CONSTRUCTING AN INTELLIGENT PATENT NETWORK ANALYSIS METHOD

Chao-Chan Wu[1] and Ching-Bang Yao[2,3*]

[1]Department of Cooperative Economics, Feng Chia University, 100, Wen-Hwa Road, Seatwen, Taichung 40724, Taiwan
Email: chaocwu@fcu.edu.tw; chaochan0829@gmail.com
[2]Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, 43, Sec. 4, Keelung Road, Taipei 106, Taiwan
*[*3]Department of Information Management, Chinese Culture University, 55, Hwa-Kang Road, Yang-Ming-Shan, Taipei 11114, Taiwan*
**Email:* yqb@faculty.pccu.edu.tw; chinbang7@yahoo.com.tw

ABSTRACT

Patent network analysis, an advanced method of patent analysis, is a useful tool for technology management. This method visually displays all the relationships among the patents and enables the analysts to intuitively comprehend the overview of a set of patents in the field of the technology being studied. Although patent network analysis possesses relative advantages different from traditional methods of patent analysis, it is subject to several crucial limitations. To overcome the drawbacks of the current method, this study proposes a novel patent analysis method, called the intelligent patent network analysis method, to make a visual network with great precision. Based on artificial intelligence techniques, the proposed method provides an automated procedure for searching patent documents, extracting patent keywords, and determining the weight of each patent keyword in order to generate a sophisticated visualization of the patent network. This study proposes a detailed procedure for generating an intelligent patent network that is helpful for improving the efficiency and quality of patent analysis. Furthermore, patents in the field of Carbon Nanotube Backlight Unit (CNT-BLU) were analyzed to verify the utility of the proposed method.

Keywords: Patent network analysis, Artificial intelligence, Ontology, Enhanced term frequency - inverse document frequency (ETF-IDF)

1 INTRODUCTION

Patents which describe main contents of technological inventions contain considerable technical knowledge. These documents are significant sources of technological data and play a critical role in the advancement and diffusion of technology (Horie, Maeno, & Ohsawa, 2007; Liu & Luo, 2007a). Furthermore, patent analysis transfers the patent data to systematic and valuable information that is helpful for managing research and development process, exploring technological trends, tracking technological development, and identifying technology plans (Liu & Luo, 2007b; Liu & Yang, 2008; Chang, Wu, & Leu, 2010). It is considered to be a useful vehicle for technology management.

Traditionally, patent bibliometric analysis has been most commonly used to implement patent analysis (Narin, 1994). Patent bibliometric analysis utilizes bibliometric data from patent documents to perform statistical analysis and citation analysis. Statistical analysis employs bibliometric data, such as number of patents, country, assignee, inventor, and so forth. Then, statistical methods are used to analyze the bibliometric data. Citations are the counts of other patents or non-patent literature cited in the patent documents. Citation analysis uses these citations in patent documents to find important patents and develop other scientific linkages. Patent bibliometric analysis, albeit easy to understand and simple to use, is limited in the scope of analysis and the richness of potential information (Yoon & Park, 2004).

To overcome these limitations, Yoon & Park (2004) suggest an advanced method of patent analysis, called patent network analysis. This method uses several patent keywords as input to produce a visual patent network. The network demonstrates the overall relationship among all patents. The analysts are thus able to comprehend the overall structure of a patent database intuitively and discover the key patents in the patent network. Although patent network analysis possesses relative advantages over traditional methods of patent analysis, it is subject to several crucial drawbacks. First, the search for patent documents to be studied relies on the subjective judgments of analysts. Second, the collection of patent documents is a time-consuming task because it requires an exhaustive search of patent databases. The current method lacks a set of systematic and convenient patent searching procedures. As a result, the dataset of patent documents being studied is not complete. Third, the relevant patent keywords used in the current method are selected by technical experts. In reality, the technical experts often use different terminologies to describe the same technology (Li, Wang, & Hong, 2009). Even though these experts have rich experience in the field of technology being studied, they have great difficulty avoiding the subjectivity involved in the extraction of patent keywords. If the keywords are not chosen properly, the visualization of patent network will be distorted. Finally, the current method assumes that the weight of each patent keyword is equal. However, the individual weights of patent keywords are different from each other. It is necessary to determine the priorities among diverse patent keywords and the relative weighted value of each keyword.

In order to resolve all of the aforementioned problems, constructing an automated technique for improving the current method is necessary. This study proposes applying artificial intelligence techniques to come up with an intelligent patent network analysis method. Artificial intelligence is usually an excellent solution when facing the abundance of current patent documents. When making a quick and effective search for the most useful and important key patents, the related techniques of artificial intelligence play a significant role. For example, these techniques can swiftly process and categorize large amounts of patent documents, automatically identify and extract keyword sets, as well as broadly and objectively select the keywords that are synonyms.

Accordingly, artificial intelligence techniques for assisting patent analysts in patent processing and analysis are in great demand. Previous study has developed a framework for automatic patent analysis method (Wu & Yao, 2012). However, the issue regarding weights of keywords was not concerned and the utility of method was not assured. Thus, this study extends previous framework to propose an intelligent patent network analysis method, and verifies the utility of this one. The proposed method is useful for making the visual patent network more substantial, which in turn improves the efficiency and effectiveness of patent analysis. That is the purpose of this study, and the details are as follows. First, in order to collect a complete dataset of patent documents, this study proposes a set of systematic patent searching procedures by introducing an ontology methodology of automatic document classification. This procedure is very convenient in terms of search time and cost. Second, this study conducts the enhanced term frequency - inverse document frequency (ETF-IDF) technique to conduct the information retrieval job to extract the patent keywords automatically from the selected patent documents. Third, the association rules, which combine the Viterbi algorithm with the Apriori algorithm, are used to determine the weighted value of each keyword. Finally, the sets of patent keywords are employed to act as the input base for generating the precise visualization of the patent network that contributes to implementing the patent analysis. In particular, the patents regarding the technological field of Carbon Nanotube Backlight Unit (CNT-BLU) are analyzed to verify the utility of the proposed method.

2 RELATED WORK

2.1 Patent network analysis method

Network analysis, by emphasizing the relationships among the social positions within a system, provides a powerful brush for painting a systematic picture of global social structures and their components (Knoke & Kuklinski, 1982). This analysis is capable of showing the structure of edges among nodes. Nodes are the given entities in the network. The relationship between nodes and the location of individual nodes in the network provide ample information and assist the analysts in realizing the overall structure. Furthermore, network analysis utilizes quantitative techniques to generate relevant indexes that clarify the characteristics of the whole network and show the position of individuals or groups in the network structure (Wasserman & Faust, 1994).

Even though network analysis was developed initially for sociological studies, it is utilized widely in other research areas (Leoncini, Maggioni, & Montresor, 1996; Cross, Borgatti, & Parker, 2001; Calero, Buter, Valdés, & Noyons, 2006; Shin, Lee, & Park, 2006). Recently, Yoon & Park (2004) applied the concept of network analysis in patent analysis and proposed patent network analysis. This method utilizes the frequency of keywords' appearance in patent documents as the input base to generate a patent network. The relationship among patents can be visually demonstrated in this analysis, and the analysts are able to comprehend the overall structure of patent network. Moreover, this method produces several meaningful indexes which can help analysts to identify the relative importance of individual patents and to explore technological trends (Chang, Wu, & Leu, 2010).

The main purpose of this study is to propose an intelligent patent network analysis method based on artificial intelligence techniques in order to develop a visually sophisticated patent network. The concept of artificial intelligence techniques will be described in the next section.

2.2 Artificial intelligence techniques

Artificial intelligence is the field of computer science focusing on enabling computers to engage in behaviors that humans consider intelligent by automatic judgment mechanic (Crevier, 1993). It attempts to achieve the goal of giving the computer human intelligence by intelligent algorithm. Today, after the advent of the computer and 50 years of research into artificial intelligence programming techniques, the dream of smart machines is becoming a reality (Yang, 2007). Researchers are creating systems that can mimic human thought, understand speech, and do countless other feats never before possible. Recently, artificial intelligence has been developed in many applied areas (Yang & Liu, 1999). A prominent branch of artificial intelligence research is the highly technical and specialized information retrieval, which can utilize techniques such as fuzzy theory, nature language processing (NLP) technique, and so on, to automatically process the abundance of information on the internet.

Among various techniques of data mining, Apriori is a classic algorithm for learning association rules which can find out the latent relations between different items (Agrawal, Imielinski, & Swami, 1993; Yang & Liu, 1999). Apriori algorithm is designed to process the abundant transactions and to operate on databases which contain transactions, such as collections of items bought by consumers or details of a website frequentation. It attempts to find the frequent subsets that have in common at least a minimum number of items, which is the cutoff or confidence threshold of the subsets. The Apriori algorithm put the association rule into practice which represents an unsupervised learning method that attempts to capture associations among groups of items. This technique can be applied to the intelligent method suggested in this study in order to quickly and automatically handle complicated patent documents.

Regarding keyword automatic identification, the term frequency - inverse document frequency (TF-IDF) methodology proposes an excellent algorithm that computes the appropriate frequency of keyword (Salton & McGill, 1983). The TF-IDF technique is usually used to weigh each word in the text document based on how unique it is. This technique captures relevant keywords, text documents, and particular categories. Our study combines the TF-IDF technique with our linguistic recognition rules, which are provided by experts in order to further select out the long word vocabularies and specialized vocabularies with a particular language purpose to give higher weighting. Then the right weightings of all keywords are automatically counted after proper adjustment through the linguistic rules. Next, the keyword set of each patent document is formed. Finally, we use the association rules to compare all keyword sets of patent documents in order to delete the unsuitable vocabularies out of the keyword set. This automatically strengthens the final suitable relevant keywords of all patent documents.

Using the above information, several artificial intelligence techniques are applied to construct our intelligent patent network analysis. The detailed methodology will be explained in the next section.

3 METHODOLOGY AND PROCEDURE

The main purpose of this study is to propose an automatically intelligent patent network analysis method. In this section, the methodology of intelligent patent network analysis presented in this study is explained. Figure 1 shows the overall procedure of the proposed method. It contains four major stages: searching and collecting patent documents, extracting patent keywords, determining the weight of each patent keyword, and generating a sophisticated visualization of the patent network. First, this study exploits the ontology of the automatic document classification process which is identified by the patent keywords agents to extract the feature subset documents. This automated technique is used to search, filter and categorize the relevant patent documents in order to collect a complete dataset of patent documents. Next, the enhanced term frequency - inverse document frequency (ETF-IDF) technique is executed to elicit the patent keywords automatically from the selected patent documents. Moreover, the Viterbi algorithm is traditionally used to detect keywords through the HMM configuration (Cho, Kim, & Lee, 2010). Each path in the decoder is a sequence of keywords and garbage elements. The decoder finds scores for all possible paths, and the one with the highest score is selected as the output for the keyword set. Therefore, through using association rules which are put to combine the Viterbi algorithm with the Apriori algorithm into practice, the intelligent system produces the weighted value of each patent keyword in every patent document and further strengthens those keywords in iteratively appearing different patent documents to derive the really appropriate keywords. Finally, the sets of weighted patent keywords are employed to serve as the input base for generating a sophisticated patent network in order to effectively implement patent analysis.

In order to assure the utility of the intelligent patent network analysis method, patents in the field of Carbon Nanotube Backlight Unit (CNT-BLU), an emerging nanotechnology, are analyzed. CNT-BLU is a new product that uses Carbon Nanotube (CNT) in the design of a back light unit for a Thin Film Transistor Liquid Crystal Displays (TFT-LCD). It has the advantages of low cost, less power consumption, no need of optical films, no toxic chemicals, and superior color performance (Kim & Yoo, 2005). The reason why CNT-BLU was selected as an example in this study is as follows. First, CNT-BLU is an emerging nanotechnology that was developed to meet urgent demands for flat panel display. Second, CNT-BLU is suitable for exploring technological trends because of its rapid technical progress. Finally, the patent dataset of CNT-BLU is a convenient size for analyzing technological information and mapping the patent network. More detailed processes for the four stages of the proposed method are described as follows.

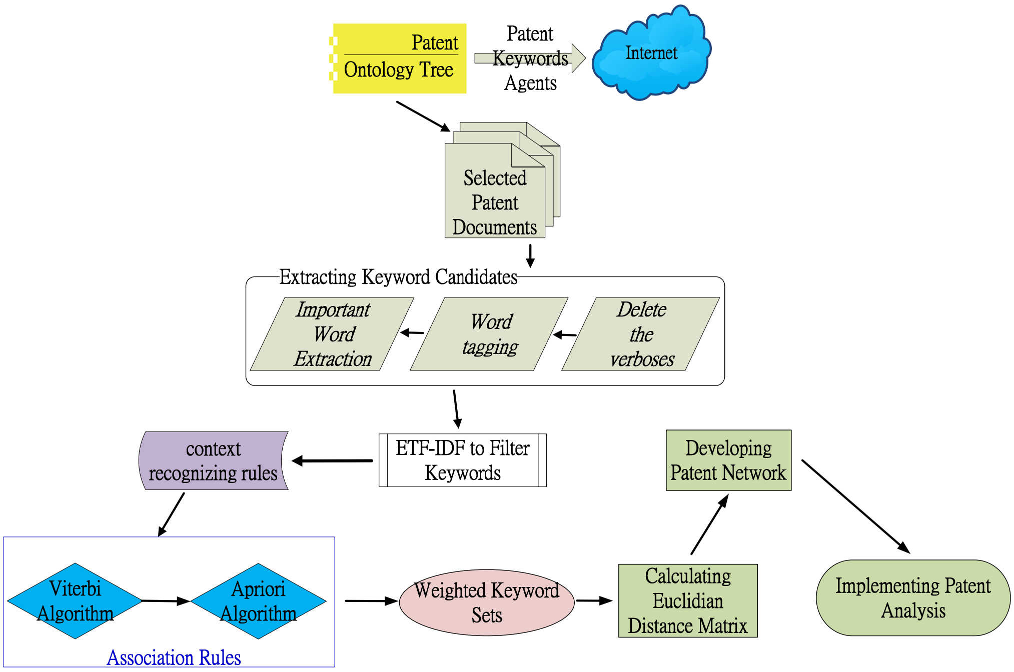

Figure 1. The overall procedure of the intelligent patent network analysis method

3.1 Selection of patent documents

Ontology is a formal representation of knowledge in artificial intelligence and knowledge management as a set of concepts including their attributes within a domain, and the relationships between those concepts (Noy & McGuinness, 2001). An ontology is used to systematically understand the entities within some domain and may be used further to automatically process the information of this domain, such as documents. Therefore, an ontology which is a "formal and definite specification of a shared epistemology" provides a shared knowledge architecture as a method that can effectively discovery and organize a domain with the definitions of objects and notions and relations to classify for much of the information on the internet to build up the semantic web (Brank, Grobelnik, Frayling, & Mladenic, 2002).

This study applies an ontology tree relevant to the field of patented technology being studied, in this case CNT-BLU, to automatically locate the relevant patent documents from the United States Patent Classification (UPC) database (United States Patent and Trademark Office, 2011), based on a keywords-based search to discover all related documents, which often cannot actually reflect the true meanings of the patent documents. The concept-based document searching method can be adopted to correctly classify the patent documents that belong to the field of technology being studied. This study uses the Protégé-2000 software (Bottou & Vapnik, 1992) to set up the ontology patent tree.

Many document retrieval technologies in the artificial intelligence field, seek to upgrade the accuracy of the document classification as an important focus (Guarino, 1998). This study combines the Salton method that automatically extracts the representative keywords from documents with the intelligent sorting document mechanism (Nowak & Wakulicz, 2005). The Salton method combines both methods of weighting by looking at both inter document frequencies and intra document frequencies. That is, by considering both the total frequency of the occurrence of a term in a document and its distribution over all documents, we can get the proper and exact term weighting values. Then, using linguistic rules, we automatically extract the representative keywords from all patent documents to further fix the proper weighting of each keyword in the keyword set. This is our improved TF-IDF algorithm (ETF-IDF). Finally, we utilize the association rule to assess the final word components in the keyword set of each patent document (Nowak & Wakulicz, 2005). By referencing the classification of the UPC to discover the category and layer of a patent document, this study is able to further filter the patent documents that are being searched. Subsequently, in order to improve the precision of the patent document classification, this study puts the resultant document through a patent classification process using a patent tree.

Through a series of searching procedures, the result reveals 97 relevant patent documents concerning CNT-BLU technology from U.S. patent numbers 6062931 to 7169005. The patent numbers and titles of these patent documents are shown in the Appendix. Because the patent numbers are too long to be usable for subsequent analysis, the patents were sorted by patent number and labeled with serial numbers from 1 to 97.

3.2 Extraction of patent keywords

3.2.1 Delete the verbose and word tagging in the patent article

After selecting the related patent documents in the specific field, as described above, the next stage extracts all possible special meaning words from these patent documents. In order to correctly process text segmentation of the English patent document, this study utilizes the stanfordLexParser-1.6 as a tool that processes English sentences. One of the great advantages of the stanfordLexParser-1.6 is that it can work well in the morphological restoration of any word and in syntactical analysis. This study introduces the stanfordLexParser-1.6 to process the three main patent contents - Abstract, Claim, and Description – in the document. The detailed steps in this stage are shown in Figure 2 and are implemented as follows:

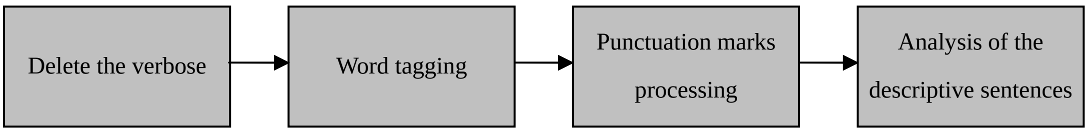

Figure 2. The steps of extracting words from patent documents

Step 1: Delete the verbose
This step segments the sentence according to different signs, ex: comma mark, full stop mark and period mark. Then, it constructs up a syntax representation tree and deletes all extra words in each sentence.

Step 2: Word tagging
In this step, the stanfordLexParser-1.6 program processes the word tagging. We added to its lexicon as references for many domain similar words to enhance the tagging result in order to get a syntax parse tree (Lyon, 1999).

Step 3: Punctuation marks processing
Because stanfordLexParser-1.6 segments sentences by punctuation marks, it can be achieved to get better results if the main different marks are dealt with and handled. Three types of punctuation marks may change the structure of sentences and should be refined in the processing to upgrade the understandings of context meanings in a sentence.

Step 4: Analysis of the descriptive sentences
The relationships of different parts-of-speech (POS) can be calculated by using their frequencies to disclose the syntax of partial structure in descriptive sentences. In particular, the POS of words are analyzed by following the major component keyword (MCK). The top-10 frequencies of the POS samples are shown in Table 1. Note that the frequency of a POS is based on the statistics of about 9000 sentences in the selected patent documents. In this study, we select only the words with the POS Na (noun), Nc (place noun), and VH (intransitive verb) for further study.

Table 1. The top-10 frequencies of parts-of-speech (POS)

(Na)*	2315
(VH)*	250
(Nc)	232
(V_2)	124
(Caa)	93
(D)	86
(Ncd)	78
(Nb)	61
(VG)	54
(FW)	43

3.2.2 Enhanced term frequency – inverse document frequency (ETF-IDF) and context recognizing rules

In this study, we focus on to amend the term frequency - inverse document frequency (TF-IDF) to strengthen those more important keywords which should have the higher weighting values. So, the ETF-IDF algorithm is upgraded from TF-IDF by considering the relative importance of each keyword in each patent document. TF-IDF is the most general weighting technology which has applied to classify the text categorizations in information retrieve. The TF-IDF function computes the weight of each vector component (each of them relating to a word of the vocabulary) of each document on the following basis. First, it incorporates the word frequency in the document. Therefore, the more a word appears in a document (e.g., its term frequency (TF) is high), the more it is estimated to be significant in this patent document. And thus, IDF measures how infrequent a word is in all patent document set and its value can be reasonably estimated.

Hence, if a word is very frequent in a document set, the IDF is not believed to be particularly representative of this document because it occurs in most patent documents, for instance, stop words and so on. On the contrary, if a word is infrequent in the document set, it is considered to be very relevant for the document in the field. Hence, by using frequency counting, the TF-IDF can identify the patent keywords and to reduce some mistakes in the filtering keywords process. Although the TF-IDF method can identify the keywords from the patent document, it cannot insure that the selected keywords are the best representative professional words. In other words, the patent keyword through our ETF-IDF filtering process can be more suitable and really keywords, so the enhanced TF-IDF algorithm is used to enhance these drawbacks of the original TF-IDF.

The ETF-IDF counts the frequency of each word in order to retrieve the meaningful words and compares a query vector with a document vector using a similarity or distance function, such as the cosine similarity function. There are several variants of TF-IDF. The following variant found by Yang & Liu (1999) was generally used in many experiments.

$$\text{Weight}_{t,d} = \log(tf_{t,d}+1)\log\frac{n}{X_t} \quad \text{if} \quad tf_{t,d} \geq 1\text{, otherwise} \qquad (1)$$
$$\text{Weight}_{t,d} = 0$$

where $tf_{t,d}$ is the frequency of word t in document d, n is the number of documents in the text collection, and x_t is the number of documents where word t occurs. Normalization to unit length is generally applied to the resulting vectors (unnecessary with KNN and the cosine similarity function).

To continue with the next step, this study discovers the real meaning of the context word and the importance of different keywords by further analyzing the syntactical relationship of the filtered words set. After several rounds, this approach can deduce the context recognizing rules that analyze the larger sets of patent documents. These context recognizing rules can help upgrade the accuracy of the selected keyword. The detailed steps are described as follows:

Step 1: Problem setting
This study addresses the problem of automatic extraction of semantic similarity relations among lexical items in relational form from which fine grained hierarchical clusters are obtained in the patent tree. In order to restrict the vocabulary and word ambiguity as well as to utilize information in abundant patent texts, this processing is confined to corpora from specific patent domains. This restriction is acceptable in the framework of Natural Language Processing (NLP) systems, which usually operate on sub-languages and are interested only in domain specific word meanings. Therefore, this process aims at developing a method applicable to every domain for which specific corpora are available in order to extract domain independent word meaning relations. Thus, this process can provide the semantic relations of the filtered keywords in relevance to thematic domains as well.

N-gram methods, which share the same perspective, focus on fast processing of large corpora and consider as context only immediately adjacent words without exploiting medium distance word dependencies (Venkataraman, 2001). Because large corpora are available only for few domains, this step aims at developing a method for processing small or medium sized corpora, exploiting as much as possible contextual information rich in semantic restrictions. The method is driven by the observation that in constrained domain corpora, the vocabulary and the syntactic structures are limited and that small or medium distance word or phrase patterns are often used to express similar facts. Stock market financial news and Modern Greek are used as domain and language test cases, respectively. Throughout the paper, examples taken from English corpora are also used.

Step 2: Context similarity estimation
Counting the number of occurrences of every semantic token found in the corpus, a frequency threshold under which no semantic clustering is attempted can be defined. Therefore, only Frequent Semantic Entities (FSE) are subjected to clustering (except the FSEs represented in the corpus by known patterns) while all but the rarest semantic tokens are used as clustering parameters. The corresponding frequency thresholds in the present experiments were set to 20 and 10 respectively in order to acquire sufficient contextual data for every FSE constraining computational time. Ideally, any word appearing at least twice in the corpus should be used as a context parameter. Definite determiners and verb auxiliaries are excluded from the processing because they have no semantic connection with their head words while pronouns are handled as semantically empty words.

Through the above processes, a total of 12 patent keywords were automatically extracted from the selected patent documents. Then, experts who work in the field of CNT-BLU further reviewed these keywords in order to confirm the correctness of automatic extraction. Consequently, all of the representative keywords with important technical features were included: “nanotube”, “backlight”, “display”, “emission”, “vacuum”, “electrode”, “cathode”, “anode”, “phosphor”, “thin film”, “binder”, and “fluorescent”.

3.3 Determination of the weight of each patent keyword

The conventional approach to detect keywords is Viterbi decoding through the HMM configuration (Cho, Kim, & Lee, 2010). Each path in the decoder is a sequence of keyword and garbage elements. The decoder finds scores for all possible paths, and the one with the highest score is selected as the output. This score is related to the joint probability of the path and the feature vectors. This scoring approach concerns the keyword spotting task. The score is a global score estimated by accumulating all likelihoods for the whole expression.

The score is not normalized with respect to the probability of the acoustic observation and thus is relative to the particular acoustic observation space (Ketabdar, Vepa, Bengio, & Bourlard, 2006). For example, it can be related to the length of the utterance, the length and number of keywords and garbage elements, the numerical range for values of evidences, etc. The values of these scores are penalized by changing keyword and garbage entrance penalties, which are effective spotting thresholds in this approach. There is no meaningful interpretation for the entrance penalty values, and they should be adjusted empirically to optimize the performance criteria. This implies that for each keyword there should be a sufficiently large development or training set. It would be ideal if we could find a reasonable threshold based on keyword characteristics, such as length, which can be known a priori or easily estimated or measured instead of adjusting in a development set.

The Apriori algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules (Agrawal, Imielinski, & Swami, 1993; Yang & Liu, 1999). In the fields of computer science and data mining, Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers or details of a website frequentation). The algorithm attempts to find subsets which are common to at least a minimum number C (the cutoff, or confidence threshold) of the itemsets.

In other words, Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time, a step known as candidate generation, and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Apriori uses breadth-first search and a hash tree structure to count candidate item sets efficiently. Through the above steps, the patent keyword set that contains the individual weighted value of each keyword is automatically derived and shown in Table 2.

Table 2. Patent keyword set in the field of CNT-BLU

keywords	weighted values	keywords	weighted values	keywords	weighted values
nanotube	0.176	vacuum	0.031	phosphor	0.027
backlight	0.158	electrode	0.086	thin film	0.101
display	0.112	cathode	0.103	binder	0.022
emission	0.063	anode	0.102	fluorescent	0.019

Note: The sum of weighted values is equal to one.

3.4 Generation of the patent network

In this stage, several techniques are employed to generate the patent network. The detailed content is described as follows:

Step 1: Counting the occurrence frequency of keywords in each patent document and then the weighted value of each keyword multiplied by the occurrence frequency to generate the weighted occurrence frequency of keywords in each patent document. The final values of each patent are integrated into keyword vectors as below:

Patent 1: $(p_{11}, p_{12}, p_{13}, \cdots, p_{1n})$

Patent 2: $(p_{21}, p_{22}, p_{23}, \cdots, p_{2n})$

$\vdots \qquad\qquad \vdots$

Patent m: $(p_{m1}, p_{m2}, p_{m3}, \cdots, p_{mn})$ (1)

For example, p_{11} is the weighted occurrence frequency of the first keyword in the Patent 1.

Step 2: Utilizing Euclidian distance to calculate the distance among the patents and to establish the relationship among patents. The Euclidian distance value (E_{ik}^{d}) between the two vectors is computed as follows:

$$E_{ik}^{d} = \sqrt{(p_{i1} - p_{k1})^2 + (p_{i2} - p_{k2})^2 + \cdots + (p_{in} - p_{kn})^2} \qquad (2)$$

Step 3: Transforming the real values of E^{d} matrix into the standardized values of E^{s} matrix in order to graph the patent network for next procedure.

$$E_{ik}^{s} = \frac{E_{ik}^{d}}{Max(E_{ik}^{d}, i = 1, \cdots, m; k = 1, \cdots, m)} \qquad (3)$$

Step 4: The cell of the E^{s} matrix must be a binary transformation, comprising 0s and 1s if it is to exceed the cut-off value q:

$$I_{ik} = \begin{cases} 1, & \text{if } E_{ik}^{s} < q \\ 0, & \text{if } E_{ik}^{s} \geq q \end{cases} \qquad (4)$$

The I matrix includes the binary value where I_{ik} equals 1 if patent i is strongly connected with patent k. I_{ik} equals 0 if patent i is weakly connected with patent k or not at all connected. That is, if the E_{ik}^{s} value is smaller than the cut-off value q, the connectivity between patent i and patent k is regarded as strong, and the I_{ik} value is set to 1. Otherwise, the connectivity is considered weak, and the I_{ik} value is set to 0. Through trying numerous cut-off values, $q = 0.10$ was chosen, which indicated that I_{ij} equaled 1 if E_{ij}^{s} was smaller than 0.10; otherwise I_{ij} equaled 0. Consequently, the binary matrix, I, was built for the implementation of the network analysis. The patent network was drawn by using UCINET 6.0 (Borgatti, Everett, & Freeman, 1999) and is shown in Figure. 3.

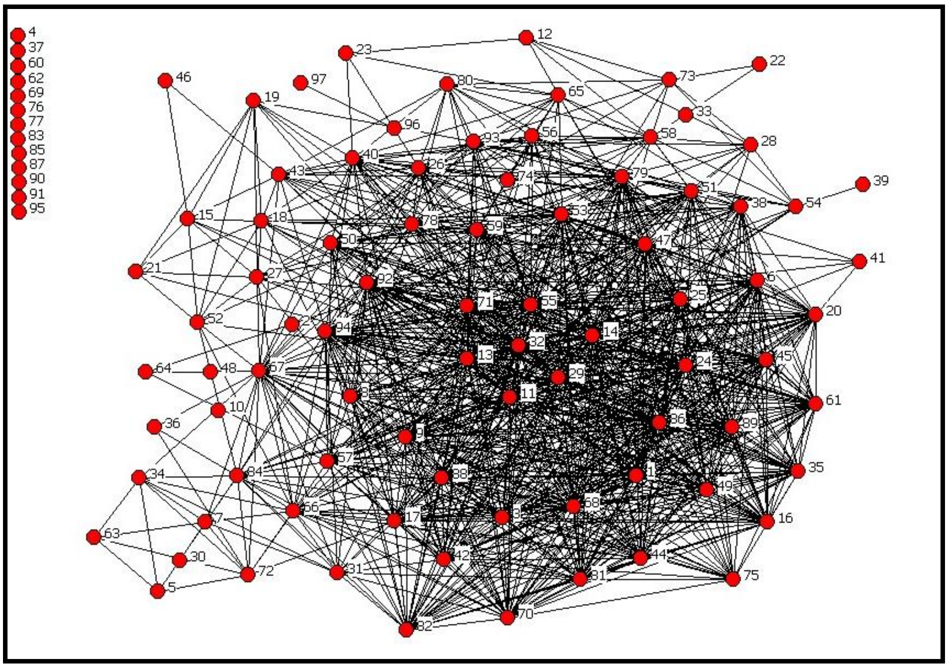

Figure 3. Patent network in the field of CNT-BLU

Figure 3 displays the overall patent network, which divides all 97 patents into interconnected and isolated sets. The interconnected set contains 84 patents and the relationship among these patents. It represents the focal point of the visual patent network and provides much information regarding the production and application of CNT-BLU. On the other hand, the isolated set includes the other 13 patents, which are quite divergent in the area of CNT-BLU. Thus, these inventions are excluded from the patent network through the above analysis process.

In the patent network, several patents that are closely located in the central position may represent the key technology in the field of CNT-BLU. In order to examine the structure of the network, the technology centrality index (TCI) can be calculated to identify the most important patents. The formula for calculating the TCI of patent i is shown below:

$$TCI_i = \frac{C_i}{n-1} \qquad (5)$$

$C_i = \sum r$, r: ties of patent i

where n denote the number of patents. This measures the relative importance of a subject patent by calculating the density of its linkage with other patents. That is, the higher the TCI, the greater the impact on other patents. The TCI can be used to identify the influential patents in the field of the technology being studied. Moreover, detailed information on these influential patents can be obtained. Technological implications can be deduced from the information as well.

Table 3 shows seven relatively important patents in the patent network with high TCI values, including No. 32, 13, 55, 29, 14, 11, and 71. The TCI values of these patents are all above 0.5 and far ahead of other patents. The core technology and developing trends in CNT-BLU were grasped by analyzing these patents in this study. Specifically, the core technologies focus on three main processes for making a CNT-BLU, including anode plate, cathode plate, and assembly of cathode and anode. Furthermore, the technological trend regarding the process of CNT-BLU manufacturing is CNT paste printing.

Table 3. TCI values of the relatively important patents in the patent network

No.	Patent number	TCI value
32	6616497	0.5313
13	6359383	0.5204
55	6803708	0.5109
29	6605589	0.5082
14	6380671	0.5058
11	6333968	0.5027
71	6903500	0.5016

4 CONCLUSIONS

This study constructs a novel patent analysis method, called the intelligent patent network analysis method, to make a precise visual network. Based on artificial intelligence techniques, this study proposes a detailed procedure for generating an intelligent patent network. First, this study utilized the concept of ontology to search and categorize relevant patent documents for collecting a complete dataset of patent documents. Second, through use of the enhanced term frequency – inverse document frequency (ETF-IDF) technique, reliable patent keywords suitable for further process analysis were extracted. Third, association rules were used to determine the weighted value of each keyword. Finally, sets of patent keywords were employed to serve as the input base for generating a sophisticated patent network. In order to assure the utility of the proposed method, the patents of CNT-BLU technology were analyzed in each stage as above. Several contributions regarding academic and practical implications are suggested as follows.

For academics, the contribution of this study is significant in terms of the methodology of patent analysis. Primarily, this study applies artificial intelligence techniques to modify current practice and proposes a rigorous method to make the visual network more sophisticated. The intelligent patent network analysis method provides a procedure for searching patent documents, extracting patent keywords, and determining the weight of each patent keyword in order to generate a precise visualization of a patent network. In this study, the effectiveness of the intelligent patent network has been verified by analyzing the patents of CNT-BLU technology. Compared with current methods, the proposed method has great improvements in terms of patent search, information extraction, visualization, and analysis.

For practical implications, the core technology and technological trends for CNT-BLU have been discovered through using the proposed method in this study. The practical application of the smart method was fully demonstrated. Thus, the intelligent patent network analysis method is valuable to the practical affairs of engineers or scientists. It enables engineers and scientists to intuitively understand the overview of a set of patents and to identify the developmental trends of critical technologies. Specifically, engineers and scientists are able to uncover significant technological information and grasp meaningful technological insights in the patent network.

Despite the above advantages, the proposed method has some challenges. For example, inevitable errors in the results of patent text categorization probably exist that would lead to the extraction of incorrect keywords. To resolve this problem, the automatic categorization results of the patent documents should be reconfirmed, that is, a mixed solution should be adopted that blends artificial intelligence and human intelligence to promote correctness and effectiveness when processing the abundant patent documents.

5 ACKNOWLEDGEMENTS

We are grateful for the financial support from the National Science Council in Taiwan (Grand No. NSC 99-2410-H-035-006) and Feng Chia University.

6 REFERENCES

Agrawal, R., Imielinski, T., & Swami, A. (1993) Mining association rules between sets of items in large databases. *In Proceedings of the 1993 ACM-SIGMOD International Conference on Management of Data*, New York, USA.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (1999) *UCINET 6.0 Version 1.00*, Harvard: Analytic Technologies Publishers.

Bottou, L., & Vapnik, V. (1992) Local learning algorithms. *Neural Computation*, 4, pp 888-900.

Brank, J., Grobelnik, M., Frayling, N., & Mladenic, D. (2002) Interaction of feature selection methods and linear classification models. *In Proceedings of 19th Conference on Machine Learning (ICML-02)*, Sydney, Australia.

Calero, C., Buter, R., Valdés, C. C., & Noyons, E. (2006) How to identify research groups using publication analysis: an example in the field of nanotechnology. *Scientometrics*, 66, pp 365-376.

Chang, P. L., Wu, C. C., & Leu, H. J. (2010) Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. *Scientometrics*, 82, pp 5-19.

Cho, Y. S., Kim, J. Y., & Lee, H. S. (2010) Efficient Viterbi scoring architecture for HMM-based speech recognition systems. *Electronic Letters*, 28, pp 2338-2340.

Crammer, K., & Singer, Y. (2002) On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning*, 2, pp 265-292.

Crevier, D. (1993) *AI: The Tumultuous History of the Search for Artificial Intelligence*, New York: Basic Books.

Cross, R., Borgatti, S. P., & Parker, A. (2001) Beyond answers: dimensions of the advice network. *Social Networks*, 23, pp 215-235.

Guarino, N. (1998) Formal ontology and information systems. *In Proceedings of the First International Conference (FOIS'98)*, Trento, Italy.

Horie, K., Maeno, Y., & Ohsawa, Y. (2007) Human-interactive annealing process with pictogram for extracting new scenarios for patent technology. *Data Science Journal*, 6, pp s132-s136.

Ketabdar, H., Vepa, J., Bengio, S., & Bourlard, H. (2006) Posterior based keyword spotting with a priori thresholds. *In Proceedings of Interspeech'2006*, Pittsburgh, USA.

Kim, Y. C., & Yoo, E. H. (2005) Printed carbon nanotube field emitters for backlight applications. *Japanese Journal of Applied Physics*, 44, pp L454-L456.

Knoke, D., & Kuklinski, J. (1982) *Network Analysis*, London: Sage Publications.

Leoncini, R., Maggioni, M. A., & Montresor, S. (1996) Intersectoral innovation flows and national technological systems: network analysis for comparing Italy and Germany. *Research Policy*, 25, pp 415-430.

Li, Y. R., Wang, L. H., & Hong, C. F. (2009) Extracting the significant-rare keywords for patent analysis. *Expert*

Systems with Applications, 36, pp 5200-5204.

Liu, C. Y., & Luo, S. Y. (2007a) Investigation of carbon nanotubes using the F-term code of Japanese patent information. *Data Science Journal*, 6, pp s255-s260.

Liu, C. Y., & Luo, S. Y. (2007b) Applying patent information to tracking a specific technology. *Data Science Journal*, 6, pp 114-120.

Liu, C. Y., & Yang, J. C. (2008) Decoding patent information using patent maps. *Data Science Journal*, 7, pp 14-22.

Lyon, M. (1999) Language related problems in the IPC and search systems using natural language. *World Patent Information*, 21, pp 89-95.

Narin, F. (1994) Patent bibliometrics. *Scientometrics*, 30, pp 147-155.

Nowak, A., & Wakulicz, A. (2005) The efficiency of the rules' classification based on the cluster analysis method and Salton's method. *Advances in Soft Computing*, 28, pp 333-338.

Noy, N. F., & McGuinness, D. L. (2001) Stanford Knowledge Systems Laboratory Technical Report, *Ontology Development 101: A Guide to Creating Your First Ontology*. California: Stanford University.

Salton, G., & McGill, M. J. (1983) *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.

Shin, J., Lee, W., & Park, Y. (2006) On the benchmarking method of patent-based knowledge flow structure: Comparison of Korea and Taiwan with USA. *Scientometrics*, 69, pp 551-574.

United States Patent and Trademark Office. Retrieved from the World Wide Web on October 29, 2011: http://www.uspto.gov/patents/resources/classification/

Venkataraman, A. (2001) A statistical model for word discovery in transcribed speech. *Computational Linguistics*, 27, pp 351-372.

Wasserman, S., & Faust, K. (1994) *Social Network Analysis: Methods and Application*, Cambridge: Cambridge University Press.

Wu, C. C., & Yao, C. B. (2012) Developing a framework for automatic patent information analysis approach. *International Journal of Digital Content Technology and its Applications*, 6, pp 435-442.

Yang, P. C. (2007) *On the automatic extracting and matching of object features between expertise corpus and oral description—A study based on the expert corpus of birds*, Master thesis, Chang Jung University, Tainan, Taiwan.

Yang, Y., & Liu, X. (1999) A re-examination of text categorization methods. *In Proceedings of ACM International Conference on Research and Development in Information Retrieval (SIGIR)*, New York, USA.

Yoon, B., & Park, Y. (2004) A text-mining-based patent network: analytical tool for high-technology trend. *Journal of High Technology Management Research*, 15, pp 37-50.

6 APPENDIX

Patent numbers and titles of CNT-BLU patents

No.	Patent Number	Title
1	6062931	Carbon nanotube emitter with triode structure
2	6097138	Field emission cold-cathode device
3	6232706	Self-oriented bundles of carbon nanotubes and method of making same
4	6239547	Electron-emitting source and method of manufacturing the same
5	6250984	Article comprising enhanced nanotube emitter structure and process for fabricating article
6	6278231	Nanostructure, electron emitting device, carbon nanotube device, and method of producing the same
7	6283812	Process for fabricating article comprising aligned truncated carbon nanotubes
8	6297592	Microwave vacuum tube device employing grid-modulated cold cathode source having nanotube emitters
9	6312303	Alignment of carbon nanotubes
10	6339281	Method for fabricating triode-structure carbon nanotube field emitter array
11	6333968	Transmission cathode for X-ray production
12	6346775	Secondary electron amplification structure employing carbon nanotube, and plasma display panel and back light using the same
13	6359383	Field emission display device equipped with nanotube emitters and method for fabricating
14	6380671	Fed having a carbon nanotube film as emitters
15	6426590	Planar color lamp with nanotube emitters and method for fabricating
16	6436221	Method of improving field emission efficiency for fabricating carbon nanotube field emitters
17	6440761	Carbon nanotube field emission array and method for fabricating the same
18	6445122	Field emission display panel having cathode and anode on the same panel substrate
19	6448709	Field emission display panel having diode structure and method for fabricating
20	6479939	Emitter material having a plurlarity of grains with interfaces in between
21	6486599	Field emission display panel equipped with two cathodes and an anode
22	6507146	Fiber-based field emission display
23	6512235	Nanotube-based electron emission device and systems using the same
24	6515415	Triode carbon nanotube field emission display using barrier rib structure and manufacturing method thereof

25	6515639	Cathode ray tube with addressable nanotubes
26	6522055	Electron-emitting source, electron-emitting module, and method of manufacturing electron-emitting source
27	6541906	Field emission display panel equipped with a dual-layer cathode and an anode on the same substrate and method for fabrication
28	6545396	Image forming device using field emission electron source arrays
29	6605589	Field emission devices using carbon nanotubes and method thereof
30	6607930	Method of fabricating a field emission device with a lateral thin-film edge emitter
31	6616495	Filming method of carbon nanotube and the field emission source using the film
32	6616497	Method of manufacturing carbon nanotube field emitter by electrophoretic deposition
33	6628053	Carbon nanotube device, manufacturing method of carbon nanotube device, and electron emitting device
34	6630772	Device comprising carbon nanotube field emitter structure and process for forming device
35	6639632	Backlight module of liquid crystal display
36	6645028	Method for improving uniformity of emission current of a field emission device
37	6645402	Electron emitting device, electron emitting source, image display, and method for producing them
38	6646382	Microminiature microwave electron source
39	6648711	Field emitter having carbon nanotube film, method of fabricating the same, and field emission display device using the field emitter
40	6652923	Electron-emitting source, electron-emitting module, and method of manufacturing electron-emitting source
41	6664722	Field emission material
42	6667572	Image display apparatus using nanotubes and method of displaying an image using nanotubes
43	6672925	Vacuum microelectronic device and method
44	6692791	Method for manufacturing a carbon nanotube field emission display
45	6700454	Integrated RF array using carbon nanotube cathodes
46	6703615	Light receiving and emitting probe and light receiving and emitting probe apparatus
47	6705910	Manufacturing method for an electron-emitting source of triode structure
48	6720728	Devices containing a carbon nanotube

49	6739932	Field emission display using carbon nanotubes and methods of making the same
50	6741017	Electron source having first and second layers
51	6741026	Field emission display including carbon nanotube film and method for fabricating the same
52	6750604	Field emission display panels incorporating cathodes having narrow nanotube emitters formed on dielectric layers
53	6774548	Carbon nanotube field emission display
54	6794814	Field emission display device having carbon nanotube emitter
55	6803708	Barrier metal layer for a carbon nanotube flat panel display
56	6806637	Flat display and method of mounting field emission type electron-emitting source
57	6812480	Triode structure field emission display device using carbon nanotubes and method of fabricating the same
58	6812634	Graphite nanofibers, electron-emitting source and method for preparing the same, display element equipped with the electron-emitting source as well as lithium ion secondary battery
59	6815877	Field emission display device with gradient distribution of electrical resistivity
60	6828722	Electron beam apparatus and image display apparatus using the electron beam apparatus
61	6838297	Nanostructure, electron emitting device, carbon nanotube device, and method of producing the same
62	6858990	Electron-emitting device, electron source, image forming apparatus, and method of manufacturing electron-emitting device and electron source
63	6882094	Diamond/diamond-like carbon coated nanotube structures for efficient electron field emission
64	6882112	Carbon nanotube field emission display
65	6885010	Carbon nanotube electron ionization sources
66	6890230	Method for activating nanotubes as field emission sources
67	6891319	Field emission display and methods of forming a field emission display
68	6897603	Catalyst for carbon nanotube growth
69	6897620	Electron emitter, drive circuit of electron emitter and method of driving electron emitter
70	6900580	Self-oriented bundles of carbon nanotubes and method of making same
71	6903500	Field emitter device comprising carbon nanotube having protective membrane
72	6911767	Field emission devices using ion bombarded carbon nanotubes

73	6917156	Fiber-based field emission display
74	6930313	Emission source having carbon nanotube, electron microscope using this emission source, and electron beam drawing device
75	6933674	Plasma display panel utilizing carbon nanotubes and method of manufacturing the front panel of the plasma display panel
76	6946800	Electron emitter, method of driving electron emitter, display and method of driving display
77	6975074	Electron emitter comprising emitter section made of dielectric material
78	7034447	Discharge lamp with conductive micro-tips
79	7041518	Low-temperature formation method for emitter tip including copper oxide nanowire or copper nanowire and display device or light source having emitter tip manufactured using the same
80	7053538	Sectioned resistor layer for a carbon nanotube electron-emitting device
81	7060356	Carbon nanotube-based device and method for making carbon nanotube-based device
82	7064474	Carbon nanotube array and field emission device using same
83	7067970	Light emitting device
84	7070472	Field emission display and methods of forming a field emission display
85	7071628	Electronic pulse generation device
86	7081030	Method for making a carbon nanotube-based field emission display
87	7083288	Illumination apparatus and image projection apparatus using the illumination apparatus
88	7115013	Method for making a carbon nanotube-based field emission display
89	7125308	Bead blast activation of carbon nanotube cathode
90	7129642	Electron emitting method of electron emitter
91	7138760	Electron emission device and electron emission display having beam-focusing structure using insulating layer
92	7147534	Patterned carbon nanotube process
93	7157848	Field emission backlight for liquid crystal television
94	7160169	Method of forming carbon nanotube emitters and field emission display (FED) including such emitters
95	S7161185	Display device and electronic device
96	7164224	Backlight having discharge tube, reflector and heat conduction member contacting discharge tube
97	7169005	Method of producing a backlight having a discharge tube containing mercury

(Article history: Received 29 March 2012, Accepted 2 September 2012, Available online 12 November 2012)