

Article

Patent Keyword Extraction for Sustainable Technology Management

Jongchan Kim ¹, Jaehyun Choi ², Sangsung Park ^{3,*} and Dongsik Jang ¹

¹ Department of Industrial Management Engineering, Korea University, Seoul 02841, Korea; ourjongchan@korea.ac.kr (J.K.); jang@korea.ac.kr (D.J.)

² Daejeon Center for Creative Economy & Innovation, Daejeon 34141, Korea; tissue@ccei.kr

³ Graduate School of Management of Technology, Korea University, Seoul 02841, Korea

* Correspondence: hanyul@korea.ac.kr; Tel.: +82-10-8700-0469; Fax: +82-2-3290-3900

Received: 4 April 2018; Accepted: 20 April 2018; Published: 22 April 2018



Abstract: Recently, sustainable growth and development has become an important issue for governments and corporations. However, maintaining sustainable development is very difficult. These difficulties can be attributed to sociocultural and political backgrounds that change over time [1]. Because of these changes, the technologies for sustainability also change, so governments and companies attempt to predict and manage technology using patent analyses, but it is very difficult to predict the rapidly changing technology markets. The best way to achieve insight into technology management in this rapidly changing market is to build a technology management direction and strategy that is flexible and adaptable to the volatile market environment through continuous monitoring and analysis. Quantitative patent analysis using text mining is an effective method for sustainable technology management. There have been many studies that have used text mining and word-based patent analyses to extract keywords and remove noise words. Because the extracted keywords are considered to have a significant effect on the further analysis, researchers need to carefully check out whether they are valid or not. However, most prior studies assume that the extracted keywords are appropriate, without evaluating their validity. Therefore, the criteria used to extract keywords needs to change. Until now, these criteria have focused on how well a patent can be classified according to its technical characteristics in the collected patent data set, typically using term frequency–inverse document frequency weights that are calculated by comparing the words in patents. However, this is not suitable when analyzing a single patent. Therefore, we need keyword selection criteria and an extraction method capable of representing the technical characteristics of a single patent without comparing them with other patents. In this study, we proposed a methodology to extract valid keywords from single patent documents using relevant papers and their authors' keywords. We evaluated the validity of the proposed method and its practical performance using a statistical verification experiment. First, by comparing the document similarity between papers and patents containing the same search terms in their titles, we verified the validity of the proposed method of extracting patent keywords using authors' keywords and the paper. We also confirmed that the proposed method improves the precision by about 17.4% over the existing method. It is expected that the outcome of this study will contribute to increasing the reliability and the validity of the research on patent analyses based on text mining and improving the quality of such studies.

Keywords: sustainable technology management; patents; keyword extraction; text mining; statistical verification

1. Introduction

Companies are striving to gain a competitive advantage through new technology development. When their efforts result in success, they can expect some profit. However, when they fail, the result can

lead to huge losses. Therefore, R&D strategies that minimize the risk of failure are necessary, as this is an essential element in the development of new technology [2]. Patent analyses have been used as tools to establish such strategies [3]. Patents are objective data that contain information about technology. Patent database services make it possible to examine patent documents in order to gain insight into technology management. The use of such databases has led to active research on patent analysis. Until now, patent analysis has mostly been analyzed by expert evaluations of patent portfolios of specific technology fields. However, qualitative patent analysis by experts are problematic. First, analysis costs and time are consumed. Second, the subjectivity of experts can be involved in their analysis results. The third is related to the second problem. The reproducibility of the analysis results is not guaranteed. Many researchers are trying to use quantitative analysis to solve the problem of qualitative analysis. Patents include various quantitative indicators such as citations, family patents, filing dates, etc. However, the biggest issue in quantitative patent analysis is the quantitative analysis of patent unstructured data, such as a description of technology.

In particular, there have been a lot of studies focusing on the quantitative analysis of unstructured data, including patents, by applying text mining. Since patent analysis using text mining is based on words included in patent documents, the process of noise term removal and keyword extraction are preceded. As the extracted keywords have a significant influence on the subsequent analytic results, it is important for researchers to check if they are valid or not. However, most keyword-based patent analyses do not evaluate the validity of these keywords [4]. In addition, although the criteria for extracting keywords may vary according to the purpose of the analysis, most studies use the keyword extraction criteria to tell how well the collected patents can be classified based on the extracted keywords, which implies the technical characteristics. In this regard, the frequency-inverse document frequency (TF-IDF) method is the most frequently used [5–8]. However, the keywords selected by this criterion can be useful for a limited range of patent analysis, but are not suitable for patent analysis without any limited scope. Therefore, it is necessary to study the keyword extraction and performance evaluation methods that are able to consider the unique characteristics of a single patent document, without limiting the range of analysis. In addition, keyword selection of a single patent document is important because of the invariant. The keywords extracted from the patent portfolio can be changed whenever the patents constituting the portfolio are changed. Therefore, it is necessary to study the keyword extraction and performance evaluation methods that can show the unique characteristics of a single patent document regardless of the patent portfolio. In this study, we propose a methodology to extract the unique keywords of a single patent and evaluate its performance. When there are specific keywords in a document designated by its authors, they can be regarded as the terms that best represent the characteristics of the document. However, a patent document does not include specific keywords chosen by its inventors. This study is aimed at developing a method of extracting keywords from patent documents based on relevant papers whose content is similar to the target patents and their authors' keywords.

In order to do this, we set up a research hypothesis and statistically tested it to determine the optimal keyword extraction method and evaluate its validity. This study is expected to be useful for research on patent analysis using text mining, and help to guarantee the validity of the research outcomes. It is also expected that this research contributes to the advancement of relevant future research in various directions.

2. Related Work and Literature Review

2.1. Patent Analysis

Consumer needs and preferences are constantly changing, owing to competition and changes in technology. When companies have difficulty in expecting success from existing products alone, they can satisfy customer needs and preferences by developing new products. However, this entails significant risk. The success of a product development strategy will allow the company to preempt the

market and gain a competitive advantage that competitors cannot easily follow [9], yielding profits and improving the reputation of the company, enabling it to survive and grow in the market. However, even if a new product is released to market, it may not satisfy consumer demand, or it may fail in the market because of a competitor's products. Thus, companies want to increase their probability of success and decrease the probability of failure by analyzing consumer needs and market trends before developing new products.

The development of a new technology is similar to that of a new product in terms of its objectives, expectations, and risks. Therefore, companies place importance on their technology and their R&D strategies. What is the most effective technology and R&D strategy? Unlike products, it is difficult to grasp the functions of technology at a glance and requires expert technical knowledge to do so [10]. Patents provide the best data for objective information about a technology. The purpose of the patent is not only to guarantee the rights of inventors of technologies but also to contribute to the development of the industry by disclosing the technical contents. For this reason, most of the technologies developed through R&D are filed and published as patents, which contain detailed information on the technology such as bibliographic information, purpose of the invention, problems to be solved, solutions, and components. Thus, patent analyses are considered to be effective and objective methods for establishing a technology and R&D strategy. Grimaldi et al. [11] have proved that patents can be applied to technology and R&D strategies. They created a new framework for evaluating the patent portfolio held by the firm and used it for strategic technology planning. The indicators used in their framework can be broadly divided into two categories. One is an evaluation index using quantitative indices, such as the number of claims and forward citations. The other is a qualitative index that is read and analyzed by an expert and expressed as a value between 0 and 1. Ernst et al. [12] demonstrated the value of patents in R&D projects through a simulation analysis. They compared the R&D project with the same project without patent protection and patent protection. As a result, they found that the project receiving the patent protection had higher expected benefits than the non-patented project. Ou Yang and Weng [13] proposed a process to utilize patents for new product design. They obtained technical information for product development design using patent citation information and discovered the product niches by constructing a performance map of patent technology combined with TRIZ theory. These studies show that patents can be effectively used in strategic technology planning, R & D projects, and new product development. The patent information used in the above patent analysis can be classified into two types. One is a quantitative patent index that logs the number of citations, filing date, number of family patents, and term of rights. The most frequently used patent information among the quantitative indicators is citations. Brem et al. [14] investigated how the existence of a dominant design affects innovative performance, radical innovation, and process innovation. In this study, the influence of dominant design and the importance of standards were confirmed. The high citations of a patent means that it is close to a standard patent and it can also be useful for a standard patent search. The other is qualitative patent information such as claims and descriptions of technology. This information has been used by experts to read and analyze.

Many large companies try to establish a technology and R&D strategy using patent analyses. They have dedicated departments for such analyses, which include patent and technical experts. These departments attempt to gain insights into technology and R&D strategies by conducting patent analyses based on expert knowledge and subjective opinions. However, as the number of global patents increases exponentially, it is becoming difficult to qualitatively analyze all patent documents. Therefore, large companies are developing and applying quantitative patent analysis methods to replace or supplement existing qualitative patent analysis methods.

On the other hand, small businesses, which compete for technological competence rather than capital, are more influenced by the success or failure of their technology and R&D processes. For those companies, it is important to establish an appropriate technology and R&D strategy, as the outcome of the R&D may influence the success or failure of them. For them, therefore, it is necessary to establish

a technology R&D strategy that differs to that of large companies. Brem et al. [15] confirmed that large firms and SMEs have different interests in the patents. In addition, SMEs are more fearful of litigation than large corporations, and the efficiency of patent protection is lower than the amount spent on patent acquisition. However, it is difficult for general SMEs to have a patent department because of cost problems. A solution to this problem could also be a quantitative patent analysis. By developing an IP expert system that allows a computer to automatically perform patent analyses instead of using experts, the time and costs of patent analyses can be greatly reduced [16]. In addition, an automated system allows for continuous analysis using the same analytical criteria that do not change over time and continue to provide objective results. In particular, as the amount of data increases, the time and cost of a qualitative analysis increases, and the accuracy of the analysis results decreases. On the other hand, a quantitative analysis does not significantly increase time and cost, and the accuracy of the analysis increases. For this reason, many studies have actively developed quantitative patent analysis methods to replace existing methods.

Liu et al. [17] proposed a patent retrieval and analysis platform that can achieve a higher search accuracy and significantly reduce the search time using quantitative indices of patents and qualitative indices using text mining. They argue that the proposed PRAP can reduce the effort required by a patent examiner to read more than 35,000 patents in a typical patent infringement search. Chen et al. [18] proposed a topic-based technology prediction approach to identify trends in technology. The case studies of 13,910 patents published in Australia between 2000 and 2014 proved that their proposed technology forecasting method was effective in predicting future trends on potential topics. Many studies have developed methodologies to quantitatively analyze the contents (title, abstract, assertion, explanation, etc.) of patents so that experts can read and analyze these studies using text mining. This study uses patent keywords extracted using text mining as its most important feature. Therefore, the validity of the keyword extraction has a great influence on the research results. If the keywords extracted during preprocessing are inappropriate, the analysis results cannot be trusted [19]. Although patent keywords have an important influence on the analysis, most studies assume that the keywords are well selected without validating the results [4]. In this study, we propose a new method for keyword extraction and verification in order to solve these problems. We intend to secure the reliability of the results of patent analyses using text mining. Also, unlike the conventional method using the patent portfolio, the proposed method can extract keywords from a single patent. In the existing method, although the same patent, the extracted keyword could be changed depending on which portfolio the patent is included in. However, since the keywords extracted from a single patent do not change, the reproducibility and consistency of the patent analysis can be improved.

2.2. Keyword Extraction

When conducting a patent analysis using text mining, we do not use every word in the patent. Since the use of meaningless words is inefficient and reduces the accuracy of the analysis results, only meaningful words are extracted and used in analyses. The extraction of patent keywords is the most important and necessary task in a patent analysis. Although many studies on keyword extraction from documents have been conducted, there have been few studies on patent keyword extraction that verify the results. Therefore, this study proposes a new method to validate and improve the appropriateness of patent keyword extraction.

Currently, the criteria for extracting patent keywords are based on how well they can be classified using extracted keywords to classify documents within a patent group to be analyzed. Xie and Miyazaki [20] emphasized the importance of the keyword search as a method to efficiently manage patents in a patent system. They suggest using recall and precision to effectively select keywords from patent components that consist of titles, abstracts, claims, and content, and propose a method to evaluate the suitability of keywords based on type-1 and type-2 errors. Noh et al. [4] examine using an orthogonal array for the keyword selection strategy considering four factors: the keyword selection algorithm, the number of keywords, usage type, and patent component.

They found the optimal combination to best classify patents according to the International Patent Classification (IPC).

The common feature of the two studies is that the method compares the words of each document using the group of patents collected to extract the keywords. Evaluating the validity of the extracted keywords is based on how well the patents in the group are classified. However, a patent analysis that does not have a limited range of analysis, such as a search for patent identification or a prior art search, requires a keyword that can represent a unique characteristic of a single document rather than a feature that can classify documents within a group. It is clear that the words that best describe the characteristics of a particular document are those selected by the authors of the document, and a paper is a representative document containing the keywords chosen by its authors. This study proposes a new method using relevant papers and their keywords to extract those keywords that provide the unique characteristics of a single patent.

2.3. Papers vs. Patents

A patent does not include its inventor's keywords, but the nature of its text, such as the format, requirements, and composition, is very similar to a paper. First, a paper is an article that argues the author's claim on a problem and is written in a form that proves the justification of his claim to it. This feature is similar to that of a patent. Second, the most important requirement of a paper is its originality. The originality of a paper that is to deny the results of previous studies and to give new results or to provide better research results corresponds to the novelty and the inventiveness of a patent, which is an important requirement of a patent. Oftentimes new ideas appear simultaneously with papers and patents [21]. Finally, the components of a paper (title, abstract, body) are the same as those of a patent. Therefore, this study suggests a keyword extraction method that provides the unique characteristics of a single patent using relevant papers and their authors' keywords. In order to prove the validity of this research hypothesis, that patent keywords can be extracted from a paper, a document similarity analysis between the patents and the papers collected through the same query is performed.

2.4. Document Similarity

In order to evaluate the validity of the proposed method, we use the cosine similarity to measure the similarity between two documents. The cosine similarity is a measure of similarity between vectors, measured using the cosine values of angles between two vectors of an inner space. In particular, it is useful when measuring the similarity of two documents in the field of text mining, where each word constitutes a dimension and the documents are represented by the frequency in which each word appears in the document [22]. Moehrle [23] found an appropriate combination of several similarity coefficients, including cosine similarity coefficients, to measure the textual similarities of patents for patent management, such as prior art analysis and infringement analysis. Shibata et al. [24] proposed a method using patent and dissertation similarities for finding potential technologies that have been developed through academic research, but that have not yet been patented. In this study, we verify the validity of our approach by calculating the similarity of patents and papers collected through the same query.

3. Patent Keyword Extraction for Sustainable Technology Management

3.1. Quantitative Analysis for Sustainable Technology Management

Recently, Sustainable growth and development has become an important issue for governments and corporations. But maintaining sustainable development is very difficult. These difficulties can be attributed to sociocultural and political backgrounds that change over time [1]. Because of these changes, the technologies for sustainability also change, so governments and companies try to predict and manage technology using patent analyses, but it is very difficult to predict the rapidly changing

technology markets. The best way to achieve insight into technology management in this rapidly changing market is to build a technology management direction and strategy that is flexible and adaptable to the volatile market environment through continuous monitoring and analysis. However, the qualitative patent analyses used so far have been one-off or short-term, owing to time and cost problems. In addition, it is almost impossible to apply the same criterion to all analysis points because they use the subjective analyses of experts. To solve this problem, many researchers have actively studied quantitative patent analysis methods [25–27]. Quantitative approaches can dramatically reduce the time and cost, and can provide objective analysis results using the same criteria at all points. Therefore, quantitative patent analyses are essential for sustainable technology management, and various methods should be developed according to the purpose of analysis. Recently, research on patent analysis methods using text mining have been actively carried out. In this study, we propose a method of extracting patent keywords to secure the validity and reliability of future studies on patent analyses using text mining.

3.2. Proposed Methodology for Keyword Extraction

The results of patent analysis using text mining are dependent on the keywords that characterize each document. The general criterion for extracting patent keywords is based on how well the chosen keywords can classify patent documents within a group of patents collected for specific analytical purposes. However, while such keywords are useful when analyzing data sets that have a range of analysis, such as a patent group, they cannot be used when the range of analysis is not fixed, such as prior art research. In order to solve this problem, this study proposes a method using author keywords and the paper itself to extract keywords that best represent the characteristics of a single patent. The proposed analysis procedure is shown in Figure 1.

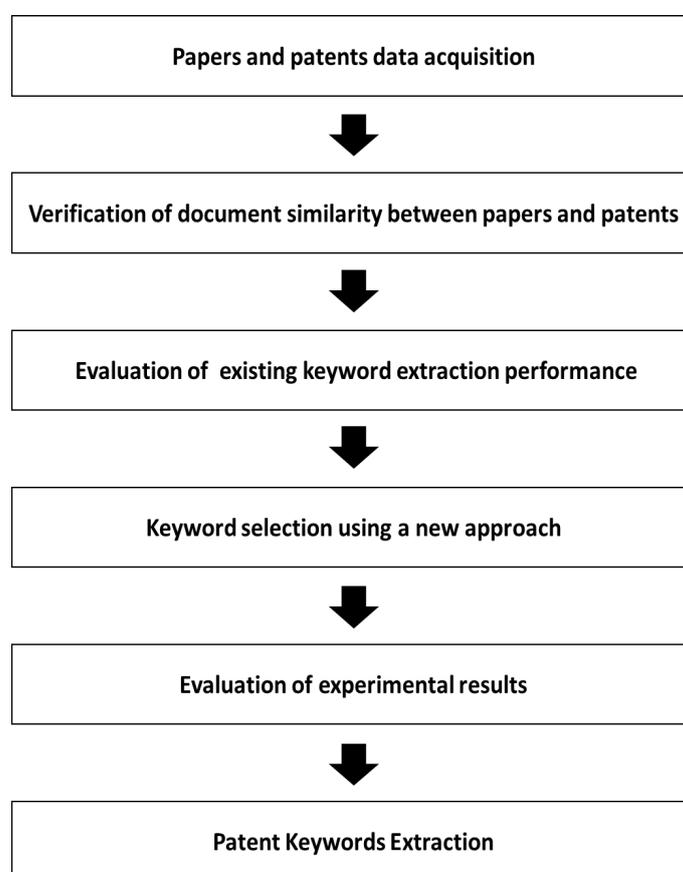


Figure 1. Procedure of proposed method.

First, analytical data are collected using the same query in the paper and patent databases. Replication and noise data are removed and the analysis accuracy is improved by selecting only valid data. Next, we extract a summary, which is a common component of both patents and papers, and then we structure the summary as a document-term matrix (DTM) using text mining [28].

Next, in order to confirm whether it is reasonable to propose a patent keyword selection method using a paper, we calculate the similarity of the DTMs between the paper and the patent using the following cosine distance [29,30].

$$\text{Cosine similarity} = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}. \quad (1)$$

In order to prove the similarity between the paper and patent, we set the analytical data as groups of paper–paper pairs, patent–patent pairs, and paper–patent pairs, and calculate the similarities of the groups. Then, an analysis of variance (ANOVA) is performed to statistically test if there are any significant differences between the three groups.

After validating the proposed method through the above test, we evaluate the performance of the TF-IDF method for patent keyword extraction, which is currently the most widely used and appropriate method for patent classification [4]. The performance of the existing method is evaluated using two indicators, precision and accuracy, and the results are compared using the proposed method.

$$\text{Precision} = \frac{\text{Number of matching keywords}}{\text{Number of extracted keywords}}. \quad (2)$$

Precision is the ratio of the actual keywords among the keyword candidates extracted by each method, and accuracy is the ratio of the extracted keyword candidates among actual keywords.

$$\text{Accuracy} = \frac{\text{Number of matching keywords}}{\text{Number of actual keywords}}. \quad (3)$$

The proposed keyword extraction method is based on the abstract and the main body of the text. However, because the body is generally too long and contains much content that is not central to the document, it is better to use the introduction and conclusion of the body of a paper. That is, we extract words appearing simultaneously in three parts of the paper: the abstract, introduction, and conclusion. Then, we calculate the precision and accuracy using these extracted words and the author keywords. Using these indicators, we evaluate the keyword extraction performance and then improve the performance using additional frequency analyses. The reliability of this study is ensured using statistical verification experiments on each analysis result using TF-IDF and simultaneous word and frequency analyses. Finally, using experiments, we determine the optimal patent keyword extraction method and conduct a case study to confirm the result.

4. Experimental Results and Analysis

4.1. Papers and Patents Data Acquisition

In this study, we propose a method of extracting patent keywords using relevant papers and their authors' keywords and the paper. However, it is necessary to objectively verify whether it is appropriate to extract patent keywords from a paper. Therefore, we verify the validity of this study by calculating the similarity between the papers and the patents we collected by an experiment. We collect papers and patents containing the same search terms in their titles from their respective databases, and then compare the similarities between papers and papers, patents and patents, and patents and papers. For this experiment, we searched for patents containing "text mining" in the title, and collected 30 US patents filed between 2000 and 2015 from the patent database WIPSON [31]. Then, we collected

30 papers from Google Scholar containing “text mining” in the titles, also published between 2000 and 2015 [32]. The results are shown in Table 1.

Table 1. Papers and patents acquisition.

Analysis Data	Database	Search Words	Period
Patents (30)	WIPSON	“Text mining”	1 January 200–31 December 2015
Papers (30)	Google Scholar	“Text mining”	1 January 200–31 December 2015

4.2. Verification of Document Similarity Between Papers and Patents

In order to prove the similarity between patents and papers, we divide the collected data into a patent data group (30), paper data group (30), and patent–paper data group (60), and then calculate the similarity within each group. The descriptive statistics of the calculated document similarity are shown in Table 2.

Table 2. Descriptive statistics of document similarity.

Group	N	Mean	SD	SE
Patents	30	0.6473	0.0694	0.0127
Papers	30	0.6179	0.0550	0.0100
Papers and Patents	60	0.6450	0.0636	0.0082
Total	120	0.6388	0.0638	0.0058

The average similarity of the 30 documents of the patent group is 0.6473, and the standard deviation and the standard error are 0.0694 and 0.0127, respectively. The average similarity of the 30 documents in the paper group is 0.6179, and the standard deviation and the standard error are 0.0550 and 0.0100, respectively. The average similarity of the 60 documents is 0.6450, and the standard deviation and the standard error are 0.0636 and 0.0082, respectively. The average similarity of 120 documents is 0.6388, and the standard deviation and the standard error are 0.0638 and 0.0058, respectively. The statistics of Table 2 show that the similarity of documents within each group is high, and the differences in similarity between the groups is not large. For the next step, we perform an analysis of variance (ANOVA) to test statistically whether there are any significant differences in terms of document similarity between the groups. Table 3 shows the results of the ANOVA tests.

Table 3. A test of the difference between document groups.

	SS	DF	MS	F	P
Between group	0.018	2	0.009	2.201	0.115
Within group	0.466	117	0.004		
Total	0.484	119			

The *F*-value of the ANOVA test is 2.201 and the *p*-value is 0.115. In other words, when collecting patents and papers using the search term “text mining” in the title, there is no significant difference in document similarity between patents and papers. Therefore, the experimental results verify the validity of the proposed method of extracting keywords using authors’ keywords and the paper.

4.3. Performance Evaluation of Existing Method of Keyword Extraction

In general, the words that can clearly distinguish different documents or classify documents of the same nature are extracted as keywords. The TF-IDF weight is widely used and its performance has been verified through research [4]. Noh et al. [4] found out that using the TF-IDF in patent abstracts is the best way to extract patent keywords through an experiment based on patent clustering and

the International Patent Classification (IPC). In this study, we propose a new approach to extract keywords that can be used for patent analyses that do not classify documents in a given data set. The keywords that can represent the unique characteristics of each patent can be extracted from a single patent document.

As described above, the keywords in each paper chosen by its authors can be regarded as the words that best represent the unique characteristics of each document. Therefore, we conduct an experiment using author keywords to confirm the performance of the method using TF-IDF weights (see Table 4).

Table 4. Verification of keyword extraction using TF-IDF.

No.	Author Keywords	The Number of Keywords	Precision (Accuracy)	TF-IDF Threshold
1	Data; Mapping; Analysis; Clustering; Techniques; Competitive; Intelligence; Assigned; Classifications; Results; Validation; Linguistic; Technology; Packaging; Patent	15	0.133	0.0760
2	Data; Visualization tools; Competitive; Intelligence; Property; Review	6	0	0.0995
3	New; Business; Areas; Technological; Strength; Patent; Information; Data; Envelopment; Analysis; Text; Mining	12	0.167	0.0608
4	Text; Mining; Feature; Extraction; Categorization; Clustering; Customer; Relationship; Management	9	0.111	0.0944
5	Text; Mining; Taxonomy; Construction; Term; Extraction	6	0	0.0680
6	Text; Mining; Natural; Language; Processing; Information; Extraction; Summarization; Image; Question; Answering; Literature-Based; Discovery; Evaluation; User; Orientation	16	0.125	0.0843
7	M&A; Target; Selection; Technology; Acquisition; Patent; Analysis; Subject-Action-Object; Similarity	9	0.222	0.0796
8	Chance; Discovery; Text; Mining; Patent; Analysis; Significant-Rare	7	0.143	0.1072
9	Text; Mining; Hong; Kong; Hotels; Competitor; Intelligence; Marketing	8	0	0.0944
10	Design; Rationale; Representation; Discovery; Text; Mining; Patent	7	0	0.0860
11	Patent; Analysis; Competitor; Company; Ranking; Social; Network	7	0.143	0.0562
12	Conjoint; Analysis; Hybrid; Approach; Morphology; Patent; Information; Technology; Forecasting	9	0.222	0.0550
13	Technological; Opportunity; Discovery; Morphology; Analysis; Text; Mining	7	0.143	0.0771
14	Patent; Analysis; Text; Mining; Visualization; Techniques; Natural; Language; Processing	9	0	0.0718
15	Patent; Content; Representation; Retrieval; Extraction; Paraphrasing; Summarization; Visualization; Navigation; Valuing; PATExpert; Classification; Translation; Documentation; Ontologies; Knowledge; Base	17	0.176	0.0673
16	Monitoring; Technology; Intelligence; Patent; Analysis; Formal; Concept; Dynamic; Lattice	9	0.222	0.0629
17	Knowledge; Discovery; Text; Mining; Patent; Databases; Linguistic; Preprocessing; Correspondence; Analysis; Cluster	11	0.182	0.0832
18	Information; Retrieval; Text; Mining; Performance; Medical; Documentation	7	0	0.0608
19	Patent; Analysis; Knowledge; Discovery; Information; Visualization; Self; Organizing; Map; Citation; Networks; Nanoscale; Science; Engineering; Nanotechnology; Technological; Innovation; International; Interactions	19	0.158	0.0393
20	Text; Mining; Knowledge; Discovery; Post; Project; Reviews; Manufacturing; Construction	9	0.222	0.0697
21	Text; Mining; Data; Visualization; Tools; Patent; Information; Intellectual; Property; Analysis; Landscape; Business	12	0.083	0.0665
22	Patent; Mining; Retrieval; Vector; Space; Model	6	0.167	0.0858
23	Chemical; Named; Entity; Recognition; Conditional; Random; Fields; Text; Mining	9	0.222	0.0912
24	Text; Mining; Word; Distribution; Zipf's; Law; STN; AnaVist; Thomson; Aureka; OmniViz; Stopwords; Patent; Mapping	14	0.071	0.0790

Table 4. Cont.

No.	Author Keywords	The Number of Keywords	Precision (Accuracy)	TF-IDF Threshold
25	Business; Intelligence; Competitive; Advantage; Data; Mining; Information; Systems; Knowledge; Discovery	10	0	0.0626
26	Open; Source; Text; Information; Mining; Analysis; Multilinguality; Automated; Media; Monitoring	10	0.200	0.1086
27	Technical; Intelligence; Bibliometrics; Foresight; Management; Rapid; Analyses; Mining; Text; Knowledge; Discovery; Databases	12	0	0.1216
28	Text; Mining; Theory; Application	4	0	0.1193
29	Text; Mining; Summarization; Feature; Extraction; Patent; Classification; Clustering	8	0.125	0.0626
30	Text; Mining; Information; Retrieval	4	0	0.1384
	Average	9.5	0.108	0.0810

For performance verification experiments, we extract the abstract of each document to generate a document-term matrix (DTM) and calculate the TF-IDF weight of each word. Next, we set the rankings based on the weights of the TF-IDFs, and select the number of author keywords in each paper. As a result, as shown in Table 4, the average number of author keywords in papers was 9.6, with an average precision of 0.108, and mean TF-IDF threshold of 0.0810. The average precision was 0.108, which is very low. This result shows that the keywords extracted using the existing keyword extraction method work well for document classification but are not suitable for determining the unique features of each document.

4.4. Statistical Exploration of a New Method for Keyword Extraction

In order to improve existing keyword extraction performance, we propose a method to extract keywords using the frequency and simultaneous appearance of words, instead of using the TF-IDF. To do so, we conducted experiments using the introduction and conclusion of the text, as well as the main content of the document and the abstract. Before conducting these experiments, we compared the performance differences for the abstract, introduction, and conclusion to verify that it was appropriate to use these parts of the paper. The results are shown in Tables 5–7.

Table 5. Descriptive statistics of performance by component.

Group	N	Mean	SD	SE
Abstract	30	0.1079	0.0864	0.0158
Introduction	30	0.2318	0.1376	0.0251
Conclusion	30	0.1980	0.1554	0.0284
Total	120	0.1792	0.1387	0.0146

In Table 5, the average precision was 0.2318 and 0.1980 when keywords were extracted from the introduction and conclusion. This shows that the performance is better than the keyword extraction from the abstract. An ANOVA was performed to test whether these results were statistically significant.

Table 6. A test of the difference between components of paper.

	SS	DF	MS	F	P
Between group	0.246	2	0.123	7.303	0.001
Within group	1.466	87	0.017		
Total	1.712	89			

Table 7. Multiple comparison using Dunnett T3.

(I) Component	(J) Component	Mean Difference	SE	P
Abstract	Introduction	−0.1239	0.0297	0.000
	Conclusion	−0.0901	0.0325	0.024
Introduction	Abstract	0.1239	0.0297	0.000
	Conclusion	0.0338	0.0379	0.754
Conclusion	Introduction	0.0901	0.0325	0.024
	Abstract	−0.0338	0.0379	0.754

In the ANOVA table in Table 6, the F -value is 7.303 and the p -value is 0.001, indicating that there is a difference in precision according to the components of the paper. Table 7 shows the post-test results using Dunnett T3.

Table 7 shows that the precision of the keywords extracted from the introduction and conclusion is higher than that of the keywords extracted from the abstract. This result is contrary to those of previous studies that find that keyword extraction performance is best when extracting keywords from an abstract [5]. Based on the results of this experiment, we extract keywords from the abstract, introduction, and conclusion.

We select words that appear simultaneously in the abstract, introduction, and conclusion as keywords of the document. Then, the results were compared against the author keywords. Table 8 shows the results of the keyword extraction using a co-word analysis.

Table 8. Keyword extraction using a co-word analysis.

No.	Author Keywords	Simultaneous Appearance Keywords in Three Components			Simultaneous Appearance Keywords in Two Components		
		Keywords	Precision	Accuracy	Keywords	Precision	Accuracy
1	15	8	0.500	0.267	18	0.222	0.267
2	6	15	0.200	0.333	15	0.000	0.000
3	12	19	0.579	0.917	27	0.000	0.000
4	9	7	0.286	0.222	18	0.111	0.222
5	6	21	0.143	0.500	26	0.077	0.333
6	16	5	0.400	0.125	24	0.167	0.250
7	9	20	0.150	0.333	33	0.121	0.444
8	7	2	1.000	0.286	18	0.167	0.429
9	7	9	0.556	0.714	196	0.000	0.000
10	7	13	0.231	0.429	18	0.111	0.286
11	7	23	0.261	0.857	18	0.000	0.000
12	9	29	0.241	0.778	57	0.035	0.222
13	7	15	0.400	0.857	30	0.033	0.143
14	9	11	0.273	0.333	19	0.158	0.333
15	16	10	0.400	0.250	19	0.105	0.125
16	9	28	0.286	0.889	37	0.000	0.000
17	11	6	0.500	0.273	19	0.105	0.182
18	7	5	0.800	0.571	18	0.056	0.143
19	19	19	0.526	0.526	39	0.051	0.105
20	9	20	0.400	0.889	32	0.000	0.000
21	12	13	0.615	0.667	23	0.043	0.083
22	6	11	0.455	0.833	18	0.000	0.000
23	9	12	0.500	0.667	18	0.111	0.222
24	14	12	0.333	0.286	20	0.000	0.000
25	10	11	0.182	0.200	24	0.292	0.700
26	10	8	0.250	0.200	14	0.429	0.600

Table 8. Cont.

No.	Author Keywords	Simultaneous Appearance Keywords in Three Components			Simultaneous Appearance Keywords in Two Components		
		Keywords	Precision	Accuracy	Keywords	Precision	Accuracy
27	12	9	0.444	0.333	59	0.068	0.333
28	4	5	0.400	0.500	11	0.090	0.250
29	8	15	0.333	0.625	27	0.000	0.000
30	4	16	0.063	0.250	20	0.050	0.250
Average	9.5	13.2	0.388	0.497	24.6	0.087	0.197

The analysis evaluates the performance of the extracted keywords using words appearing in two parts and three parts of the paper. The average number of words appearing at the same time in two parts was 24.6, which is not as good as the conventional TF-IDF weighting method. However, the average number of the words appearing at the same time in three parts was 13.2, with a precision of 0.388 and accuracy of 0.497, which is significantly better than the conventional TF-IDF weighting method. In addition, if the existing method does not know the actual number of keywords, they have to be determined subjectively. The advantage of the proposed method is that the number of keywords can be determined objectively using the simultaneous occurrence word.

Table 9 shows the result of the keyword extraction using a frequency analysis and its performance.

Table 9. Keyword extraction using frequency analysis.

No	Author Keywords	Keywords	Accuracy	Precision
1	15	21	0.533	0.381
2	6	7	0.833	0.714
3	12	12	0.500	0.500
4	9	12	0.333	0.250
5	6	7	0.667	0.571
6	16	17	0.250	0.235
7	9	9	0.333	0.333
8	7	9	0.571	0.444
9	7	8	0.286	0.250
10	7	7	0.143	0.143
11	7	9	0.429	0.333
12	9	9	0.444	0.444
13	7	7	0.286	0.286
14	9	11	0.444	0.364
15	16	18	0.313	0.278
16	9	10	0.778	0.700
17	11	17	0.455	0.294
18	7	12	0.714	0.417
19	19	21	0.421	0.391
20	9	11	0.444	0.364
21	12	14	0.583	0.500
22	6	6	0.333	0.333
23	9	9	0.222	0.222
24	14	20	0.286	0.200
25	10	10	0.400	0.400
26	10	16	0.300	0.188
27	12	15	0.333	0.267
28	4	6	0.500	0.333
29	8	8	0.375	0.375
30	4	4	0.250	0.250
Average	9.5	11.4	0.425	0.359

The average number of keywords extracted by the frequency analysis was 11.4, with precision of 0.359, and accuracy of 0.425. We extracted keywords using the TF-IDF, co-word analysis, and frequency analysis to search for optimal keyword extraction methods, and evaluated the performance by comparing these against the actual keywords. Table 10 shows a summary of the experimental results obtained by the three methods.

Table 10. Performance comparison of three methods.

	TF-IDF	Co-Word Analysis	Frequency Analysis
Number of keywords	9.5	13.2	11.4
Precision (%)	10.8	38.8	35.9
Accuracy (%)	10.8	49.7	42.5

The results show that the co-word analysis has the highest precision of 38.88 and, unlike the other two methods, the number of keywords can be extracted objectively. Therefore, the experiments show that this method performs best when extracting unique keywords from a single patent. However, the number of keywords extracted through co-word analysis is larger than the number of actual keywords. This means that words extracted using the co-word analysis contain relatively more noise. Then, we applied a frequency analysis to improve the keyword extraction performance and to minimize the noise. First, words were selected based on an analysis of the co-words. In order to remove the noise among the selected words, we rank the frequency of the words. Then, we extract the keywords with the highest frequency, up to the number of words selected through the co-words analysis, and the remaining words are removed as noise.

Table 11 shows the result of the keyword extraction using the combination of the co-word and frequency analysis proposed in this study.

Table 11. Keywords extraction using co-words and frequency.

No.	Keywords	Precision	Accuracy
1	8	0.500	0.267
2	8	0.375	0.500
3	14	0.786	0.917
4	5	0.400	0.222
5	13	0.231	0.500
6	3	0.667	0.125
7	10	0.300	0.333
8	2	1.000	0.286
9	5	0.800	0.571
10	6	0.167	0.143
11	17	0.353	0.857
12	12	0.583	0.778
13	7	0.571	0.571
14	5	0.600	0.333
15	8	0.500	0.250
16	17	0.471	0.889
17	5	0.600	0.273
18	4	1.000	0.571
19	12	0.667	0.421
20	9	0.778	0.778
21	8	0.875	0.583
22	7	0.571	0.667
23	5	0.600	0.333
24	7	0.571	0.286

Table 11. Cont.

No.	Keywords	Precision	Accuracy
25	3	0.667	0.200
26	5	0.400	0.200
27	5	0.800	0.333
28	4	0.500	0.500
29	10	0.400	0.500
30	7	0.143	0.250
Average	7.7	0.562	0.448

The experimental results show that the average number of keywords extracted is 7.7, with a precision of 56.2%, and accuracy of 44.8%. Table 12 shows the results compared with the existing experimental results.

Table 12. Performance comparison of four methods.

	TF-IDF	Co-word Analysis	Frequency Analysis	Combine Analysis
Number of keywords	9.5	13.2	11.4	7.7
Precision (%)	10.8	38.8	35.9	56.2
Accuracy (%)	10.8	49.7	42.5	44.8

As a result of removing noise using a frequency analysis, the average number of keywords that were extracted is reduced to 7.7. In addition, the precision (the probability of actual keywords) increased significantly from 38.8% to 56.2%. However, in the noise removal process, it was observed that the average accuracy dropped from 49.7% to 44.8% due to actual keywords being misclassified as noise. The precision increased by 17.4% and the accuracy decreased by 3.9%, suggesting that the keyword extraction performance improved. We have statistically verified that these experimental results are meaningful (see Table 13).

Table 13. T-test between co-words analysis and combined analysis.

	Difference of Mean	T	P-Value
Precision	−0.174	−3.203	0.002
Accuracy	0.051	0.825	0.413

Significance level: 0.05.

Table 13 shows that the precision T-statistic is -3.203 and the p -value is 0.002 at the 5% level of significance. There is a statistically significant difference between the two methods. That is, the combined analysis shows that precision has improved. On the other hand, the T-statistic of the accuracy is 0.825 and the p -value is 0.413, and are not statistically significant. That is, we cannot say that the accuracy is reduced by the combined analysis. Therefore, we confirm that the combination of a co-word analysis and frequency analysis is an optimal method for extracting unique keywords from a single document.

4.5. Patent Keywords Extraction

Finally, a case study is conducted using the keyword extraction method proposed in this paper. We extracted the keywords of the registered patent, “Category and term polarity mutual annotation for aspect-based sentiment analysis” in the United States [33]. This patent relates to a technique for classifying an aspect into a predefined category by performing an aspect-based sentiment analysis from the text for opinion mining. In the co-word analysis step, 14 co-words (analysis, aspect, classification, differ, express, feature, negative, opinion, mining, positive, product, sentiment, service,

text) were extracted. First, in the frequency analysis step, 15 words representing the number of high-frequency words, which is the number of co-words, are extracted to remove the noise. Next, in the frequency analysis step, 15 high-frequency words (same number as the number of co-words) are extracted to remove noise. As a result, a total of eight patent keywords (aspect, classification, express, feature, opinion, mining, sentiment, text) were extracted. Considering the content of the patent, the extracted keywords represent the unique characteristics of the patent. The patent keywords extracted through this study can be used for various patent analyses as a patent index. We could not extract keywords from a single patent document through the existing patent keyword extraction method. The conventional method is to extract keywords by comparing words between patents within a patent portfolio. There is a serious problem with this method. In the existing method, although the same patent, the extracted keyword could be changed depending on which portfolio the patent is included in. However, since the keywords extracted from a single patent do not change, the reproducibility and over-fitting problems of patent analysis can be improved.

5. Conclusions

Governments and corporations are striving to gain a competitive advantage through new technology development. However, these efforts result in significant losses if they fail. Therefore, technology and R&D strategies that minimize the risk of failure are essential elements of new technology development. We utilize patent analyses to establish such a strategy. Patents are objective data that contain information about the technology and now a vast number of patents are stored in the database, providing insights for technology management using patents. Therefore, research on patent analysis methodologies is being actively carried out. In particular, research focuses on the quantitative analysis of unstructured data, including patents, by applying text mining. Patent analysis using text mining is based on words and are preceded by noise word removal and keyword extraction. The extracted keywords have a significant influence on the subsequent analysis. Therefore, the results of the analysis cannot be relied on without verifying the results of the keyword extraction. However, most prior studies do not verify the selected keywords [3]. In addition, although the criteria for extracting keywords may vary depending on the purpose of the analysis, most studies use keywords to classify the collected patents by their technical characteristics. The method using TF-IDF is the most used [4–7]. However, keywords based on this criterion may be useful for conducting patent analyses with a limited range of analysis, but are not suitable for analyses without limitations on the range. Therefore, it is necessary to study the keyword extraction and performance evaluation methods that can show the unique characteristics of a single patent document for patent analysis that does not limit the range of the analysis. In this study, we proposed a method of extracting unique keywords from a single patent using a paper. The best keywords are those chosen by the author(s), as specified in the paper. We proposed a new method using a paper to extract keywords with unique characteristics of a single patent. In order to secure the validity of the proposed method, we divided the collected data into a patent data group (30), paper data group (30), and patent–paper data group (60), and calculated the document similarity within each group. The document similarities of the three groups was 0.6473, 0.6179, and 0.6450, respectively. There were no statistically significant differences between the three groups. Therefore, we verified the validity of the method of extracting keywords using the author keywords and paper using an experiment. Next, we evaluated the performance of the existing method using TF-IDF weights and the abstract, which is known as the best way to extract patent keywords. As a result, the average TF-IDF threshold was 0.0810 and the precision was 0.108. In other words, the probability that the keywords extracted using the TF-IDF weight matches the actual author keywords is 10.8%, which is very low. This result shows that the keywords extracted by the classification performance between the collected documents are not suitable for representing the unique characteristics of each document. In this study, we proposed a keyword selection method using a frequency analysis and a co-word analysis to improve and verify the performance. Experiments were carried out using the introduction and conclusion of the text, which contains the main content of

the document, as well as the abstracts, which are generally used for keyword extraction, in order to perform a co-word analysis in a single document. We used precision and accuracy as two indicators to evaluate performance. Precision is the ratio of the actual keywords among the keyword candidates extracted by each method, and accuracy is the ratio of the extracted keyword candidates among the actual keywords. The experimental results showed that the co-word analysis has the best performance, with an average keyword count of 13.2, precision of 38.8%, and an accuracy of 49.7%. However, since the average number of actual keywords is 9.5, noise is included. We performed a noise-reduction process using a frequency analysis to improve the performance of the analysis. Finally, we performed a combined analysis using a co-word analysis and a frequency analysis. The average keyword count was 7.7, with a precision of 56.2%, and accuracy of 44.8%, which improved the performance of keyword extraction. It is also expected that other keywords that cannot be explained quantitatively, but are not consistent with the author keywords, can be used for keyword-based analysis as meaningful words to explain the characteristics of the paper. Finally, a case study was conducted using the keyword extraction method proposed in this paper. We extracted the keywords of the registered patent, "Category and term polarity mutual annotation for aspect-based sentiment analysis" in the United States [27]. The extracted keywords were aspect, classification, express, feature, opinion, mining, sentiment, and text. The patent keywords extracted through this study can be used for various patent analysis as a patent index. In particular, it is expected to be very useful for word-based patent analyses.

Recently, sustainable growth and development has become an important issue for governments and corporations. However, maintaining sustainable development is very difficult. These difficulties can be attributed to sociocultural and political backgrounds that change over time [1]. Because of these changes, the technologies for sustainability also change, so governments and Companies are trying to predict and manage technology using patent analyses, but it is very difficult to predict the rapidly changing technology markets. The best way to achieve insight into technology management in this rapidly changing market is to build a technology management direction and strategy that is flexible and adaptable to the volatile market environment through continuous monitoring and analysis. Quantitative patent analysis using text mining is an effective method for sustainable technology management. In this study, we set up a research hypothesis to extract the unique keywords of a patent, and examined the validity of the optimal keyword extraction method using a statistical test procedure. First, by comparing the document similarity between papers and patents containing the same search terms in their titles, we verified the validity of the proposed method of extracting patent keywords using author keywords and the paper. As a result of this experiment, we have proved document similarity between a paper and a patent, which will serve as a basis for future research on the paper and the patent. Second, we have recognized the necessity of patent keyword extraction and performance verification according to current research trends, in which word-based patent analyses using text mining is actively conducted. Although patent keywords have an important influence on the analysis, most studies assume that the keywords are well selected, without validating the results [3]. Therefore, we proposed a new method for extracting and verifying keywords and secured the reliability of patent analysis results based on text mining. Third, the common feature of recent studies on keyword extraction was that the method compares the words of each document using the group of patents collected to extract the keywords. Evaluating the validity of the extracted keywords is based on how well the patents in the group are classified [20]. In the existing method, although the same patent, the extracted keyword could be changed depending on which portfolio the patent is included in. However, since the keywords extracted from a single patent do not change, the reproducibility and over-fitting problems of patent analysis can be improved. The proposed method extracts patent keywords representing the unique characteristics of a single patent, unlike the conventional methods based on the classification performance of a patent group. We also confirmed that the proposed method improves the precision by about 17.4% over the existing method. The results of this study can be used as a basic study of word-based patent analysis research, and it is expected that it will be helpful to develop the research in various directions, as well as to verify the validity of the research results.

In addition, we can extract the patent keyword automatically as a new field of the existing patent database using the method proposed in this study, and develop research methodologies based on this to construct a complete patent expert system in the future. There is a need for more research on how the proposed method can be applied to a company. In particular, the benefits that can be gained from patents depend on the size of the firm. We need to make sure that our method fits for big or small firms in future studies. In addition, it is necessary to apply quantitative patent analysis methods to reduce the difference in profits that a small company and a large company can obtain from a patent. In addition, we extracted keywords as a quantitative patent index in a single patent document, but it is not enough to obtain various insights into technology management with only patent keywords. In future work, we will apply artificial intelligence and machine learning algorithms to keyword-based patent analysis. This enables us to perform patent analysis such as emerging technology forecasts, automatic technology classification, and technology valuation, and gain various insights for sustainable technology management.

Acknowledgments: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF, Daejeon, South Korea) funded by the Ministry of Education (NRF-2015R1D1A1A01059742). This research was supported by the BK 21 Plus (Big Data in Manufacturing and Logistics Systems, Korea University, Seoul, South Korea).

Author Contributions: Jongchan Kim designed this research and collected the data set for the experiment. Jaehyun Choi and Dongsik Jang analyzed the data to show the validity of this study. Sangsung Park wrote the paper and performed the entire research steps. In addition, all authors have cooperated with each other in revising the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brent, A.C.; Pretorius, M.W. Sustainable development: A conceptual framework for the technology management field of knowledge and a departure for further research. *South Afr. J. Ind. Eng.* **2008**, *19*, 31–52. [[CrossRef](#)]
2. Lee, S.; Yoon, B.; Park, Y. An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation* **2009**, *29*, 481–497. [[CrossRef](#)]
3. Holger, E. Patent information for strategy technology management. *World Patent Inf.* **2003**, *25*, 233–242.
4. Noh, H.; Jo, Y.; Lee, S. Keyword selection and processing strategy for applying text mining to patent analysis. *Expert Syst. Appl.* **2015**, *42*, 4348–4360. [[CrossRef](#)]
5. Yan, L.; Leuo, W.; Chao, H. Extracting the significant-rare keywords for patent analysis. *Expert Syst. Appl.* **2009**, *36*, 5200–5204.
6. Magerman, T.; Van Looy, B.; Song, X. Exploring the feasibility and accuracy of latent semantic analysis based text mining techniques to detect similarity between patent documents and scientific publications. *Scientometrics* **2010**, *82*, 289–306. [[CrossRef](#)]
7. Chang, L.; Bok, K.; June, S. Novelty-focused patent mapping for technology opportunity analysis. *Technol. Forecast. Soc. Chang.* **2015**, *90*, 355–365.
8. Aviv, S.; Jussi, K.; Chil, J.; Jae, L. Analyzing multilingual knowledge innovation in patents. *Expert Syst. Appl.* **2013**, *40*, 7010–7023.
9. Dorothy, B. Core capabilities and core rigidities: A paradox in managing new product development. *Strateg. Manag. J.* **1992**, *13*, 111–125.
10. Kim, Y.; Suh, J.; Park, S. Visualization of patent analysis for emerging technology. *Expert Syst. Appl.* **2008**, *34*, 1804–1812. [[CrossRef](#)]
11. Grimaldi, M.; Cricelli, L.; Di Giovanni, M.; Rogo, F. The patent portfolio value analysis: A new framework to leverage patent information for strategic technology planning. *Technol. Forecast. Soc. Chang.* **2015**, *94*, 286–302. [[CrossRef](#)]
12. Holger, E.; Legler, S.; Lichtenthaler, U. Determinants of patent value: Insights from a simulation analysis. *Technol. Forecast. Soc. Chang.* **2010**, *77*, 1–19.
13. OuYang, K.; Weng, C.S. A new comprehensive patent analysis approach for new product design in mechanical engineering. *Technol. Forecast. Soc. Chang.* **2011**, *78*, 1183–1199. [[CrossRef](#)]

14. Alexander, B.; Petra, A.N.; Emma, L.H. Open innovation and intellectual property rights. *J. Mech. Des. Manag. Decis.* **2011**, *55*, 1285–1306.
15. Alexander, B.; Petra, A.N.; Gerd, S. Innovation and de facto standardization: The influence of dominant design on innovative performance, radical innovation, and process innovation. *Technovation* **2016**, *50–51*, 79–88.
16. Tseng, Y.; Lin, C.; Lin, Y. Text mining techniques for patent analysis. *Inf. Process. Manag.* **2007**, *43*, 1216–1247. [[CrossRef](#)]
17. Liu, S.; Liao, H.; Pi, S.; Hu, J. Development of a patent retrieval and analysis platform—A hybrid approach. *Expert Syst. Appl.* **2011**, *38*, 7864–7868. [[CrossRef](#)]
18. Chen, H.; Zhang, G.; Zhu, D.; Lu, J. Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. *Technol. Forecast. Soc. Chang.* **2017**, *119*, 39–52. [[CrossRef](#)]
19. Cheong, H.; Chiu, I.; Shu, L.; Stone, R.; McAdams, D. Biologically meaningful keywords for functional terms of the functional basis. *J. Mech. Des.* **2011**, *133*, 1–11. [[CrossRef](#)]
20. Xie, Z.; Miyazaki, K. Evaluating the effectiveness of keyword search strategy for patent identification. *World Patent Inf.* **2013**, *35*, 20–30. [[CrossRef](#)]
21. Fiona, M. Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering. *Res. Policy* **2002**, *31*, 1389–1403.
22. Singhal, A. Modern Information Retrieval: A Brief Overview. *Bull. IEEE Comput. Soc. Tech. Comm. Data Eng.* **2001**, *24*, 35–43.
23. Moehrl, M. Measures for textual patent similarities: A guided way to select appropriate approaches. *Scientometrics* **2010**, *85*, 95–109. [[CrossRef](#)]
24. Shibata, N.; Kajikawa, Y.; Sakata, I. Detecting potential technological fronts by comparing scientific papers and patents. *Foresight* **2011**, *13*, 51–60. [[CrossRef](#)]
25. Kim, S.; Jang, D.; Jun, S.; Park, S. A Novel Forecasting Methodology for Sustainable Management of Defense Technology. *Sustainability* **2015**, *7*, 16720–16736. [[CrossRef](#)]
26. Choi, J.; Jang, D.; Jun, S.; Park, S. A predictive model of technology transfer using patent analysis. *Sustainability* **2015**, *7*, 16175–16195. [[CrossRef](#)]
27. Choi, J.; Jun, S.; Park, S. A patent analysis for sustainable technology management. *Sustainability* **2016**, *8*, 688. [[CrossRef](#)]
28. Kim, J.; Lee, J.; Kim, G.; Park, S.; Jang, D. A hybrid method of analyzing patents for sustainable technology management in humanoid robot industry. *Sustainability* **2016**, *8*, 474. [[CrossRef](#)]
29. Ye, J. Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Math. Comput. Model.* **2011**, *53*, 91–97. [[CrossRef](#)]
30. Gerstenkorn, T.; Manko, J. Correlation of intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **1991**, *44*, 39–43. [[CrossRef](#)]
31. WIPSON. 2017. Available online: <http://www.wipson.com/service/mai/main.wips> (accessed on 11 May 2017).
32. Google. 2017. Available online: <http://scholar.google.co.kr> (accessed on 11 May 2017).
33. Xerox Corporation. Category and Term Polarity Mutual Annotation for Aspect-Based Sentiment Analysis. US Patent No. 9690772, 5 December 2014.

