IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received September 26, 2017, accepted November 1, 2017, date of publication November 8, 2017, date of current version December 22, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2771371

Generating Information Relation Matrix Using Semantic Patent Mining for Technology Planning: A Case of Nano-Sensor

WANWOOK KI AND KWANGSOO KIM
Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang 37673, South Korea
Corresponding author: Kwangsoo Kim (kskim@postech.ac.kr)

This work was supported by the National Research Foundation of Korea under Grant NRF-2016R1A2B4008381.

ABSTRACT For the purposes of technology planning and research and development strategy development, we present a semi-automated method that extracts text information from patent data, uses natural language processing to extract the key technical information of the patent, and then visualizes this information in a matrix form. We tried to support qualitative analysis of patent contents by extracting functions, components, and contexts, which are the most important information about inventions. We validated the method by applying it to patent data related to nanosensors. The matrix can emphasize technical information that have not been exploited in patents, and thereby identify development opportunities.

INDEX TERMS Information relation matrix, natural language processing, patent analysis, patent matrix, semantic patent mining, technical information, technology planning, text mining.

I. INTRODUCTION

Technology management has become an essential component in operation of enterprises. Creation of new business ideas and diversification of businesses in an industrial field require a detailed understanding of systematic technology management in the field. Enterprises manage their technologies in the form of patents, utility models, and trade secrets. Patents have the important function of claiming technology rights and assuring progress and novelty in a technology. Patents have a standard format, so by reviewing patents, planners in a business can grasp in detail the technologies its field, and use this information to guide development of technology. As a result, many researchers have tried to use patent analysis to understand, predict, and manage the contents of modern technology.

A. PATENT ANALYSIS

Governments and private companies devote resources and effort into their long-term technological development and innovation. However, a single organization cannot always respond quickly to changes in customer demands. Reduction in the technology lifecycle, and change in technology paradigms have increased the demand for management and analysis of intellectual property rights including patents.

Patent analysis is the most objective indicator that can grasp the technical competitiveness of a country or enterprise. Patents are often considered to be the best resource for timely recognition of technological changes [1]. Patents analysis can provide information on specific conditions that relate to technology or to market-related developments that help decisionmakers to track competitor activity and innovation trends [2]. Patent analysis can also produce much useful technical information, and it can provide reliable information that can guide evaluation of technology and prediction of technological trends [3], [4].

B. WEAKNESS OF EXISTING ANALYSIS METHOD

Various patent-analysis methods have been developed, but they do not include qualitative techniques and algorithms for empirical analysis. The current patent analyses are mostly quantitative and based on bibliographic information. Many countries have developed patent maps, mostly based on the analysis of structured data from bibliographic fields in patents [5], [6]. However, although conventional patent analysis using bibliographic information is simple and easy to understand, it has limited ability to describe technology or to guide creative utilization of the information gathered [7]; therefore, bibliographic information alone cannot accurately extract the technical contents and semantic meaning of a patent. On the other hand, some researchers conducted a portfolio analysis using patent contents [27] and analyzed the requirements of technology from extracting the frequency of patent data [28]. There was also a case study of problem-solution linkage through technical patent text analysis [29]. However, these studies also failed to effectively elicit the key information of the patent that users really wanted.

C. NEEDS FOR TECHNICAL INFORMATION MINING

Many organizations, including companies, should understand their technology competitiveness, but also identify their competencies. For this reason, patent analysis has been used to identify competitors' status, technology level, and competitiveness [8]. The process has also been used to make decisions about technology planning and R&D investments [9], [10], and to monitor technology changes in competitors in related industries [11].

The processes of reading and analyzing patents, and of developing strategies, are difficult and expensive tasks. One particular disadvantage is that the process requires patent experts with knowledge of the specific domain, so the analysis cost may increase exponentially if a large set of patent data is to be analyzed. Furthermore, patent analysis based on expert opinion may be subjective, and is therefore not always accurate [12], and the time required to extract the desired information cannot be easily forecast.

Moreover, even when patent analysis extracts bibliographic information, the technical characteristics and information structure of patents can be difficult to derive. To solve this problem, automation technology that uses data mining is required. Data mining is widely used to access and extract useful information in databases [13], [14]. Text mining is a popular data mining technique to process a huge amount of unstructured textual document patent information; this method can extract text information from patents [15], and patent mining based on Natural Language Processing (NLP) enables extraction of effective technical information from a patent set.

D. INFORMATION RELATION MATRIX

A tool must be developed to analyze the extracted technical information effectively, and patent analysis results with complex structures should be visualized simply. A careful analysis of the technical information in patents is expressed visually as a patent map or matrix, which simplifies the task of understanding complex patent information [16], and also highlights various elements of knowledge about technologies and competitive positions [17]. Moreover, a patent map or matrix can provide details about technologies, relationships among them, business trends, and clues to create new solutions in the industry; the visualizations thereby support decisionmaking for technology development and investment [18], [19].

A patent contains various kinds of information; but its descriptions of technical function, context, and component information are the most fundamental and essential. By utilizing these three information areas, patents that contain technical information desired by the user can be expressed in a matrix form. The Information Relation Matrix (IRM) describes relationships among technical information in the patent.

II. RELATED WORK

Barriers to analysis of patent documents are so high that only superficial theoretical methods have been developed. However, the development of text mining and NLP technology has enabled extraction and analysis of the main contents of the text within a patent. Although this process provides an opportunity to obtain various information necessary for technical planning, the technical information in patents has not been defined and systematized, although these processes would help technology management.

A. EXISTING PATENT ANALYSIS METHOD

A patent is a structured technical document that uses vocabulary that is specialized to a specific domain. Therefore, non-specialists can have difficulty understanding the technical information of patents. Even a technical expert may have difficulty finding patent information that matches a stated purpose.

In the past, quantitative patent analysis based on bibliographic information was dominant. Bibliographic data analysis is defined as the measurement of text and information; this process helps to explore, organize and analyze large amounts of historical data to help researchers and practitioners to identify technology trends and patterns, to guide decisionmaking [20]. Other common bibliometric tools use data such as applicants, inventors, international patent classification (IPC) code, and citations [21]. Citation analysis is the most widely-used analytical technique, because the number of citations of a patent represents its relative importance [22].

Analytical techniques for extracting text and then processing it into semantic information are being developed. This qualitative semantic analysis has become possible because of the development of text mining technology capable of analyzing large amounts of unstructured data [15]. Text mining has been used to classify and cluster useful patent information; in this way some researchers have tried to overcome the limitation of existing patent analysis methods [23]. Text mining using patent keywords has been applied to featured patents, and a method of mapping patents to two-dimensional maps according to keyword similarity between patents has been proposed [24]. A patent map based on text mining for the technology domain of carbon nanotubes was created by using a series of text mining techniques including summary extraction, feature selection, cluster generation, and information mapping [25], [26].

Semantic descriptive information can be extracted by counting the frequency of a certain word. The larger number of times a keyword appears in a key literature, the more important the keyword is [27]. A patent network or matrix can also be constructed by identifying co-occurrences of technical information in patents. The patent contains a variety of unstructured textual information, which can be more useful than formalized information [28]. Moreover, texts in patent contain many technical information, mainly composed of sentences that improve performance or solve specific problems [29]. Subject-Action-Object (SAO) structure can be extracted to grasp the essential information from the sentences in patents. The method presented here will extract technical information in the form of use SAO structures from patents.

B. SUBJECT-ACTION-OBJECT STRUCTURES FOR TECHNOLOGY ANALYSIS

Technical information in a patent is described as text in sections such as title, abstract, background, description, and claim. SAO-based patent analysis analyzes sentences in the patent to extract technical information [30]. Most of the sentences have the form of subject, action (verb) and object, which can identify important information such as the function, effects, characteristics, solutions, components, and context of the technology. Indeed, the 'function', which is the task or action that a system or technology can perform, can be expressed in the form of Action-Object (AO) [30]. Objects, tools, methods, and systems of the inventions specified in the patent can be expressed in the form of Subject (S) in a sentence that contains a description of the technology [30]. For example, in the sentence, 'Nanosensor detects micro-signal' is composed of 'Subject ('Nanosensor'), Action ('detects') and Object ('micro-signal'); 'Nanosensor' represents the solution or tool described in this technical sentence, and 'detect micro-signal' clearly means the function that the 'nanosensor' performs. In addition, if subject and object are components or subsystems of a technical system, the action included in that sentence may represent the structural relationship between the components [31]. For example, in the sentence, 'Nanosensor includes carbon reactor', 'Nanosensor' and 'carbon reactor' mean the components described in this technical sentence and 'includes' can mean a structural and partitive relationship. A SAO structure can also represent a problem-solution relationship, in which the subject represents a solution and AO represents a required function or problem [32].

Other fields such as inventor profiling [32], patent-infringement identification [33], technology monitoring [34], technological-opportunity analysis [32], technology structuring [36], technology evaluation [37], and technology-trend identification [38] have already adopted SAO-based patent analysis.

C. THE CONSTRUCTION OF TECHNOLOGY MATRIX AND MAP

In practical patent analysis, patent maps or matrices (PMs) have been used as useful tools for visual analysis and for management of technical information. However, because of the complex and diverse variables and relationships involved in a patent, the explanatory capacity and operational efficiency of PMs remain limited [21]. Nevertheless, a PM can still be a useful tool for comparing different types of information simultaneously. Above all, its function is very clear and the construction method is very simple, so a PM can enable intuitive judgments. Patents include various types of information, so a PM is very useful for finding and analyzing patents based on each need, by helping to identify co-occurrences and frequency of each kind of information in a patent. Another advantage of the PM is that it provides crucial indicators to convey domain-specific knowledge to R&D managers and executives [39].Several authors have developed similarity matrices of SAO structure. One used semantic measurement based on 'WordNet' to cluster phrases and AOs [40]. Another devised four types of maps (trend, query, aggregation, and zooming) for text-based visualization by filtering, pivoting, or slicing data based on bibliographic information [25].Typical usage of the patent map can be divided into four areas (Table 1) [24], and can be categorized into several types depending on the application [41] (Table 2).

TABLE 1. Typical usage of PM.

Functional area	Specific purpose
Administration	Identify industry trend or specific technology trend Understand competitor's technology asset and strategy
R&D	Select new product theme
	Understand portfolio of technological asset
	Identify technology vacuum
Personnel	Identify portfolio of R&D human resources
Patent management	Identify patent conflicts
	Investigate life cycle and scope of patents

TABLE 2. Classification of PM from 'Korea Invention Promotion Association' in 1999 [24], [41].

Type	Purpose	Main variables
Technical PM	Understand core technology	Patent classification
	Identify technology vacuum	Year
	Catch overall stream of technology	Righter
Management PM	Identify specific technology trend	Applicant
	Select new product theme	References
	Understand competitor's technology and strategy	Inventor
		Year
Claim PM	Understand patent conflicts	Area of the right
	Investigate life cycle and applicability of patents	Patent classification Reference

III. PROPOSED METHOD

The existing patent analysis system searches for keywords in patent contents, and most of the classification methods are based on IPC codes. However, this analysis does not help to

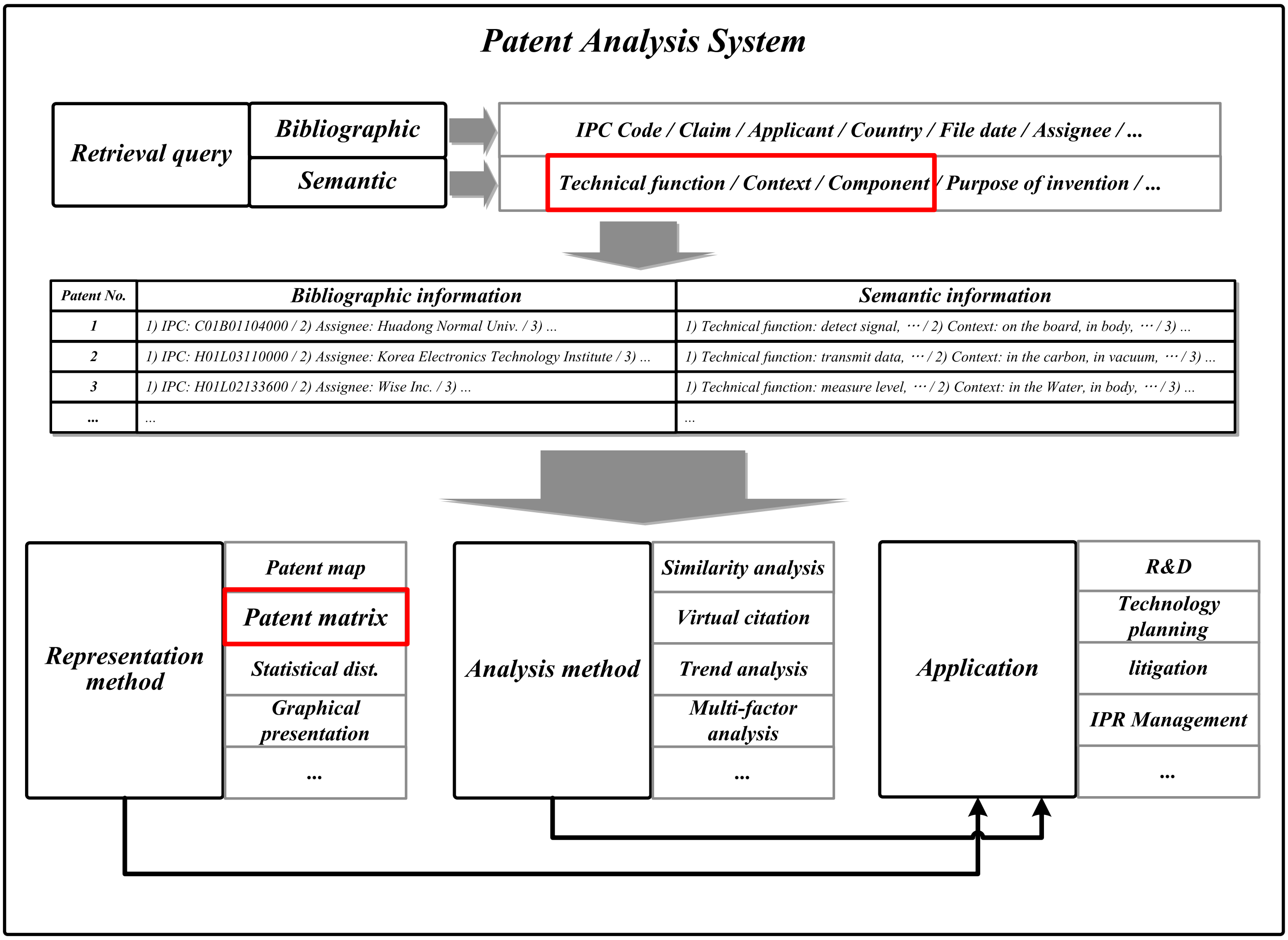

FIGURE 1. Suggested patent analysis system.

grasp the critical information in contents of patents. At best, it can identify the main topic trend of the patent or the frequency distribution of a certain word. Most analysis of patents that is based on bibliographic information has been and is still a necessary part of patent analysis. However, to improve the limitations of such quantitative bibliographic analysis, a function should selectively provide information according to user requirements. People who search for patents, frequently want to know what a particular patent describes; i.e., most users who search for patents will want results that include the main technical functions, context, component, purpose, and possibility of convergence with other fields.

The results of patent analysis can be expressed in various ways. Patent map, patent matrix and patent road map are mainly used, and graphical methods are being developed. The extracted patent information, enables similarity checks, construction of virtual citations, trend analysis by field, and multi-factor analysis. In this research, we focused on understanding the type of patent information and the relationship between information in one patent and other patents by expressing patent technical information effectively. By selecting the desired type of technical information, a user can confirm whether the specific patent includes the required technical information, and can detect relations within this information. Relations in information can clarify the current state of the patented technology, and application of vacancy analysis can identify opportunities and ideas for new technology development or technology planning. This research considered technical function, context, and component of patents; these topics were selected by interviewing patent experts and work-site operators. In the next section we describe the method used to build an IRM based on these three types of technical information.

The proposed method (Fig. 1) to build an IRM extracts technical information, including functions in the form of 'verb + noun' pairs, contexts in the form of 'preposition + noun', and components in the form of 'inclusion verb + noun' from sentences of patents. The extracted technical information is tagged with the identity of the patent from which it was extracted.

These three factors can be considered as features of patents, and as axes in the matrix. Therefore, an IRM can be organized

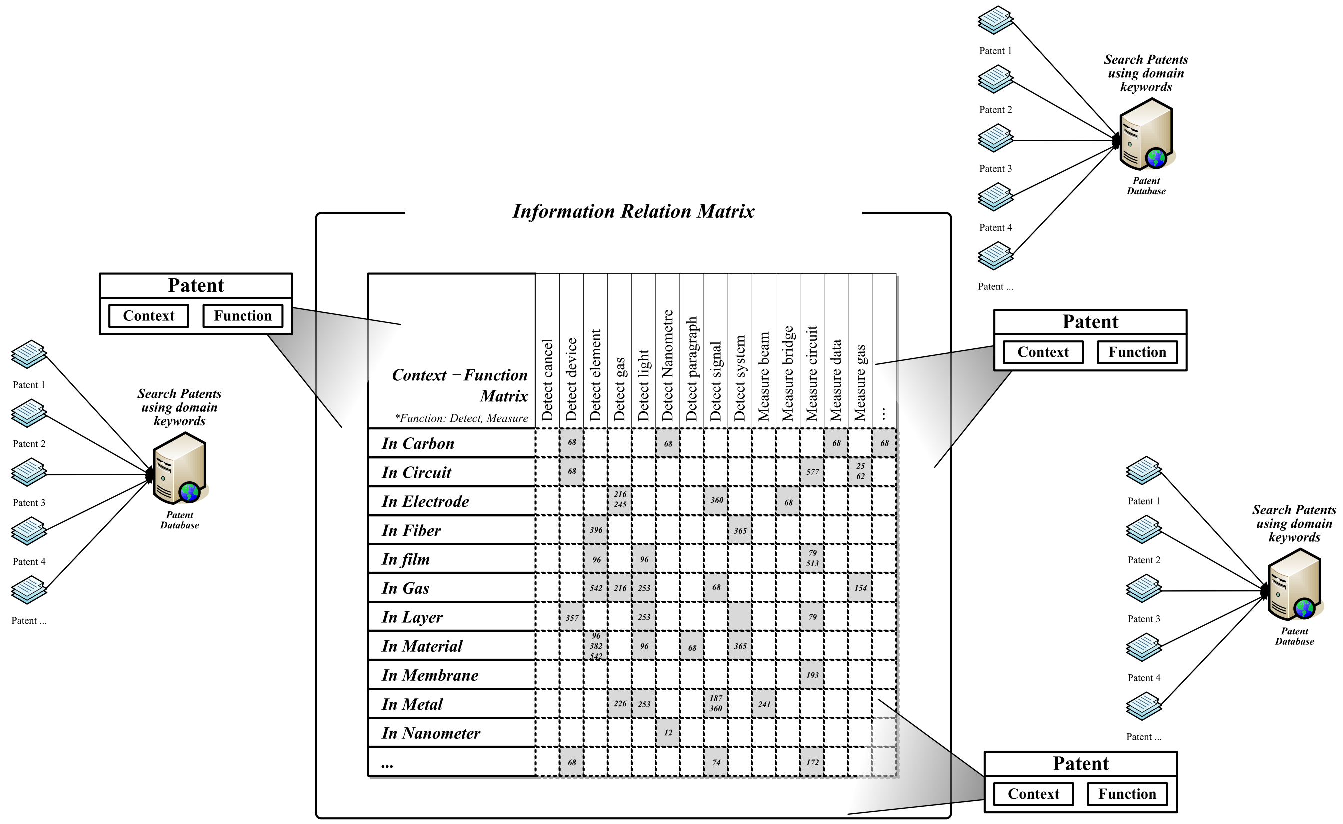

FIGURE 2. Information relation matrix (IRM).

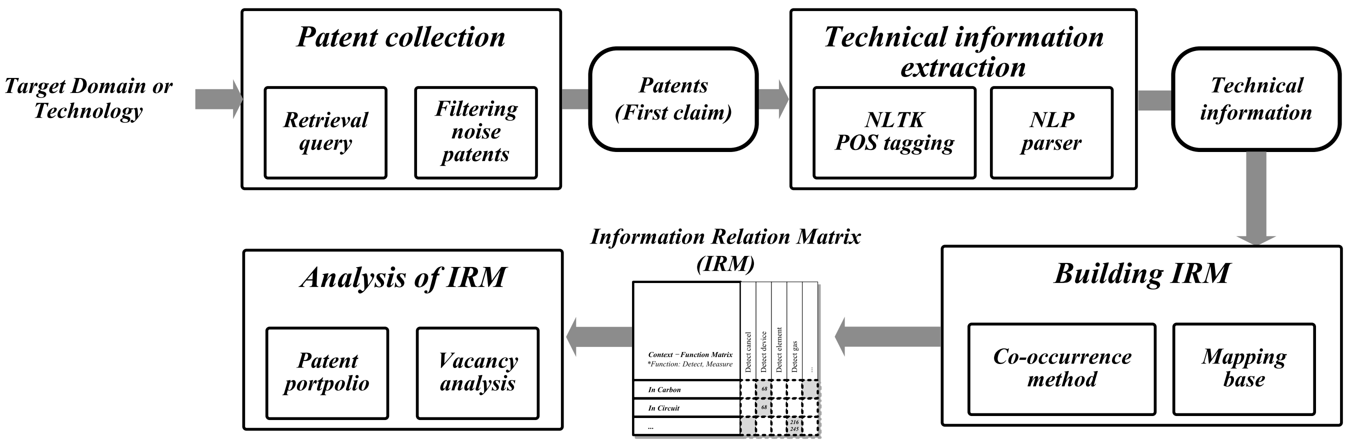

FIGURE 3. A procedure to generate an IRM.

using extracted technical information (Fig. 2). In this Section, the first three subsections describe a procedure to generate an IRM using NLP, and last subsection demonstrate how to analyze and use IRM in the technological and business aspects.

The suggested method consists of collecting patents, extracting technical information, building an IRM, and analyzing the IRM of the whole patent set (Fig. 3).

A. PATENT DATA COLLECTION

The first step in generating an IRM is to collect patents. A target domain or technology should be selected and the patents within the domain should be collected by using IPC or keyword retrieval from free patent databases such as the United States Patent and Trademark Office (USPTO), or from commercial patent databases such as LexixNexis PatentStrategiesTM, Thomson Innovation or WIPS. To maximize the credibility of the analysis, various databases should be considered. In this paper, LexixNexis PatentStrategiesTM, is mainly used, and USPTO database is used to compensate for the defects in the main database.

A Retrieval query is composed of bibliographic information such as IPC, applicants, application date and assignee, and textual information related to a target domain or

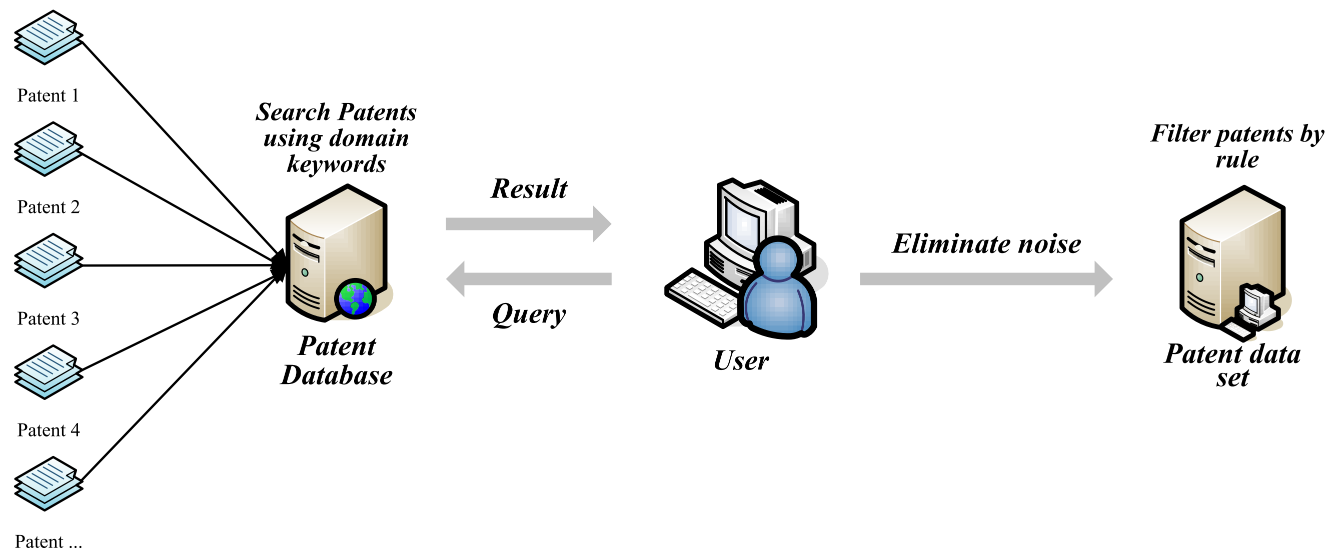

FIGURE 4. A procedure to collect patent data.

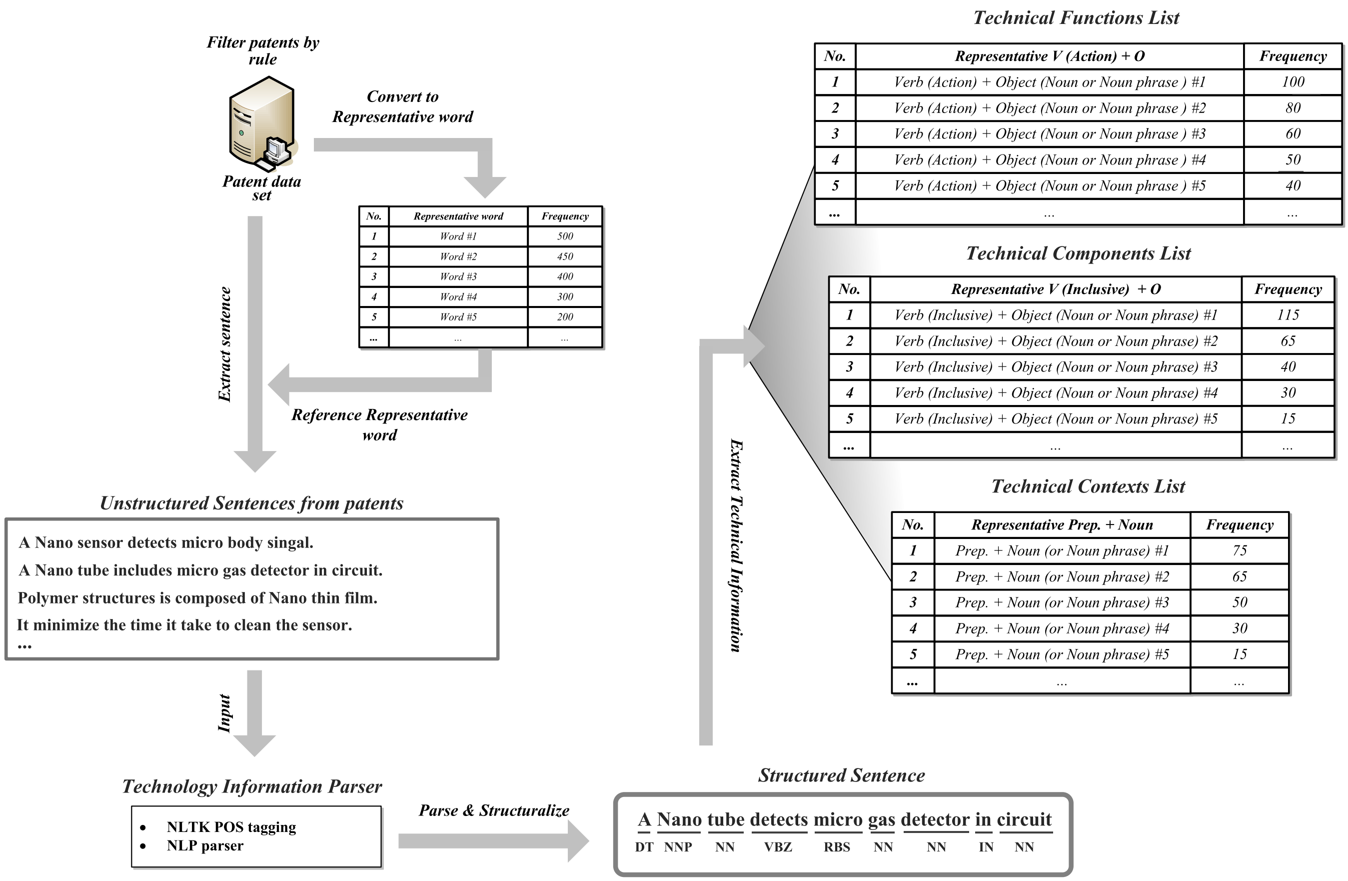

FIGURE 5. A procedure to extract technical information.

technology. Bibliographic information is a distinct part, but it cannot have detailed technological information, so textual information should be used for qualitative patent analysis.

In this work, irrelevant or redundant patents were eliminated by applying a filtering rule. When patents were extracted in certain conditions, some redundant

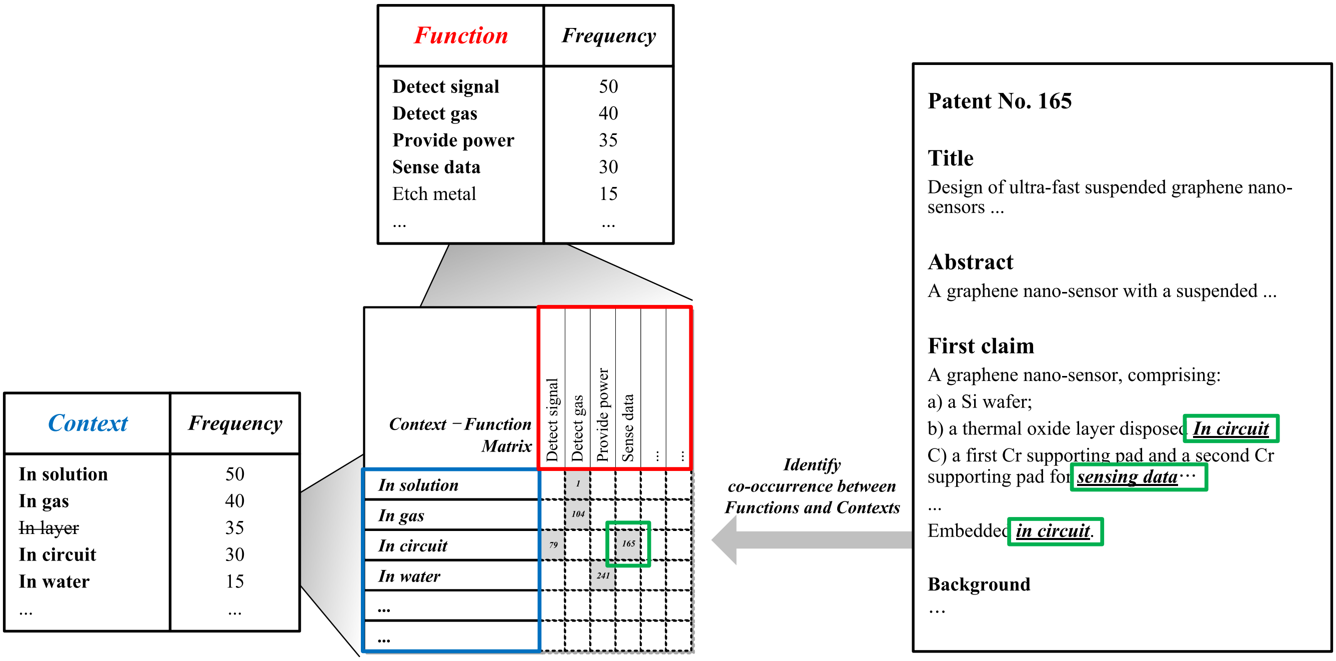

FIGURE 6. A procedure to build IRM.

patents or applicants can be extracted as a result of errors in the database. These problems can be fixed by sorting data using a data tool such as MS Excel. Among various sections in a patent, some are narratives under major five headings: title, abstract, background summary, detailed description, and claims. Some researchers argue that the human-generated abstract is the most significant and concise textual information for the relevant invention [42]. Others use the 'claims' section to identify technological points that the inventors want to protect legally; these authors puts emphasis on detailed description to analyze technological structures and findings [12]. To analyze technical information in this paper, we use the 'claims' section, and in it, we use only the first claim, because it generally expresses the most important and detailed information, and contains the overall technical description.

B. TECHNICAL INFORMATION EXTRACTION

The second step is to extract essential technical information from a patent data set. We use SAO structures to extract technical information including function, context, and component. First we used the NLTK package in Python to extract all text in patents, then used a part-of-speech (POS) tagger to split it. Then we use NLP grammatical techniques to extract technical information from each sentence.

The process of classifying words into their parts of speech and labeling them accordingly is known as POS tagging, or simply tagging [43]. Words are split into roots and affixes (tokenized) then each word (token) is tagged by NLTK POS tagger. For example, the tokenizing function in NLTK splits "A Nanosensor detects micro-body signal", can be split to ['A', 'Nano', 'sensor', 'detects', 'micro', 'body', 'signal']. The tokenized sentence is POS tagged to [('A', 'DT'), ('Nano', 'NN'), ('sensor', 'NN'), ('detects', 'VBZ'), ('micro', 'NN'), ('body', 'NN'), ('signal', 'NN')] by the POS-tagging function. Here, 'DT' means determiner, 'NN' means noun, 'VBZ' means Verb, 3rd person singular present. Each tagged token is composed of two elements as ['token', 'POS'].

For technical patent analysis based on NLP, the most important task is to extract essential technology information based on the user's needs. During R&D or technology planning, numerous patents must be scanned and analyzed. However, because technology is changing rapidly, development of an efficient strategy or tactic to perform these tasks is a difficult task. A better solution may be simply to find fundamental information that describes only the relevant contents of the technique in the patent, and to do this in a way that reduces the necessary human effort.

AO structure is commonly used to represent function of technology. 'Verb + noun (or noun phrase)' pairs can be used to defined functions of product or mechanical system [44]. Function is essential formation in a patent, but the context in which specific function operates is also important, because the context can affect the use and specification of a technology. In this research, 'preparation + noun (or noun phrase)' pairs are used to extract context. Most structures of context are expressed in this form, such as, 'in circuit' or 'on the polymer'. SAO structure also states partitive relationships among products or technologies [45]. If the action word is a partitive verb such as 'have', 'compose', 'include', and 'be made of', the component of a subject may include the object as a component [36]; if so, 'partitive verb + noun (or noun phrase)' pairs may be used to identify components. In the preceding example, [('detects', 'VBZ'), ('micro', 'NN'), ('body', 'NN'), ('signal', 'NN')] consist of an action verb and a noun, so it can be function. For new example, in 'A Nanotube includes micro gas detector in circuit', [('Nanotube', 'NN'), ('includes', 'VBZ'), ('micro gas detector', 'NN')] is partitive relationship and Nanotube and micro gas detector may be components in certain technology description. Finally, [('in', 'IN'), ('circuit', 'NN')] can represent context, because these words are composed of preposition and noun, and mean the location in which the function operates.

C. BUILDING IRM (INFORMATION RELATION MATRIX)

The third step is to build IRM by organizing relations extracted from patent semantic analysis results. The matrix is a tool that can effectively solve a given problem by comparing two types of information. Patent analysis using the matrix is very simple, but it can reduce the amount of time and effort used to read full patent documents, and can be used effectively in mid-to-long term technical planning and technology development of an organization. The other advantage of patent analysis using the matrix is that a user can see information based on her specific needs at a glance. Using the results of the previous process, all technical information can be tagged in each patent. Extracting and comparing all technical information in a patent is a difficult task; therefore, we used Python to code an algorithm that compares and analyzes all technical information in each patent. A semi-automated algorithm can be easily made by using open source packages and libraries. After the technical information is extracted, representative technical information that is to be represented in the IRM should be identified. From this information, about 20 specific points are selected, based on frequency and technical suitability (Table 3).

Representative functions, components and contexts can be used as axes in the matrix, and three types of matrices can be created using binary relation in three factors. In one matrix, when the patent has two specific factors from technical information list, the patent can be mapped in the matrix. By identifying co-occurrence in technical information simply, the resources required to compare documents one by one can be greatly reduced and the existence of the patent that is sought can be quickly checked.

D. ANALYSIS OF IRM

The fourth step is to analyze the IRM. Two methods can be used for this process. The first is vacancy analysis; i.e., identification of areas of technology in which the patent is not applied. The biggest advantage of IRM is that it facilitates this process. Given this information, a user can identify technology areas that include a product group but not the desired technology function, and technical areas in which the basic technology function is developed but not commercialized as a product. This analysis can also generate ideas for technology and product innovation, and support disruptive innovation processes by patent analysis that compares different technologies.

TABLE 3. Criteria of selecting representative technical information.

Technical information	Example	Frequency	Technical suitability
Function	**Detect signal**	50	O
	Detect gas	40	O
	Provide power	35	O
	Etch metal	30	X
	…	…	…
Context	**In solution**	50	O
	In gas	20	O
	On layer	15	X
	In circuit	10	O
	…	…	…
Component	Array	45	X
	Bridge	30	O
	Carbon	25	O
	Diameter	15	X
	…	…	…

The second method is technology competitiveness analysis. Within the technology area of interest, the IRM allows the user to determine the level of the company's technology and competitors. After identifying its patent set, the contents and level of the competitor's applied patent can be quantified in several ways, such as by counting citations, family size, and identifying claims of priority. Furthermore, patent infringement and other legal affairs can be identified by referring to competitors filing date, expiration date, and direction of citations.

IV. CASE STUDY: NANOSENSOR

This section illustrates the proposed method by using a case study of 'Nanosensor'. Nanosensors are devices that use nanoparticles to obtain surgical, biological, or chemical information. Uses of nanosensors mainly include medical sensing, and providing gateways to construction of other Nanoproducts, such as computer chips that work at the nanoscale, and Nanorobots [46]. Research into technologies that use nanosensors is expanding rapidly.

As a first step, all national patents related to Nanosensors were collected from the LexixNexis PatentStrategiesTM database (https://app.lexisnexispatentstrategies.com). The collected patents were published from January 1, 2000 to December 31, 2016. Relevant patents were found by using a patent query to find keywords related to 'Nano' and 'sensor' in Title, Abstract and Claim sections, which are the sections in which the terms 'Nano' and 'sensor' are detected most frequently. Patents obtained using this approach would be more relevant than patents obtained by searches that use complex expressions such as hypernyms, hyponyms, and synonyms. After eliminating irrelevant patents, 583 patents were used to build and analyze an IRM of Nanosensor research (Figure 7).

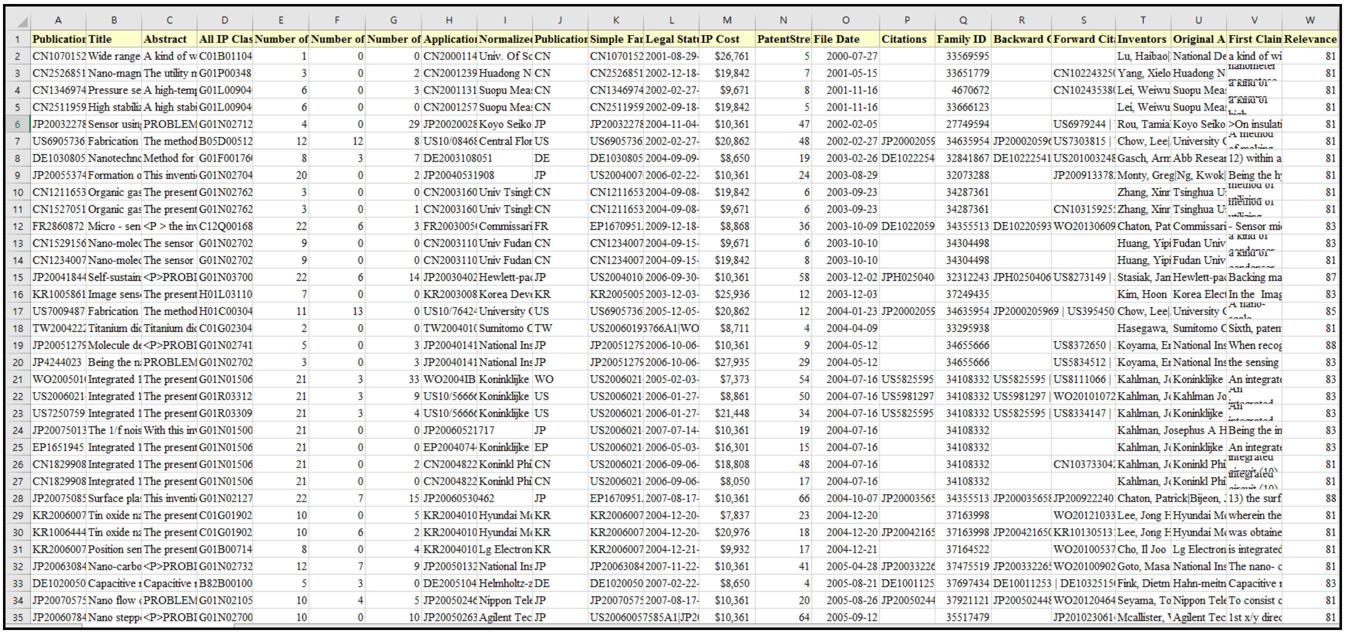

FIGURE 7. Patents set related to Nanosensor.

FIGURE 8. Derived patent text information (function).

After the filtered patent data set was derived, a pre-defined thesaurus was used to combine synonyms in the dataset into representative words. For example, ‘nanoparticles’ was converted to ‘nanoparticle’, ‘CNT’ was converted to ‘Carbon Nanotube’, and ‘Bio Nanosensor’ and ‘Bio NS’ were converted to ‘Nanobio sensor’. To extract technical information from the derived patent set, the algorithm to extract technical information (Section 3.2) was used. In this research, we used Python NLTK package to develop a technical information parser. First, the text information in patents is read line by line, then using the tokenizing function provided by NLTK, the information is split into morpheme units and then POS-tagged. The textual information derived using the proposed grammatical rule was classified according to its purpose, and imported into Excel. Then technical information such as ‘function’, ‘component’, and ‘context’ (Table 4) related to the Nanosensor were extracted using on the algorithm suggested in the previous section.

After tagging the technical information in each patent, the frequency (Freq) and technical suitability (TS) of each piece of technical information was judged, then used to guide selection of x-axis and y-axis index to be included in the IRM.

TABLE 4. Technical information extracted from patents.

Patent No.	Function	Context	Component
1	Obtaining data	At sensor	Circuit
	Communicating sensor	On substrate	Tread thickness
	Monitoring article	In array	Insulating layer
	Processing data	At circuit	Sensor
	…	…	…
2	Provided electrode	On insulating	Substrate
	Patterning metal	On electrode	Nano-pore
	Patterning graphene	In layer	Graphene
	Producing Nanosensor	In insulating	Catalyst layer
	…	…	…
3	Control nanoparticle	At bottom	Nanoparticle
	Detecting molecule	In receptor	Nanosensor
	Absorbed surface	On surface	Disease diagnosis
	Removing nanoparticle	In nanostructure	Device
	…	…	…
…	…	…	…

We excluded terms that are too general, or not technically relevant to the nanosensor. The suitability of the technology was judged based on the information described in a white paper on nanosensor technology and on domestic and foreign technical reports. The information was sorted in order of decreasing frequency, then factors that did not fit the technical suitability were removed one by one.

In the case of 'function' information, only 'detect' and 'measure' were selected (Table 5). The 'detect' function was extracted using two keywords, 'sense' and 'detect', then 'detect' was used as the representative word. The 'measure' function was extracted using three keywords, 'measure', 'gauge', and 'check', then 'measure' was used as the representative word. Because all verb forms (present, past tense, gerund, present participle, and past participle) had been converted to a basic form during data preprocessing, all results of this extraction process are expressed in the basic form. Because only two of the functions described in the 'first claim' were extracted (Table 5), the frequency is very small. However, the problem of small sample size will be solved naturally by broadening the scope of the function. The ability to detect and measure materials, signals, and light in confined spaces is a major requirement in recent industry, and as equipment and devices become more sophisticated and denser, the need for increasingly fine detection and measurement within confined spaces is becoming critical. The frequencies of functions 'detect element', 'detect signal', and 'detect light' were highest (Table 5). This result suggests that these are the most widely-used functions of applied Nanosensor patents. In contrast, the functions 'detect nanometer' and 'measure gas' occur at relatively low frequency, but they were selected because of their technical suitability and importance.

From a hardware standpoint, nanosensors belong to the category 'devices or components'. The combination of the sensor and its peripheral devices may vary depending on the purpose and environment of use, so finding a component that meets the needs of the user is very important in R&D and product planning. Components such as 'layer', 'sensor', 'metal' and 'surface' were most frequent (Table 6); they are very general and basic components of the sensor; materials such as 'semiconductor', 'silicon', and 'carbon' are also main constituents of nanosensors. In contrast, 'circuit' and 'baseplate' occur at low frequency, but they are considered major components because they are essential elements of electronic devices.

Context provides information about where the technical function in a patent is implemented. Even with the same functions implemented using the same components, the contents and usage of the technology may vary depending on the operating environment. In this research, for convenience, all prepositions ('in', 'on', 'at', and 'beneath') of all context information extracted from the patent set were represented by 'in'. Nanosensors are very compact systems; therefore, they can be used in a wide range of industries, and depending on the usage, various application technologies are currently under development. The top five contexts are too general (Table 7) and therefore lack technical suitability. The major contexts such as 'metal', 'solution', and 'electrode' that have close effects on sensors were selected as the most important top technical information. Moreover, many contexts such as 'silicon', 'carbon', 'polymer', and 'fiber' as material type were also extracted. Because nanosensors constitute a convergence technology and their uses vary, the frequency of occurrence cannot be the only criterion of importance. Therefore, contexts such as 'membrane', 'nanometre', and 'resin', which are covered only in some patents, have been selected.

Context - Function IRM (Figure 9) is a tool that shows intuitively which context and function in a patent are interrelated. This IRM reveals three groups (A, B, C) that have no patents. Group A lacks patents that cover major sensing functions in 'oxide', 'polymer' and 'resin'; this observation may suggest an opportunity to develop techniques for sensing functions in specific materials. The semiconductor process or the inorganic material reaction process may be carried out in an environment where the oxygen density is higher than a certain density. Therefore, even in such an environment, it is necessary to be able to communicate information of product and status. In this context, fine nanosensor technology and devices that work well within a certain oxygen concentration will be needed, and semiconductor business or sensor manufacturing companies may plan new business based on these technical requirements. Group B lacks nanosensor technology to measure 'gas', 'data', and 'level' in various contexts. Although technical function has been developed in 'circuit' or 'semiconductor', sensor technology with these measurement functions should be developed for use in various environments. Group C lacks major sensing and measuring functions in 'vacuum' and 'wire' contexts. Nanosensors also include data-measurement capabilities for

TABLE 5. Function list for building IRM.

No.	Function name	Freq	TS	No.	Function name	Freq	TS
1	**detect element**	21	O	14	**measure circuit**	17	O
2	**detect signal**	16	O	15	**measure resistance**	13	O
3	**detect light**	15	O	16	measure head	9	X
4	detect method	14	X	17	**measure data**	8	O
5	detect target	14	X	18	measure diagonals	8	X
6	**detect gas**	12	O	19	**measure level**	8	O
7	detect field	12	X	20	**measure pressure**	8	O
8	detect surface	11	X	21	measure change	7	X
9	**detect cancel**	10	O	22	measure sensor	7	X
10	**detect nanometre**	10	O	23	measure subject	6	X
11	**detect device**	9	O	24	**measure gas**	6	O
12	detect presence	8	X	25	measure characteristic	6	X
13	detect change	7	X	26	measure method	5	X

TABLE 6. Component list for IRM.

No.	Component name	Freq	TS	No.	Component name	Freq	TS
1	**layer**	290	O	17	nano	43	X
2	**sensor**	182	O	18	**channel**	40	O
3	**metal**	123	O	19	hydrogen	38	X
4	**surface**	112	O	20	**catalyst**	36	O
5	**structure**	111	O	21	Pd	36	X
6	**nanostructure**	107	O	22	**polymer**	27	O
7	**semiconductor**	92	O	23	spectrum	25	X
8	**nanofibers**	82	O	24	**system**	25	O
9	**gas**	73	O	25	**bridge**	24	O
10	**film**	64	O	26	solution	24	X
11	**wire**	60	O	27	**water**	24	O
12	**silicon**	58	O	28	**circuit**	23	O
13	**carbon**	56	O	29	resin	23	X
14	oxide	52	X	30	**baseplate**	22	O
15	**nanoparticles**	51	O	31	diameter	22	X
16	array	49	X	32	insulating	20	X

data transfer and monitoring within and between systems. The sensors can also measure the amount of data and signal transmitted, and the level of the material monitored.

Patent No.68 (CN101460321, Bridgestone Firestone North Am) covers techniques for measuring the amount of data or levels within contexts such as ‘carbon’, ‘solution’, and ‘water’ (Fig. 9). The first part of the abstract of this patent specifies ‘A sensor system for obtaining an elastomeric article includes at least one wireless sensor’. This means that this patent and is about Nanosensor technology related to a wireless sensor that can transmit and receive data and can be operated in various contexts. The company, Bridgestone Firestone North Am, has registered the patent in four countries, including the United States, and has commercialized technology for Nanodata sensors.

Developed nanosensors have many functions to detect ‘element’, ‘gas’ and ‘light’ in general (Fig. 10), but this IRM reveals three groups (D – G) that lack patents. Group D shows that the functions to detect ‘cancel’ and ‘device’ were not detected in the patent set. Group E shows that the development of detection and measurement functions using ‘catalyst’, ‘channel’ and ‘circuit’ level is low. ‘Nanoparticles’ and ‘Nanofibers’ are used as the main components of Nanosensors, but Group F shows that few measurement technologies use them as major components. Group G shows that the sensing technology using ‘silicon’, ‘structure’, ‘water’, or ‘wire’

TABLE 7. Context list for building IRM.

No.	Context name	Freq	TS	No.	Context name	Freq	TS
1	in surface	277	X	17	**in film**	41	O
2	in substrate	240	X	18	in steps	38	X
3	in sensor	149	X	19	in item	37	X
4	in method	98	X	20	**in water**	36	O
5	in temperature	95	X	21	in particle	33	X
6	**in metal**	95	O	22	**in polymer**	30	O
7	**in solution**	84	O	23	**in semiconductor**	29	O
8	**in electrode**	78	O	24	**in fiber**	24	O
9	in layer	77	X	25	**in circuit**	22	O
10	**in material**	71	O	26	**in oxide**	17	O
11	in range	66	X	27	**in vacuum**	14	O
12	in top	64	X	28	**in wire**	12	O
13	in side	63	X	29	**in membrane**	11	O
14	**in silicon**	62	O	30	**in nanometre**	11	O
15	in contact	60	X	31	**in resin**	5	O
16	**in carbon**	42	O				

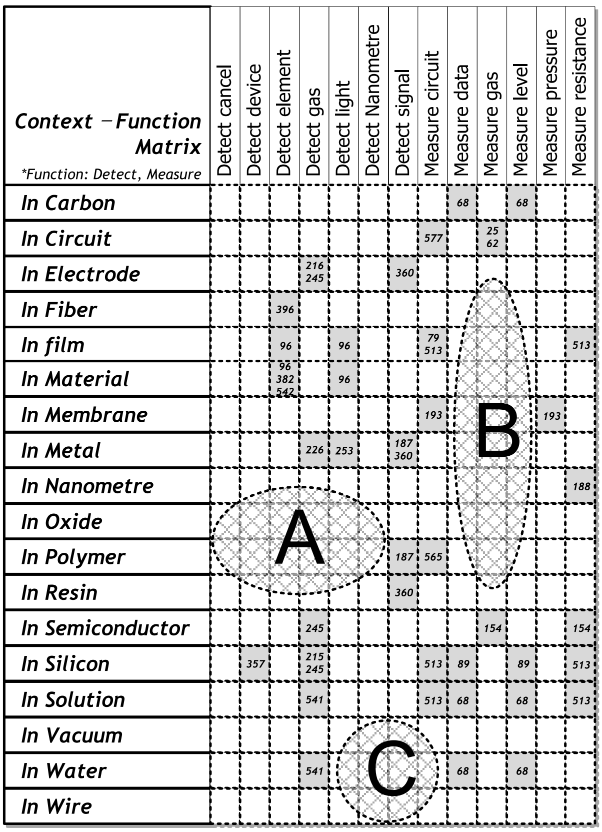

FIGURE 9. Context-Function IRM.

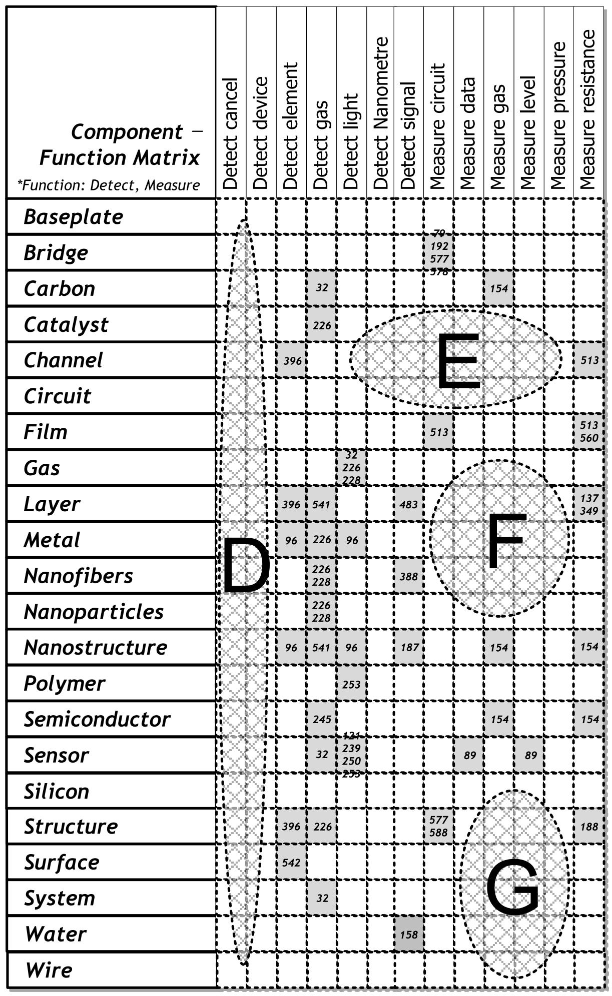

FIGURE 10. Component-Function IRM.

is rare. A number of sensors are needed for tidal power generation and seabed resource exploration. The main components needed for this are structures and wires, which may require components that come in contact with water or silicon components that can withstand water. The fact (opportunity) that there is no technology available for sensing components or underwater sensing systems using these materials yet will enable someone to develop a variety of business strategies to address them.

Context – Component Matrix	Baseplate	Bridge	Carbon	Catalyst	Channel	Circuit	Film	Gas	Layer	Metal	Nanofibers	Nanoparticles	Nanostructure	polymer	Semiconductor	Sensor	Silicon	Structure	Surface	System	Water	Wire
In Carbon			127 154				38		244				154			113	113 244	331			499	
In Circuit					11	1,6 7,9 11 ...							11			6,7 9,11 24			61 104			199
In Electrode	107 197			226	109 110 344 484		156	226 228	2,43 109 110 ...	107 109 110 ...	226 228	77 226 228 378		423		61 344 481	244 423	226	447	447	86 499	43
In Fiber											433								213 395 415			
In film					109 110		156 512 575	552	109 110	109 110	388		512					95 171 331	447 542 552	447	329	
In Material						1,6 7 9	156 391		28 42 ...	96		301 376	95 541	233		6,7 9 ...		579 95 171	28 321 ...	60 321 324		43 60
In Membrane		79 192																79 192	79 192			
In Metal	265		154	226 235 252	109 110	199	401	226 228	43 109 ...	10 96 ...	226 228 378	138 148 ...	96 154 378	10	240	115 239 250	496	226 286 ...	131 235 ...	32 324 447		43 199 252
In Nanometre							156		27										27		158	
In Oxide				235			38	228		235	228	228 235 376			168 323		5		235 323			
In Polymer			212 336			345			16	10				10 16 253 426		253		171 342 564			158	16
In Resin					344 484											344						
In Semiconductor				45	81 85				137 244 349	153			153		168 323	113	45 81 ...		323 338			
In Silicon							140 512		244 434 444			376	512			113	113 244 282		338			
In Solution	267		127	226	81 85		156 267 512	226 228	582 43 131 ...	10 131 ...	226 228	226 228 376	371 512	10	168		81 85 282	226 286 ...	131 267 ...	321 324	86 158 329 499	43
In Vacuum																						
In Water								552	114	286 291 416						114		286 291	53 552		53 478 492 499	
In Wire																			281			281

FIGURE 11. Context-Component IRM.

In this IRM the patents that include many technical functions for detecting 'gas', are No. 226 (CN101893494, Wuhan Univ.) and No. 228 (JP2012052864, Chiba Univ.). Both have nanosensor technology that includes all major components such as 'Nanofibers' and 'Nanostructure'. Patent No. 226 can also be included in 'chemical sensor' technology that uses 'catalyst'. Given that all of these are patents filed by university institutions, the development can be assumed to be at the laboratory development level rather than at the commercialization level.

The Context-Component IRM (Fig. 11) shows fewer application voids than the other two matrices. Because this IRM represents a relation between general context and component rather than limited 'detect' and 'measure' functions, we judge that it detects many cases that include each type of technical information. This IRM shows three underutilized areas (H, I, J). Group H group suggests that none of the sensors that operate in contexts such as 'fiber' and 'film' are composed of 'carbon' or 'catalyst'. Moreover, Groups I and J show that sensor technology in 'vacuum', 'water' and 'wire' is still in development and that available sensors are composed of a limited set of components. In a limited space, such as in vacuum or underwater, electronic devices often fail to function properly. Through groups I and J, it can be seen that there are very few components operated in this limited environment. Therefore, a company that conducts business in these area will be able to outsource a company that has the ability to produce products that can cope with it, or solve the problems through collaboration with a university or research institute that is conducting such research.

The most-frequently tagged patents are No. 226 (CN101893494, Wuhan Univ., 18 times) and No. 235 (CN101973510, Tianjin Univ., 16 times) (Fig. 11). The number of tags means the number of co-occurrences of context and component, so these two patents used technologies composed of various components in various contexts. Especially, No. 226 is a nanosensor-related technology that is composed of a wide variety of components within the 'electrode', 'metal', and 'solution' contexts. This result indicates that this technology can use various material components to construct the sensor and its peripheral system, and that the operating environment varies. In addition, these two universities have a total of 12 patents in Nanosensors, which is a technology that can detect various inputs such as oxide, light, glucose, and gas. Moreover, the number of forward and backward citations of the two patents is above average; i.e., the technical maturity is somewhat high.

V. CONCLUSION

We present a semi-automatic method to extract text information of patent data set, then to use NLP technology based on grammar rules to extract the major technical information, then to visualize it in a matrix form.

The proposed method extracts only the technical information that meets the needs of users in a large amount of patent data, and shows the interrelationships among types of technical information simply and intuitively. To verify the effectiveness of this method, it was applied to 583 patents related to Nanosensors. The method successfully extracted semantic technical information qualitatively. This method will assist the technology or management decision-makers of governments and industry to establish medium and long-term technical plans and R&D directions.

This paper proposed a semi-automated method based on Python to process a large amount of patent data, but further research should develop a patent analysis solution based on -,machine learning to automate patent data mining, technical information extraction, and analysis using IRM, and to retrain the analyzed results to increase accuracy. Research to develop a web-based patent information search engine would also be valuable.

REFERENCES

[1] K. Brockhoff, "Competitor technology intelligence in German companies," *Ind. Marketing Manage.*, vol. 20, no. 2, pp. 91–98, 1991.

[2] R. Cantrell, "Patent intelligence—Information to compete before products are launched," *Competitive Intell. Rev.*, vol. 7, no. 1, pp. 65–69, 1996.

[3] B. Yoon, "On the development of a technology intelligence tool for identifying technology opportunity," *Expert Syst. Appl.*, vol. 35, no. 1, pp. 124–135, 2008.

[4] J. Yoon and K. Kim, "TrendPerceptor: A property–function based technology intelligence system for identifying technology trends from patents," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2927–2938, 2012.

[5] C. Camus and R. Brancaleon, "Intellectual assets management: From patents to knowledge," *World Patent Inf.*, vol. 25, no. 2, pp. 155–159, 2003.

[6] S. Morris, C. DeYong, Z. Wu, S. Salman, and D. Yemenu, "DIVA: A visualization system for exploring document databases for technology forecasting," *Comput. Ind. Eng.*, vol. 43, no. 4, pp. 841–862, 2002.

[7] S. Lee, B. Yoon, and Y. Park, "An approach to discovering new technology opportunities: Keyword-based patent map approach," *Technovation*, vol. 29, no. 6, pp. 481–497, 2009.

[8] F. Narin, E. Noma, and R. Perry, "Patents as indicators of corporate technological strength," *Res. Policy*, vol. 16, nos. 2–4, pp. 143–155, 1987.

[9] M. E. Mogee, "Using patent data for technology analysis and planning," *Res.-Technol. Manage.*, vol. 34, no. 4, pp. 43–49, 1991.

[10] M. Hirschey and V. J. Richardson, "Valuation effects of patent quality: A comparison for Japanese and US firms," *Pacific-Basin Finance J.*, vol. 9, no. 1, pp. 65–82, 2001.

[11] D. Archibugi and M. Planta, "Measuring technological change through patents and innovation surveys," *Technovation*, vol. 16, no. 9, pp. 451–468 and 519, 1996.

[12] A. Fujii, M. Iwayama, and N. Kando, "Introduction to the special issue on patent processing," *Inf. Process. Manage.*, vol. 43, no. 5, pp. 1149–1153, 2007.

[13] G. Piatetsky-Shapiro, *Advances in Knowledge Discovery and Data Mining*, vol. 21. U. M. Fayyad, P. Smyth, and R. Uthurusamy, Eds. Menlo Park, CA, USA: AAAI Press, 1996.

[14] G. Piatetsky-Shapiro and W. Frawley, *Knowledge Discovery in Databases*. Cambridge, MA, USA: MIT Press, 1991.

[15] R. N. Kostoff, "Text mining using database tomography and bibliometrics: A review," *Technol. Forecast. Social Change*, vol. 68, no. 3, pp. 223–253, 2001.

[16] *WIPO-MOST Intermediate Training Course on Practical Intellectual Property Issues in Business*, document, WIPO, Patent Map With Exercises, 2003. [Online]. Available: http://www.wipo.int/edocs/mdocs/sme/en/wipo_ip_bis_ge_03/wipo_ip_bis_ge_03_16-main1.pdf

[17] B. P. Abraham and S. D. Moitra, "Innovation assessment through patent analysis," *Technovation*, vol. 21, no. 4, pp. 245–252, Apr. 2001.

[18] R. S. Campbell, "Patent trends as a technological forecasting tool," *World Patent Inf.*, vol. 5, no. 3, pp. 137–143, 1983.

[19] S. Jung, *Importance of Using Patent Information, WIPO-Most Intermediate Training Course on Practical Intellectual Property Issues in Business*, document, World Intellectual Property Organization (WIPO), Geneva, Switzerland, 2003, pp. 10–14.

[20] M. Norton, *Introductory Concepts in Information Science*. Medford Township, NJ, USA: Information Today, Inc., 2000.

[21] T. U. Daim, G. Rueda, H. Martin, and P. Gerdsri, "Forecasting emerging technologies: Use of bibliometrics and patent analysis," *Technol. Forecast. Social Change*, vol. 73, no. 8, pp. 981–1012, Oct. 2006.

[22] M. M. S. Karki, "Patent citation analysis: A policy analysis tool," *World Patent Inf.*, vol. 19, no. 4, pp. 269–272, 1997.

[23] M. Fattori, G. Pedrazzi, and R. Turra, "Text mining applied to patent mapping: A practical business case," *World Patent Inf.*, vol. 25, no. 4, pp. 335–342, 2003.

[24] B.-U. Yoon, C.-B. Yoon, and Y.-T. Park, "On the development and application of a self-organizing feature map-based patent map," *R&D Manage.*, vol. 32, no. 4, pp. 291–300, 2002.

[25] Y.-H. Tseng, C.-J. Lin, and Y.-I. Lin, "Text mining techniques for patent analysis," *Inf. Process. Manage.*, vol. 43, no. 5, pp. 1216–1247, 2007.

[26] Y.-H. Tseng, "Patent surrogate extraction and evaluation in the context of patent mapping," *J. Inf. Sci.*, vol. 33, no. 6, pp. 718–736, 2007.

[27] A. J. C. Trappey, C. V. Trappey, U. H. Govindarajan, J. J. Sun, and A. C. Chuang, "A review of technology standards and patent portfolios for enabling cyber-physical systems in advanced manufacturing," *IEEE Access*, vol. 4, pp. 7356–7382, 2016.

[28] V. Huddar, B. K. Desiraju, V. Rajan, S. Bhattacharya, S. Roy, and C. K. Reddy, "Predicting complications in critical care using heterogeneous clinical data," *IEEE Access*, vol. 4, pp. 7988–8001, 2016.

[29] A. A. Amamou, S. Kelouwani, L. Boulon, and K. Agbossou, "A comprehensive review of solutions and strategies for cold start of automotive proton exchange membrane fuel cells," *IEEE Access*, vol. 4, pp. 4989–5002, 2016.

[30] H. Park, K. Kim, S. Choi, J. Yoon, "A patent intelligence system for strategic technology planning," *Expert Syst. Appl.*, vol. 40, no. 7, pp. 2373–2390, 2013.
[31] G. Cascini, A. Fantechi, and E. Spinicci, "Natural language processing of patents and technical documentation," in *Proc. Int. Workshop Document Anal. Syst.*, 2004, pp. 89–92.
[32] M. G. Moehrle, L. Walter, A. Geritz, and S. Müller, "Patent-based inventor profiles as a basis for human resource decisions in research and development," vol. 35, no. 5, pp. 513–524, 2005.
[33] I. Bergmann, "Evaluating the risk of patent infringement by means of semantic patent analysis: The case of DNA chips," *R&D Manage.*, vol. 38, no. 5, pp. 550–562, 2008.
[34] J. M. Gerken and M. G. Moehrle, "A new instrument for technology monitoring: Novelty in patents measured by semantic patent analysis," *Scientometrics*, vol. 91, no. 3, pp. 645–670, 2012.
[35] J. Yoon and K. Kim, "Detecting signals of new technological opportunities using semantic patent analysis and outlier detection," *Scientometrics*, vol. 90, no. 2, pp. 445–461, 2012.
[36] S. Choi, "An SAO-based text mining approach to building a technology tree for technology planning," *Expert Syst. Appl.*, vol. 39, no. 13, pp. 11443–11455, 2012.
[37] H. Park, J. J. Ree, and K. Kim, "Identification of promising patents for technology transfers using TRIZ evolution trends," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 736–743, 2013.
[38] J. Yoon, H. Park, and K. Kim, "Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis," *Scientometrics*, vol. 94, no. 1, pp. 313–331, 2013.
[39] O. Granstrand, *The Economics and Management of Intellectual Property*. Cheltenham, U.K.: Edward Elgar Pub, Feb. 2000.
[40] T. Simpson, and T. Dao, "WordNet-based semantic similarity measurement," The Code Project., Tech. Rep., 2005.
[41] *A Guide of Patent Map*. Korea Invention promotion Assoc., Daejeon, South Korea, 1999.
[42] L. Chen, N. Tokuda, H. Adachi, "A patent document retrieval system addressing both semantic and syntactic properties," in *Proc. ACL-Workshop Patent corpus Process.*, Sapporo, Japan, Jul. 2003, pp. 1–6, doi: 10.3115/1119303.1119304.
[43] W. Wagner, "Steven Bird, Ewan Klein and Edward Loper: Natural language processing with python, analyzing text with the natural language toolkit," *Lang. Resour. Eval.*, vol. 44, no. 4, pp. 421–424, 2010.
[44] J. M. Hirtz, R. B. Stone, S. Szykman, D. A. McAdams, and K. L. Wood, "Evolving a functional basis for engineering design," in *Proc. ASME Design Eng. Tech. Conf. (DETC)*, Pittsburgh, PA, USA, 2001, pp. 3–5.
[45] G. Cascini, D. Lucchesi, and P. Rissone, "Automatic patents functional analysis through semantic processing," in *Proc. 12th ADM Int. Conf.*, Rimini, Italy, 2001.
[46] L. E. Foster, *Nanotechnology: Science, Innovation, and Opportunity*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.

WANWOOK KI received the B.E. degree in systems management engineering from Sungkyunkwan University, South Korea, in 2012, and the M.S. degree in industrial and management engineering from the Pohang University of Science and Technology, South Korea, in 2014, where he is currently pursuing the Ph.D. degree.

He has authored four conference papers and one journal paper, and he holds one patent. His research interests include patent data mining and analysis, technology planning, creative invention education, and systems engineering.

Dr. Ki was a recipient of the Campus Patent Strategy Universiade Participation Prize in 2013 and the Korean Society of Systems Engineering Best Paper Award in 2013.

KWANGSOO KIM received the B.E. and M.S. degrees in industrial engineering from Seoul National University, South Korea, in 1977 and 1979, respectively, and the Ph.D. degree from the University of Central Florida, Orlando, FL, USA, in 1985.

He has authored over 70 international journal papers, and he holds five patents. His research interests include patent mining, technology planning, business planning, future planning, and creative invention education.

He has been a Professor with the Department of Industrial Management Engineering, Pohang University of Science and Technology (POSTECH), for 30 years and has been planning and operating POSTECH Creative Entrepreneurship Omphalos (PCEO) providing gifted education for inventions as the director of PCEO for eight years.