9/10

2017_19pages - 2021.11.03

→ useful framework - pointing word

→ patent → MD&A

Article

Developing a Methodology of Structuring and Layering Technological Information in Patent Documents through Natural Language Processing

Taeyeoun Roh [iD], **Yujin Jeong and Byungun Yoon** * [iD]

Department of Industrial & Systems Engineering, School of Engineering, Dongguk University, 26, Pil-dong 3-ga, Chung-gu, Seoul 100-715, Korea; arongi1320@naver.com (T.R.); withss501@naver.com (Y.J.)
* Correspondence: postman3@dongguk.edu

Abstract: Since patents contain various types of objective technological information, they are used to identify the characteristics of technology fields. Text mining in patent analysis is employed in various fields such as trend analysis and technology classification, and knowledge flow among technologies. However, since keyword-based text mining has the limitation whereby, when screening useful keywords, it frequently omits meaningful keywords, analyzers therefore need to repeat the careful scrutiny of the derived keywords to clarify the meaning of keywords. In this research, we structure meaningful keyword sets related to technological information from patent documents; then we layer the keywords, depending on the level of information. This research involves two steps. First, the characteristics of technological information are analyzed by reviewing the patent law and investigating the description of patent documents. Second, the technological information is structured by considering the information types, and the keywords in each type are layered through natural language processing. Consequently, the structured and layered keyword set does not omit useful keywords and the analyzer can easily understand the meaning of each keyword.

Keywords: text mining; NLP; technological information; patent analysis; text structure

1. Introduction

A patent is objective and proven technological information, which contains one or more unique technical features that cannot be duplicated in other patents. Patents have been used as essential information for effective management strategies [1,2]. Structured data such as the issue date and the number of citations, as well as unstructured data such as summaries and claims included in the patent, can be useful in analyzing competitive markets and technology trends [3]. In many studies on structured data, the features or trends of technology development have been obtained by using patent search and international patent classification (IPC) [4–6]. Currently, considerable research is underway that includes unstructured data rather than only structured data to produce more meaningful results. In order to analyze a large number of patents, it is necessary to analyze the contents of unstructured information such as titles, summaries, and claims [7]. In the most commonly used keyword-oriented analysis, not all keywords are extracted, but keywords are analyzed according to specific criteria [4] and the linkage between technological fields [8].

Keyword-oriented text mining cannot accurately convey the meanings of individual words. Therefore, analysis on tagging parts of speech through natural language processing (NLP) is utilized to recognize information at the sentence level rather than at the keyword level. For example, subject–action–object (SAO) methodology is a representative method that analyzes the subject, verb, and object in a sentence as one structure. The SAO methodology interprets an object as a problem, and the subject and action as a way to solve problems. The SAO methodology can derive more

detailed meaning by reflecting the semantic relations among keywords compared to the traditional keyword analysis methods. Using SAO methodology, patent analysis has been conducted from a more semantic point of view, such as for patent infringement and similarity of technology [9] and technology roadmaps [10].

The text described in the patent document must contain various types of technological information, such as functions, components, and operating methods, in accordance with the patent law. However, the existing keyword-based analysis and the SAO methodology analyze the structure of extracted keywords based on the frequency of occurrence and part-of-speech, respectively. Thus, various types of technological information pertinent to the patent are often missing from the documents, and a secondary analysis is necessary. In addition, the existing keyword-based analysis and the SAO methodology have limitations since the inherent information of each patent such as the functions, components, and operating methods cannot be deduced.

Therefore, the purpose of this study is to derive a set of keywords that contain technological information of patented texts without omission in order to overcome the limitations of the text mining used in existing patent analysis. To achieve this purpose, in this research, information about phrases that form parts of speech is used, in addition to the dictionary meaning of words through Natural Language Processing (NLP). The technological information included in the patent exists in a formulated description form based on the part-of-speech, and the sentences described in the title, summary, and claims can be structured according to the type of technological information. In addition, words included in phrases, and words modifying phrases can be identified by using information on phrases. Keywords can then be selected according to the quality of information using the importance of words in a sentence rather than the part-of-speech and frequency. The keywords are structured according to the type of technological information and it is possible to understand the meaning of the keywords without secondary interpretation of the keyword set by using the hierarchical keyword set based on the importance of the word. Various levels of information can be selectively extracted from the detailed information. Since the keyword set obtained through this study can be used to analyze a large number of patents without further use of expert opinion or searching the technical field, this methodology can support the process of searching new fields for technology development. In addition, since the technological information of the keyword set is presented for each type, it is possible to classify the technology without additional clustering in terms of the function and the component.

In Section 2, we discuss the limitations of the existing methods through a literature review on existing patent analysis and text mining. We also present the specific research objectives and explain NLP, a key technology for achieving the objectives of this research. In Section 3, we present the basic concept and the detailed process of the proposed approach. Section 4 shows the application of the proposed methodology to a real case, the user interface field, to derive and verify the results. In Section 5, the limitations and areas for future research are discussed.

2. Background

2.1. Patent Analysis

Patent analysis is performed to understand the nature of the technology and industry such as detailed properties of technology and industrial trends [11]. In addition, patent analysis can extract various information in the patent, through classification, visualization, and clustering analysis [7]. The information contained in the patent can be divided into structured information and unstructured information. Structured information includes the information quantified in the patent database such as patent number, registration date, number of citation, and number of claims. In addition, the technical classification codes defined differently for each country are included to indicate the technical field of a patent. Unstructured information includes information described in the text such as the title, summary, and claims.

Lee et al. [12] and Altuntas et al. [8] analyzed the technology convergence, innovation, and relationship among technologies by investigating the citation network between technology fields. Jeong et al. [4] and Su [13] presented the trends of technology development among structured information by analyzing the number of patents registered each year. Based on IPC codes, Kim [14] discovered core technology in environmental ecology based on data envelopment analysis (DEA) and association rule mining. In addition, Kang et al. [15] proposed a convergence index to explore promising convergence technologies using structured information. Yun [5], Lim et al. [6] and Niemann et al. [16] evaluated the importance of the technology and patenting patterns by using the number of citations.

Unstructured information mainly focuses on the analysis of text information by converting text information into quantitative information using text mining. Noh et al. [17] selected keywords that have the highest text mining efficiency among the titles, abstracts, and claims. Huang et al. [18] and Lee et al. [19] selected the important keywords shown in the summary and claims, and searched for patents and technologies with high similarity using IPC codes. Lee et al. [20] evaluated the novelty of patents by using the similarity between major keywords. Ko et al. [21] analyzed the degree of technology convergence in the technology field. Lee and Sohn [22] identified shale gas development by analyzing the abstracts of patents.

As mentioned above, existing studies have utilized structured information and unstructured information to interpret various types of technological information. However, most of the studies analyze patents by integrating keywords that are shown in terms of technology, and the keywords are not interpreted in terms of the respective patents. In addition, although the patent has specific technological information related to features, functions, methods, components etc., it might only reflect certain areas of configuration at the keyword level. Thus, in such approaches, a secondary analysis should be performed in order to clearly present the meaning of keywords.

In order to overcome the limitations of existing research, this study defines the type of technological information in advance and extracts the unique information of each patent by structuring the information according to the type of technological information. We then propose a method that can interpret both the patent information of one patent as well as the technological field.

2.2. Text Mining

Text mining extracts meaningful information from unstructured data, and is utilized in many research fields because it can express a large amount of text. Text mining can be divided into keyword-based analysis and word-based analysis [23]. Keyword-based analysis methods are based on the frequency of occurrence, such as the method of using the Term Frequency-Inverse Document Frequency (*TF-IDF*) value which is an index to judge word importance in the document, Latent Semantic Analysis (LSA), and Latent Dirichlet Allocation (LDA) [24]. SAO is well known as a representative method in the word-based analysis method.

The method of using the occurrence frequency involves determining the number of times that the keyword appears in the analysis target [25]. Since the *TF-IDF* value is calculated by using the number of keywords appearing in the document, it is the relative value of the degree of importance of words. In the method of using the occurrence frequency and the *TF-IDF* value, parts of the keywords are selected and analyzed by evaluating the importance of the keyword without using all the keywords appearing in the document. Kim et al. [26] and Min et al. [27] selected future promising areas by analyzing the time series of words appearing in papers, news, and policy research reports. Choi et al. [2] predicted promising technologies by investigating the network of keywords. Kim et al. [28] used this approach to create a patent development map in technology fields using the patent keywords.

LSA is a method used to understand the linkage relationship between unknown documents based on the occurrence frequency of keywords. A matrix of the occurrence frequency of keywords in each document is constructed and the similarities between documents are compared, using singular value decomposition. The advantage of this method is that multiple keywords other than a single keyword

can be compared together and their latent relationship is discovered. Ghazizadeh et al. [29] proposed a future service direction by clustering users' complaints through LSA. LDA is an analytical method that can derive topics from texts and extract keywords corresponding to each topic [30]. Since the relation between topics and keywords can be derived using the probability of the keywords for each subject based on the Dirichlet probability distribution, it is possible to derive a set of keywords that reflects the meaning contained in the word in comparison with other text mining methodologies. Based on the topic modeling methodology, Kwon et al. [31] analyzed social impacts of emerging technology using LSA. Park and Song [32] used the abstract of a paper to understand the research trends based on LDA. Jang et al. [33] discovered technology opportunity in heterogeneous technology fields using LDA. Jin et al. [34] used Twitter data to select topics and search for issue changes through the network analysis of thematic keywords. Gao and Eldin [35] interpreted topics as one cluster and derived independent meaning for each topic using only the relationship between words in the topic. Furthermore, Guo et al. [36] compared various topic modeling methods and detected user interests in microblog platform.

SAO analysis interprets subject-action-object as a structure, in contrast to analyzing the meaning of one word in existing text mining. In this case, the verb and object are interpreted as a way to solve the subject, and are analyzed by focusing on the function of each word. Park et al. [9] analyzed patent infringement and technology similarity, and Yoon and Kim [23] selected the promising technology and applied it to technical planning. Wang et al. [24] utilized the SAO structure to construct the seven layers of the technology roadmap.

In order to increase the accuracy of the analysis, various studies are conducted to extract more meaningful and accurate keywords. Lee and Kim [37] and Rose et al. [38] suggested keyword extracting methodology based on *TF-IDF* index and the number of frequency. Hulth [39] used natural language processing to extract keyword based on the linguistic meaning of the words.

However, previous text mining based on keyword extraction has four limitations. First, many meaningful keywords are excluded without considering the meaning and function of each word. This process does not reflect the characteristics of patents that have different independent information, as shown in Figure 1. Second, in order to analyze the results of a large amount of patent analysis, further analysis is required to more specifically confirm the keyword and the patent. Therefore, this approach is limited since a large amount of patent analysis hinders the ability to make efficient and quick decisions. Third, keyword extracting methodology based on topic modeling has another problem. Since they are aimed at constructing a group of words and analyzing the words in each group, they are not aimed at the meaning of each word itself. Fourth, since SAO analysis extracting constructed information by noun and verb, other parts of speech cannot be extracted.
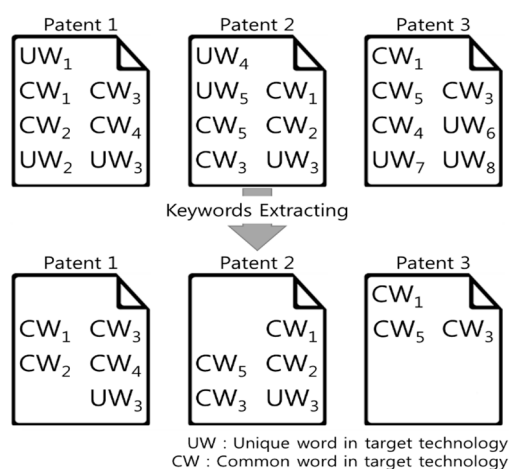
Figure 1. Errors in extracting keyword.

In order to overcome this problem, in this study, we classify the technological information of the patent according to the description form, and extract the information after structuring the information of the patent terms of technological information. This makes it possible to extract the technological information of the patent without omission of useful keyword and specific part of speech. In addition, meaning of extracted keyword in one patent is clarified, showing which technological information it contained. Furthermore, groups of the patent keywords can be representing technology fields as previous approaches do.

2.3. Natural Language Processing (NLP)

Natural Language Processing (NLP) refers to the process of converting a natural language described by humans into a machine language understood by a computer. Besides the dictionary meaning of the words in the sentence, NLP provides various types of information through the existing learning or stored algorithm. Among these types of information, part of speech (POS), which tags the part of speech of each word, can be used to check the grammatical meaning of each word in the sentence, and can decompose a sentence into several phrases and clauses. In addition, one complicated sentence can be decomposed into plural complete sentences according to a predefined sentence decomposition standard.

NLP has mainly been used to classify sentence patterns based on parts of speech or to extract sentence features. Nasukawa and Yi [40] and Yi et al. [41] proposed an algorithm to judge positive and negative statements, and Jin and Xiong [42] suggested an algorithm to classify the types of sentences and translate Chinese into English. Yang and Seo [43] classified the type of claims on the basis of the part of speech and extracted the information on the keyword type.

In this study, not only the meanings of keywords but also the information on phrases are utilized through NLP. In addition, this research classifies the sentences according to the type of technological information they provide by using the form of phrases and the parts of speech that each sentence contains. Then, the level of the information of the keyword is classified using the degree of influence of each keyword in the sentence, and the keyword is extracted by layering the keywords in a structure.

3. Methodology

3.1. Basic Concept

The main purpose of this research is extracting useful keywords from patents and interpreting them in terms of technological information. The main limitation of the previous studies is that they do not aim at the meaning of the keyword itself. Thus, they omit useful keywords, and extracted keywords cannot be interpreted easily. In order to solve this limitation, we construct a set of keywords by structuring the sentences included in a patent by using descriptive information and layering the keywords in order to clearly interpret the meaning of extracted keywords without missing information unique to each patent. In this study, as shown in Figure 2, the technological information of patents is structured and layered by two parallel processes of technological information analysis, patent text structuring/layering. First, in the technological information analysis stage, the analysis of existing studies and various patents is incorporated to classify the technology types, and each sentence is structured according to each type of description and the parts of speech. In this case, when specific words such as pronouns and parts of speech appear, we define pointing words as an indicator that identifies the technical information contained in each sentence. Using pointing words, we can determine the type of technological information contained in a sentence when a certain word appears in the sentence, depending on the meaning of the pointing words. Then, the extracting rule is defined, which is a method of extracting keywords in a hierarchical manner based on the degree of importance of each keyword in technological information. Second, text in patent is preprocessed to apply an extracting rule which is results of the first step, the information of the part of speech and phrase in patented text is tagged through the NLP process to apply pointing words and extracting

rule defined in the technological information analysis stage. We then classify and structure the types of technological information of sentences through pointing words. In addition, technological information in patents is layered using NLP. As results of the structuring & layering process, each type of technological information is derived in specific meaning, and information is layered by their importance and impacts.
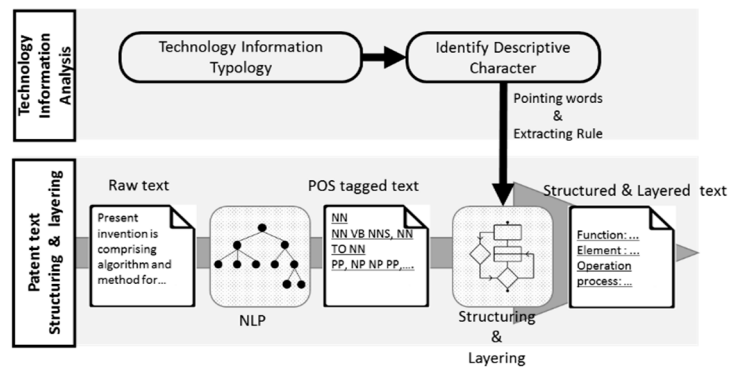
Figure 2. Research concept.

3.2. Overall Process

3.2.1. Technological Information Analysis

(1)   Definition of technological information type

A patent describes the structure, method, function, substance, or combination of these, in order to protect the invention via the Patent Act that stipulates that a person having ordinary knowledge in the technical field can easily carry out the invention. In addition, the application for the patent should summarize the above matters, and the claims shall be clearly and concisely stated in the Act [44]. The patent also has a claim that specifies the type of technological information it includes. Claims are the most important factor in securing legal protection and uniqueness by explicitly claiming the patent owner's legal rights. In addition, according to 35 U.S.C § 112 [45], 'a component may be expressed as means or steps for performing a specific function other than a description of a material or an act'. Accordingly, when writing the patent specification, the uniqueness of the patent is secured by describing its various unique components.

In accordance with these regulations, unlike general documentation, patents must describe specific technological information. Therefore, all statements described in the patent can be interpreted to include specific technological information. In previous research, the keywords were extracted first and the meaning of each keyword was interpreted from expert opinions and prior research in the technical field, in terms of technological information such as keywords related to function and keywords related to the application field of the patent. However, in this study, we need to clarify the meaning of the keyword from a technological information perspective without additional analysis by structuring and extracting the patent information. In this process, the type of technological information in the patent is defined by referring to the international technology classification, the domestic technology industry classification system, and the existing research that interprets the technological information of the patent from various viewpoints.

(2)   Identification of description type according to technological information type

Technological information has different descriptions depending on the form. For example, the function of a patent is described in the form of a verb or a gerund, and a component is expressed in the form of a noun. In this way, the form described by the type of technological information

is understood and only the desired information is extracted according to the type of technological information. Thus, this research defines the type of technological information, and uses pointing words to extract only specific information. The pointing word is an indicator that can identify which sentences have technical information and which contents of technical information are included in the sentences. By using pointing words, the sentences appearing in the title, type, and claim can be structured according to the type of technological information.

For example, interface, and system, which are nouns described in the title of patent number US8570295 [46] shown in Table 1, are nouns commonly used in the patent itself or as nouns referring to the whole patent. The first sentence in the abstract of the document in which nouns are subjects is the explanation of the patent. In this case, 'interface' and 'system' can be defined as pointing words that each word explains the patent itself. In other words, if 'interface' or 'system' appeared in a sentence, the sentence contains technological information about the patent itself. It is to be understood that the words 'include' and 'comprise' described in the abstract and claim parts both mean 'inclusion', while the following words, 'layer' and 'fluid channel' correspond to a component. In this case, if 'include' and 'comprise' occur, they can be defined as pointing words in which the component is described. If two defined pointing words are used, it is a sentence that contains technological information that describes the patent itself when it contains only 'interface', and 'system'. If it contains both 'include' and 'comprise', it can be classified into sentences expressing technological information about elements.

Table 1. Index of patent US8570295.

US8570295	
Title	User interface system
Abstract	The user interface system of the preferred embodiment includes: ...
Claim	A user interface comprising: a layer comprising ..., a fluid channel, and a tactile surface, ...

The sentences that are classified according to the type of technological information and included as pointing words have technological information for the different parts of speech. In addition, the technological information is not a single word, but contains a more detailed description in terms of phrases. Therefore, in this study, a 'depth' concept is defined and used as a measure of the degree of importance of keywords in each technological information.

Figure 3 shows the result of tagging information of phrases and the parts of speech in the sentence, "sensors recognize gesture moving horizontally" through NLP. Assuming that 'sensors' is one of the constituent elements of the patent, it can be seen that the most important information is the 'sensors', which provide additional information on horizontal movement, which is a feature of the sensors and gestures. The information in the Depth 1 can be defined as 'sensors' and the keywords in the Depth 2 become 'sensors' and 'recognize'. Thus, the keywords of Depth $n + 1$ can be defined as a set of keywords appearing at a level of the nth and keywords appearing at a level of $n + 1$th. Therefore, the extraction rule that extracts layered keywords using the depth concept is defined in the sentence classified by types of technological information.
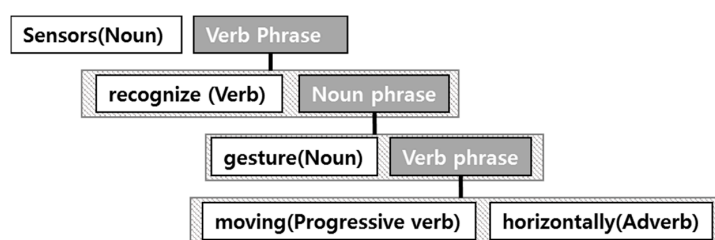

Figure 3. Part of speech (POS) tagged sentence.

3.2.2. Patent Structuring and Layering

The sentences in the title, summary, and claims are converted into a tagged form of words in parts of speech and phrases through the NLP process. The sentences described in the title, summary, and claims are classified and structured according to the types of technological information given by the pointing words defined through the patent analysis and the existing prior studies. The technological information is then layered according to depth, using the extracting rule defined for each type of technological information.

(1)   Structuring technological information in patent documents

We use pointing words defined on the basis of descriptive type of technological information. The sentences in the title, summary, and claims are structured according to technology types based on whether or not pointing words appear. As described in Section 3.2.1, the type of technological information is not classified by only one pointing word, but by multiple pointing words and the texts included in the titles, abstracts, and claims are structured according to the types of technological information in a sentence, as shown in Figure 4. For example, if a sentence has a noun that refers to a patent, and a verb that means inclusion, it includes the elements that the patent itself contains. On the other hand, when a sentence has a noun that refers to the patent as well as a verb that is related to inclusion, it can be seen that one of the components of the patent contains elements other than the patent itself. In a sentence, a noun that is not the name of a patent can be considered as a component of the patent.
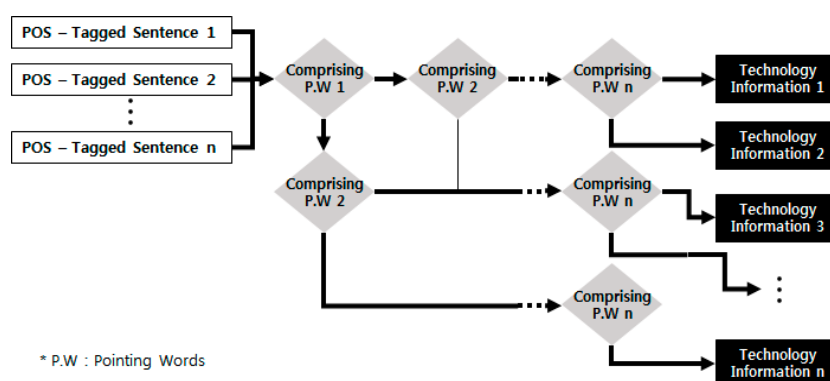
Figure 4. Text structuring based on pointing words.

A patent is characterized by a long, complex sentence, rather than a short sentence. If the pointing words appear once in a complex sentence, errors will most likely occur. Therefore, in this study, using the depth concept throughout the sentence, the error can be prevented only by including the lowest depth, that is, the pointing words as the key information, as shown in Figure 4.

(2)   Layering keywords by technological information

Keywords are structured depth using the extraction rule defined for each structured sentence according to technology types. For example, the sentence, "device includes: a sensor recognizes a gesture moving horizontally", is classified as a sentence containing a component, and indicates that the noun is technological information corresponding to a component. Therefore, the phrase containing "a sensor" can be extracted as technological information and layered, as shown in Figure 5. In other words, a lot of information is divided into various layers, enabling to choose how many information extract to interpret a patent. For example, if we want formal information about a sentence, we choose Depth 1 and interpret a main topic of the sentence. If we want more information than Depth 1, we can

analyze Depth 2 and Depth 3. Each depth means that Depth 2 contains information about the main topic and their verb, and Depth 3 contains more information than Depth 2, which is an object of the verb.
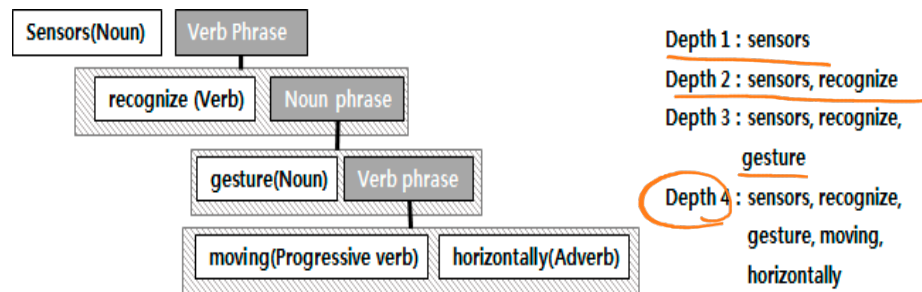

Figure 5. Layered sentence by depth.

3.3. Efficiency Verification

The verification process consists of two stages. First, we validate the efficiency of the patent-level keyword sets by extracting the technological information for each type of patent quantitatively and qualitatively. Then, by analyzing one patent through structured and layered keywords, we confirm whether the level of patent analysis has been analyzed, not the level of technology field analysis. First, we compare the average *TF-IDF* value of a set of keywords to verify the efficiency of the set of keywords derived from a higher technology level in order to verify the set of keywords by a quantitative method. For the qualitative verification, we verify the importance of the keywords obtained based on the technical classification system of the technical field and the advice of experts and verify the efficiency of the method by judging whether the meaningful keywords are present in the existing method and the proposed method. The *TF-IDF* value indicates the degree of importance of each keyword in the document. As the *TF-IDF* value increases, the keyword in the document becomes more important. In this paper, we evaluate the importance of keywords in a set of keywords, rather than the importance of the keywords in the documents. The formula for obtaining the *TF-IDF* value is as follows.

$$TF = \frac{keyword\ frequency\ \in keyword\ set}{all\ word\ freqyency\ \in keyword\ set} \tag{1}$$

$$IDF = \log_2 \frac{number\ of\ patent}{number\ of\ patent\ containing\ keyword} \tag{2}$$

$$TF - IDF = TF \times IDF \tag{3}$$

Therefore, when the average *TF-IDF* value of the keyword set is high, it can be interpreted that the importance of the keywords belonging to the keyword set is high. The average *TF-IDF* values of a set of keywords obtained by the keywords extracted through the conventional method and the proposed method are compared. A comparison of two sets of keywords suggests that a set of keywords with a high *TF-IDF* value contains important keywords in each patent.

This research evaluates the quality of the keywords obtained by the proposed method based on the classification system of the existing technical field and the consultation of experts. Although the keywords evaluated to be meaningful keywords are not revealed by the existing methods, the result shows that the keywords appear when extracted through the proposed method, confirming that important keywords are extracted in the interpretation of patents and technical fields. Finally, it is confirmed that the patent can be obtained by structuring and layering the information at the level of one patent rather than at the level of technology, based on the keywords of the technical field extracted from one patent.

4. Data Analysis

Based on the suggested framework, the extracting rule is defined by both literature review and linguistic analysis about texts included in patent documents. Then, the defined rule is applied to a target technology and then technological information is structured and layered. Finally, the results are verified in qualitative and quantitative ways.

4.1. Selection of Analysis Target

The user interface (UI) field was used in this study in order to select various types of technological information in patents. UI is a technology that enables communication between people and machines. It is a technology field that connects many smart devices and smart contents with which we are in contact [47]. UI technology is attracting attention as 'human-centered' technology, leading the ICT market [48]. Even if products have the same function, a user can select a product that provides the high efficiency and convenience of the UI. Thus, the UI technology should be considered as one of core factors in developing a user-oriented product. The UI field is used in various technologies and industries, and can be divided into detailed fields according to the method in which information is provided to the user and the input method. In the method of providing information, a user-experience interface is provided as a method of using all five senses instead of using only sight. As a method of inputting information, the intention of the user can be input through various methods such as a touch screen, motion recognition, biometric signal recognition etc., other than a keyboard and a mouse input method. In addition to its features, it has various applications such as personal computers, navigation, and medical devices. Therefore, it can be seen that the UI field has various keywords according to the type of technological information for the purpose of this study.

In this study, the patents of UI are analyzed, and keywords are structured and layered according to types of technological information. In order to collect the patents in the UI field, a patent database of United States Patent and Trademark Office (USPTO) was applied. A total of 500 patents are collected by using the search expression (1) "User Interface" in the title and summary from 2011 to 2015, excluding sentences that have particular symbols such as "", [], and <> that could cause an error in the NLP process.

4.2. Technical Analysis

(1)　Definition of types of technological information

In this study, four types of technological information were defined by reviewing the patent classification, industry classification technology, and information in the patent in previous studies. First, patents are classified by international standards of the IPC code that have the five classification criteria for a patent, including industry types, the patent form, component, functions, and detailed features. Industrial technical classification has been based on the type of industry, application, function, and operation method.

In terms of the use of various technological information, Huang et al. [17] utilized the SAO structure, interpreting the action as patented functions and the object as the objectives of patents. Lee et al. [18] classified patents by analyzing patent claims in terms of the patent rights. Kim and Choi [49] demonstrated that the meaning of each keyword has a greater effect on the technical classification than positions of keywords such as title, summary, and claims by analyzing the effect of keywords on classification based on Japanese patents.

From literature reviews on technological information, types of technological information can be defined as functions, applications, components, and operation methods. Functions refer to a key feature of the patent. The application is an object where a patent is performed or operated, referring to the areas in which the patent is carried out. Components refer to a generic name of various types of components, such as the physical and functional components included in the patent, and can be

divided into non-legal components and legal components. The method of operation refers to the specific method for achieving a specific function described in the patent.

The description of information types proposed in this study can be classified by the patent laws, enabling classification of all sentences contained in the patent document. According to the patent law, the subject of sentences should be a noun that is previously defined because every word in the patent documents must be clearly explained to the subject. A patent has specific nouns that indicate the type of patent, such as a device or method, or general nouns such as 'handle' and 'glasses' in the title of the patent to identify the patent. A sentence that has such nouns as a subject explains a patent itself. Thus, all sentences that have a subject which is a noun to indicate a patent, explain the functions and applications of the patent.

In terms of components, non-legal and legal components are clarified by a verb that includes the meaning of the components in the patent claims and summary. Since the simple identification of components is not able to ensure the uniqueness of each patent, the features of each component in the patent document are defined to ensure uniqueness. Therefore, if the subject of a sentence is a component, it can be said that the sentence describes methods of operation for implementing a key feature of the patent.

Finally, a sentence has a subject that is a noun, indicating the patent itself or its functions. A sentence that explains a function deals with the detailed explanation of the critical features and operating methods. However, according to patent law, a patent should provide novelty compared with the functions of existing patents; therefore, it is extremely unusual to describe the detailed features through the new sentence. Therefore, a sentence in which a subject is a function can be interpreted as unique information for providing an operating method of the patent as well.

(2) Identifying narrative forms of technological information

The technological information described in the patent has a typical description form according to its type. When a sentence is described as a particular form, it can be classified as a sentence that contains the information of the description. By extracting the technological information that is described in the form of classified sentences, it is possible to extract each type of technological information. This research defines pointing words in order to classify the sentences, considering the type of technological information. Pointing words are words that can determine a type of information technology, which is included in the sentence based on the occurrence or absence of the specific words. Pointing words consist of nouns that represent a patent (Representing Noun; RN), verbs that have common meaning (General Verb; GV), verbs that appear in front of components (Component Verb; CV), and nouns that mean components of patents (Component Noun; CN).

RN is a noun related to the form of patents such as device, method, system, algorithm, program, apparatus, and invention. Since its frequency is quite high, yet the *TF-IDF* values are very low, the keyword has information that has little impact on the meaning of the whole patent. Further, in addition to the form of a noun related to the patent, patents may be briefly expressed as a noun, such as 'display' and 'handle'. Since these nouns are in the lowest level in the title of the patent, the terms related to the form of the patent and nouns that are in the lowest level of the noun can be defined as RN. Although RNs that can be obtained in two ways have little significance in the patent, they explain the patent itself. A sentence in which RNs appear can be interpreted as a description of the patent itself. GV consists of verbs such as 'provide', 'suggest', 'mean', 'relate', and 'describe' that have high frequency, yet the *TF-IDF* values are very low. GV are verbs that, when interpreted in Korean, are not interpreted as a specific meaning, and a general description is given such as "the patent provides (or proposes, means) the ~~", before describing the key information. Thus, since a GV does not give special information in the sentence, the subsequent phrases and nouns contain more key features. CV refers to verbs such as 'consist', 'include', 'compose', 'form', and 'involve' that have high frequency yet very low *TF-IDF* values. Nouns and noun phrases that are contained behind the aforementioned verbs can be construed as a component of the patent. CN can be interpreted as

a component of patents because it is a noun or noun phrase appearing after CV. Although CN is not the key factor of the patent, the CN can implement the key features or functional components of the patent due to the interaction between the CNs. Therefore, a sentence in which the CN is a subject can be interpreted as the method used. The type of technological information can be determined by using the aforementioned four pointing words included in their position in the title, summary, and claims. In addition, since the technological information has a particular part of speech according to its type, extracting rules can be defined to extract the desired technological information by utilizing NLP as shown in Table 2.

Table 2. Described form and extracting rule of technological information.

Technological Information Type	Described Form					Extracting Rule
	Described Item	Pointing Words				
		RN	GV	CV	CN	
Function	Title	O	X	X	X	Verb, Gerund phrase, Verb phrase with preposition
	Abstract	O	X	X	X	Verb phrase, Gerund phrase
	Abstract	O	O	X	X	Gerund phrase with preposition, supine
Object	Title, Abstract	O	O	X	X	Noun phrase with preposition
Unauthorized Component	Abstract	O	X	X	X	Noun phrase, gerund phrase, supine after GV
Authorized Component	Claim	O	X	O	X	Noun phrase, Gerund phrase, supine after 'Comprising: ' with 'a, an'
Method of operation	Abstract, Claim	X	X	X	O	Verb phrase in taking a component as a subject
	Abstract	O	X	X	O	Verb phrase after 'the + RN or CN'

Statements containing information about the function of the patent can be classified into two types. The first type is a sentence that includes RN in the title and summary, but does not contain a CV and GV. In such a case, the extracted verb or gerund phrase can be interpreted as a function. Another form contains the RN, GV, CV, and a preposition before a verb phrase or 'to' infinitive. In such a case, the extracted verb or gerund phrase can be interpreted as a function used to extract the verb phrase or 'to' infinitive subsequent to prepositions as its adverbial usage.

A sentence that expresses the application of the patent includes the RN in the title and summary, and the noun extracted after the preposition can be interpreted as an application. In terms of components, the description forms of legal and non-legal components differ. First, a sentence containing a legal constituent element includes all the sentences described in the claim. It is possible to extract noun phrases, gerund phrases, and 'to' infinitive phrases including the articles 'a/an' described after the word "comprising" and to interpret them as a legal constituent in noun usage. Sentences containing non-legal components appear in the abstract with RN and CV. It is possible to extract noun phrases, gerund phrases, and 'to' infinitive phrases including 'a/an' described after CV, and interpret them as non-legal constituents according to noun usage.

The sentence containing the method of operation appears in two forms based on the presence or absence of the article. First, the verb phrase can be interpreted as the method of operation in the sentence that has the CN, rather than the RN, as the subject. Second, if "The + RN" is a subject of a sentence, it can be interpreted as an operation method.

4.3. Patent Structuring and Layering

The texts included in the title, summary, and claim according to the sentence are separated and each sentence is tagged with POS. Then, based on the tagged information, the text is structured according to the type of technological information, and the structured information is layered according to its depth.

4.3.1. Structuring Technological Information according to the Types

Sentences that extracting rules can be applied to are explored in the title, summary, and claims of patents. Although the title and claims are described only in one sentence, a summary generally has multiple sentences. This paper analyzes 500 sentences with titles and claims, respectively, and 951 sentences with a summary. The patent documents therefore contain a total of 1951 sentences. The number of sentences classified by type of technological information by the pointing words is shown in Table 3.

Table 3. Sentence Ratio.

Technological Information	Number of Sentences	Ratio
Function	411	21%
Object	484	9%
M.O *	64	3%
Function, Object	550	28%
Function, Component	207	11%
Function, Object, Component	271	14%
Object, M.O	53	3%
None	221	11%

* Method of operation.

4.3.2. Keyword Layering by Technological Information

In the title, summary, and claims, sentences are structured according to function, application, object, component, and operation method, considering the depth in which information is extracted at a desired level. Depth 1 contains the keywords that can be obtained at the highest level. Functions include verbs, nouns for applications, nouns or gerunds for components, and verbs for operating methods. The lower depth comprises words that can further describe the words in Depth 1. Table 4 shows the results of extracting the keywords by changing the level of depth from Depth 1 that can obtain only keywords to Depth 4 that can obtain additional information.

Table 4. The number of extracted keyword.

	All	Depth			
		1	2	3	4
Total		1044	1442	1581	1656
Function		550	699	809	874
Object	4243	282	389	436	471
U.C *		454	535	596	643
A.C **		415	966	4063	1117
M.O ***		203	251	290	322

* Unauthorized component; ** Authorized component; *** Method of operation.

4.4. Verification

In order to verify all the keywords obtained from the technical field in this study, we quantitatively and qualitatively compared our proposed method with the method of extracting the top 20% of the existing *TF-IDF* values. In addition, we analyzed the function, application object, component, and operation method of the patent by using the structured keyword which could not be obtained by the conventional method.

4.4.1. Keyword Set Verification

We compared the average *TF-IDF* of the keyword set using the existing keyword extraction method and the structuring and layering (S & L) proposed in this study while changing the ratio of the keywords extracted from each keyword set. Figure 6 shows that as the number of keywords increases, the average value of *TF-IDF* decreases. In addition, it can be confirmed that the S & L method has a higher average value of *TF-IDF* than the conventional method. In this case, as the depth decreases, the average *TF-IDF* value increases. Therefore, it can be said that the keyword with low depth has information that is unique to each patent. On the other hand, as the depth increases, the *TF-IDF* value becomes lower. Therefore, as a keyword that has higher depth is extracted, each keyword appears in various patents rather than in one patent.
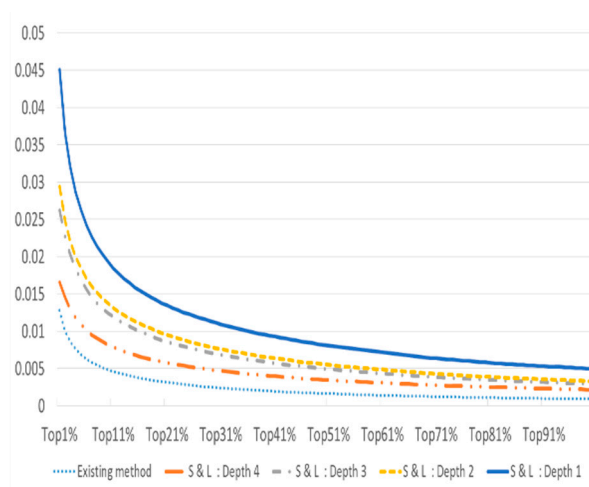
Figure 6. Average of Term Frequency-Inverse Document Frequency (*TF-IDF*) index per extracted ratio.

Table 5 shows the results obtained by comparing a set of keywords through the existing method and S & L. It demonstrates that it is more useful to extract the Depth 1 and 2 keywords using S & L than the method that extracts only the *TF-IDF* values corresponding to the upper 20% of the existing keywords. In addition, the rate that the keyword obtained through S & L is extracted in the method which depends on the *TF-IDF* value is calculated in order to determine whether the problem in which significant keywords are excluded has been resolved. Table 6 shows that only 45.31% of the extracted keywords are extracted from the top 20% of the extracted keywords. It can be seen that the existing method extracts only part of the significant keywords extracted through S & L.

Table 5. Average of *TF-IDF* index per extracted ratio.

Extracting Method	Ratio	Average of *TF-IDF*
Existing Method	20%	0.003228
Existing Method	100%	0.000917
S & L: Depth 1	100%	0.004991
S & L: Depth 2	100%	0.003322
S & L: Depth 3	100%	0.002959
S & L: Depth 4	100%	0.002166

Table 6. Extracted keyword ratio in existing method.

Extracting Ratio	Extracted Keyword (Ratio)	
	S & L: Depth 1	S & L: Depth 4
20%	475 (45.31%)	849 (51.27%)
40%	699 (66.95%)	1624 (98.07%)
60%	829 (79.12%)	1654 (99.88%)
80%	910 (87.16%)	1656 (100%)
100%	1044 (100%)	1656 (100%)

4.4.2. Keyword Verification

Through the keyword set verification, it can be confirmed that S & L is a set of keywords that is quantitatively superior to the existing keyword extraction method. For the qualitative analysis of the extracted keywords, valid keywords were selected from S & L's classified functions, applicable object, non-legal constituent element, legal constituent element, and operation method based on expert's advice and related research [6,50]. At this time, since the rank of extracted keywords differs according to the depth, the depth that can select the most significant keyword according to technological information type is derived. Table 7 shows that the main functions of the user interface technology are represented by display, detect, and control, and their applications are graphical display, gesture, information, and application. Elements constituting each patent include functions used to perform detect, control, and receive, as well as physical components such as display, processor, surface, and sensor. In addition, it can be seen that the main functions of the patent are implemented through the actions of detect, associate, integrate, and configure.

Table 7. Extracted keyword index per technological information.

	S & L: Depth 1				S & L: Depth 4					
Rank	Function		Object		Unauthorized Component		Authorized Component		Method of Operation	
	Words	Ranked in E.M *	Words	Ranked in E.M	Words	Ranked in E.M	Words	Ranked in E.M	Words	Ranked in E.M
1	Display	14	Graphical	47	Display	14	Display	14	Area	7
2	Detect	340	Display	14	Detect	340	Professor	125	Display	14
3	Control	12	Gesture	11	Surface	10	Detect	340	Associate	**1689**
4	Configure	**1691**	Information	16	Control	12	Configure	**1691**	Detect	340
5	Determine	356	Data	**2691**	Receive	218	Select	340	Integrate	**1555**
6	Manage	**3691**	Application	1	Touch	4	Receive	218	Configure	**1691**
7	receive	218	location	40	sensor	76	control	12	Gesture	11

* Existing Method.

In the existing keyword extraction method, the upper 20% of the *TF-IDF* value is extracted. The total number of keywords to be analyzed is 4243, and when 20% is extracted, only the top 849 keywords are extracted and other keywords are excluded. Table 7 shows that the main functions of the patents in the User Interface field are display, detect, control, configure, and manage. At this time, it can be seen that "configure" and "manage" have *TF-IDF* values as high as the 1691th and 3691th keywords in the entire keyword set, and thus cannot be extracted by the conventional method. In the same way, it can be seen that important keywords in technological information are not extracted by the conventional method.

4.4.3. Verification of Patent Interpretation Method

Keyword extraction through S & L can be used not only in the entire patent data set, but also when extracting keywords from a single patent. Table 8 shows the results obtained by applying US857029 with the information level of Depth 4 in the S & L approach. Analysis of the results in

Table 8 shows that this patent has the function of inputting keywords in various ways and is used in computers. In addition, it consists of display function and the physical components of the keyboard. Legally, the way that users type letters and text on the keyboard is protected. A function of inputting characters by sensing various touches of the keyboard is implemented.

Table 8. Keyword index of US8570295.

US8570295	
Title	Touch screen device, method, and graphical user interface for inserting a character from an alternate keyboard
Function	insert character alternate keyboard
Object	computer-implemented method use
Unauthorized component	display soft keyboard
Authorize component	plurality character-insertion key select soft keyboard different single move break text input area correspond
Method of operation	key select soft keyboard different contact detect response Movement lift off character

5. Conclusions

In this study, the sentence was structured according to types of technological information by analyzing the descriptive form of technological information described in the title, summary, and claims. Each keyword is then layered based on the degree of importance in the technological information, and the technological information is structured according to type. The core keywords and a set of keywords layered by keywords are then obtained. We confirmed that the keyword set extracted by the level based on the depth from the sentence that is classified according to the descriptive information provides more significant keywords than a set using only part of the existing keyword based on the average *TF-IDF* value of the keywords in a quantitative manner. In addition, when extracting some keywords, it is possible to determine whether or not the keywords included in the verified set of keywords are detected with a significant set of keywords, so that the proposed method can extract the meaningful keywords without missing them.

Based on the preliminary research on the UI field and the consultation of experts, it is found that the proposed method can extract the keywords that are important in function, application object, non-legal component, component, and operation method, but not extracted through the existing method. It is verified that the same word can be interpreted differently depending on the type of technological information. In addition, it is possible to analyze each patent according to the type of technological information by extracting the keywords of each patent without defining the characteristics of patents based on extracted keywords from the technical field level. Thus, we can confirm that the proposed approach can derive a more detailed level of information than the existing text mining techniques.

The proposed methodology has three theoretical contributions. First, keywords of patents are extracted by linguistic criteria. Most previous keyword extraction methodology is based on their meaning or the number of occurrence. However, this research suggests that another criterion can be used to extract meaningful keywords in patent. Second, the proposed methodology can be used not only to derive the characteristics of the technology field but also to derive the characteristics of each patent according to the type of technological information. Third, it is advantageous to extract more significant keywords using fewer keywords by applying the depth after classifying the sentences in terms of analysis efficiency, because the method does not screen valid keywords after extracting all keywords. In addition, according to the intention of a researcher, it is possible to obtain a flexible set of keywords according to the purpose of analysis by varying the depth.

The proposed methodology can be utilized in industry fields by various ways. First, when the technical field has a lack of prior knowledge, or when it is difficult to interpret the extracted keywords

in the field of fusion technology, the meaning of each keyword can be clarified and the characteristics of the technical field can be derived. Second, when presenting specific technology opportunities such as development purpose, target, development direction in technology, and product development process, time series analysis can be applied to type-specific technological information to explore more detailed technology opportunities. Third, using features that are structured according to types, technological information can be extracted from the patent level, and the technology can be classified based on the new application target, functions, and components by comparing patents. Moreover, in order to analyze the possibility of patent infringement, the technological information of the two patents to be analyzed can be compared from various viewpoints.

However, this research has two limitations. First, the results of the research are greatly affected by the quality of NLP. The patent documents were structured in this study by relying on the parts of speech tagged by the NLP. In other words, if parts of speech tagging process do not work well, each sentence in patents cannot be structured, technological information cannot be extracted in the structured form. Second, the suggested methodology cannot be applied in all technology domains. In the case of the User Interface technology that analyzed in this research, the technological information of the four types of technology, function, application object, component, and operation method is evenly included. However, excellent results cannot be obtained in a technical field that requires technological information other than the four types, for example, when a patent is expressed through an algorithm or a chemical formula.

In order to overcome the limitations of this study, it is necessary to utilize a better quality NLP, such as NLP using deep learning. In addition, in the analysis of other technical fields, various types of technological information need to be defined according to technical fields such as operation sequence and interaction in addition to the four types of technological information through sufficient literature survey. Moreover, the proposed approach needs to be generalized in any types of documents and technologies by reflecting the unique characteristics of documents and technologies.

Author Contributions: T.R. designed the study, outlined the methodology, conducted the data analysis, and wrote the manuscript. Y.J. interpreted the results, and wrote the manuscript. B.Y. implemented the research, designed the study, outlined the methodology, and helped draft the paper. All authors have read and approved the final manuscript.

References

1. Ernst, H. Patent information for strategic technology management. *World Pat. Inf.* **2003**, *25*, 233–242. [CrossRef]
2. Choi, J.; Kim, H.; Im, N. Keyword Network Analysis for Technology Forecasting. *J. Intell. Inf. Syst.* **2011**, *17*, 227–240.
3. Liu, S.J.; Shyu, J. Strategic planning for technology development with patent analysis. *Int. J. Technol. Manag.* **1997**, *13*, 661–680. [CrossRef]
4. Jeong, E.S.; Kim, Y.G.; Lee, S.C.; Kim, Y.T.; Chang, Y.B. Identifying Emerging Free Technologies by PCT Patent Analysis. *J. Korea Inst. Electron. Commun. Sci.* **2014**, *9*, 111–122. [CrossRef]
5. Yun, J.H. Patent Information Analysis: Tools for Systematic R & D Planning. *Ind. Eng. Mag.* **2011**, *18*, 23–28.
6. Lim, C.; Yun, D.; Park, I.; Park, G.; Koh, S.; Yoon, B. Exploring Prospective Research Areas in UI/UX through the Analysis of Patents. *Korean Manag. Sci. Rev.* **2015**, *32*, 1–18. [CrossRef]
7. Tseng, Y.H.; Lin, C.J.; Lin, Y.I. Text mining techniques for patent analysis. *Inf. Process. Manag.* **2007**, *43*, 1216–1247. [CrossRef]
8. Altuntas, S.; Dereli, T.; Kusiak, A. Analysis of patent documents with weighted association rules. *Technol. Forecast. Soc. Chang.* **2015**, *92*, 249–262. [CrossRef]
9. Park, H.; Yoon, J.; Kim, K. Identifying patent infringement using SAO based semantic technological similarities. *Scientometrics* **2011**, *90*, 515–529. [CrossRef]

10. Wang, X.; Qiu, P.; Zhu, D.; Mitkova, L.; Lei, M.; Porter, A.L. Identification of technology development trends based on subject–action–object analysis: The case of dye-sensitized solar cells. *Technol. Forecast. Soc. Chang.* **2015**, *98*, 24–46. [CrossRef]

11. Campbell, R.S. Patent trends as a technological forecasting tool. *World Pat. Inf.* **1983**, *5*, 137–143. [CrossRef]

12. Lee, C.; Kang, B.; Shin, J. Novelty-focused patent mapping for technology opportunity analysis. *Technol. Forecast. Soc. Chang.* **2015**, *90*, 355–365. [CrossRef]

13. Su, H. Global Interdependence of Collaborative R&D-Typology and Association of International Co-Patenting. *Sustainability* **2017**, *9*, 541.

14. Kim, C. A patent analysis method for identifying core technologies: Data mining and multi-criteria decision making approach. *J. Korea Saf. Manag. Sci.* **2014**, *16*, 213–220. [CrossRef]

15. Kang, H.J.; Um, M.J.; Kim, D.M. A study on forecast of the promising fusion technology by US patent analysis. *J. Technol. Innov.* **2006**, *14*, 93–116.

16. Niemann, H.; Moehrle, M.G.; Frischkorn, J. Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application. *Technol. Forecast. Soc. Chang.* **2017**, *115*, 210–220. [CrossRef]

17. Noh, H.; Jo, Y.; Lee, S. Keyword selection and processing strategy for applying text mining to patent analysis. *Exp. Syst. Appl.* **2015**, *42*, 4348–4360. [CrossRef]

18. Huang, L.; Shang, L.; Wang, K.; Porter, A.L.; Zhang, Y. Identifying target for technology mergers and acquisitions using patent information and semantic analysis. In Proceedings of the 2015 Portland International Conference on Management of Engineering and Technology (PICMET), Portland, OR, USA, 2–6 August 2015.

19. Lee, C.; Song, B.; Park, Y. How to assess patent infringement risks: A semantic patent claim analysis using dependency relationships. *Technol. Anal. Strateg. Manag.* **2013**, *25*, 23–38. [CrossRef]

20. Lee, W.S.; Han, E.J.; Sohn, S.Y. Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. *Technol. Forecast. Soc. Chang.* **2015**, *100*, 317–329. [CrossRef]

21. Ko, N.; Yoon, J.; Seo, W. Analyzing interdisciplinarity of technology fusion using knowledge flows of patents. *Exp. Syst. Appl.* **2014**, *41*, 1955–1963. [CrossRef]

22. Lee, W.J.; Sohn, S.Y. Patent analysis to identify shale gas development in China and the United States. *Energy Policy* **2014**, *74*, 111–115. [CrossRef]

23. Yoon, J.; Kim, K. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics* **2011**, *88*, 213–228.

24. Wang, H.; Cheng, D.; Chen, C.; Wu, Y.; Lo, C.; Lin, H. A Novel Real-Time Speech Summarizer System for the Learning of Sustainability. *Sustainability* **2015**, *7*, 3885–3899. [CrossRef]

25. Guo, X.; Sun, H.; Zhou, T.; Wang, L.; Qu, Z.; Zang, J. SAW Classification Algorithm for Chinese Text Classification. *Sustainability* **2015**, *7*, 2338–2352. [CrossRef]

26. Kim, H.J.; Jo, N.O.; Shin, K.S. Text Mining-Based Emerging Trend Analysis for the Aviation Industry. *J. Intell. Inf. Syst.* **2015**, *21*, 65–82. [CrossRef]

27. Min, K.Y.; Kim, H.T.; Ji, Y.G. A Pilot Study on Applying Text Mining Tools to Analyzing Steel Industry Trends: A Case Study of the Steel Industry for the Company "P". *J. Soc. e-Bus. Stud.* **2014**, *19*, 51–64. [CrossRef]

28. Kim, M.; Park, Y.; Yoon, J. Generating patent development maps for technology monitoring using semantic patent-topic analysis. *Comput. Ind. Eng.* **2016**, *98*, 289–299. [CrossRef]

29. Ghazizadeh, M.; McDonald, A.D.; Lee, J.D. Text Mining to Decipher Free-Response Consumer Complaints Insights From the NHTSA Vehicle Owner's Complaint Database. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2014**, *56*, 1189–1203. [CrossRef] [PubMed]

30. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

31. Kwon, H.; Kim, J.; Park, Y. Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology. *Technovation* **2017**, *60–61*, 15–28. [CrossRef]

32. Park, J.; Song, M. A study on the Research Trends in Library & Information Science in Korea using Topic Modeling. *J. Korean Soc. Inf. Manag.* **2013**, *30*, 7–32.

33. Jang, H.; Roh, T.; Yoon, B. User needs-based technology opportunities in heterogeneous fields using opinion mining and patent analysis. *J. Korean Inst. Ind. Eng.* **2017**, *43*, 39–48. [CrossRef]

34. Jin, S.A.; Heo, G.E.; Jeong, Y.K.; Song, M. Topic-Network based Topic Shift Detection on Twitter. *J. Korean Soc. Inf. Manag.* **2013**, *30*, 285–302. [CrossRef]

35.　Gao, L.; Eldin, N. Employers' Expectations: A Probabilistic Text Mining Model. *Procedia Eng.* **2014**, *85*, 175–182. [CrossRef]

36.　Guo, H.; Chen, Y. User interest detecting by text mining technology for microblog platform. *Arab. J. Sci. Eng.* **2016**, *41*, 3177–3186. [CrossRef]

37.　Lee, S.; Kim, H.J. News keyword extraction for topic tracking. In Proceedings of the Fourth International Conference on Networked Computing and Advanced Information Management, NCM′08, Gyeongju, Korea, 2–4 September 2008; pp. 554–559.

38.　Rose, S.; Engel, D.; Cramer, N.; Cowley, W. Automatic Keyword Extraction from Individual Documents. 2010. Available online: https://pdfs.semanticscholar.org/5a58/00deb6461b3d022c8465e5286908de9f8d4e.pdf (accessed on 16 November 2017).

39.　Hulth, A. Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2003; pp. 216–223.

40.　Nasukawa, T.; Yi, J. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd International Conference on Knowledge Capture*; ACM: New York, NY, USA, 2003; pp. 70–77.

41.　Yi, J.; Nasukawa, T.; Bunescu, R.; Niblack, W. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In Proceedings of the Third IEEE International Conference on Data Minning, Melbourne, FL, USA, 22 November 2003; pp. 427–434.

42.　Jin, Y.; Xiong, W. A sentence degeneration model and its application in Chinese-English patent machine translation. In Proceedings of the 2011 7th International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE), Tokushima, Japan, 27–29 November 2011; pp. 421–424.

43.　Yang, S.Y.; Soo, V.W. Extract conceptual graphs from plain texts in patent claims. *Eng. Appl. Artif. Intell.* **2012**, *25*, 874–887. [CrossRef]

44.　Korea Patent Law. 2016. Available online: www.kipo.go.kr (accessed on 6 November 2017).

45.　United States Patent ACT, 35 U.S.C 112 (2011). Available online: www.uspto.gov (accessed on 6 November 2017).

46.　Cisesla, M.C.; Yairi, M.B. User Interface System. U.S. Patent 85,702,795, 29 October 2013.

47.　Kim, S.J.; Cho, D.E. Technology trends for UX/UI of smart Contents. *Korea Contents Assoc. Rev.* **2016**, *14*, 29–33. [CrossRef]

48.　Sohn, K.S. Technology convergence's present and future. *Ind. Eng. Mag.* **2012**, *19*, 28–33.

49.　Kim, J.H.; Choi, K.S. Patent document categorization based on semantic structural information. *Inf. Process. Manag.* **2007**, *43*, 1200–1215. [CrossRef]

50.　Park, I.; Park, G.; Yoon, B.; Koh, S. Exploring Promising Technology in ICT Sector Using Patent Network and Promising Index Based on Patent Information. *ETRI J.* **2016**, *38*, 405–415. [CrossRef]