.

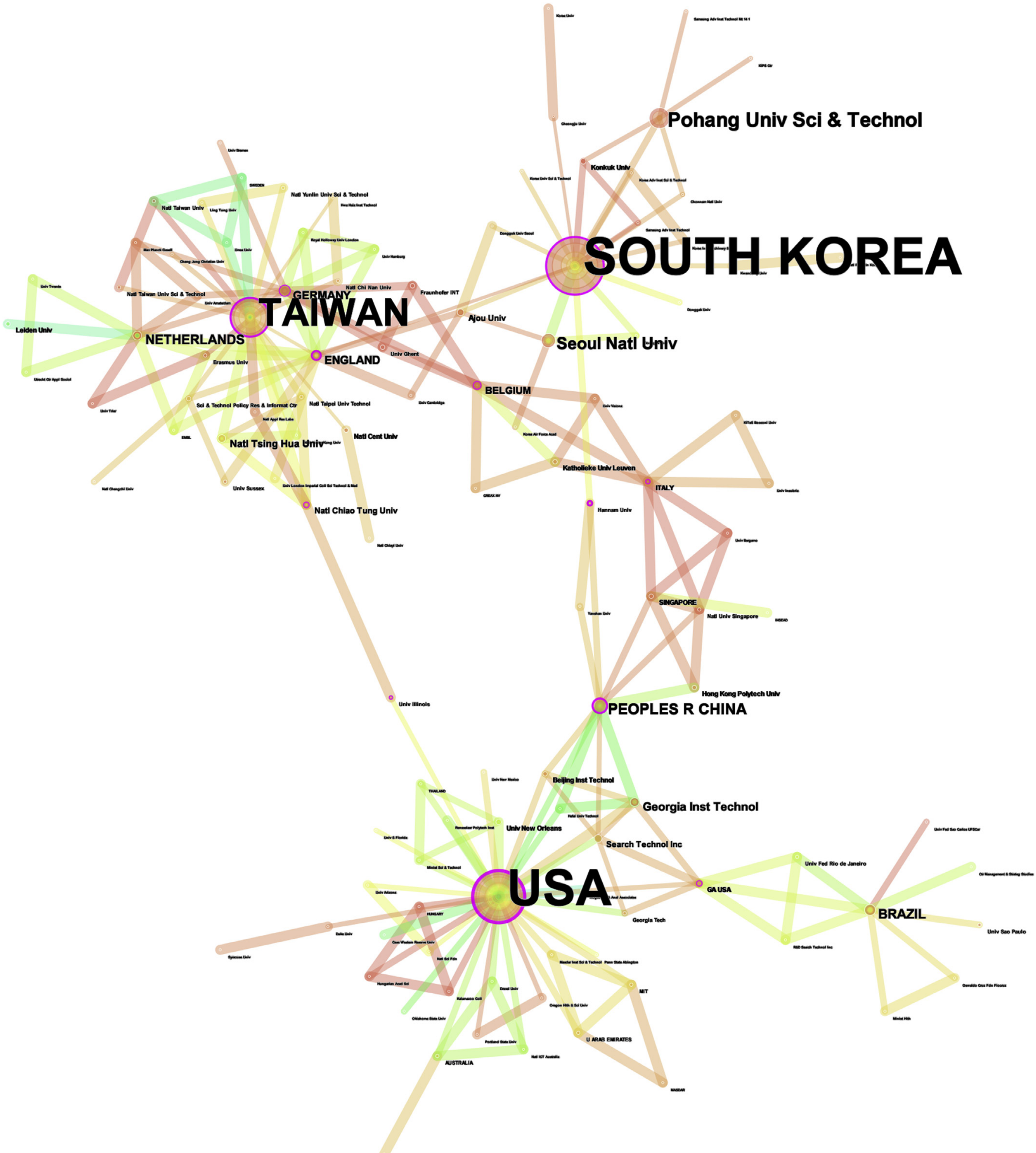

Fig. 3. Country-Institute analysis.

As shown in Fig. 3, South Korea, the USA, Taiwan, and Japan have had the most participation in 'patent mining' with 30, 29, 22, and 12 papers, respectively. Moreover, Pohang University, Seoul National University, University of Tokyo, and Georgia Tech are the most active universities in 'patent mining' with 11, 9, 6, and 6 papers, respectively.

According to eigenvector centrality, the ten most effective papers are introduced in Table 1. Interestingly, South Korean authors have dominated in this ranking. Also, the majority of these papers were published after 2008. Given the results shown in Table 1, patent citation analysis is the dominant methodology applied for different purposes including policy making, and technology

Table 1
10 most effective papers in ‘patent mining’ based on ‘Eigenvector centrality’.

Author(s)	Title	Publication year	Method	Application
Tijssen [49]	Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows	2001	Patent citation analysis	Policy Making
Lee, PC; Su, HN; Wu, FS [50]	Quantitative mapping of patented technology - The case of electrical conducting polymer nanocomposite	2010	Patent citation analysis	Technology Forecasting
Yoon, Byungun [51]	A systematic approach for identifying technology opportunities: Keyword-based morphology analysis	2005	Patents morphology analysis	Technology Forecasting
Gerken, Jan M [52].	A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis	2012	Semantic patent analysis	Technology monitoring
Lee, C, Cho, Y, Seol, H, Park, Y [53]	A stochastic patent citation analysis approach to assessing future technological impacts	2012	Stochastic patent citation analysis	Technology Forecasting
Jeon, J, Lee, C, Park, Y [6]	How to Use Patent Information to Search Potential Technology Partners in Open Innovation	2011	Network analysis Text mining	Technology partner selection
Kuan, CH, Huang, MH, Chen, DZ [54]	Capturing and Tracking Performance of Patent Portfolio Using h-Complement Area Centroid	2013	Patent citation analysis	Patent portfolio performance analysis and forecasting
Yoon, B [55]	On the development of a technology intelligence tool for identifying technology opportunity	2008	Morphology analysis Clustering analysis Network analysis	Identify technology opportunities
Lee, C, Jeon, J, Park, Y [56]	Monitoring trends of technological changes based on the dynamic patent lattice: A modified formal concept analysis approach	2011	Dynamic patent lattice (based on Formal concept analysis)	Technology monitoring
Geum, Y, Lee, S, Yoon, B, Park, Y [57]	Identifying and evaluating strategic partners for collaborative R&D: Index-based approach using patents and publications	2013	Patent analysis Publication Analysis	Technology partner selection

forecasting, Also, morphology analysis and semantic analysis are two other growing methodologies that are being applied to technology forecasting and opportunity identification. The results also show that technology forecasting is the main area in patent mining research; however, technology partner selection is another growing domain in patent mining research.

‘Sceintometrics’ and ‘Technological Forecasting and Social Change’ have published the most papers in ‘patent mining’; see Fig. 4. Moreover, the first five journals contain more than 50% of ‘patent mining’ papers. Fig. 1 shows that more than 70% of ‘patent mining’ papers have been published after 2010. This indicates ‘patent mining’ publications have progressively accelerated in recent years.

4.2. Keywords network analysis

In total, 343 keywords are extracted and organized in the network shown in Fig. 6. The timeline view of the network, shown in Fig. 5, helps recognize chronologically the most applied keywords which are innovation, science, indicators, citation analysis, technology, research-and-development, bibliometrics, knowledge, patent mining, patent analysis, text mining, patent citations, citations, industry, information, nanotechnology. The timeline view shows ‘citation analysis’ has been used as the first methodology for patent mining purposes, but ‘patent mining’ has been more recently applied by the authors. This transformation shows researchers are increasingly interested in applying text mining tools for deeper patent content analysis. Nonetheless, it doesn't mean citation analysis is obsolete.

Since keywords are good representatives for research purposes, cluster analysis can lead us to the evolution of patent mining and patent analysis. To do so, the minimum spanning tree method is applied to the keywords network, and eventually 14 clusters are recognized. To comprehend the evolution of patent mining, the articles related to the keywords are extracted and their abstracts are evaluated based on three main aspects including ‘Application’ in technology management, ‘Methodology’ and ‘Domain of usage’; key phrases addressing the three aspects are inferred and elicited by reviewing the abstracts and the content of the papers. The results are reflected in Appendix 2.

5. Discussion

In the prior section, clustering 343 keywords revealed 14 clusters, and three main aspects including application (A), methodology (M), and domain (D) are considered. To illuminate the three aspects of the clusters, the main points extracted from the abstracts are recognized, summarized and illustrated chronologically in Appendix 2.

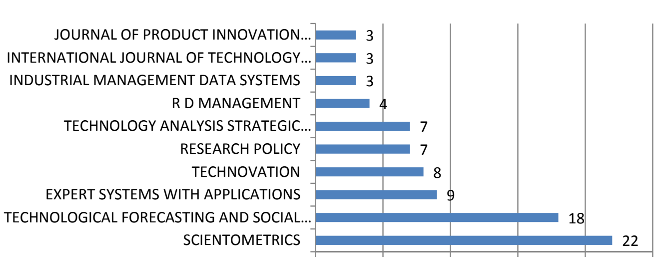

Fig. 4. 10 top journals publishing ‘patent mining’ articles.

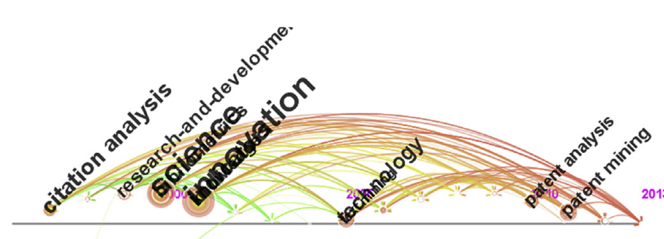

Fig. 5. Time line view of the keywords.

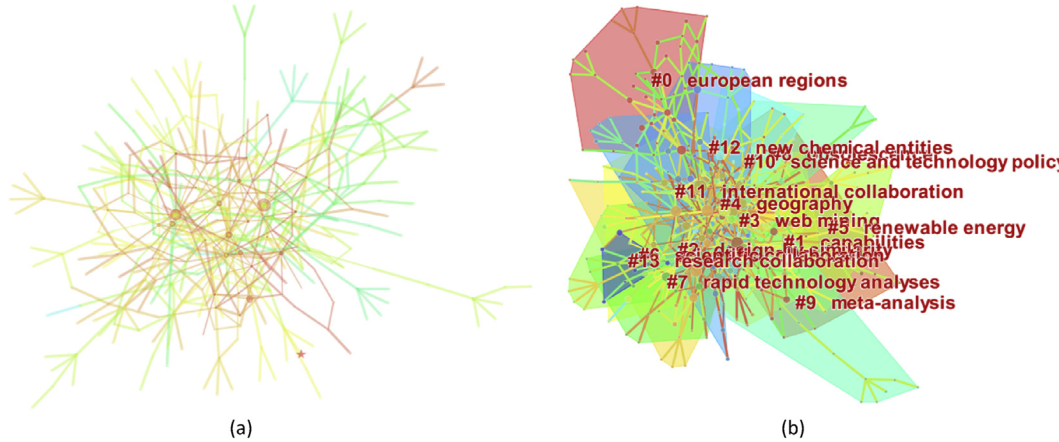

Fig. 6. a) Keywords network, b) keywords network clustered.

5.1. Patent analysis evolution

To reveal the evolution of patent analysis methodologies, they are mapped based on the mean value of the publishing year of each cluster. The map shows there are three main stages in this evolution (see Fig. 7). In the first stage, bibliometrics analysis and citation analysis are the main methodologies applied to analyze the metadata of patents. In the second stage, cluster analysis and network analysis are applied to extract more information and knowledge from metadata and citations. In the third stage, patent mining emerged as a method of analysis through the application of text mining and other complimentary methodologies such as semantic analysis and ontology analysis. Each of these stages is discussed in turn.

In the first stage (clusters 13, 0, and 4), the researchers started analyzing metadata and citations of patents to discover patterns and gaps in technologies. Different types of patent data are analyzed bibliometrically to examine different purposes, particularly for national [58] or regional studies [59,11,60]. For instance, 'forward patent citation' is used to examine the quality of university technology across European regions [59], or, in another study, citations and co-inventors are represented as channels of knowledge flows from G-5 countries to Latin American countries [11].

In stage 2, researchers applied *cluster analysis* and *network analysis* to the papers, in order to enable more detailed and in depth investigations. Cluster analysis (clusters 7, 9 and 12) groups patents

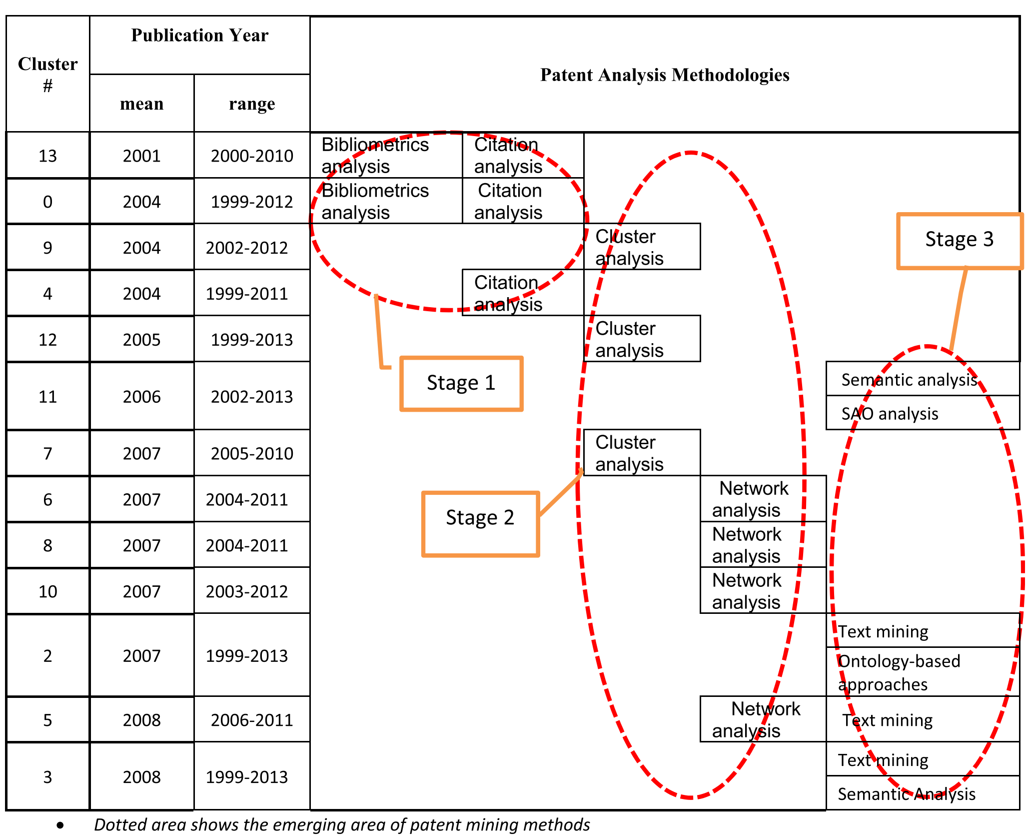

Fig. 7. The map of patent analysis methodologies.

into similar categories, whereas network analysis (clusters 6, 8, 10, and 5) studies the structure of patent networks or citations networks.

In order to clarify how authors of papers in stage 2 applied cluster analysis, we extracted key phrases from clusters 7, 9, and 12, which are available in Appendix 2. Our analysis revealed that the majority of the researchers utilized keywords extracted from patents to cluster patents based on their contents. For instance, Trappey et al. clustered key phrases to group patents that define key innovations [61] and to forecast RFID technologies [2]. In other research, Jun et al. applied a matrix map and the K-medoids clustering method for vacant technology forecasting [62]. In addition to keywords, citations are another facet of patents used for cluster analysis. For instance, Lee et al. applied network analysis and cluster analysis on patent citations to explore technology evolution in electrically conducting polymer nanocomposite [63]. As mentioned, many scholars apply cluster analysis and network analysis together to benefit from both. Cluster analysis helps focus on more specific groups or classes of patents recognized in network analysis, so it helps researchers to analyze subjects such as technology forecasting or technology trend analysis more efficiently.

In order to clarify how authors of papers in stage 2 applied network analysis, we extracted key phrases from clusters 6, 8, 10, and 5, which are also available in Appendix 2. Our investigation revealed that network analysis has enabled scholars to discover the relations between patents and to interpret the content of patents more deeply and efficiently. In the majority of this type of research, patent citations are the most common aspect used for network analysis. This method, often referred to as patent citation network analysis, is applied for a variety of reasons. For instance, it is applied to study technology trend analysis [50,63,18], to detect emerging knowledge domains [64,65], to understand the structure of research in a field of study [66], to explore technology diffusion [67], and to analyze other issues in the technology management field including technology identification [12] and technology transfer [9]. In addition to citation, other aspects of patents are used for network analysis. For instance, the 'co-inventors network' is used to explore knowledge spillover in Latin America [11], or the 'research grants network' is utilized to analyze interdisciplinary research relationships [10].

We reviewed key phrases of clusters 2, 3, 5 and 11 to investigate the third stage, in *patent mining* has already emerged. Our analysis indicates that text mining has enabled researchers to gain access to technical information in the content of patents and to discover the latent knowledge within patents by complementary methodologies, which are semantic analysis and ontology-based approaches. Since patent mining is the focal point of our research, stage three is scrutinized and discussed in section 5.2.

5.2. Patent mining evolution

In the third stage of patent analysis evolution, scholars noticed that just relying on and analyzing citations and the other bibliographic aspects is not enough, since there is a huge amount of knowledge and information in the patent content that has not been considered in prior analyses. As text mining methods have progressed, the scholars started developing content-based approaches [68] by applying text mining methods to extract knowledge and information from patent content through 'patent mining' [19]. This movement is progressing as scholars struggle to create synergies by applying text mining methods and other analytical methods such as network analysis and cluster analysis to develop more efficient patent mining methods. Reviewing clusters 2, 3, and 11 reveals how patent mining methods have been developed and applied over recent years. These novel approaches are based on 'information retrieval', and 'pattern recognition and analysis', as shown in Fig. 8.

5.2.1. Information retrieval

Two main streams of literature discuss how to extract text content more accurately and more efficiently: 1) applying lexical approaches and 2) applying corpus approaches.[1] In the lexical approach, natural language processing capabilities including syntax tagging, word stemming, and stop-word elimination allow us to distinguish words in sentences based on their syntactic features. Since lexical approaches recognize semantic patterns, they can mine patents more accurately and determine important keywords. Therefore, in comparison to traditional methods like TF-IDF, the scholars don't lose synonyms or polysemy and thereby extract all important keywords. In the corpus approach, experts provide a predefined collection of main concepts addressing the text content. These collections are made in either unstructured forms like dictionaries or in structured forms such as ontologies or morphologies. Having these collections create a more efficient keyword extraction and content analysis in later stages of analysis.

Natural language processing (NLP) is able to recognize 'lexical' and 'semantic' relation between words. This capability enables the scholars to follow two different strategies for pattern recognition: 1) semantic analysis and 2) ontology-based analysis. Semantic analysis recognizes important keywords and their relationships, as well as semantic patterns. In other words, semantic patterns are identified based on the meaning of words and their roles in a sentence. The Subject-Action-Object (SAO) approach and its peer approach, the property-function approach, are two semantic analytical approaches, which have been developed to extract textual patterns for purposes of patent analysis. A SAO structure is composed of Subject (noun phrase), Action (verb phrase), and Object (noun phrase) that can be extracted by using natural language processing (NLP) of textual patent information [69]. For example, a sample SAO structure such as 'fire ignites oil' comprises the subject ('fire'), the action ('ignite'), and the object ('oil'). Similarly, in the property-function approach, the property is an adjective describing a specific character of a product and the function is a verb referring to an action of the product [17]. Numerous studies can be fulfilled by applying SAO or property-function approaches for different purposes such as technology roadmapping [15], technology trend identification [16.18], technology monitoring [52], and strategic planning [68]. In addition, some scholars try to find syntactic patterns by applying heuristic algorithms. For instance, Wang and Cheung [70] have developed a method containing heuristics rules to detect simple syntactic patterns. This method enables users to search for patents related to a potential new invention and to provide the relationship and patterns among a group of patents.

In ontology-based analysis, the concepts of patent content are modeled based on their properties, relationships, constraints and behavior [71]. Therefore, to extract terms used for the same concept [72], it is required to deploy NLP methods. Ontology-based analysis provides a framework for interacting with application systems,

[1] Basically, there are three main approaches to extract keywords from texts: 1) corpus approach, 2) lexical approach, and 3) statistical approach. In the corpus approach, it is often required to have a dictionary or predefined corpora developed by subject matter of experts (SME's). In the lexical approach, natural language processing (NLP) is utilized to find out semantic relations among the keywords, since it assumes that the relation between keywords and semantic context of documents determines important keywords. In statistical approach, term frequency is a proxy for the importance of a keyword. The Inverse Document Frequency (TF-IDF) method is one of the first developed and broadly applied statistical methods [75].

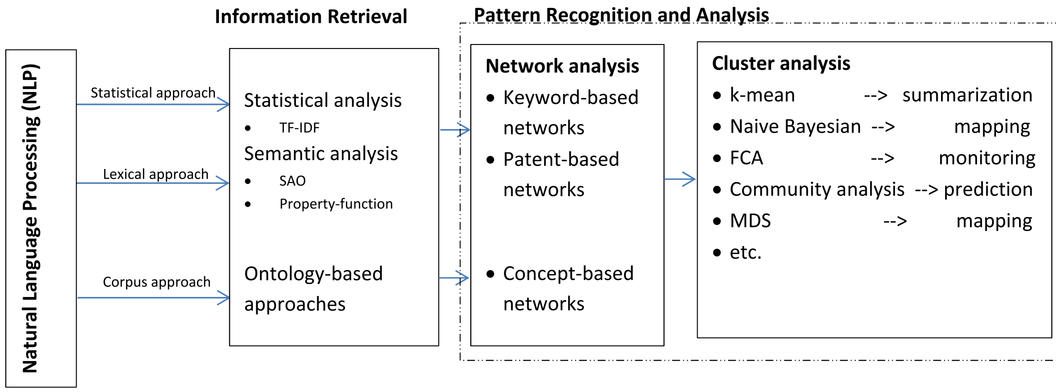

Fig. 8. Patent mining evolution.

improving the communication model between humans and machines [73], and providing information with a knowledge domain [74]. Amy Trappey et al. [75] examined the performance of an ontology-based approach combined with TF-IDF approach in terms of compression ratio, retention ratio, and classification accuracy of the summarization results. They figured out that the ontology-based approach does not provide significant improvement in the compression ratio, but it does produce an 11% improvement for the retention ratio and a 14% improvement for the classification accuracy. In another research, Amy Trappey et al. [76] have developed a knowledge management approach and applied an ontology-based artificial neural network (ANN) to search and classify patent corpora. They combined term frequencies and the concept probabilities of key phrases as the ANN inputs. These approaches have produced significant improvement in classification accuracy.

5.2.2. Pattern recognition and analysis

After given keywords have been extracted and keyword vectors have been prepared, it is time to process keyword vector elements, in order to provide information. In doing so, network analysis and cluster analysis are the two main methodologies that are applied. To process keyword vectors, it is necessary to take these steps: 1) calculating a similarity function and preparing a similarity matrix, 2) transforming the similarity matrix into an adjacency matrix is accomplished by applying a cut-off threshold value, 3) applying basic network analysis measures, and 4) applying cluster analysis algorithms. In pattern analysis, it is not necessary to do all four of these steps. Some scholars only apply similarity calculations, step 1, to do an analysis. For instance, Jeon et al. [6] only applied similarity calculations to search for potential partners for collaboration purposes in open innovation, or Chen et al. [77] applied similarities for cross-language patent matching. Some scholars apply only basic network analysis, consisting of steps 1 to 3. For instance, Choi et al. [15] applied basic network analysis measures to develop roadmaps. And finally, in the most advanced analysis, some scholars apply all of the four steps to create a cluster analysis. For example, Yoon and Kim [78] applied a clustering method, *k* nearest neighbors (*k*-NNs), in their methodology to identify technological opportunities in the network.

5.2.2.1. Network analysis. Network analysis allows researchers to create a set of connected nodes with shared properties and to analyze them based on their network structure and their relationships. There are three types of networks applicable in patent mining: 1) patents-based networks, 2) keyword-based networks, and 3) concept-based networks. The nodes of the networks are patents, keywords, and concepts, respectively, and the relationships are created based on how similar the nodes are. In keyword-based networks, keywords are actors that are connected to a network, and their relationships are specified, if a word-pair has been repeated in an extracted sentence or in a semantic pattern.

Building a patent-based network requires two steps: 1) determine the similarity between the patents based upon their keyword vectors, and 2) convert the similarities to 0 or 1 by applying a predetermined cut-off value. There are different approaches to determining similarity: 1) syntactical and 2) semantic. In a syntactical approach, a network is built based on the role of an extracted keyword in their related sentences. For instance, Choi et al. [15] created a keyword-based network, named Product-Function-Technology (PFT) map, based on the SAO approach. They used network analysis to show how products and technologies are related to functions in order to develop a roadmap. Also, they utilized degree analysis to determine the technological trend, and used centrality measure to illustrate how a core function changes over time. Similarly, Choi and et al. applied this approach to technology trend identification [16,18]. It is worthy to explain that Choi et al. did not use a cut-off value in their research, in order to obtain all possible keywords in the network.

In a semantic approach, semantic similarity between patents is computed based on the similarity of the pairs of words. Tokenizing, stemming, tagging, and determining synonyms are the main steps to figuring out the words [78]. As mentioned above, the similarity matrix is converted to an adjacency matrix by applying a cut-off value to produce the relationships in the network. However, a cut-off value is a task-based and case dependent variable; Lee et al. [74] suggested an empirical method to optimize the cut-off threshold value of similarity.

In concept-based networks, concepts are extracted based on an ontology created by domain experts. Like for patent-based networks, it is necessary to extract keywords addressing the concepts and to apply a similarity measurement in order to develop relationships. For instance, Trappey et al. [75] used a combined ontology based and TF-IDF concept clustering approach to extract, cluster, and integrate the content of a patent to derive a summary and a cluster tree diagram of key terms.

5.2.2.2. Cluster analysis. Given similarity values, cluster analysis is the best methodology to figure out groups of patents, keywords, or concepts that are similar to each other, but different from others in other groups. Cluster analysis contains various algorithms with significantly different views of what makes a cluster and how to efficiently catch a cluster. In recent years, the scholars have tried to apply different clustering methods for various purposes in patent mining. For example, the ‘k-mean algorithm’ for patent summarization [75,4], the ‘naive Bayesian algorithm’ for patent mapping [79], ‘formal concept analysis’ (FCA) for technology monitoring [56], ‘community structure analysis’ for technology prediction

[64,14], 'multi-dimensional scaling' (MDS) for patent mapping [78].

6. Conclusion

Through advances in text mining tools and methods, the popularity of patent mining approaches among technology management scholars has been growing rapidly. In order to provide a comprehensive literature review and analyze the evolution of patent analysis and patent mining, keyword-based network analysis has been applied for the first time, which constitutes the primary contribution of this research. To do so, 143 papers were extracted from the Web of Science database. The next step consisted of using CiteSpace, a bibliometrics tool, to process the data on papers including title, abstract, keywords, and citations. Bibliometrics analysis and cluster analysis are the two additional methodologies that have been applied in this research.

Bibliometric analysis revealed some interesting information, which is described below:

- Among patent mining authors, South Korea has produced four scholars from Pohang University Science & Technology who have had the most impact on the topic of patent mining. Their efforts have resulted in the publication of eleven papers.
- Interestingly, applying Eigenvector centrality shows that the top ten authors are all from South Korea. Six of the authors, Kastoff, Porter, Narin, Jaffe, Yoon, and Trajtenberg are the most cited authors with 37, 31, 30, 28, 26, and 25 citations, respectively.
- Scholars from four countries, South Korea, the USA, Taiwan, and Japan have written the bulk of the 143 papers.
- Researchers working in Pohang University, Seoul National University, University of Tokyo, and Georgia Tech have produced the most diligent and prolific scholars that publish patent mining papers.
- The most applied keywords are 'Innovation', 'Science', 'Indicators', 'Citation Analysis', 'Technology', 'Research-And-Development', 'Bibliometrics', 'Knowledge', 'Patent Mining', 'Patent Analysis', 'Text Mining', 'Patent Citations', 'Citations', 'Industry', 'Information', and 'Nanotechnology'.
- The journals 'Sceintometrics' and 'Technological Forecasting and Social Change' have published the greatest number of 'patent mining' papers, and the first five journals contain more than 50% of 'patent mining' papers. Moreover, Fig. 1 shows that more than 70% of 'patent mining' papers have been published since 2010. This is an indication that the use of 'patent mining' has accelerated over recent years.

In regard to cluster analysis, the papers are divided into 13 clusters. The most important aspects of the clusters are methodology, application, and domain, which are reflected in Appendix 2. To figure out the pattern of evolution among the papers, cluster analysis is applied to the built network based on the keywords extracted from the abstracts. Since patent mining is part of the patent analysis area, related keywords are used in the search query. The evolution is recognized in two main stages: 1) patent analysis evolution and 2) patent mining evolution. As shown in Fig. 7, the scholars started applying measures, indicators, and statistics to analyze bibliometrics data and citations, and then started applying network analysis and cluster analysis to find more complicated patterns in the bibliometrics and citations. After merging and enhancing text mining tools, the scholars commenced utilizing text mining methods to extract keywords, concepts, and generally semantic patterns. In this phase, network analysis has become the main methodology to create relationships, while cluster analysis still shows a remarkable capability to detect patterns to solve different problems. Different clustering methods are used for different purposes; k-mean method for summarization, naïve Bayesian for mapping, FCA for monitoring, community analysis for prediction, and MDS for mapping. These are some examples of how scholars use cluster analysis in patent mining. Scholars are continually trying to be innovative by applying different cluster methods to other technology management issues.

Appendix 1. List of the papers

Authors	Title	Journal	year
Chen, YL; Chiu, YT	Cross-language patent matching via an international patent classification-based concept bridge	Journal Of Information Science	2013
Thorleuchter, D; Van den Poel, D	Weak signal identification with semantic web mining	Expert Systems With Applications	2013
Carvalho, MM; Fleury, A; Lopes, AP	An overview of the literature on technology roadmapping (TRM): Contributions and trends	Technological Forecasting And Social Change	2013
Kiriyama, E; Kajikawa, Y; Fujita, K; Iwata, S	A lead for transvaluation of global nuclear energy research and funded projects in Japan	Applied Energy	2013
Montecchi, T; Russo, D; Liu, Y	Searching in Cooperative Patent Classification: Comparison between keyword and concept-based search	Advanced Engineering Informatics	2013
Kuan, CH; Huang, MH; Chen, DZ	Capturing and Tracking Performance of Patent Portfolio Using h-Complement Area Centroid	Ieee Transactions On Engineering Management	2013
Thorleuchter, D; Van den Poel, D	Web mining based extraction of problem solution ideas	Expert Systems With Applications	2013
Scopel, F; Gregolin, JAR; de Faria, LIL	Technological Trends in the Use of Sisal in Composites Through Patent Mining	Polimeros-Ciencia E Tecnologia	2013
Hopkins, MM; Siepel, J	Just how difficult can it be counting up R&D funding for emerging technologies (and is tech mining with proxy measures going to be any better)?	Technology Analysis & Strategic Management	2013
Gomila, JMV; Marro, FP	Combining tech-mining and semantic-TRIZ for a faster and better technology analysis: a case in energy storage systems	Technology Analysis & Strategic Management	2013
Geum, Y; Lee, S; Yoon, B; Park, Y	Identifying and evaluating strategic partners for collaborative R&D: Index-based approach using patents and publications	Technovation	2013
Block, J; Miller, D; Jaskiewicz, P; Spiegel, F	Economic and Technological Importance of Innovations in Large Family and Founder Firms: An Analysis of Patent Data	Family Business Review	2013
Kodama, H; Watatani, K; Sengoku, S	Competency-based assessment of academic interdisciplinary research and implication to university management	Research Evaluation	2013
Park, H; Kim, K; Choi, S; Yoon, J	A patent intelligence system for strategic technology planning	Expert Systems With Applications	2013

(*continued*)

Authors	Title	Journal	year
Erdi, P; Makovi, K; Somogyvari, Z; Strandburg, K; Tobochnik, J; Volf, P; Zalanyi, L	Prediction of emerging technologies based on analysis of the US patent citation network	Scientometrics	2013
Arora, SK; Porter, AL; Youtie, J; Shapira, P	Capturing new developments in an emerging technology: an updated search strategy for identifying nanotechnology research outputs	Scientometrics	2013
Barirani, A; Agard, B; Beaudry, C	Discovering and assessing fields of expertise in nanomedicine: a patent co-citation network perspective	Scientometrics	2013
Park, H; Ree, JJ; Kim, K	Identification of promising patents for technology transfers using TRIZ evolution trends	Expert Systems With Applications	2013
Yoon, J; Park, H; Kim, K	Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis	Scientometrics	2013
Choi, S; Kim, H; Yoon, J; Kim, K; Lee, JY	An SAO-based text-mining approach for technology roadmapping using patent information	R & D Management	2013
Liang, Y; Liu, Y; Kwong, CK; Lee, WB	Learning the “Whys”: Discovering design rationale using text mining - An algorithm perspective	Computer-Aided Design	2012
Choi, S; Park, H; Kang, D; Lee, JY; Kim, K	An SAO-based text mining approach to building a technology tree for technology planning	Expert Systems With Applications	2012
Popp, D; Newell, R	Where does energy R&D come from? Examining crowding out from energy R&D	Energy Economics	2012
Leydesdorff, L; Bornmann, L	Mapping (USPTO) Patent Data Using Overlays to Google Maps	Journal Of The American Society For Information Science And Technology	2012
Hung, WC	Measuring the use of public research in firm R&D in the Hsinchu Science Park	Scientometrics	2012
Yoon, J; Kim, K	An analysis of property-function based patent networks for strategic R&D planning in fast-moving industries: The case of silicon-based thin film solar cells	Expert Systems With Applications	2012
Garechana, G; Rio, R; Cilleruelo, E; Gavilanes, J	Tracking the evolution of waste recycling research using overlay maps of science	Waste Management	2012
Gerken, JM; Moehrle, MG	A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis	Scientometrics	2012
Park, Y; Lee, S; Lee, S	Patent analysis for promoting technology transfer in multi-technology industries: the Korean aerospace industry case	Journal Of Technology Transfer	2012
Thongpapanl, N	The changing landscape of technology and innovation management: An updated ranking of journals in the field	Technovation	2012
Acosta, M; Coronado, D; Martinez, MA	Spatial differences in the quality of university patenting: Do regions matter?	Research Policy	2012
Prickett, P; Aparicio, I	The development of a modified TRIZ Technical System ontology	Computers In Industry	2012
Magazzini, L; Pammolli, F; Riccaboni, M	Learning from Failures or Failing to Learn? Lessons from Pharmaceutical R&D	European Management Review	2012
Behkami, NA; Daim, TU	Research Forecasting for Health Information Technology (HIT), using technology intelligence	Technological Forecasting And Social Change	2012
Yoon, J; Kim, K	TrendPerceptor: A property-function based technology intelligence system for identifying technology trends from patents	Expert Systems With Applications	2012
Yoon, J; Kim, K	Detecting signals of new technological opportunities using semantic patent analysis and outlier detection	Scientometrics	2012
Park, H; Yoon, J; Kim, K	Identifying patent infringement using SAO based semantic technological similarities	Scientometrics	2012
Jun, S; Park, SS; Jang, DS	Technology forecasting using matrix map and patent clustering	Industrial Management & Data Systems	2012
Lee, S; Mortara, L; Kerr, C; Phaal, R; Probert, D	Analysis of document-mining techniques and tools for technology intelligence: discovering knowledge from technical documents	International Journal Of Technology Management	2012
Guo, Y; Ma, TT; Porter, AL; Huang, L	Text mining of information resources to inform Forecasting Innovation Pathways	Technology Analysis & Strategic Management	2012
Otsuka, K	University patenting and knowledge spillover in Japan: panel-data analysis with citation data	Applied Economics Letters	2012
Lee, C; Cho, Y; Seol, H; Park, Y	A stochastic patent citation analysis approach to assessing future technological impacts	Technological Forecasting And Social Change	2012
Cho, TS; Shih, HY	Patent citation network analysis of core and emerging technologies in Taiwan: 1997–2008	Scientometrics	2011
Wang, WM; Cheung, CF	A Semantic-based Intellectual Property Management System (SIPMS) for supporting patent analysis	Engineering Applications Of Artificial Intelligence	2011
Jin, JH; Park, SC; Pyon, CU	Finding research trend of convergence technology based on Korean R&D network	Expert Systems With Applications	2011
Liu, JS; Kuan, CH; Cha, SC; Chuang, WL; Gau, GJ; Jeng, JY	Photovoltaic technology development: A perspective from patent growth analysis	Solar Energy Materials And Solar Cells	2011
Ruegg, R; Thomas, P	Tracing from applied research programs to downstream innovations: value in multiple techniques	Research Evaluation	2011
Jeon, J; Lee, C; Park, Y	How to Use Patent Information to Search Potential Technology Partners in Open Innovation	Journal Of Intellectual Property Rights	2011
Choi, S; Yoon, J; Kim, K; Lee, JY; Kim, CH	SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells	Scientometrics	2011
Shibata, N; Kajikawa, A; Sakata, I	Measuring Relatedness Between Communities in a Citation Network	Journal Of The American Society For Information Science And Technology	2011

(*continued on next page*)

(*continued*)

Authors	Title	Journal	year
Yoon, J; Kim, K	Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks	Scientometrics	2011
Verhaegen, PA; D'hondt, J; Vandevenne, D; Dewulf, S; Duflou, JR	Identifying candidates for design-by-analogy	Computers In Industry	2011
Chen, YL; Chiu, YT	An IPC-based vector space model for patent retrieval	Information Processing & Management	2011
Lee, C; Jeon, J; Park, Y	Monitoring trends of technological changes based on the dynamic patent lattice: A modified formal concept analysis approach	Technological Forecasting And Social Change	2011
Porter, AL; Newman, NC	Mining external R&D	Technovation	2011
Woon, WL; Zeineldin, H; Madnick, S	Bibliometric analysis of distributed generation	Technological Forecasting And Social Change	2011
Yoon, J; Choi, S; Kim, K	Invention property-function network analysis of patents: a case of silicon-based thin film solar cells	Scientometrics	2011
Porter, AL; Guo, Y; Chiavatta, D	Tech mining: text mining and visualization tools, as applied to nanoenhanced solar cells	Wiley Interdisciplinary Reviews-Data Mining And Knowledge Discovery	2011
Montobbio, F; Sterzi, V	Inventing together: exploring the nature of international knowledge spillovers in Latin America	Journal Of Evolutionary Economics	2011
Duan, CH	Mapping the intellectual structure of modern technology management	Technology Analysis & Strategic Management	2011
Klincewicz, K; Miyazaki, K	Sectoral systems of innovation in Asia. The case of software research activities	International Journal Of Technology Management	2011
Trappey, CV; Wu, HY; Taghaboni-Dutta, F; Trappey, AJC	Using patent data for technology forecasting: China RFID patent analysis	Advanced Engineering Informatics	2011
Beyhan, B; Cetindamar, D	No escape from the dominant theories: The analysis of intellectual pillars of technology management in developing countries	Technological Forecasting And Social Change	2011
Schoeneck, DJ; Porter, AL; Kostoff, RN; Berger, EM	Assessment of Brazil's research literature	Technology Analysis & Strategic Management	2011
Lee, PC; Su, HN; Chan, TY	Assessment of ontology-based knowledge network formation by Vector-Space Model	Scientometrics	2010
Yoon, B; Lee, S; Lee, G	Development and application of a keyword-based knowledge map for effective R&D planning	Scientometrics	2010
Hung, CL; Chou, JCL; Roan, HW	Evaluating a national science and technology program using the human capital and relational asset perspectives	Evaluation And Program Planning	2010
Wang, MY; Chang, DS; Kao, CH	Identifying technology trends for R&D planning using TRIZ and text mining	R & D Management	2010
Partridge, H; Menzies, V; Lee, J; Munro, C	The contemporary librarian: Skills, knowledge and attributes required in a world of emerging technologies	Library & Information Science Research	2010
Rogers, JD	Citation analysis of nanotechnology at the field level: implications of R&D evaluation	Research Evaluation	2010
Wang, JC; Chiang, CH; Lin, SW	Network structure of innovation: can brokerage or closure predict patent quality?	Scientometrics	2010
Shibata, N; Kajikawa, Y; Sakata, I	Extracting the commercialization gap between science and technology - Case study of a solar cell	Technological Forecasting And Social Change	2010
Seol, SS; Jin, FZ; Kwon, S	An Analysis of Chinese Studies on the Management of Science, Technology and Innovation	Asian Journal Of Technology Innovation	2010
Trappey, CV; Trappey, AJC; Wu, CY	CLUSTERING PATENTS USING NON-EXHAUSTIVE OVERLAPS	Journal Of Systems Science And Systems Engineering	2010
Yang, CH; Park, HW; Heo, J	A network analysis of interdisciplinary research relationships: the Korean government's R&D grant program	Scientometrics	2010
Islam, N; Miyazaki, K	An empirical analysis of nanotechnology research domains	Technovation	2010
Guo, Y; Huang, L; Porter, AL	The research profiling method applied to nano-enhanced, thin-film solar cells	R & D Management	2010
Lee, PC; Su, HN; Wu, FS	Quantitative mapping of patented technology - The case of electrical conducting polymer nanocomposite	Technological Forecasting And Social Change	2010
Lee, YG	Sectoral strategic differences of technological development between electronics and chemistry: a historical view from analyses of Korean-invented US patents during the period of 1989–1992	Scientometrics	2010
Cheng, YH; Kuan, FY; Chuang, SC; Ken, Y	Profitability decided by patent quality? An empirical study of the US semiconductor industry	Scientometrics	2010
Criscuolo, P	Inter-firm reverse technology transfer: the home country effect of R&D internationalization	Industrial And Corporate Change	2009
Woon, WL; Madnick, S	Asymmetric information distances for automated taxonomy construction	Knowledge And Information Systems	2009
Morel, CM; Serruya, SJ; Penna, GO; Guimaraes, R	Co-authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programs on Neglected Diseases	Plos Neglected Tropical Diseases	2009
Choi, C; Park, Y	Monitoring the organic structure of technology based on the patent development paths	Technological Forecasting And Social Change	2009
Trappey, AJC; Trappey, CV; Wu, CY	AUTOMATIC PATENT DOCUMENT SUMMARIZATION FOR COLLABORATIVE KNOWLEDGE SYSTEMS AND SERVICES	Journal Of Systems Science And Systems Engineering	2009
Bose, R	Advanced analytics: opportunities and challenges	Industrial Management & Data Systems	2009
Chang, SB; Lai, KK; Chang, SM	Exploring technology diffusion and classification of business methods: Using the patent citation network	Technological Forecasting And Social Change	2009

(*continued*)

Authors	Title	Journal	year
Vasconcellos, E; Bruno, MAC; Campanario, MD; Noffs, SL	A new graphic format to facilitate the understanding of technological innovation models: the seesaw of competitiveness	Technology Analysis & Strategic Management	2009
Kajikawa, Y; Takeda, Y	Structure of research on biomass and bio-fuels: A citation-based approach	Technological Forecasting And Social Change	2008
Shibata, N; Kajikawa, Y; Takeda, Y; Matsushima, K	Detecting emerging research fronts based on topological measures in citation networks of scientific publications	Technovation	2008
Yoon, B	On the development of a technology intelligence tool for identifying technology opportunity	Expert Systems With Applications	2008
Deng, Y	The value of knowledge spillovers in the US semiconductor industry	International Journal Of Industrial Organization	2008
Kajikawa, Y; Yoshikawa, J; Takeda, Y; Matsushima, K	Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy	Technological Forecasting And Social Change	2008
Page, AL; Schirr, GR	Growth and development of a body of knowledge: 16 years of new product development research, 1989–2004	Journal Of Product Innovation Management	2008
Seol, SS; Park, JM	Knowledge sources of innovation studies in Korea: A citation analysis	Scientometrics	2008
Trappey, AJC; Trappey, CV	An R&D knowledge management method for patent document summarization	Industrial Management & Data Systems	2008
Li, X; Chen, HC; Dang, Y; Lin, YL; Larson, CA; Roco, MC	A longitudinal analysis of nanotechnology literature: 1976–2004	Journal Of Nanoparticle Research	2008
McMillan, GS	Mapping the invisible colleges of R&D Management	R & D Management	2008
Leten, B; Belderbos, R; Van Looy, B	Technological diversification, coherence, and performance of firms	Journal Of Product Innovation Management	2007
Senthilkumaran, P; Amudhavalli, A	Mapping of spices research in Asian countries	Scientometrics	2007
Alencar, MSM; Porter, AL; Antunes, AMS	Nanopatenting patterns in relation to product life cycle	Technological Forecasting And Social Change	2007
Miyazaki, K; Islam, N	Nanotechnology systems of innovation - An analysis of industry and academia research activities	Technovation	2007
Singh, J	Asymmetry of knowledge spillovers between MNCs and host country firms	Journal Of International Business Studies	2007
Yoon, B; Park, Y	Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information	Ieee Transactions On Engineering Management	2007
Mogoutov, A; Kahane, B	Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking	Research Policy	2007
Shin, J; Park, Y	Building the national ICT frontier: The case of Korea	Information Economics And Policy	2007
Bhattacharya, S; Garg, KC; Sharma, SC; Dutt, B	Indian patenting activity in international and domestic patentsystem: Contemporary scenario	Current Science	2007
Nagaoka, S	Assessing the R&D management of a firm in terms of speed and science linkage: Evidence from the US patents	Journal Of Economics & Management Strategy	2007
Rebholz-Schuhman, D; Cameron, G; Clark, D; van Mulligen, E; Coatrieux, JL; Barbolla, ED; Martin-Sanchez, F; Milanesi, L; Porro, I; Beltrame, F; Tollis, I; Van der Lei, J	SYMBIOmatics: Synergies in Medical Informatics and Bioinformatics - exploring current scientific literature for emerging topics	Bmc Bioinformatics	2007
Lin, BW; Chen, CJ; Wu, HL	Predicting citations to biotechnology patents based on the information from the patent documents	International Journal Of Technology Management	2007
Trumbach, CC; Payne, D	Identifying synonymous concepts in preparation for technology mining	Journal Of Information Science	2007
Merino, MTG; Do Carmo, MLP; Alvarez, MVS	25 Years of Technovation: Characterisation and evolution of the Journal	Technovation	2006
Trumbach, CC; Payne, D; Kongthon, A	Technology mining for small firms: Knowledge prospecting for competitive advantage	Technological Forecasting And Social Change	2006
Becker, HA; Sanders, K	Innovations in meta-analysis and social impact analysis relevant for tech mining 10.1016/j.techfore.2006.01.008	Technological Forecasting And Social Change	2006
Santo, MD; Coelho, GM; dos Santos, DM; Filho, LF	Text mining as a valuable tool in foresight exercises: A study on nanotechnology	Technological Forecasting And Social Change	2006
Hou, JL; Yang, ST	Technology-mining model concerning operation characteristics of technology and service providers	International Journal Of Production Research	2006
Lange, T	The imprecise science of evaluating scholarly performance - Utilizing broad quality categories for an assessment of business and management journals	Evaluation Review	2006
Trumbach, CC	Addressing the information needs of technology managers: Making derived information usable	Technology Analysis & Strategic Management	2006
Pilkington, A; Teichert, T	Management of technology: themes, concepts and relationships	Technovation	2006
de Buenaga, M; Mana, M; Gachet, D; Mata, J	The SINAMED and ISIS projects: Applying text mining techniques to improve access to a medical digital library	Research And Advanced Technology For Digital Libraries	2006
Antunes, AMD; Mangueira, ACS	The importance of the observatory of industrial activities and trends in science, technology and innovation	Quimica Nova	2005
Porter, AL	QTIP: Quick technology intelligence processes	Technological Forecasting And Social Change	2005
McCain, KW; Verner, JM; Hislop, GW; Evanco, W; Cole, V	The use of bibliometric and Knowledge Elicitation techniques to map a knowledge domain: Software Engineering in the 1990s	Scientometrics	2005
Coronado, D; Acosta, M	The effects of scientific regional opportunities in science-technology flows: Evidence from scientific literature in firms patent data	Annals Of Regional Science	2005

(*continued on next page*)

(*continued*)

Authors	Title	Journal	year
Schlogl, C	Information and knowledge management: Dimensions and approaches	Information Research-An International Electronic Journal	2005
Fung, MK	Are knowledge spillovers driving the convergence of productivity among firms?	Economica	2005
Lai, KK; Wu, SJ	Using the patent co-citation approach to establish a new patent classification system	Information Processing & Management	2005
Yoon, B; Park, Y	A systematic approach for identifying technology opportunities: Keyword-based morphology analysis	Technological Forecasting And Social Change	2005
Linton, JD; Thongpapanl, N	PERSPECTIVE: Ranking the technology innovation management journals	Journal Of Product Innovation Management	2004
Acosta, M; Coronado, D	Science-technology flows in Spanish regions - An analysis of scientific citations in patents	Research Policy	2003
Morris, S; DeYong, C; Wu, Z; Salman, S; Yemenu, D	DIVA: a visualization system for exploring document databases for technology forecasting	Computers & Industrial Engineering	2002
Zhu, DH; Porter, AL	Automated extraction and visualization of information for technological intelligence and forecasting	Technological Forecasting And Social Change	2002
Tijssen, RJW	Science dependence of technologies: evidence from inventions and their inventors	Research Policy	2002
Deeds, DL	The role of R&D intensity, technical development and absorptive capacity in creating entrepreneurial wealth in high technology start-ups	Journal Of Engineering And Technology Management	2001
Tijssen, RJW	Global and domestic utilization of industrial relevant science: patent citation analysis of science-technology interactions and knowledge flows	Research Policy	2001
Melin, G; Danell, R; Persson, O	A bibliometric mapping of the scientific landscape on Taiwan	Issues & Studies	2000
De Moya-Anegon, F; Herrero-Solana, V	Science in America Latina: A comparison of bibliometric and scientific-technical indicators	Scientometrics	1999
Tijssen, RJW; van Wijk, E	In search of the European Paradox: an international comparison of Europe's scientific performance and knowledge flows in information and communication technologies research	Research Policy	1999
Mogee, ME; Kolar, RG	Patent citation analysis of new chemical entities claimed as pharmaceuticals	Expert Opinion On Therapeutic Patents	1998
Davidse, RJ; VanRaan, AFJ	Out of particles: Impact of CERN, DESY and SLAC research to fields other than physics	Scientometrics	1997
Ingwersen, P; Christensen, FH	Data sea isolation for bibliometric online analyses of research publications: Fundamental methodological issues	Journal Of The American Society For Information Science	1997
Penan, H	R&D strategy in a techno-economic network: Alzheimer's disease therapeutic strategies	Research Policy	1996
DELAMOTHE, J	THE REVISION OF INTERNATIONAL SCIENCE INDICATORS - THE FRASCATI MANUAL	Technology In Society	1992

Appendix 2. Keyword cluster analysis results

Cluster#	Size (keyword)	Silhou-ette	Year (mean)	Top terms (tf*idf)	Evolutionary aspect (Methodology/ Application/Domain)	Inferred key pharase
13	7	0.954	2001	(7.85) research collaboration; (7.28) taiwan r&d; (7.28) scientific networks; (7.28) co-authorship; (7.28) asia	• Bibliometrics analysis (M) • Citation analysis (M) • Science/technology landscaping (A)	• bibliometrics citation analysis and mapping techniques are applied to find main actors and publication patterns in university level in Taiwan to figure out the structure of the national research and development systems [58]. • scientific publications as representative of scientific researches and patents as representative of technology are examined by citation analysis to discover gaps between science and technology in solar systems and consequently reveal industrial commercialization opportunities [80].
0	40	0.851	2004	(11.8) european regions; (11.8) multinational firms; (9.15) r& d; (9.15) bioscience megacentres; (9.15) bayh-dole act	• Bibliometrics analysis (M) • Citation analysis (M) • knowledge flow patterns (A) • regional studies(D)	• patent citation analysis to analyze the quality of university technology across European regions [59]. • patent citation analysis to study three channels of knowledge flows from G-5 countries to Latin American Countries [11].

(*continued*)

Cluster#	Size (keyword)	Silhou-ette	Year (mean)	Top terms (tf*idf)	Evolutionary aspect (Methodology/ Application/Domain)	Inferred key phrase
4	32	0.797	2005	(10.7) geography; (9.15) bibliometric techniques; (9.15) factor analysis; (9.09) capabilities; (8.04) technovation	• Citation analysis (M)	• To explore the role of R&D intensity, technical development and absorptive capacity in creating entrepreneurial wealth [81], • To study the effects of scientific regional opportunities in science-technology flows [60], • To study technological coherence in technology diversification strategies of firms [82]
9	20	0.846	2004	(14.46) meta-analysis; (11.8) think tanks; (11.8) solidarity at work; (11.8) social impact assessment; (11.8) value transfer	• Cluster analysis (M)	• Applying a clustering algorithm on key phrases to group patents defining related key innovations [61] • Introducing DIVA, a visualization software program, for mapping patents and scientific literatures by applying cluster analysis [83]
12	17	0.822	2005	(9.15) new chemical entities; (9.15) references; (9.15) technology cycle time; (8.04) information distance; (8.04) patent mapping	• Cluster analysis (M) • Technology forecasting (A)	• Applying patent content clustering and technology life cycle forecasting for RFID technology forecasting in China [2] • Studying the evolution of electrical conducting polymer nanocomposite technology by applying patent citation network analysis [63]
7	22	0.754	2007	(10) rapid technology analyses; (8.62) technical intelligence; (8.04) business intelligence; (8.04) cartography; (8.04) software systems	• Cluster analysis (M) • Technology Forecasting (A) • Patent classification (A)	• Cluster analysis ˆ Cluster analysis clustering method based on supportive vector clustering for vacant technology forecasting [62] ˆ Developing a clustering algorithm to group patents defining related key innovations [61] • Technology Forecasting ˆ Applying different methodologies such as Morphology analysis [51], matrix map and K-medoids clustering method [62] • Patent classification ˆ Creating a patent classification system to replace IPC or UPC by applying co-citation analysis of bibliometrics [84]
6	22	0.871	2007	(10) scientific collaboration; (10) social networks; (10) mycobacterium-tuberculosis; (10) coauthorship networks; (10) institutional collaboration	• Network analysis(M) • Nano-technology (D)	• Network analysis ˆ Examining network topologies of interdisciplinary research relationships in science and technology (S&T) and investigates the relational linkages between the interdisciplinary relations and the quality of research performance by applying network analysis [10] ˆ Applying patent citation network analysis to examine patents quality [85] ˆ Applying network analysis on patent citations and co-inventorships to study three channels of knowledge flows from G-5 countries to Latin American Countries [11] • Nano-technology ˆ Making a classification to understand the whole research spectrum in Nano-technology providing an insight into horizontal comparisons between the research domains using tech mining [86] ˆ Assessing the status of Nano-technology research by applying bibliographic, content map, and citation network analysis [87] ˆ Comparing the positions of national nanotechnology development efforts based on analyses of patenting from 1994 to 2005 [46]
8	20	0.807	2007	(8.04) obsolescence; (8.04) competitive technological intelligence; (8.04) innovation indicators; (8.04) pareto/nbd model; (8.04) technology maps	• Network analysis (M) • Trend technology analysis	• Network analysis ˆ Trend technology analysis by applying patent citation network and patent citation map [50] ˆ Establishing the indicators for finding basic patents and measure the relationship of these basic patents and classifying the basic patents and explaining the groups of technology diffusion through patent citation analysis [67] ˆ Using patent citation as a proxy of interaction between technologies, a national ICT network is developed [12] ˆ Applying patent citation network analysis to examine patents quality [85] • Trend technology analysis ˆ Trend technology analysis by applying patent citation network and patent citation map [50] ˆ Applying Formal Concept Analysis (FCA) based approach to developing a dynamic patent lattice to monitor trends of technological changes [56]

(*continued on next page*)

(*continued*)

Cluster#	Size (keyword)	Silhou-ette	Year (mean)	Top terms (tf*idf)	Evolutionary aspect (Methodology/ Application/Domain)	Inferred key phrase
10	20	0.739	2007	(8.04) science and technology policy; (8.04) innovation system; (8.04) bibliometrics and tech mining; (7.38) nanotechnology; (6.49) datamining	• Network analysis (M) • Nanotechnology (D)	• Network analysis ˆ developing a systematic method to analyze the transferability of technology via patent citation analysis [9] • Nanotechnology ˆ Proposing a strategy based on an automated lexical modular methodology to track nanotechnology evolutions [88] ˆ Profiling the research patent via 'tech mining' to capture key technological attributes, leading actors and networks [89]
5	25	0.883	2008	(14.91) renewable energy; (13.67) sustainable energy; (11.8) biodiesel production; (11.8) biomass; (11.8) biofuels	• Network analysis (M) • Trend technology analysis (A)	ˆ Using citation network analysis with topological clustering method to detect emerging knowledge domains [64] ˆ Applying citation network analysis to know the current structure of biomass and bio-fuel cell researches [66] ˆ Tracking emerging research domains in energy research by using citation network analysis [65] ˆ Trend technology analysis by applying patent citation network and patent citation map [50] ˆ Using property-function approach to solve the limitations of co-word analysis and applying social network analysis to identify technological trends [18]
11	19	0.864	2006	(8.6) international collaboration; (8.6) mathematics; (8.04) world science; (6.49) information flow; (6.49) diversity	• Text mining (M) ˆ Semantic analysis ˆ SAO analysis • Technology identification (A)	• Introducing a patent intelligence system for strategic planning by applying Subject-Action-Object (SAO) analysis and semantic analysis to construct patent maps and patent networks [68] • Identifying inventions of high novelty in patent data by applying specific form of semantic analysis [52]
2	33	0.746	2007	(10) design-by-similarity; (10) product aspect; (10) design-by-analogy; (10) design methodology; (10) wordnet	• Text mining (M) • Ontology-based approaches (M)	• Text mining ˆ Using the Subject-Action-Object (SAO) approach for technology roadmapping [15] and technology trend identification [16] ˆ Using property-function approach to solve the limitations of co-word analysis and applying social network analysis to identify technological trends [18] ˆ Applying Formal Concept Analysis (FCA) based approach to developing a dynamic patent lattice to monitor trends of technological changes [56] • Ontology-based approaches ˆ Applying Vector-Space Model (VSM) in four ontology knowledge based networks for determining probability of network tie formation between network actors (journal papers and research projects) [74] ˆ Using a combined ontology-based and TF-IDF concept clustering approach to automatically and systematically extract information from patent documents [75]
3	32	0.737	2008	(11.29) web mining; (10.7) latent semantic indexing; (10) acquisition; (10) weak signal; (10) cross impact analysis	• Text mining (M) • Semantic Analysis (M)	• Text mining ˆ Developing a systematic approach to searching potential technology partners via applying co-occurrence vectors and text mining [6] ˆ Investigating data, text, and web mining technologies in terms of how they are used and the issues that they are related to [90] ˆ Developing an automatic patent summarization method combining the concepts of key phrases recognition and significant information density [4] • Semantic Analysis ˆ Identifying inventions of high novelty in patent data by applying specific form of semantic analysis [52] ˆ Assisting the inventors in patent analysis and providing support in technological innovations by employing semantic analysis and text mining techniques [79] ˆ Constructing a cross-language mediator by applying latent semantic indexing to extract concepts [77]

(*continued*)

Cluster#	Size (keyword)	Silhouette	Year (mean)	Top terms (tf*idf)	Evolutionary aspect (Methodology/ Application/Domain)	Inferred key pharase
1	34	0.877	2006	(10.79) capabilities; (10.7) operations management; (10.7) nonparametric methods; (10.7) compatible research quality; (10.7) tenure and promotion practices	• R&D management (A)	• Analyzing the performance of a patent assignee's patent portfolio by applying the centroid of a h-complement from the assignee's citation distribution [54] • Introducing a R&D network to make up the missing aspect of the traditional approaches about using multi-sources and to find out the trend of convergence technology R&D in Korea by constructing a weighted network between experts by using meta-data mapping and the network folding technique [91] • Applying co-citation analysis to explore relationship between a high technology venture's R&D intensity, technical capabilities and absorptive capacity [81]

References

[1] J. Delorme, Dissemination of patent information, World Pat. Inf. 4 (4) (Jan. 1982) 155–158.
[2] C.V. Trappey, H.-Y. Wu, F. Taghaboni-Dutta, A.J.C. Trappey, Using patent data for technology forecasting: China RFID patent analysis, Adv. Eng. Inf. 25 (1) (2011) 53–64.
[3] L. Akers, The future of patent information——a user with a view, World Pat. Inf. 25 (4) (Dec. 2003) 303–312.
[4] A.J.C. Trappey, C.V. Trappey, An R&D knowledge management method for patent document summarization, Ind. Manag. Data Syst. 108 (1–2) (2008) 245–257.
[5] A.J.C. Trappey, S.C.I. Lin, A.C.L. Wang, Using neural network categorization method to develop an innovative knowledge management technology for patent document classification, in: Proceedings of the Ninth International Conference on Computer Supported Cooperative Work in Design, Vol. 2, 2005, pp. 830–835.
[6] J. Jeon, C. Lee, Y. Park, How to use patent information to search potential technology partners in open innovation, J. Intellect. Prop. Rights 16 (5) (2011) 385–393.
[7] F. Narin, E. Noma, R. Perry, Patents as indicators of corporate technological strength, Res. Policy 16 (2–4) (Aug. 1987) 143–155.
[8] M.E. Mogee, R.G. Kolar, International patent analysis as a tool for corporate technology analysis and planning, Technol. Anal. Strateg. Manag. 6 (4) (1994) 485–504.
[9] Y. Park, S. Lee, Patent analysis for promoting technology transfer in multi-technology industries: the Korean aerospace industry case, J. Technol. Transf. 37 (3) (2012) 355–374.
[10] C.H. Yang, H.W. Park, J. Heo, A network analysis of interdisciplinary research relationships: the Korean government's R&D grant program, Scientometrics 83 (1) (2010).
[11] F. Montobbio, V. Sterzi, Inventing together: exploring the nature of international knowledge spillovers in Latin America, J. Evol. Econ. 21 (1) (2011) 53–89.
[12] J. Shin, Y. Park, Building the national ICT frontier: the case of Korea, Inf. Econ. Policy 19 (2) (2007) 249–277.
[13] H.P. Bulsara, S. Gandhi, P.D. Porey, Commercialization of technology innovations and patents: issues and challenges, Asia-Pacific Tech. Monit. (2010) 12–18.
[14] J. Choi, Y. Hwang, Patent keyword network analysis for improving technology development efficiency, Technol. Forecast. Soc. Chang. 83 (2014) 170–182.
[15] S. Choi, H. Kim, J. Yoon, K. Kim, J. Lee, An SAO-based text-mining approach for technology roadmapping using patent information, R&D Manag. 43 (1) (2013) 52–74.
[16] S. Choi, J. Yoon, K. Kim, C.-H. Kim, 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) (2011).
[17] J. Yoon, K. Kim, TrendPerceptor: a property–function based technology intelligence system for identifying technology trends from patents, Expert Syst. Appl. 39 (3) (Feb. 2012) 2927–2938.
[18] J. Yoon, S. Choi, K. Kim, Invention property-function network analysis of patents: a case of silicon-based thin film solar cells, Scientometrics 86 (3) (2011) 687–703.
[19] A.L. Porter, S.W. Cunningham, Tech Mining;: Exploiting New Technologies for Competitive Advantage, Wiley, N.J, 2005.
[20] A. Abbas, L. Zhang, S.U. Khan, A literature review on the state-of-the-art in patent analysis, World Pat. Inf. 37 (Jan. 2014) 3–13.
[21] D. Chiavetta, A. Porter, Tech mining for innovation management, Technol. Anal. Strateg. Manag. 25 (6) (Jul. 2013) 617–618.
[22] C. Chen, CiteSpac, 2014 [Online]. Available: http://cluster.cis.drexel.edu/~cchen/citespace/.
[23] C. Chen, CiteSpace II;: detecting and visualizing emerging trends, J. Am. Soc. Inf. Sci. Technol. 57 (3) (2006) 359–377.
[24] C. Chen, J. Zhang, M.S. Vogeley, Visual analysis of scientific discoveries and knowledge diffusion, in: The 12th International Conference on Scientometrics and Informetrics, 2009, pp. 14–17.
[25] C. Chen, Information Visualization, Springer, Berlin Heidelberg, 2004.
[26] Y. Tonta, H.R. Darvish, Diffusion of latent semantic analysis as a research tool: a social network analysis approach, J. Informetr. 4 (2) (Apr. 2010) 166–174.
[27] M.K. Dhami, H. Olsson, Evol. Interpers. Confl. Paradigm 3 (7) (2008) 547–569.
[28] H.J.Z.C.W. Xu-kun, The Information Visualization Analysis of the Study in International S & T Policy, 2008, pp. 1–9.
[29] Y.Y. Yang, T. Klose, J. Lippy, C.S. Barcelon-Yang, L. Zhang, Leveraging text analytics in patent analysis to empower business decisions – a competitive differentiation of kinase assay technology platforms by I2E text mining software, World Pat. Inf. 39 (Dec. 2014) 24–34.
[30] R. Eito-Brun, Knowledge dissemination patterns in the information retrieval industry: a case study for automatic classification techniques, World Pat. Inf. 39 (Dec. 2014) 50–57.
[31] P. Masiakowski, S. Wang, "Integration of software tools in patent analysis, World Pat. Inf. 35 (2) (Jun. 2013) 97–104.
[32] S. Spangler, C. Ying, J. Kreulen, S. Boyer, T. Griffin, A. Alba, L. Kato, A. Lelescu, S. Yan, Exploratory analytics on patent data sets using the SIMPLE platform, World Pat. Inf. 33 (4) (Dec. 2011) 328–339.
[33] Y.Y. Yang, L. Akers, C.B. Yang, T. Klose, S. Pavlek, Enhancing patent landscape analysis with visualization output, World Pat. Inf. 32 (3) (Sep. 2010) 203–220.
[34] Y. Yang, L. Akers, T. Klose, C. Barcelon Yang, Text mining and visualization tools – impressions of emerging capabilities, World Pat. Inf. 30 (4) (Dec. 2008) 280–293.
[35] B. Buchanan, Unlocking the value of patent data: patent informatics services at the Uk intellectual property office (Uk-IPO), World Pat. Inf. 30 (4) (Dec. 2008) 335–337.
[36] J. Eldridge, Data visualisation tools—a perspective from the pharmaceutical industry, World Pat. Inf. 28 (1) (Mar. 2006) 43–49.
[37] I. McCulloh, H. Armstrong, A. Johnson, Social Network Analysis with Applications, John Wiley, 2013.
[38] J. Yoon, K. Kim, Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks, Scientometrics 88 (1) (2011) 213–228.
[39] J. Yoon, H. Park, K. Kim, Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis, Scientometrics 94 (1) (2013) 313–331.
[40] H. Park, J. Yoon, K. Kim, Identifying patent infringement using SAO based semantic technological similarities, Scientometrics 90 (2) (2012) 515–529.
[41] S. Choi, H. Park, D. Kang, J.Y. Lee, K. Kim, An SAO-based text mining approach to building a technology tree for technology planning, Expert Syst. Appl. 39 (13) (2012) 11443–11455.
[42] A. Porter, N. Newman, Mining External R&D, Technovation, January, 2011.
[43] A.L. Porter, QTIP: quick technology intelligence processes, Technol. Forecast. Soc. Change 72 (9) (2005) 1070–1081.
[44] A.L. Porter, Y. Guo, D. Chiavatta, Tech mining: text mining and visualization tools, as applied to nanoenhanced solar cells, WILEY Interdiscip. Rev. Min. Knowl. Discov. 1 (2) (2011).
[45] Y. Guo, L. Huang, A.L. Porter, The research profiling method applied to nano-enhanced, thin-film solar cells, R&D Manag. 40 (2) (2010).

[46] M.S.M. Alencar, A.L. Porter, A.M.S. Antunes, Nanopatenting patterns in relation to product life cycle, Technol. Forecast. Soc. Change 40 (1661–1680) (2007).
[47] D. Zhu, A.L. Porter, Automated extraction and visualization of information for technological intelligence and forecasting, Technol. Forecast. Soc. Change 69 (5) (2002) 495–506.
[48] Y. Guo, T. Ma, A.L. Porter, L. Huang, Text mining of information resources to inform Forecasting Innovation Pathways, Technol. Anal. Strateg. Manag. 24 (8) (2012) 843–861.
[49] R.J. Tijssen, Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows, Res. Policy 30 (1) (Jan. 2001) 35–54.
[50] P.-C. Lee, H.-N. Su, F.-S. Wu, Quantitative mapping of patented technology — The case of electrical conducting polymer nanocomposite, Technol. Forecast. Soc. Change 77 (3) (Mar. 2010) 466–478.
[51] B. Yoon, Y. Park, A systematic approach for identifying technology opportunities: keyword-based morphology analysis, Technol. Forecast. Soc. Change 72 (2) (Feb. 2005) 145–160.
[52] J.M. Gerken, A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis, Scientometrics 91 (3) (2012) 645.
[53] C. Lee, Y. Cho, H. Seol, Y. Park, A stochastic patent citation analysis approach to assessing future technological impacts, Technol. Forecast. Soc. Change 79 (1) (2012) 16–29.
[54] C. Kuan, M. Huang, D. Chen, Capturing and tracking performance of patent portfolio using h -complement area centroid, IEEE Trans. Eng. Manag. 60 (3) (2013) 496–505.
[55] B. Yoon, On the development of a technology intelligence tool for identifying technology opportunity, Expert Syst. Appl. 35 (1–2) (2008) 124–135.
[56] C. Lee, J. Jeon, Y. Park, Monitoring trends of technological changes based on the dynamic patent lattice: a modified formal concept analysis approach, Technol. Forecast. Soc. Change 78 (4) (May 2011) 690–702.
[57] Y. Geum, S. Lee, B. Yoon, Y. Park, Identifying and evaluating strategic partners for collaborative R&D: index-based approach using patents and publications, Technovation 33 (6–7) (2013) 211–224.
[58] G. Melin, R. Danell, A bibliometric mapping of the scientific landscape on Taiwan, Issues Stud. 36 (5) (2000) 61–82.
[59] M. Acosta, D. Coronado, M. Angeles Martinez, Spatial differences in the quality of university patenting: do regions matter? Res. Policy 41 (4) (2012) 692–703.
[60] D. Coronado, M. Acosta, The effects of scientific regional opportunities in science-technology flows: evidence from scientific literature in firms patent data, Ann. Reg. Sci. 39 (3) (2005) 495–522.
[61] C.V. Trappey, A.J.C. Trappey, C.-Y. Wu, Clustering patents using non-exhaustive overlaps, J. Syst. Sci. Syst. Eng. 19 (2) (2010) 162–181.
[62] S. Jun, S.S. Park, D.S. Jang, Technology forecasting using matrix map and patent clustering, Ind. Manag. Data Syst. 112 (5–6) (2012) 786–807.
[63] P.-C. Lee, H.-N. Su, F.-S. Wu, Quantitative mapping of patented technology - the case of electrical conducting polymer nanocomposite, Technol. Forecast. Soc. Change 77 (3) (2010) 466–478.
[64] N. Shibata, Y. Kajikawa, Y. Takeda, K. Matsushima, Detecting emerging research fronts based on topological measures in citation networks of scientific publications, Technovation 28 (11) (2008) 758–775.
[65] Y. Kajikawa, J. Yoshikawa, Y. Takeda, K. Matsushima, Tracking emerging technologies in energy research: toward a roadmap for sustainable energy, Technol. Forecast. Soc. Change 75 (6) (Jul. 2008) 771–782.
[66] Y. Kajikawa, Y. Takeda, Structure of research on biomass and bio-fuels: a citation-based approach, Technol. Forecast. Soc. Change 75 (9) (2008) 1349–1359.
[67] S.-B. Chang, K.-K. Lai, S.-M. Chang, Exploring technology diffusion and classification of business methods: using the patent citation network, Technol. Forecast. Soc. Change 76 (1) (2009) 107–117.
[68] H. Park, K. Kim, S. Choi, J. Yoon, A patent intelligence system for strategic technology planning, Expert Syst. Appl. 40 (7) (2013) 2373–2390.
[69] G. Cascini, A. Fantechi, E. Spinicci, Natural Language Processing of Patents and Technical Documentation, Springer, Berlin Heidelberg, 2004.
[70] W.M. Wang, C.F. Cheung, A semantic-based intellectual property management system (SIPMS) for supporting patent analysis, Eng. Appl. Artif. Intell. 24 (8) (2011) 1510–1520.
[71] N.F. Noy, D.L. McGuinness, Ontology Development 101: a Guide to Creating Your First Ontology, 2001.
[72] M. Bermudez-Edo, M. Noguera, N. Hurtado-Torres, M.V. Hurtado, J.L. Garrido, Analyzing a firm's international portfolio of technological knowledge: a declarative ontology-based OWL approach for patent documents, Adv. Eng. Inf. 27 (3) (Aug. 2013) 358–365.
[73] S. Weng, H. Chang, Using ontology network analysis for research document recommendation, Expert Syst. Appl. 34 (3) (2008) 1857–1869.
[74] P.-C. Lee, H.-N. Su, T.-Y. Chan, Assessment of ontology-based knowledge network formation by Vector-Space Model, Scientometrics 85 (3) (2010) 689–703.
[75] A.J.C. Trappey, C.V. Trappey, C.-Y. Wu, Automatic patent document summarization for collaborative knowledge systems and services, J. Syst. Sci. Syst. Eng. 18 (1) (2009) 71–94.
[76] A.J.C. Trappey, C.V. Trappey, T.-A. Chiang, Y.-H. Huang, Ontology-based neural network for patent knowledge management in design collaboration, Int. J. Prod. Res. 51 (7) (Apr. 2013) 1992–2005.
[77] Y.-L. Chen, Y.-T. Chiu, Cross-language patent matching via an international patent classification-based concept bridge, J. Inf. Sci. 39 (6) (2013) 737–753.
[78] J. Yoon, K. Kim, Detecting signals of new technological opportunities using semantic patent analysis and outlier detection, Scientometrics 90 (2) (Oct. 2011) 445–461.
[79] W.M. Wang, C.F. Cheung, A semantic-based intellectual property management system (SIPMS) for supporting patent analysis, Eng. Appl. Artif. Intell. 24 (2011) 1510–1520.
[80] N. Shibata, Y. Kajikawa, I. Sakata, Extracting the commercialization gap between science and technology - case study of a solar cell, Technol. Forecast. Soc. Change 77 (7) (2010) 1147–1155.
[81] D. Deeds, The role of R&D intensity, technical development and absorptive capacity in creating entrepreneurial wealth in high technology start-ups, J. Eng. Technol. Manag. 18 (1) (2001) 29–47.
[82] B. Leten, R. Belderbos, B. Van Looy, Technological diversification, coherence, and performance of firms, J. Prod. Innov. Manag. 24 (6) (2007) 567–579.
[83] S. Morris, C. DeYong, S. Salman, D. Yemenu, DIVA: a visualization system for exploring document databases for technology forecasting, Comput. Ind. Eng. 43 (4) (2002) 841–862.
[84] K. Lai, S. Wu, Using the patent co-citation approach to establish a new patent classification system, Inf. Process. Manag. 41 (2) (2005) 313–330.
[85] J.-C. Wang, C. Chiang, S.-W. Lin, Network structure of innovation: can brokerage or closure predict patent quality? Scientometrics 84 (3) (2010) 735–748.
[86] N. Islam, K. Miyazaki, An empirical analysis of nanotechnology research domains, Technovation 30 (4) (2010) 229–237.
[87] X. Li, H. Chen, Y. Dang, Y. Lin, C.A. Larson, M.C. Roco, A longitudinal analysis of nanotechnology literature: 1976-2004, J. Nanoparticle Res. 10 (2008) 3–22.
[88] A. Mogoutov, B. Kahane, Data search strategy for science and technology emergence: a scalable and evolutionary query for nanotechnology tracking, Res. Policy 36 (6) (2007) 893–903.
[89] Y. Guo, L. Huang, A.L. Porter, The research profiling method applied to nano-enhanced, thin-film solar cells, R&D Manag. 40 (2) (2010) 195–208.
[90] R. Bose, Advanced analytics: opportunities and challenges, Ind. Manag. Data Syst. 109 (1–2) (2009) 155–172.
[91] J.H. Jin, S.C. Park, C.U. Pyon, Finding research trend of convergence technology based on Korean R&D network, Expert Syst. Appl. 38 (12) (2011) 15159–15171.