Article

Introducing Patents with Indirect Connection (PIC) for Establishing Patent Strategies

Juhyun Lee [1], Sangsung Park [2] and Jiho Kang [3,*]

1   Department of Industrial Management Engineering, Korea University, Seoul 02841, Korea; leeju@korea.ac.kr
2   Department of Big Data and Statistics, Cheongju University, Chungbuk 28503, Korea; hanyul@cju.ac.kr
3   Machine Learning Big Data Institute, Korea University, Seoul 02841, Korea
*   Correspondence: kangmae@korea.ac.kr; Tel.: +82-2-3290-3900

Abstract: A patent system requires novelty and progressiveness so that new patents do not infringe on the rights of prior art. Patent investigation including a prior art search is essential to the process of commercialization of technology. In general, patent investigation has been conducted by experts based on their qualitative judgement. However, the number of patents has increased so fast that it has become difficult to handle the quantitative burdens of the search with a conventional approach. There have been previous studies dealing with patent investigation to find similar technologies. They had limitations as they did not utilize the citation relationship and similarity between patents in a comprehensive way. In addition, they could not properly reflect the sequential citation relationship of patents though this is effective in discovering similar patents. In this study, we propose an efficient methodology to discover similar technologies by comprehensively considering the similarity and citation relationship between patents. In particular, we intended to reflect the citation sequence and indirect citation relationship in the process of searching for similar patents. For this, we introduced the concept of "patents with indirect connections" (PICs) and devised an algorithm to efficiently detect patent pairs having such a relationship. The proposed methodology of this study contributes to preventing patent litigation in advance by discovering patents with such potential risks. It is expected that this method will provide patent applicants with the opportunity to establish appropriate strategies against competitors with similar technologies. In order to examine the practical applicability of the proposed method, Korean patents related to machine learning and deep learning were collected. As a result of the experiment, it was possible to identify 24 pairs of similar patents without a direct citation relationship and derive appropriate counter strategies.

Keywords: prior art search; patent infringement prevention; finding similar patents; patent big data; patent strategy; patent litigation; patent network analysis

1. Introduction

Sustainable growth and development are very important goals for companies, but they are difficult to achieve [1]. It is technology that makes these goals possible by giving companies a competitive advantage in the marketplace [2]. According to Benz (2011), technology is created through the application of knowledge and plays an important role in sustainable growth [3–5]. Thus, it is inevitable that there is fierce competition among companies for secure superior technologies to gain competitive power in the market [6–10]. In a highly competitive market environment, there is a need for an institutional device that can safely protect the right to technology created as a result of research and development (R&D). It is a patent system that guarantees applicants the legal rights to a technology. It promotes the development of the industry by letting companies disclose the contents of their technology to the public. As compensation for this, they are guaranteed an exclusive right to implement this technology for a certain period of time.

A patent without novelty is likely to cause legal disputes and social losses. A patent is registered after examination by the patent offices of each country and its rights are

granted to the applicant. In order for a patent to be registered, it requires novelty and progressiveness as well as industrial applicability. Among the requirements for patent registration is novelty, meaning that the rights claimed by an applied patent are sufficiently differentiated without infringing the rights of any prior art. If a patent without novelty infringes on the scope of the rights belonging to prior art, there is a high possibility of conflict between the patent owners. In addition, this could lead to legal litigation, resulting in financial losses for both sides and impeding industrial development. Therefore, prior art investigation is an essential prerequisite for research and development (R&D) and patent application. It plays an important role in preventing such problems in advance.

It is common that companies conduct a prior art search before R&D or patent filing and reflect the results in their management strategy. However, the direction of strategy differs depending on companies' positions and the existence of similar technologies [11]. Companies trying to apply for new patents use this investigation process to prevent potential disputes with prior art owners. If similar prior art is found, they might attempt to invalidate the rights of existing patents or differentiate the claims of new patents from them. On the other hand, the owners of existing patents can also carry out patent investigation to monitor whether any following patents infringe on their scope of rights. If an infringement is occurring, they can file a lawsuit to claim compensation. Another possible alternative is that companies compromise with each other through cross–licensing. Therefore, it is obvious that patent investigation including a prior art search is a very important procedure for allowing applicants to determine the direction of their patent management strategy.

The main purpose of the prior art search is to investigate whether a technology similar to a patent to be applied exists. If there are live patents with an overlapping scope of rights in the market and they cannot be found them in time, it will be difficult to avoid conflicts between their owners. For this reason, there have been a lot of studies dealing with the methodology of prior art search [12–14]. The value of these studies lies in how effectively and efficiently similar technologies can be discovered. Some scholars have proposed a method to search for similar technologies based on the citation relationship of patents [15–17]. These studies have the advantage of being able to effectively find similar patents connected to each other in the patent citation network. However, since direct connection in the citation network does not always guarantee high similarity between patents, we need to expand the scope of the search to include patents with indirect connection. Another group tried to search for prior art based on the similarity of the text in documents such as patents and papers [18–21]. The advantage of this method is that it can quantitatively assess the degree of similarity. However, there are also disadvantages in that it is difficult to limit the scope of the search for prior art and to reflect the changes in terminology used over time. The motivation for conducting this study is the recognition that the above limitations can be improved if there is a methodology of patent investigation that utilizes both the document similarity evaluation using the bibliographic information and citation information of patents. Even if Yaghtin et al. (2019) recognized a significant correlation between the citation information and the degree of similarity between patent documents, there are few studies that have applied both methods to identify core patents and prior art in a comprehensive way. Even when both pieces of information were used to search for similar prior art, the citation sequence or indirect citation relationship could not be reflected.

For the sustainable growth of companies and industries, the methodology of finding prior art and similar technologies should be able to answer the following questions.

1. What patents may pose a potential threat to my organization?
2. Which of our technologies could be involved in lawsuits?
3. What are the prior technologies that can serve as a driving force for competitive advantage when converged with our technology?

The case corresponding to the first question occurs when a patent of a competitor is likely to infringe on the rights of a company's existing intellectual properties. In this case, the risk of potential loss can be eliminated by claiming the legal rights through patent litigation or licensing agreements. The second is the case where it is determined that the

patent to be filed by an organization is similar to the prior art. In this case, it may be necessary to amend the claims or insist on the invalidation of the preceding patent so as not to infringe on the rights of the prior art. The final question is to find a technology that can generate synergy through fusion with the patent to be applied.

In order to effectively and efficiently discover similar technologies at risk of potential legal disputes, this study proposes a methodology for prior art searching with new principles. To be specific, the proposed method detects similar technologies by utilizing both the citation relationship and the similarity between patents. In particular, in order to overcome the limitations of previous research utilizing the citation relationship, we defined a special relationship between patents that may appear in the citation network as "patents with indirect connection (PIC)", which is useful in finding similar technologies and improves the search efficiency. The proposed method takes into account the sequential citation relationship among patents based on PIC. Patents tend to re-cite documents cited by similar prior patents, in order that the sequential citation relationship can be helpful to make it efficient in discovering similar technologies. In this study, the algorithm used to identify similar patents generates a citation network and a similarity network by using the citation and bibliographic information of patents. It also includes the process of integrating the two networks into one numeric matrix, from which we can detect patent groups that are similar to each other and have a high potential for rights conflict. In addition, the result is represented as a visualized network to allow users to easily find the pairs of patents corresponding to PICs. We expect the proposed methodology to provide patent applicants with an opportunity to prepare for potential patent disputes by making it easier to find similar technologies.

The rest of this article is organized as follows. In Section 2, the related works of this study is explained. Section 3 describes the theoretical background for network analysis using patent citation information. In Section 4, the proposed methodology for finding similar technologies is explained in detail. Section 5 deals with an experiment to verify the applicability of the proposed methodology. Section 6 discusses the disadvantages as well as the strengths of the proposed method. Finally, Section 7 proposes future research to improve the shortcomings discussed in the previous section.

2. Related Works

2.1. Studies on Finding Core Patents and Prior Art

When there are lots of patents owned by competitors in a specific industry, it is necessary for companies entering the market to establish a counter strategy to overcome the barriers to entry. Existing companies also need to constantly monitor whether there is a possibility of patent rights conflict with other applicants when filing new patents or implementing existing technologies, which should be reflected in their management strategies. Establishing such strategies requires companies to find core patents and prior art. Patent investigation, which searches for core patents and similar prior art, is a process that must be preceded in establishing a company's technology management strategy. A core patent is not only unique and likely to be used for mass production, but a major target of patent disputes and licensing [22]. Prior art must be investigated to prevent the infringement of rights and to prove the novelty of a new patent. Identifying these kinds of patents based on qualitative analysis is quite time-consuming and costly. In recent years, therefore, research has been widely conducted to effectively search for core patents and prior art from patent data and utilize the results for establishing counter strategies.

Applicants cite prior art to claim the novelty and differentiation of their patents. They also use the family patent system to secure patent rights in several countries. Such information helps to find similar patents and build strategies to respond to them [23–27]. Su et al. (2011) proposed the concept of a patent priority network (PPN) using family patent information, which is applied when searching for valuable patents. They also defined a critical chain and a significant chain to detect the possibility of a patent dispute. Kim et al. (2015) conducted a study to extract core patents by using information such as

citation and family patents. In order to visualize the results, they represented a matrix composed of patent documents and international patent classification (IPC) codes. Yoon & Choi (2012) and Kwon et al. (2018) carried out studies to derive core patents by indexing quantitative information such as the number of forward citations, conducting a matrix analysis based on it. They visualized patents in a two-dimensional matrix and proposed a method of constructing a counter strategy according to the characteristics of the patents in each quadrant. Kang et al. (2017) collected patents in direct citation with target patents to develop the invalidation logic of core patents. The study also proposed a method for selecting candidate patents likely to be used for the invalidation. Furthermore, there have been lots of prior studies which have applied co-citation information into prior art search [28,29]. In particular, Shibata et al. (2008) clarified the concept of inter-citation as well as co-citation to derive insights from the citation relationship between documents. Yaghtin et al. (2019) implied that the existence of a co-citation relationship had a significant correlation with the degree of similarity between patent documents.

Patents contain textual information such as an abstract and claims, as well as various numeric information, both of which can be effectively used to evaluate the degree of similarity between patents and find prior arts corresponding to a target technology [30–33]. The method of prior art search proposed by Chen et al. (2011) improved the search efficiency by using the similarity matrix of documents. They applied text mining techniques to reflect similar words and synonyms when searching for prior art. Dejean et al. (2013) conducted a study to derive prior art candidates by applying an agglomerative hierarchical clustering (AHC) algorithm. Jeong et al. (2017) proposed a method of recommending prior art to be used for invalidation logic development by calculating the similarity between two arbitrary patents based on information entropy and topic modeling.

As a result of reviewing the literature, previous studies used citation relationships and textual information to find core patents and prior art in a domain of interest. Some of them identified a significant correlation between the citation information and the similarity of the document through empirical experiments [29]. However, there was a limitation in that they mostly did not use both sets of information to identify core patents and prior art in a comprehensive way. Although there have been some studies that recommend prior art candidates based on citation relationships and similarity between patents, there is still a problem that the citation sequence of the patents is not considered.

2.2. Development of Counter Strategies

When filing a new patent application, it is necessary to be careful not to infringe on the rights of prior art. In order to avoid a conflict of rights with prior art discovered through investigation, it is required that companies consider the following strategies [34]:

1. Developing non-infringement logic: Discovering loopholes in existing patents' claims.
2. Developing invalidation logic: Prior art searches that could deny the novelty or progressiveness of the claims included in existing patents.
3. Design of circumvention: Alternative technology design to avoid infringing on the rights of existing patents.
4. Cross license: Negotiation through contracts with patent owners where there is potential for patent rights conflict.

We can classify the first two as defensive strategies and the last two as aggressive strategies. Grindley et al. (1997) defined that the defensive strategies are to freely innovate and commercialize technology in a market where competitors possess a lot of prior art [35]. Developing non-infringement and invalidation logics can be used when a lawsuit for the infringement of rights is filed against a later patent. In this situation, defendants may attempt to invalidate the patents owned by the plaintiff by examining patents filed earlier than them. They can also try to claim non-infringement by logically explaining the difference of their invention from that of the plaintiff. The design of circumvention and cross licenses can also be possible alternatives to reduce the risk of conflict. Applicants should make an effort to write a claim with novelty so as not to infringe on the scope of

the rights of prior art. If it is difficult to invent in such a way, it is better to try to cooperate with the holders of prior patents. Lippman & Rumelt (1982) maintained that the aggressive strategies were to prevent their technology from being imitated and to attain a monopolistic advantage in the marketplace [36]. For example, first movers and fast followers might try to monitor new competitors' patent activity in order to protect their own patents and prevent potential losses. According to Arora and Andrea (2003), new companies that lack commercialization capabilities tend to become active negotiators and try licensing with others with relatively good capabilities and more experience [37]. As such, patent strategies can vary depending on the purpose and the size and position of a company [38,39].

3. Backgrounds

A patent is a document to protect the scope of legal rights on a technology. It is required that prior art is cited when filing a patent application so as not to infringe upon their legal rights. In the context that patents with superior technical characteristics are more likely to be cited from other patents, there have been many citation-related studies [40–42]. A citation network analysis of patents is a representative case in this research field. Figure 1 is an example of citation network analysis of patents.
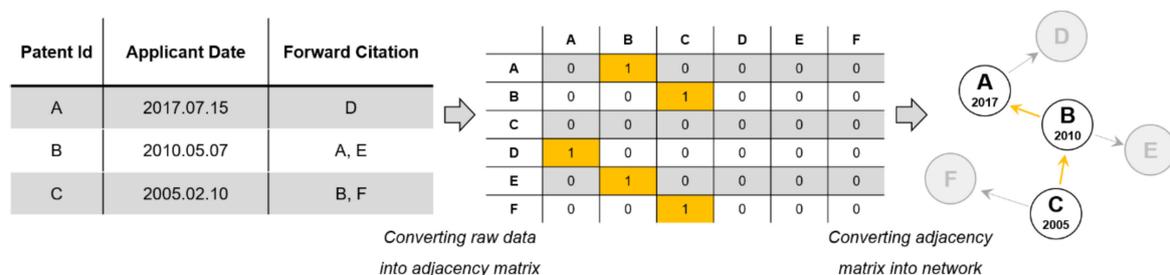

Figure 1. Conceptual diagram of citation patent network.

Citation patent networks help in understanding the trend of technological development. In Figure 1, patent A filed in 2017 is cited by patent D. Patent B, filed in 2010, is cited by patents A and E. Patent C filed in 2005 is cited by patents B and F. This relationship can be expressed as the citation adjacency matrix (CAM).

The patent document can be converted into a vector based on its term frequency. Then, it is possible to evaluate the degree of similarity between two patents. The most representative is the cosine similarity index, which measures how closely the direction of two vectors coincide [43–45]. If A and B are both N-dimensional vectors, the cosine similarity between the two can be obtained by Equation (1):

$$Cosine\ similarity(A,\ B) = \frac{\sum_{i=1}^{N} A_i \times B_i}{\sqrt{\sum_{i=1}^{N} (A_i)^2} \sqrt{\sum_{i=1}^{N} (B_i)^2}} \tag{1}$$

If the two vectors are in exactly the same direction, the value is equal to 1. On the other hand, when the value equals −1 they are in the completely opposite direction. Therefore, the cosine similarity between two documents is calculated as a value between −1 and 1. Figure 2 shows the process of creating a similarity network using the similarity measure.
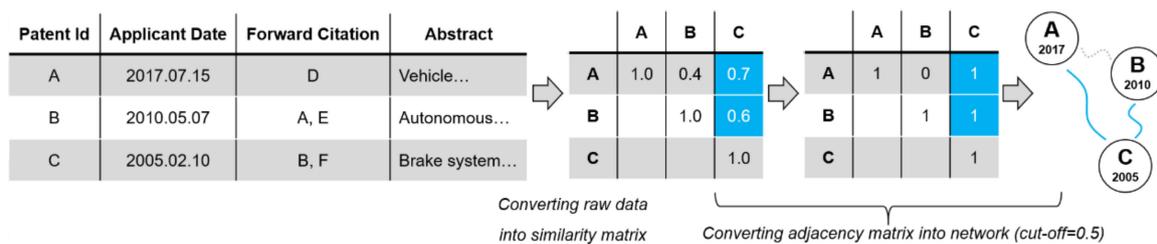
Figure 2. Conceptual diagram of similarity network.

To make this, pairs of documents whose similarity value is greater than a preset threshold value are identified. In the example, the threshold value is set as 0.5. The similarity network used in this study is the result of visualizing a similarity adjacency matrix (SAM) constructed based on the similarity between documents.

4. Proposed Methodology

This study proposes a method of searching for patent groups likely to have overlapping scopes of rights by using the citation relationship and similarity between them. Figure 3 shows the task flow of the proposed methodology. First, patent documents matching the purpose of the analysis are collected. The text in the collected patent is preprocessed and converted to document-term matrix (DTM) through lexical analysis. Next, we draw the citation network by using the citation information of the collected patents. Then, the text similarity between each patent is calculated based on the contents of the representative claims.
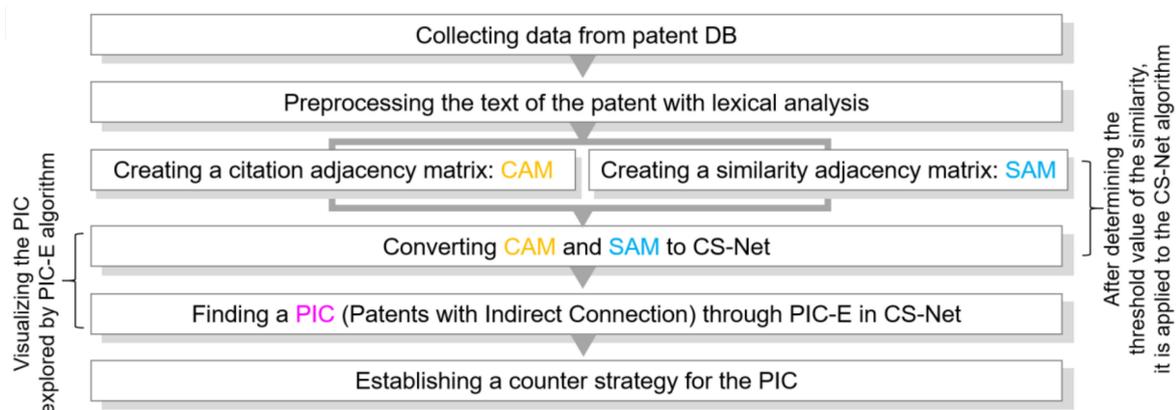

Figure 3. Task flow of proposed methodology.

The completed citation network and similarity network are integrated into a citation and similarity network (CS-Net). How to configure the CS-Net through combining citation and similarity networks is described in Section 4.2. It is effective in finding patents with a special relationship that we define as "patents with indirect connection (PIC)", because it considers citation relationship and similarity between patents in a comprehensive way. PIC refers to a pair of patents considered similar to each other as they are indirectly connected in CS-Net. In order to make it easy to discover PICs in CS-Net, which is a large network composed of collected patent big data, we propose an algorithm named PIC-explorer (PIC-E). The details of the PIC-E algorithm is described in Section 4.3. Even though a pair of patents corresponding to PIC are not directly linked in a citation network, there is a possibility that they have a similar scope of rights. Therefore, once there are PICs detected by PIC-E, it is necessary to examine the possibility of patent infringement and establish an appropriate response strategy.

4.1. PIC: A Pair of Patents that Can Be Found by Similarity and Citation Information between Technologies

This study aims to find sets of patents dealing with similar technologies, but that are not directly linked to each other in the citation network, as shown in Figure 4. For example, suppose (i) patent B is a cited patent of C and it is re-cited by patent A; and (ii) patent A and patent C are similar to each other.
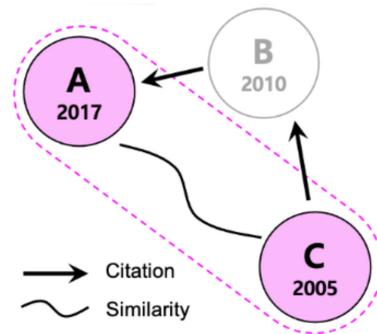

Figure 4. Conceptual diagram of PIC.

In this case, the filing years of patents A and C are 2017 and 2005, respectively. It is highly likely that C, whose filing date is earlier than A, is a prior art of A. In this study, the relationship between A and C is defined as PIC. Considering the relationship, the applicant of the preceding patent C needs to investigate if patent A has infringed the rights of patent C. On the other hand, the applicant of the succeeding patent A may need to develop a differentiation or invalidation logic in order not to infringe C's scope of rights. Besides, new market entrants in this domain need to carefully scrutinize the claims of existing patents and establish a filing strategy to circumvent their scope of rights. Therefore, it is helpful for them to analyze the patents constituting the PICs.

4.2. CS-Net: A Method of Merging the Citation Network and the Similarity Network

Let n be the number of documents collected. The size of the document similarity matrix is then n by n. When there is no citation among the collected documents, the size of the citation matrix is also n by n.

However, if there are k documents not included in the collected patents, the size of the citation matrix is $(n + k)$ by $(n + k)$. Therefore, the size of the citation matrix is always greater than or equal to that of the document similarity matrix. Thus, the addition of the two matrices cannot be done by the general sum of matrices. Table 1 shows the pseudo code of the CS-Net algorithm for merging the two networks. The adjacency matrix to build CS-Net is the sum of the CAM and the SAM. Since it is the sum of two matrices composed of 0 and 1, each element constituting the adjacency matrix is one of 0, 1, and 2.

Table 1. Algorithm of adjacency matrix for CS-Net.

Input : CAM = citation adjacency matrix (m × m), m = n + k
SAM = similarity adjacency matrix (n × n), m ≥ n
Output : CS = adjacency matrix (m × m)
Initialize: CS = zero matrix (m × m)
FOR all the elements in the CAM
IF the order of elements ≤ n **then**
Summation each element of CAM and SAM
ELSE the value of CAM is used as it is
END

Figure 5 shows the conceptual diagram of CS-Net. In the example, both n and k are three. The sizes of the citation matrix and the similarity matrix are 6 by 6 and 3 by 3, respectively. Since both matrices contain patents A, B, and C, values corresponding to the same patent pair are added together. In this example, not only is there a direct citation relationship between patents A and B, and B and C, but A and C are similar. As a result, a pair of PIC (A and C) can be found from a network consisting of six nodes.
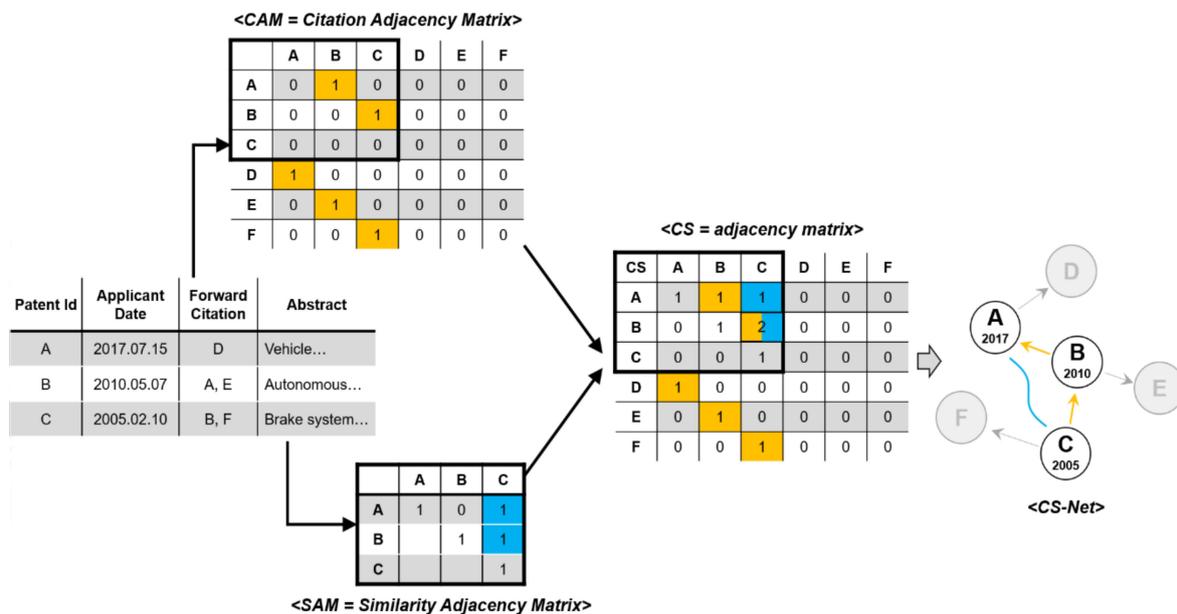

Figure 5. Conceptual diagram of CS-Net.

4.3. PIC-E: A Method of Exploring PIC from CS-Net

Table 2 describes the pseudo code of PIC-E, an algorithm to search for PICs in CS-Net. The input of PIC-E is the CS matrix obtained by applying the CS-Net algorithm with the citation matrix and the similarity matrix. x_{ij} denotes an element in the ith row and jth column of the CS matrix, and has a value of 0, 1, or 2. The first condition that x_{ij} has a non-zero value is that there is a citation relationship between the ith patent P_i and the jth patent P_j. The other condition is that the similarity value between the two patents is greater than or equal to the preset threshold. The value of x_{ij} is 2 when both conditions are satisfied, and 1 when only one of the conditions is satisfied. Figure 5 is an example of CS where m is 6. In this example, patents A and C are P_1 and P_3, respectively. Seeing the value of x_{13} in the CS matrix equals to 1, one of the two conditions mentioned above is satisfied. The following step is to compare $Date_1$ and $Date_3$, which are the filing dates of P_1 and P_3. Since $Date_1$ is later than $Date_3$, Diff representing the time difference is positive. Therefore, the later patent P_1 corresponds to P_L, and P_3 is P_E. F_1 denotes a set of forward citations of the prior patent P_E. In Figure 5, patents B and F are included in F_1 because they are forward citations of P_E (patent C). F_2 represents the forward citations of the patents belonging to F_1. If P_L (patent A) whose filing date is later than P_E (patent C) is included in F_2, the relationship between P_E and P_L is PIC.

Table 2. Algorithm of PIC-explorer for PIC.

Input	: CS = CS-Net (CAM, SAM) (m × m)
	Date$_i$ = Application date of ith patent P$_i$
Output	: PIC
Initialize	: PIC as a list

FOR x$_{ij}$ is the element in the **CS** (i,j = 1, 2, …, m and i ≠ j) **do**

    IF x$_{ij}$ ≥ 1 **then**

        DEFINE Diff = Date$_i$–Date$_j$

        IF Diff ≤ 0 **then**

            the prior patent P$_i$ is P$_E$ and the later patent P$_j$ is P$_L$ (means that P$_i$ was filed earlier than P$_j$)

        ELSE P$_i$ is P$_L$, P$_j$ is P$_E$

        DEFINE   F$_1$ = set of forward citations of P$_E$

                     F$_2$ = set of forward citations of P$_{E'}$, one of F$_1$

        IF P$_L$ exists in F$_2$ **then**

            SAVE (P$_i$, P$_j$) to PIC

        ELSE Pass

    ELSE Pass

END

CS-Net is able to visualize both the citation relationship and the similarity information of the collected patent big data. In other words, it is easy to grasp the citation flow of a patent and the process of similar technology development through CS-Net. However, since CS-Net is a very large network, it can be reconstructed through PIC-E by selecting only the patents with a PIC relationship. As a result, we can efficiently visualize big data and find the patents with a high risk of patent disputes.

5. Experimental Study

This section conducts experiments to confirm the practical applicability of the proposed method. For the experiment, we collected 1484 patents related to machine learning and deep learning published by the Korean Intellectual Property Office (KIPO). These technologies have recently been widely applied in robotics, specifically to the part that plays the role of the brain in robots. Based on the time of analysis, the number of patents cited more than once from other patents is 771. The largest number of times a patent was cited from another patent was 38.

Khaiii (Kakao Hangul Analyzer III) is a morpheme analyzer that learned the Korean corpus called 'Sejong' provided by the National Institute of the Korean Language with a deep learning structure [46–48]. We use Khaiii as a tokenizer and part-of-speech (POS) tagger for preprocessing text in the collected Korean patents. Tokenization and POS-tagging were performed for the representative claims in the patent documents, and only nouns were extracted [49,50]. In addition, DTM was constructed by calculating the term frequency–inverse document frequency (TF–IDF) weights of each extracted noun [51,52]. Based on this, a similarity matrix was created based on the cosine similarity.

The next step was to explore PICs from the CS-Net via PIC-E. As a result, a total of 24 pairs of PICs consisting of 48 patent nodes were identified. Figure 6 shows the CS-Net which expresses the 48 patent nodes corresponding to the PICs in a dark color. The nodes expressed in a light color are patents that are similar to or have a citation relationship with the patents belonging to the PICs. Patents corresponding to the PIC are marked with a number denoted by h, and the two patent nodes constituting a pair of PICs are assigned the same number (h = 1, 2, . . . , 24). For example, the two patents of the 3rd PIC are both labeled 3. In addition, the edge representing the PIC relationship in the network is indicated by a thick blue line. Through the visualization results, nodes 5, 7, 13, and 17 each formed

an independent network. On the other hand, the others appear to have a similar or citation relationship as they constitute one connected network.
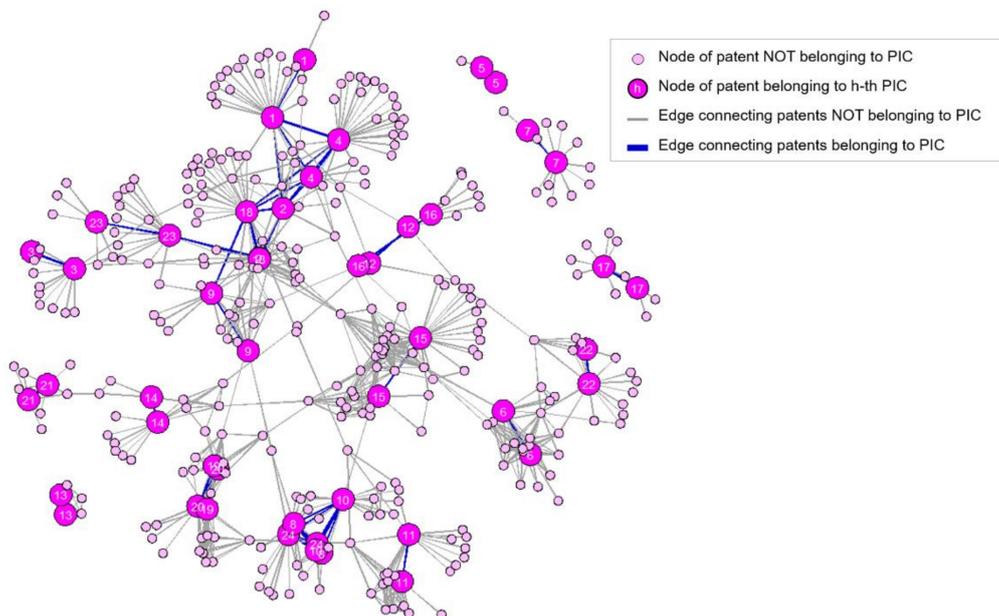

Figure 6. Visualization of PIC and related patents in CS-Net.

The patent pairs that formed the PIC relationship are likely to deal with similar technologies. Table 1 in the Appendix A shows the cosine similarity (Sim) and applicant year (Year) of patents corresponding to PICs, and the title of the patent document. Among the PIC-related patent pairs, the technology with the largest similarity is related to a vehicle vision system equipped with an artificial intelligence chip. The second-largest pair of patents dealt with an energy management system using machine learning. Considering the PIC relationship, it is possible to establish a strategy for prior art applicants to claim infringement from new patents. Conversely, applicants of later patents may develop logic to circumvent or invalidate the scope of the prior art.

6. Discussion

Industrial applicability, novelty, and progressiveness are essential elements of a patent. Novelty is crucial because patents legally protect rights instead of disclosing technology. Considering prior art, researchers try to improve and develop advanced technologies. This process is the ideal goal of the patent system. In this context, a prior art search is essential for rights protection and technological advancement. Researchers plan to develop new technologies through a prior art analysis. If similar prior art exists, they attempt to make their invention different or more advanced. Without the investigation of prior art, the risk of potential patent litigation increases.

Previous studies used various methods to improve the efficiency of searching for prior art. Most studies used a citation relationship to search for similar patents. Similar prior art and technologies that infringed on the rights of other patents were searched with the information. There were also related studies that evaluated and reflected the degree of similarity of patent documents using the text information. However, previous studies did not comprehensively consider the correlation between citation information and document similarity, and the sequence of citations.

This study proposes a prior art search method that considers both citation information and document similarity. It is designed based on the characteristic that patents cited by other patents tend to be re-cited by patents similar to them. If there is no citation relationship among similar patents, it is necessary to question whether the rights of preceding patents are infringed. Therefore, the first purpose of this method is to find prior art whose scope of

rights may have been infringed by later patents. The second purpose is to monitor later patents so that they do not infringe on the scope of the rights of previous patents.

Our research still has some limitations as described below. First, it is difficult to reflect new technologies because they have relatively few opportunities to be cited by other patents. The proposed method is designed considering the tendency that similar patents are likely to cite the same patents. Therefore, recently developed technologies may be less likely to be detected by this method. The second limitation is on the depth of sequential citations. We have focused on the indirect relationship between two patents. In some cases, however, similarities between patents in direct citation relationship may be large. Furthermore, sufficient consideration is required for the case where the length of the citation sequence is long.

7. Conclusions

The purpose of the patent system is to promote technological advancement and industrial development. According to the purpose of the patent system, a new patent requires novelty and progressiveness compared to prior art. When a new patent infringes the rights of prior art, it is inevitable that companies spend lots of time and money in resolving patent disputes. In order to prevent patent disputes, this paper proposed a method of establishing a counter strategy using citation relationships and similarities of prior art.

The proposed method was tested with patents related to machine learning and deep learning to confirm the practical applicability of the method. As a result of the experiment, a total of 48 patents were similar, but there was no direct citation relationship. In addition, some of the patents in the indirect citation relationship were judged to have a possibility of dispute because their claims are similar to each other. The similar patent pairs differ in the time of filing, so it is possible to prepare a strategy for judging the infringement of rights and a strategy for developing a non-infringement or invalidation logic. This methodology is expected to be widely used to search for prior art or to monitor the occurrence of rights infringement in domains that form a complex citation relationship, such as in the field of robotics.

In the future, it is necessary to study the counter strategy that has expanded from the patent level to the company level. If this is possible, it can search for competitors. In addition, research on algorithms that can reflect new technologies is needed. To this end, not only the citation information of a patent but also family patents may be used. Finally, a method that considers a deeper citation relationship is needed. Such research can be utilized to analyze the direction of technology development and search for basic technologies. Basic technologies are patents that form the basis of a technical field, and once they are identified, the flow can be easily understood. Therefore, it is expected that this method will be used for efficient patent big data analysis.

Appendix A

Table 1 in the Appendix A is a list of PICs derived from the experiment in Section 5. The first column of the table is the index of the PIC. Sim is the value of cosine similarity (Sim) of the two patents belonging to the PIC in descending order. Year refers to the filing year of each patent. Among the two patents corresponding to each PIC, the older one is written at the top.

Table 1. PICs explored in the proposed methodology.

PIC	Sim	Year	Title of Patent Document
1	0.913	2011	The apparatus and method of multi-lane car plate detection and recognition
		2013	The One shot camera for artificial intelligence function by using neuromorphic chip
2	0.818	2006	Real time predicting system for energy management system using machine learning
		2017	Predicting system for energy management system
3	0.781	2016	The intelligent disclosure of public records management system based machine learning
		2016	System for classifying and opening information based on natural language
4	0.702	2010	Apparatus for analysis of mobile big data
		2017	Device for analyzing mobile data using data mining and method thereof
5	0.677	2007	Lotto lottery numbers mixing system for using data mining and service method thereof
		2009	System and method of recommendation number of lotto lottery number for providing lotto lottery for increasing winning ration using data mining
6	0.675	2007	Grid-based hybrid data mining device and method thereof
		2015	Simulation-based computational grid resource management device using ontology and method thereof
7	0.655	2009	Semantic information based grid management system and method for grid computing
		2015	Simulation-based computational grid resource management device using ontology and method thereof
8	0.650	2012	Storeroom environment state management system and method of base ontology
		2018	System and method for smart refrigerator management based on situation-awareness
9	0.613	2016	Method for mining weighted erasable by using underestimated constraint-based pruning technique
		2017	Method of miming top-k important patterns
10	0.597	2014	System and method for searching contents using ontology
		2016	Apparatus and method for frequent sub-graph component mining in graph data
11	0.568	2009	Apparatus and method for generating a reconstituted ontology based on the conceptual structure
		2012	Browsing system and method of information using ontology
12	0.566	2010	Method for mining maximal weighted frequent patterns
		2016	Method for mining weighted erasable by using underestimated constraint-based pruning technique
13	0.563	2007	System and method for providing context cognition to control home network service
		2015	Personalized home automation service providing method based on ontology and service providing system using ontology based on context awareness
14	0.549	2016	Intelligent video surveillance system for school zone
		2017	Method for counting vehicles based on image recognition and apparatus using the same
15	0.549	2014	Method and apparatus for usability test based on big data
		2018	Automatic task classification based upon machine learning
16	0.518	2007	Modeling method and apparatus for multi-ontology
		2010	System and method for retrieving/classifying web ontology
17	0.515	2012	System and method for processing ontology models, and its program recorded recording medium
		2014	Apparatus and method for converting English ontology to Korean ontology

Table 1. *Cont.*

PIC	Sim	Year	Title of Patent Document
18	0.494	2013	Pattern mining method for searching tree on top-down traversal for considering weight in a data stream
		2016	Method for mining weighted erasable by using an underestimated constraint-based pruning technique
19	0.456	2000	Study system and method for foreign language
		2013	System for assessing improvement of basic skills in education
20	0.434	2008	English learning method and apparatus thereof
		2010	Method and system for learning English using word order map
21	0.431	2003	Single-pass mining of frequent simultaneous event groups for stream data, an apparatus for single-pass mining of frequent simultaneous event groups for stream data
		2007	System and mechanism for discovering temporal relation rules from interval data
22	0.423	2009	Apparatus and method for generating a reconstituted ontology based on the conceptual structure
		2011	Web ontology editing and operating system
23	0.413	2006	Clustering system and method using search result documents
		2015	Analysis system for environment research using environmental geographical information and textmining among big data
24	0.406	2007	System for recommending personalized meaning-based web-document and its method
		2010	Method for calculating similarity between document elements

References

1. Brent, A.; Pretorius, M. Sustainable development: A conceptual framework for the management of knowledge and a departure for further research. *S. Afr. J. Ind. Eng.* **2008**, *19*, 31–52. [CrossRef]
2. Park, S. Development of a Categorized Checklist for Valuation of Patent Technology. *J. Intellect. Prop.* **2007**, *2*, 30–56. [CrossRef]
3. Betz, F. *Managing Technological Innovation: Competitive Advantage from Change*, 3rd ed.; Wiley–Interscience: Hoboken, NJ, USA, 2011. [CrossRef]
4. Choi, J.; Jun, S.; Park, S. A patent analysis for sustainable technology management. *Sustainability* **2016**, *8*, 688. [CrossRef]
5. Schilling, M. *Strategic Management of Technological Innovation*, 5th ed.; McGraw-Hill Education: New York, NY, USA, 2016. [CrossRef]
6. Storey, C.; Easingwood, C. Types of new product performance: Evidence from the consumer financial sector. *J. Bus. Res.* **1999**, *85*, 275–287. [CrossRef]
7. Roberts, J. Developing new rules for new markets. *J. Acad. Mark. Sci.* **2000**, *28*, 31–44. [CrossRef]
8. Menor, L.; Tatikonda, M.; Sampson, S. New service development: Areas for exploitation and exploration. *J. Acad. Mark. Sci.* **2002**, *20*, 135–157. [CrossRef]
9. Tseng, C. Technology development and knowledge spillover in Africa: Evidence using patent and citation data. *Int. J. Technol. Manag.* **2009**, *45*, 50–61. [CrossRef]
10. Kim, C.; Lee, H. A database–centred approach to the development of new mobile service concepts. *Int. J. Mob. Commun.* **2012**, *10*, 248–264. [CrossRef]
11. Lee, S.; Park, S.; Jung, E. A Study on the Analysis of Competitiveness of Corporations by Comparing of Patent Citations Based on Data Mining. *J. Korean Inst. Intell. Syst.* **2019**, *29*, 452–457. [CrossRef]
12. Lai, K.; Wu, S. Using the patent co–citation approach to establish a new patent classification system. *Inf. Process. Manag.* **2005**, *41*, 313–330. [CrossRef]
13. Gipp, B.; Beel, J. Citation Proximity Analysis (CPA)–A New Approach for Identifying Related Work Based on Co–Citation Analysis. 2009. Available online: http://nbn-resolving.de/urn:nbn:de:bsz:352-0-285851 (accessed on 13 January 2021).
14. Ritchie, A. Citation Context Analysis for Information Retrieval. 2009. Available online: https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-744.pdf (accessed on 13 January 2021).
15. Gurulingappa, H.; Mueller, B.; Klinger, R.; Mevissen, H.; Hofmann–Apitus, M.; Fluck, J.; Friedrich, C. Prior Art Search in Chemistry Patents on Semantic Concepts and Cocitation Analysis. 2010. Available online: https://trec.nist.gov/pubs/trec19/papers/fraunhofer-scai.chem.rev.pdf (accessed on 13 January 2021).
16. Zhao, H. Sharding for literature search via cutting citation graphs. In Proceedings of the IEEE International Conference on Big Data, Washington, DC, USA, 27–30 October 2014; pp. 77–79. [CrossRef]
17. Rodriguez, A.; Kim, B.; Turkoz, M.; Lee, J.M.; Coh, B.Y.; Jeong, M.K. New multi-stage similarity measure for calculation of pairwise patent similarity in a patent citation network. *Scientometrics* **2015**, *103*, 565–581. [CrossRef]

18. No, H.; An, Y.; Park, Y. A structured approach to explore knowledge flows through technology–based business methods by integrating patent citation analysis and text mining. *Technol. Forecast. Soc. Chang.* **2015**, *97*, 181–192. [CrossRef]

19. Nakamura, H.; Suzuki, S.; Sakata, I.; Kajikawa, Y. Knowledge combination modeling: The measurement of knowledge similarity between different technological domains. *Technol. Forecast. Soc. Chang.* **2015**, *94*, 187–201. [CrossRef]

20. Kim, J.; Park, S. A method of Establishing Patent Strategy using Self–Organizing Map. *J. Korean Inst. Intell. Syst.* **2018**, *28*, 422–427. [CrossRef]

21. Zhu, D. Bibliometric analysis of patent infringement retrieval model based on self–organizing map neural network algorithm. *Libr. Hi Tech.* **2019**, *38*, 479–491. [CrossRef]

22. Korean Intellectual Property Office. Studies on the effect of IP strategies on the Survival and Performance of firms. In *Korean Institute of Intellectual Property*; Korean Intellectual Property Office: Seoul, Korea, 2015.

23. Su, F.; Yang, W.; Lai, K. A heuristic procedure to identify the most valuable chain of patent priority network. *Technol. Forecast. Soc. Chang.* **2011**, *78*, 319–331. [CrossRef]

24. Kim, H.; Kim, J.; Lee, J.; Park, S.; Jang, D. A Novel Methodology for Extracting Core Technology and Patents by IP Mining. *J. Intell. Syst.* **2015**, *25*, 392–397. [CrossRef]

25. Yoon, J.; Choi, S. Planning Future Technology Strategies Using Patent Information Analysis and Scenario Planning: The Case of Fuel Cells. *J. Inf. Sci. Theory Pract.* **2012**, *43*, 169–197. [CrossRef]

26. Kwon, W.; Lee, J.; Kang, J.; Park, S.; Jang, D. Patent Information Analysis Using Quantitative Patent Index. In Proceedings of the Korean Institute of Intelligent Systems, Seoul, Korea, 19–21 April 2018; pp. 9–10.

27. Kang, J.; Kim, J.; Lee, J.; Park, S.; Jang, D. Methodology of Prior Art Search Based on Hierarchical Citation Analysis. *J. Korean Inst. Intell. Syst.* **2017**, *28*, 72–78. [CrossRef]

28. Shibata, N.; Kajikawa, Y.; Takeda, Y.; Matsushima, K. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation* **2008**, *28*, 758–775. [CrossRef]

29. Yaghtin, M.; Sotudeh, H.; Mohammadi, M.; Mirzabeigi, M.; Fakhrahmad, S. A Correlation Study of Co–Opinion and Co–Citation Similarity Measures. 2019. Available online: https://ijism.ricest.ac.ir/index.php/ijism/article/view/1517/366 (accessed on 13 January 2021).

30. Jui, C.; Trappey, A.; Fu, C. Method of Claim–Based Technology Analysis for Strategic Innovation Management–Using TPP–Relates as Case Examples. 2016. Available online: http://nopr.niscair.res.in/handle/123456789/35372 (accessed on 13 January 2021).

31. Chen, C.; Chen, R.; Wang, D.; Dai, T. GA–based Dissimilarity Visualization Engine for Design Patent Map Systems. In Proceedings of the IEEE International Conference on Hybrid Intelligent Systems, Melacca, Malaysia, 5–8 December 2011; pp. 595–600. [CrossRef]

32. Dejean, S.; Faessel, N.; Marty, L.; Mothe, J.; Sadala, S.; Thiam, S. Analysis of Patents for Prior Art Candidate Search. 2013. Available online: ftp://ftp.irit.fr/IRIT/SIG/2013_IMMM_DFMMST.pdf (accessed on 13 January 2021).

33. Jeong, B.; Ko, N.; Kyung, J.; Choi, D.; Yoon, J. Development of a Patent Prior Art Search System for Invalidation Analysis of Barrier Patents. 2017. Available online: https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=ART002249787 (accessed on 13 January 2021).

34. Korean Intellectual Property Office. Patent–oriented R&D innovation strategy. In *R&D Patent Center*; Korean Intellectual Property Office: Seoul, Korea, 2012.

35. Grindley, P.; Teece, D. Managing Intellectual Capital: Licensing and Cross–Licensing in Semiconductors and Electronics. 1997. Available online: https://journals.sagepub.com/doi/pdf/10.2307/41165885?casa_token=AHfMaSfSAZUAAAAA:hvc2RUlk4NsgNmpZ5ylCQn1ad_T4QvxpSZCUINwghmAxzo-dg42blzt8lUt6GrNeUXOeW_lpxDJnGw (accessed on 13 January 2021).

36. Lippman, S.; Rumelt, R. Uncertain imitability: An analysis of interfirm differences in efficiency under competition. *Bell J. Econ.* **1982**, *13*, 418–438. [CrossRef]

37. Arora, A.; Fosfuri, A. Licensing the market for technology. *J. Econ. Behav. Organ.* **2003**, *52*, 277–295. [CrossRef]

38. Rherrad, I.; Gallaud, D. Exploring appropriation strategies: Evidence from French high–tech firms. *Int. J. Technol. Transf. Commercialis.* **2009**, *8*, 316–339. [CrossRef]

39. Davis, L.; Kjær, K. Patent Strategies of Small Danish High–Tech Firms. 2003. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.5905&rep=rep1&type=pdf (accessed on 13 January 2021).

40. Choi, J.; Kim, H.; Im, N. Keyword Network Analysis for Technology Forecasting. *J. Intell. Inf. Syst.* **2011**, *17*, 227–240. [CrossRef]

41. Gui, B.; Ju, Y.; Liu, Y. Mapping technological development using patent citation trees: An analysis of bogie technology. *Technol. Anal. Strateg. Manag.* **2019**, *31*, 213–226. [CrossRef]

42. Park, S. A Study on Patent Big Data Visualization Using Inference model–based Performance Indicator Network. *J. Korean Inst. Intell. Syst.* **2020**, *30*, 74–79. [CrossRef]

43. McGill, M.; Koll, M.; Noreault, T. An Evaluation of Factors Affecting Document Ranking by Information Retrieval Systems. In *School of Information Studies*; 1978. Available online: https://files.eric.ed.gov/fulltext/ED188587.pdf (accessed on 13 January 2021).

44. Salton, G.; McGill, M. Introduction to Modern Information Retrieval. 1986. Available online: https://sigir.org/files/museum/introduction_to_modern_information_retrieval/frontmatter.pdf (accessed on 13 January 2021).

45. Zhang, J.; Korfhage, R. A distance and angle similarity measure. *J. Am. Soc. Inf. Sci.* **1990**, *50*, 772–778. [CrossRef]

46. Khaiii. Github. 2018. Available online: https://github.com/kakao/khaiii (accessed on 3 November 2020).

47. Han, G.; Baek, S.; Lim, J. Open Sourced and Collaborative Method to Fix Errors of Sejong Morphologically Annotated Corpora. 2017. Available online: https://www.koreascience.or.kr/article/CFKO201731951960133.pdf (accessed on 13 January 2021).
48. Lee, Y.; Kim, S.; Hong, H.; Kim, J. Comparison and Evaluation of Morphological Analyzer for Patent Documents. In Proceedings of the Korean Institute of Information Technology, Daejeon, Korea, 13–15 June 2019; pp. 264–265.
49. Hong, J.; Cha, J. Error Correction of Sejong Morphological Annotation Corpora using Part–of–Speech Tagger and Frequency Information. 2013. Available online: https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART001789163 (accessed on 13 January 2021).
50. Shim, K. Morpheme Restoration for Syllable–based Korean POS Tagging. 2013. Available online: https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE02112476 (accessed on 13 January 2021).
51. Aizawa, A. An information–theoretic perspective of tf–idf measures. *Inf. Process. Manag.* **2003**, *39*, 45–65. [CrossRef]
52. Wu, H.; Luk, R.; Wong, K.; Kwok, K. Interpreting tf–idf term weights as making relevance decisions. *ACM Trans. Inf. Syst.* **2008**, *26*, 1–37. [CrossRef]