Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 35 (2008) 124–135
Expert Systems with Applications
www.elsevier.com/locate/eswa

On the development of a technology intelligence tool for identifying technology opportunity

Byungun Yoon *

Department of Industrial & Systems Engineering, School of Engineering, Dongguk University, 3-26 Pil-dong, Chung-gu, Seoul 100-715, Korea

Abstract

Technology intelligence tools have come to be regarded as vital components in planning for technology development and formulating technology strategies. However, most such tools currently focus on providing graphical frameworks and databases to support the process of technology analysis. *Techpioneer*, the proposed tool in this paper, aims to offer decisive information in order to identify technology opportunities. To this end, the system uses textual information from technological document databases and applies morphology analysis to derive promising alternatives and conjoint analysis to evaluate their priority. In addition, the method used in developing a technology dictionary is presented, employing clustering and network analysis. This system also has the ability to communicate with experts in order to estimate the value of existing patents, which is inevitable for the priority-setting of alternatives, construct a morphological matrix and so on. This paper presents the system architecture and functions of this tool and moreover, illustrates the prototype implementation and case study of the same.

Keywords: Morphology analysis; Text mining; Technology intelligence tool; Technology opportunity analysis

1. Introduction

Technology development mostly depends on the creativity of technologists or product designers. Since unprecedented opportunities in developing new technology must be explored by domain experts, their novel, innovative ideas are fundamental for successful technology development (Arai, 2006; Yang & Liu, 2006). Creativity can be defined as the ability to discern new relationships, examine subjects from new perspectives and form new concepts from existing notions (Couger, 1995; Evans, 1990). However, many factors such as cognitive, environmental and personality variables affect the achievement of creativity (Eysenck, 1995). Furthermore, researchers have found that creativity is more dependent on an idea-nurturing environment than an individual genius (Gatignon, Tushman, Smith, & Anderson, 2002). Therefore, creativity-related work needs to be enhanced by a systematic approach which facilitates an idea generation process and provides valuable information.

As a method to improve the performance of technology development with creativity, technology intelligence has been introduced to identify potential alternatives for new technology and reduce the probability of failure in the face of technological discontinuities (Cooper & Schendel, 1976; Utterback & Brown, 1972). This notion includes technology monitoring, technology assessment, technology forecasting and so on (Lichtenthaler, 2004). Technology intelligence has several advantages in comparison with an expert-based approach in technology management. Firstly, it can deal with massive volumes of information which cannot be analyzed by humans alone. The lack of information usage might bear a biased output in technology analysis. Secondly, technology intelligence tools can generate significant amounts of information which humans cannot produce. They can visualize the relationship between technology and companies, and analyze the characteristics of technology with statistical analysis. Finally,

* Tel.: +82 2 2260 8659; fax: +82 2 2269 2212.
E-mail address: postman3@dongguk.edu

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.06.022

technology intelligence is capable of managing updated information which is created in the internet as well as within companies. Systems which support a decision-making process with technology intelligence can handle real-time data and respond to the requirement of users quickly.

Recently, these systems have been actively developed to assist researchers and practitioners to make strategic technology plans. TRIZ, the Soviet inventive problem-solving system, has been used extensively to resolve conflicts of parameters in products and technology, applying 40 fundamental solutions to perceived problems (Kobayashi, 2006). TRIZ is built on the systematic study of over two million patents by encapsulating the best practices of the world's most successful patents. However, most recent studies in technology intelligence tools tackle the application of data mining techniques to technology management. Diva, a visualization system for exploring document databases, helps to perform bibliometric analysis of scientific literature and patents for technology forecasting (Morrisa, DeYongb, Wua, Salmanb, & Yemenub, 2002). In addition, intelligence systems such as VantagePoint and Aureka can support analysts to discover relationships or trends in technology by providing clustering, mapping, searching techniques and so on (Zhu & Porter, 2002; Trippe, 2003). However, in general, existing tools are no more intelligent as their brand names. Although they can show the trends in technological keywords and the relationship between patents, this information hardly provides the potential for a technology breakthrough. Advanced technology intelligent tools need to identify promising opportunities for new technology development by investigating the detailed characteristics of technology, rather than the historical change of keywords in their documents.

As a remedy, this paper proposes a supporting system that uses text mining and morphology analysis in order to understand the trend in the morphology of technology and excavate potential technology opportunities from documents. *TechPioneer*, as discussed earlier, has a communication function with domain experts, which enables reflecting the expertise knowledge in defining the structure of a technological feature and estimating the value of a patent. Additionally, it can support the process of generating a technology dictionary that consists of groups of technological keywords. Therefore, this tool takes a hybrid approach which is supported by both computational algorithms and experts, providing valuable information such as the trends of keywords, the morphology of existing technology, the structure of technology and promising opportunities of technology development.

This paper is organized as follows. Section 2 reviews the background of technology forecasting based on text mining. Section 3 describes the systemic architecture of *TechPioneer,* and its functions are presented in Section 4. Section 5 explains the prototype implementation and case study. Section 6 discusses the limitations and future research of this paper.

2. Text mining-based technology forecasting

Text mining can be defined as a knowledge discovery process which automatically extracts useful information and analyzes significant patterns from large amounts of textual data (Atkinson, 2002). The ability of this technique to elicit unknown, beneficial information allows itself to be distinguishable from search engines or database management systems. Many researchers have carried out advanced text mining studies, theoretically and practically, which can be classified into four categories. Firstly, text mining can be employed to cluster documents on the basis of their similarity. K-means and hierarchical clustering are used to group textual data (Dhillon & Modha, 2001) and on-line clustering has been suggested in response to requirements for readiness and effectiveness of clustering in an internet era (Cutting, Karger, Pedersen, & Tukey, 1992). Secondly, many studies deal with the subject that classifies documents into a category by a predefined pattern. Various techniques such as a machine learning and k-nearest neighbor (k-NN) technique can be applied for automatic text classification (Cohen & Hirsh, 1998; Sebastiani, 2002). The third category is to study information extraction from texts. This research examines the subject of documents (Clifton, 2004) and summarizes the contents of them in an automatic way (Mani & Maybury, 1999). Finally, various approaches to visualizing the relationship among documents have been presented in order to produce information regarding the linkage among the contents of documents as well as their authors (Rohrer, Ebert, & Sibert, 1998).

The advantage of text mining has recently been acknowledged by technology management. Profiles of research projects are identified by visualizing keywords of documents (Porter, Kongthon, & Lu, 2002) and the trend in technological growth is investigated by tracking the historical change of keywords (Watts & Porter, 1997). However, many researchers have recognized the potential of text mining in technology forecasting. Zhu and Porter (2002) introduce a family of maps that help convey emphases, players and patterns in a technology trajectory through text mining. Yoon and Park (2004) propose a keyword-based morphology analysis to identify the detailed configurations of promising technology. Kostoff, Boylan, and Simons (2004) suggest a systematic approach to exploring disruptive technology that is realistic and operable by taking advantage of text mining literature. Van Raan and van Leeuwen (2002) assess the economic and technological value of research on the basis of keywords and forecast promising research to influence technology landscape.

3. System architecture

3.1. Basic concepts

TechPioneer is a technology intelligence tool that supports the process of technology forecasting by identifying new technology opportunities systematically. Most

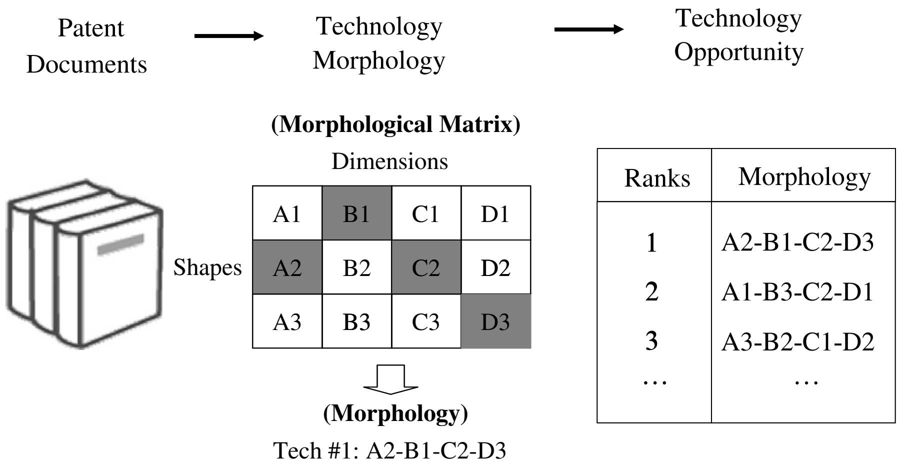

Fig. 1. Conversion of technological information.

traditional technology forecasting generally depends on the intuition of domain experts and qualitative data, which might yield problematic issues such as over-investment of time and cost, as well as the lack of consideration for large volumes of useful information. Therefore, this tool needs to be designed to utilize quantitative data, employing various systematic methodologies that can be computerized, for example text mining, morphology analysis and conjoint analysis. Patent documents are used as a principal data source in order to derive the morphology of existing patents and identify any undeveloped configuration of a technology. *TechPioneer* is a semi-automated tool because a computerized process to extract keywords and investigate the morphology of technology has to be bolstered by the knowledge of domain experts. While systemic processes such as keyword extraction or priority calculation can be practised by a computer system, the modeling and definition of morphology must reflect the experienced opinion of experts. Fig. 1 explains the core concept of this research, which converts textual information into technology opportunity by analyzing technological morphology.

3.2. Architecture

The architecture of this tool consists of three modules – text mining module, morphology analysis module and decision-making module. Each module that includes subordinate engines and processes is conducted sequentially and communicates with specific databases to retrieve and store data. The text mining module is developed for data preprocessing in order to carry out the following main processes. In this module, documents that are collected from external databases such as US Patent Trade Organization (USPTO), Europe Patent Organization (EPO) and other international patent organizations' databases are transformed into a patent keyword vector by a search engine, a keyword extractor and a keyword vector builder. In the morphology analysis module, a keyword vector of each patent is converted into a morphological configuration by matching the keyword vector with a morphological structure. The profiles of new technology opportunities are excavated by listing the morphology of existing patents and discovering unprecedented morphology. In the decision-making module, promising profiles from all alternatives are identified by evaluating the value of existing patents and estimating the future value of alternatives. Fig. 2 shows the architecture of *TechPioneer* in detail.

3.3. Text mining module (TMM)

Although many data sources have been analyzed for technology forecasting, patent information is regarded as one of the most critical data sources to explain the technological evolution over the whole territory of technology. Moreover, it is public information that has high accessibility and whose value is guaranteed by government privilege legally. Therefore, patent documents are chosen as principal data to explore technology opportunity. However, since patent documents in a text format are unstructured data, it is demanding for researchers to investigate all patent documents of interest and to identify the configuration of each patent. Thus, the format of all patent documents needs to be changed into structured data – keyword vectors. In order to facilitate this process, the text mining module (TMM) requires three engines – a search engine, a keyword extractor and a keyword vector builder. In addition, a technology dictionary must be prepared to support the procedure of keyword extraction and a keyword vector database is necessary to store outputs that will be utilized in the subsequent modules.

3.3.1. Search engine

Patent documents must be initially collected to execute the main analysis of technology opportunity. To this end, this system can connect with public or commercial databases, including various databases of international patent organizations. If a company has its own patent database in advance, appropriate patent documents for

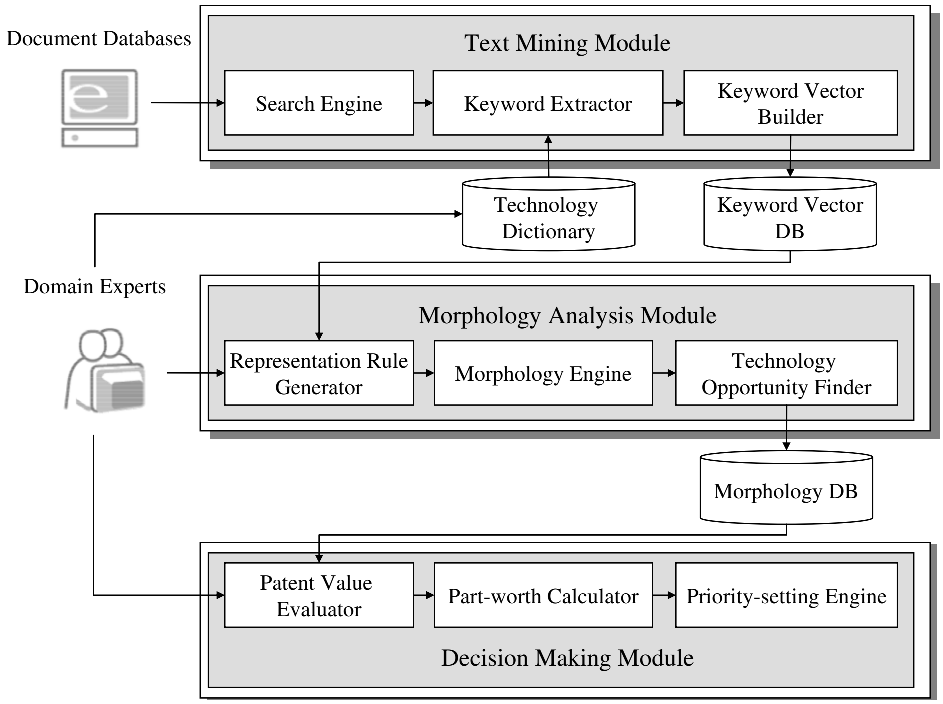

Fig. 2. Architecture of *TechPioneer*.

technology opportunity analysis can be searched from the database. This engine provides a combination search of keywords, reference period and patent classification to improve the performance of searching. In particular, a search by reference period is imperative, in that, the dynamic analysis of technology opportunity can be conducted to achieve the high accuracy of forecasting.

3.3.2. Keyword extractor

After patent documents of interest are gathered from patent database by the search engine, keywords must be extracted by a keyword extractor that calculates the occurrence frequency of words and refers to a technology dictionary. The technology dictionary as an ontology of technology has to be developed with the aid of domain experts and can be recursively updated by the outputs of keyword analysis. Namely, words that are closely associated with corresponding technology and appear frequently in patent documents can be selected as keywords. Peripheral words such as conjunctions and articles must be naturally eliminated because they do not need to be included in a technology dictionary. Therefore, only technology-related words can remain as keywords to develop a technology dictionary and examine the characteristics of technology. Fig. 3 depicts the procedure of extracting keywords[1].

[1] This flow chart represents the formal procedure of extracting keywords after a technology dictionary is developed. In the case of developing the technology dictionary, the process of this step must be located in the last step after the extraction of keywords.

3.3.3. Keyword vector builder

A keyword vector is composed of a data field that displays the frequency in which the extracted keywords occur in a patent document. In this procedure, various configurations of a word must be considered to accurately count the frequency of the word. In other words, since a word can have various forms such as adjective, noun and verb, and the frequency of diversely formed words must be aggregated to reduce the complexity and confusion of analysis. In addition, the frequency of similar semantic words needs to be summed to exactly identify the characteristics of technology. Even if a technology dictionary which is formed by a statistical analysis in this system generates groups of similar keywords, it might have unclearly divided categories of words. Therefore, the involvement of experts is essential in defining mutually exclusive categories and calculating the precise frequency of keywords.

3.3.4. Technology dictionary

The objectives of developing a technology dictionary are two-fold. First of all, in the procedure for extracting keywords, a technology dictionary can play a critical role in deriving important keywords from those words obtained by analyzing their occurrence frequency. Furthermore, it helps to identify the morphology of technology by matching keywords into associated morphology, as the hierarchical structure of keywords is useful to understand the relationship between keywords and relevant morphology. A technology dictionary provides a topology of keywords, which means that the relationship among keywords is

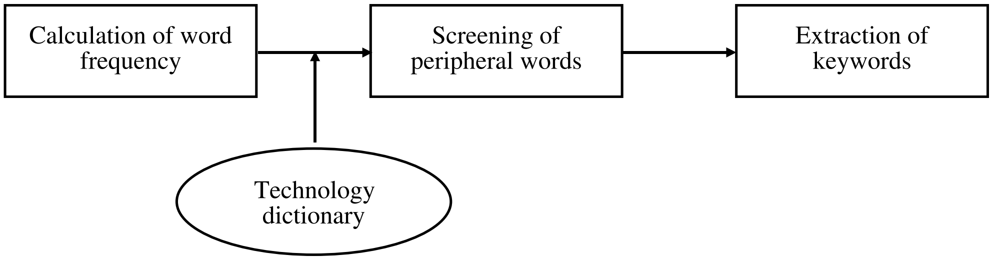

Fig. 3. Procedure of keyword extraction.

arranged in a tree format. It can be developed by the following process.

1. Patent documents are collected in a technological area of interest. A technology dictionary needs to include the whole sub-subclasses of the targeted technology. For example, even if technology opportunity in wide viewing angle (WVA) technology area is examined, the technology dictionary must be developed in the entire areas of TFT-LCD technology. This enables the extensive reflection of the relationship among technology.
2. The second step is to extract words from documents through text mining and decide on technology-related words by separating peripheral words. This process is subject to a semi-automated process. While the process of obtaining words is executed by text mining software, the screening of technology-related words must be finally conducted by experts.
3. A proximity matrix is constructed by analyzing the co-occurrence frequency of words. Two approaches can be considered to measure the relationship between words. In general, the number of times that two words occur together over the whole documents can be used to analyze their relationship. However, this approach might distort the relationship because the imbalance in the use of words usually exists, which means that some documents might include a high number of specific words but most other documents might not even contain those words. Therefore, the number of documents in which two words co-exist can reflect the authentic relationship among words and thereby, this paper uses the latter approach.
4. The fourth step derives the network of words through network analysis. With a proximity matrix, the network of keywords is drawn and then the subgroup of keywords is identified. A clique in network analysis is defined as a maximally complete sub-graph where all actors have linkage with all the other actors in the clique. Cliques can be used as a basis for generating a technology dictionary because they group intensively related keywords.
5. Finally, cliques are grouped by hierarchical clustering and the development of technology dictionary can be completed by selecting the level of the clustering. As a consequence, a technology dictionary is composed of groups of keywords on the basis of co-occurrence frequency.

Fig. 4 shows the example of a technology dictionary which consists of a hierarchical structure and where the

1. Hardware	**2. Software**	**3. Network**
1.1. Main Board	2.1. Application Software	3.1. Circuit Switching
1.2. Micro Processor	2.2. Data Processing Software	3.2. Packet Switching
1.3. Memory	2.3. Software Certification	3.3. ATM
1.4. Input/output	2.4. Computer Recognition	3.4. Routing

1.4.1. Keyboard – KW1, KW2, KW3,…
1.4.2. Mouse – KW11, KW12, KW13, …
1.4.3. Touch Screen – KW21, KW22, KW23,…

Fig. 4. Example of technology dictionary (computer-related technology).

group of keywords is located in leaf nodes. In Fig. 4, keywords are involved in the third level such as 'Keyboard', 'Mouse' and 'Touch Screen'.

3.4. Morphology analysis module (MAM)

All configurations of existing technology can be identified by mapping keywords of existing patents to related morphology. For this, the morphological structure of corresponding technology must be defined in advance and extracted keywords must be matched with pre-defined morphology. A representation rule generator helps domain experts define the morphology of each technology by providing a graphical framework of structuring the characteristics of each technology. A morphology engine examines the morphology of existing patents by analyzing the keyword vector of each patent with the representation rule. After listing a whole set of morphology of traditional patents, undeveloped configurations of technology are suggested by a technology opportunity finder. A morphology matrix of a technology relating to a mouse (a computer device) is exemplified in Table 1, consisting of several dimensions such as connection, method, material and movement control. The morphology of a mouse can be identified by combining a specific level in each shape. For example, a common mouse can be described as combining a wireless connection, an optical method, a plastic material and a control by a scroll.

This matrix is comprised of four dimensions that can each respectively be decomposed into two or three shapes. The keyword vector enables the identification of the morphological shape of patents. For instance, if a specific technology includes more keywords relating to 'wireless' than 'cord', the shape of the patent in the 'connection' dimension is described by 'wireless'. The shapes in the other dimensions are also decided in the same way, providing the complete combination of shapes. Consequently, undeveloped configurations in the morphological matrix can be listed by excluding the existing morphology from all possible configurations.

3.4.1. Representation rule generator

This engine conducts the process to generate a rule of converting extracted keywords into morphology. To this end, domain experts must construct a morphological matrix and match the keywords with the matrix. The representation rule generator requires experts to define dimensions and shapes in order to describe the morphology of each patent.

Table 1
Example of morphological matrix (a mouse)

Shapes	Dimensions			
	Connection	Method	Material	Movement control
1	Wireless	Optical	Plastic	Scroll
2	Cord	Mechanical	Metal	Trackball
3		Mixed		Joystick

However, although more detailed dimensions and shapes make the configurations of patents be more specific, they might yield many similar alternatives which are not worth being separated. Therefore, a morphological matrix needs to be defined in a concise but exclusive way. In addition, experts must identify a contradictory relationship that can exist between a pair of levels, as morphology analysis derives configurations of patent by combining the shapes in each dimension, and a pair of levels might be incompatible due to a technical problem. After a morphological matrix is built, representation rules that clarify which particular level in the matrix each keyword is located in are generated. Categories in a technology dictionary can be a reference to execute this process because groups of keywords provide a basis to explain the characteristics of a patent.

3.4.2. Morphology engine

With a representation rule, the configurations of all patents can be identified by a morphology engine. Keyword vectors which are generated in the TMM are transformed into a morphological form through the representation rule. This process is operated on the basis of the sum of frequency of keywords through the following algorithm.

$w_{ijk} = v_l$, $M_{ij} = \text{Sum}\ (w_{ijk})$
If $M_{ij*} > M_{i\cdot}$, then $j = j^*$ in ith dimension
Where $i = i$th dimension, $j = j$th shape, $k = k$th keyword in (i,j) morphology,
$l = l$th keyword in a keyword vector which is associated with kth keyword

Data fields in a keyword vector must be rearranged into a morphology vector. This vector is composed of the sum of frequency of keywords that belong to a particular morphology, which indicates that the number of a data field in a morphology vector amounts to (i^*j). For this, keywords which are matched with a specific level in a morphological matrix are grouped as similar keywords (w_{ijk}). Thus, frequencies of all keywords which belong to jth shape in ith dimension are summed, which becomes the value of each shape in the dimension. Consequently, a shape which has the maximum value in the occurrence frequency of keywords is chosen as an appropriate configuration of ith dimension in a patent. Fig. 5 presents the example of identifying the morphology of a patent by converting its keyword vector into morphology which combines the selected shapes.

3.4.3. Technology opportunity finder

A technology opportunity finder has the set of all possible configurations of patents in advance before the morphology engine identifies the existing configurations. Although the total number of alternatives can be derived by multiplying the number of shapes in each dimension, the contradictory relationship must be considered to reflect

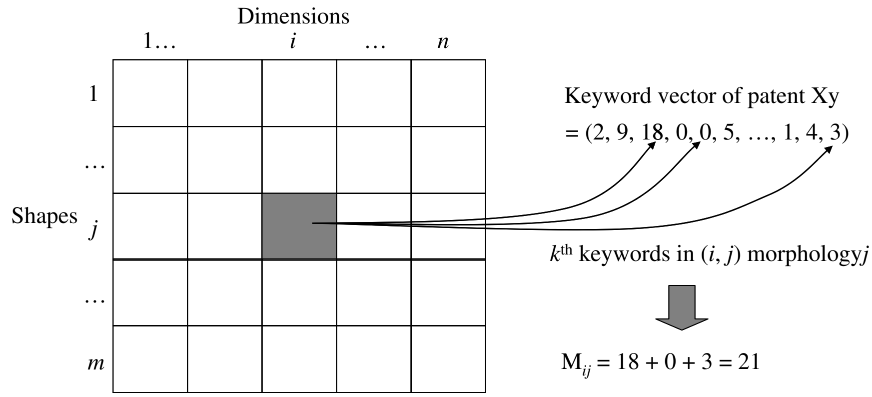

Fig. 5. Example of the process of identifying morphology from a keyword vector.

the feasibility of technology development. If a pair of shapes in two dimensions is incompatible, the number of alternatives which are eliminated in a set of opportunities can be calculated by multiplying the number of shapes in the remaining $(i-2)$ dimensions. Therefore, alternatives to be considered can be dramatically reduced by investigating incompatible pairs of shapes. This engine displays the unprecedented configurations by excluding existing configurations from all feasible configurations.

3.5. Decision-making module (DMM)

This tool ultimately aims at supporting a decision of managers to develop promising technology. Although a supporting tool generally offers substantial information for managerial decisions and facilitates an analytic process with functions, such as graphical frameworks and computerized algorithms, it more importantly needs to focus on aiding a decision-making process through a collaboration function. *TechPioneer* supplies an evaluation function with which experts can make a scoring for perceived alternatives. However, since the number of alternatives might be too enormous to explore exhaustively, in cases where a morphological matrix might consist of many dimensions and shapes, it is impossible for participants to deal with all alternatives in detail. Therefore, conjoint analysis can be employed to analyze the priority of them without all the alternatives being scrutinized manually.

3.5.1. Patent value evaluator

It might be difficult to evaluate the future value of undeveloped technology. However, if the value of existing technology can be used to estimate the value of unprecedented technology, the reliability and validity of valuation will be improved. To this end, this tool adopts conjoint analysis because its advantages are that the value of a subject is decomposed into part-worths of shapes, and by evaluating several configurations, the value of all the other configurations can be also calculated on the basis of the part-worth. For this, a patent value evaluator in this system requires analysts to assess the value of existing patents by rating or ranking their configurations. This process is similar to that of a negotiation tool which is conducted iteratively until they reach a final agreement. This engine facilitates this evaluating process to derive the agreed value of existing patents by providing the information such as the average of ratings for each configuration in each negotiation phase.

3.5.2. Part-worth calculator

A part-worth calculator performs a core process of conjoint analysis, in that the value of shapes which compose the configuration of a patent is derived from the results of evaluation of experts, and all alternatives can be prioritized on the basis of the value. Conjoint analysis can determine the contribution of each shape to the overall value of a profile and establish a predictive model for new combinations. To this end, OLS is adopted as a conjoint measurement model to derive the priority of alternatives. Although various methodologies such as OLS, LINMAP and MANOVA have been applied, the OLS approach is a straightforward means to obtain alternative forms of respondent utilities[2]. The objective of OLS is to produce a set of additive part-worth utilities that are drawn from the evaluation of each expert for each shape. The part-worth model is used as a utility preference model in this paper. In conjoint analysis, the contribution of a dimension (or shape) to the total utility is called a part-worth and the total utility of a profile is equal to the sum of their part-worths. Consequently, the part-worth calculator offers the value of dimensions and shapes to a following priority-setting engine. In addition, analysts can obtain the information about critical factors to develop advanced technology by grasping dimensions and shapes of high impact. The following equation explains the method of calculating the utility of a configuration of technology.

[2] OLS (ordinary least square); LINMAP (linear programming technique for multidimensional analysis for preference); MANOVA (multivariate analysis of variance).

$$U = \sum X_i \beta_i$$

Where U: utility of a profile, X_i: the value of dimension i,
β_i: weight parameter of dimension i

3.5.3. Priority-setting engine

The feasible configurations need to be ranked in order of their value. The priority-setting engine calls the list of feasible configurations from the technology opportunity finder and the part-worths of each shape from the part-worth calculator. The value of each configuration is estimated by summing the part-worths of shapes which it combines. All alternatives are sorted in order of their total utility. This engine displays the ranks and total utility of all feasible configurations, which enables analysts to discuss the results of priority setting and finally conclude a plan to develop technology.

4. System functions

4.1. Keyword trend analysis

Since keywords reflect a technological feature that is highlighted in a specific period, the trend of keywords can indicate the change of technological focuses. The TMM of *TechPioneer* creates the keyword vector of each patent and stores the data in a keyword vector database. The occurrence frequency of each keyword is examined according to a time period by analyzing patents that belong to the specific period. Consequently, emerging, novel keywords and declining, mature keywords are identified by generating the changing pattern of keywords in a timetable, which helps users grasp the transition of technological interests historically.

4.2. Morphology analysis

The proposed system can analyze the structural, procedural and functional morphology of specific technology. Thus, such morphological structures assist analysts to understand underlying characteristics of technology, from which current concepts can be elaborated into a breakthrough effectively. Furthermore, since this system provides a function to identify the morphology of technology, all morphology of existing technology can be investigated exhaustively. The historical change of morphology can be also analyzed by scrutinizing the dynamic pattern of technological morphology.

4.3. Critical technology attributes

It is important to clarify influential technological attributes in developing new technology or planning R&D projects. In this paper, a morphological structure is suggested in order to define the configuration of technology and conjoint analysis enables the identification of critical technology attributes with this structure by evaluating the value of alternatives. The part-worths of all dimensions and shapes are calculated, providing the ranks of significant attributes. This function helps researchers discern valuable dimensions and shapes that take key parts in the value of each technology and devise strategic plans for technology development.

4.4. Promising technology

The main objective of this tool is to suggest opportunities of new technology development. *TechPioneer* extracts potential configurations of new technology which have a high possibility of creating technological advancement and economical benefits. The priority of technology opportunities, which are based on the value of existing patents, allows analysts to intensively discuss the importance and feasibility of developing the perceived technology. Although the result is not a final optimal solution for technology development, it can be considered as useful information by researchers who otherwise should have started with unstructured documents and been mainly dependent on their experience.

5. Prototype implementation and case study

On the basis of the aforementioned system architecture, a software prototype of this system was implemented in order to demonstrate the usability of the suggested process and practical tool. The prototype provides promising technology opportunities by analyzing technological documents and employing diverse data mining techniques. Visual Basic 7.0 is a primary programming language for system implementation in this paper and Microsoft Access serves as a database to store the results of analysis and as a mean of communication with the engines.

5.1. Data description

TechPioneer is operated by investigating patent documents and evaluating the value of alternatives. Therefore, patent documents and patent value are basic information to activate functions of this tool. Although patent documents are gathered from the patent databases of patent organizations in countries that have a national intellectual property system, patent value can be estimated through various ways according to the objective of valuation. In this paper, the rating of experts plays a role in evaluating the patent value because it must be assessed under synthetic analysis, considering both economical and technological factors. Thus, TFT-LCD patent documents are collected to demonstrate this approach. The period of the patents ranges from 1997 to 2006 and the total number of documents is equal to 308.

5.2. Information extraction from documents

Although the role of text mining varies from retrieval of documents to knowledge management, the recent trend of related research is to extract significant information from a large amount of documents. This paper is dedicated to propose an intelligence tool to identify the configurations of patents from their documents. Therefore, text mining derives keywords from all patents by calculating the occurrence frequency of words. With these keywords, a technology dictionary is developed by analyzing their relationship. In this research, 282 keywords are extracted from 308 patent documents, producing 13 keyword categories which can be named by examining the characteristics of grouped keywords. These categories are created by clustering cliques which are obtained by network analysis. 2092 cliques in the TFT-LCD patents are discovered in the network of keywords, which means that a keyword can be involved in more than two cliques. On the basis of the relationship that a pair of keywords is included in a same clique, 282 keywords are clustered into 13 categories.

5.3. Technology configuration identification

The configurations of the collected 308 patents are identified with their keywords. For this, first of all a morphological matrix must be defined by domain experts. This matrix is also linked to keywords with an aid of a technology dictionary. Fig. 6 shows a screen of *Techpioneer* which provides the information regarding the morphological matrix of the TFT-LCD technology and keywords that belong to each shape. In addition, the morphology engine in this tool examines the configurations of existing patents by combining shapes which are decided to have the highest sum of the frequency of keywords in each dimension. In this paper, 67 configurations are drawn from 308 patents by aggregating the same configurations.

6. Technology opportunity analysis

New opportunities for technology development are explored by excluding existing configurations from all possible alternatives. Since 67 existing configurations must be eliminated from 432 possible alternatives, the remaining 365 configurations can be considered as new opportunities. Although this list can be offered intact to analysts, the priority of alternatives enables them to conclude a plan for new technology development effectively. To this end, *TechPioneer* provides the high ranked configurations and the part-worths of shapes in all dimensions. Domain experts in TFT-LCD technology area evaluate the value of existing patents and the part-worth calculator estimates the part-worths of all shapes and dimensions. Fig. 7 presents those values in the bottom, right-hand side of the screen. A graph which depicts the proposition of each shape (or dimension) in the value of all shapes (or all dimensions) helps analysts understand the importance of each attribute visually.

The estimated value of new opportunities is derived by summing the part-worths of shapes which they take as their configurations. A configuration which includes the highest value of shapes ranks high in the list of alternatives. Consequently, all new opportunities are enumerated in order of value as shown in Fig. 8. In addition, the competitiveness of a company can be examined by comparing the sum of the value of patents that the company holds with that of its competitors. Therefore, if a company develops new

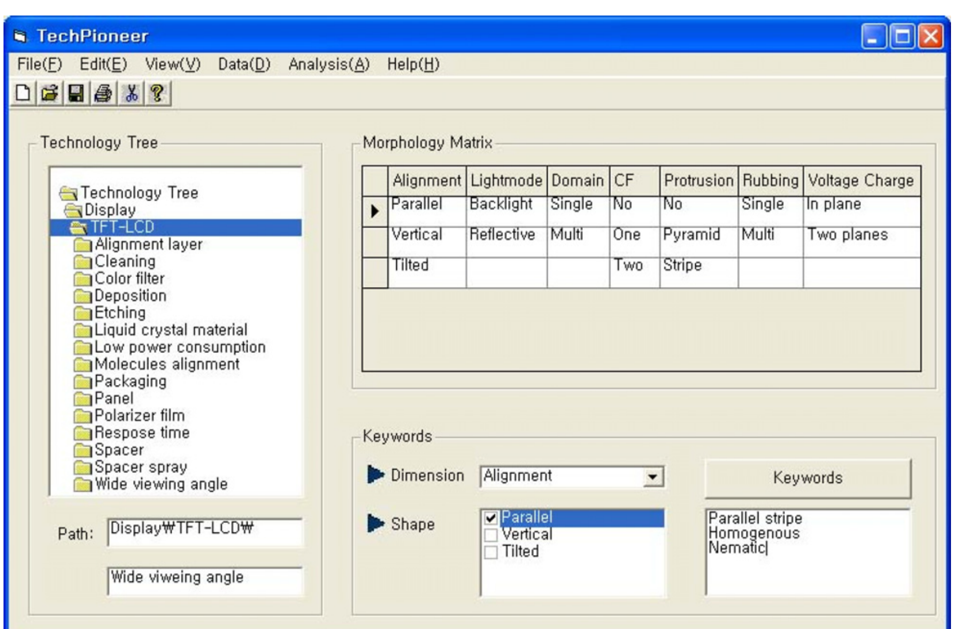

Fig. 6. Screen of morphological matrix and keywords.

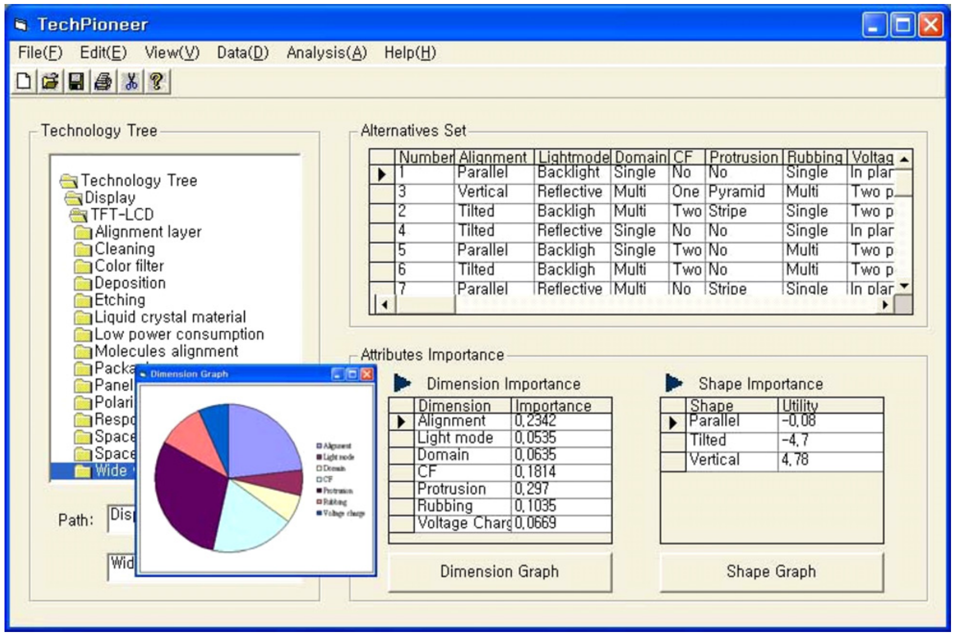

Fig. 7. Screen of alternatives list and attributes importance.

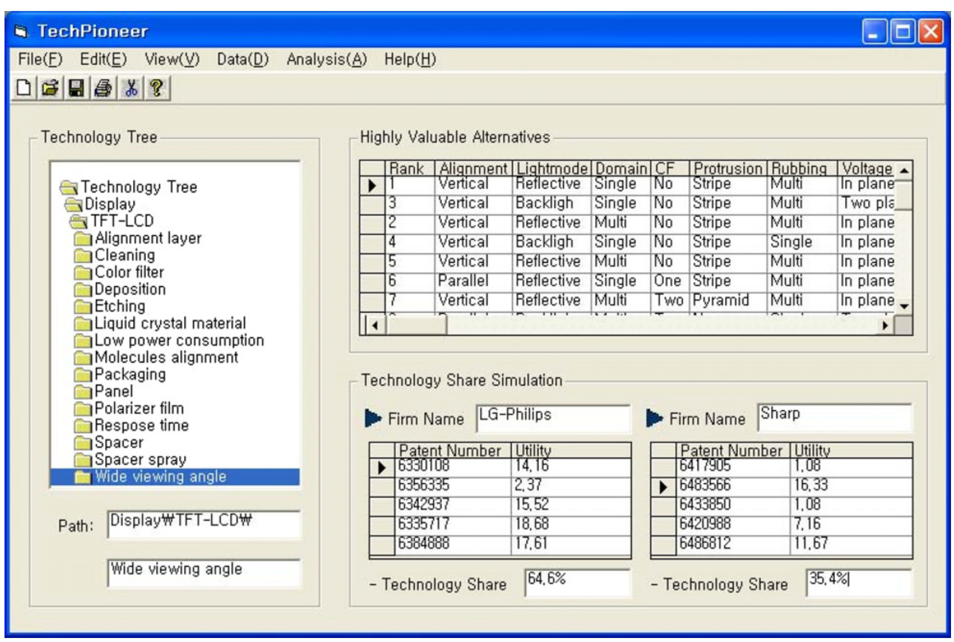

Fig. 8. Screen of priority of alternatives.

technology, the enhanced technological capability of the company can be forecasted by simulating the value of the technology.

7. Summary and conclusions

TechPioneer, the proposed system, aims at supporting the decision-making of managers, researchers and technologists in planning technology development. One of the most overriding points in this system is to identify potential configurations of technology which might dominate a future market and technology regime though novel, innovative ideas. To this end, the new system uses textual information from public and private databases and applies the concept of morphology to observe the trend of technology in detail and forecast promising configurations of technology. The involvement of domain experts is inevitable to reflect the underlying features of a specific technology in defining a morphological matrix and filtering extracted words toward noteworthy keywords. In particular, they must play a crucial role in prioritizing perceived alternatives by evaluating the value of existing patents.

The main advantage of this system is to provide a systematic approach to devising a plan for technology development. Most of the current methods and systems associated with technology development mainly depend on individual expertise of researchers and practitioners. This systematic process presents a standard framework to explore new technology opportunities, which means that the quality of outputs does not rely solely on the ability of participants. In addition, *TechPioneer* can use textual information which has been hardly utilized to forecast the future of technology. Technological documents such as patent documents and technical reports have been paid no attention to, in examining the characteristics of technology, whereas, bibliometric information such as an issue date, authors (applicants) and citation data has served as a principal data source in research for technology management. Another advantage is the ability of this system to communicate with experts. Intelligent systems must reflect the expertise knowledge of researchers and practitioners to enhance the reliability of outputs. Nevertheless, most systems are operated in an automatic way like a 'black box'. This system has to receive input data from domain experts in defining a morphological matrix and evaluating the value of alternatives.

These benefits allow analysts to apply this system to various areas related to technology management. First of all, the main application is to plan new technology development for both a short-term and a long-term period. Since the set of possible alternatives can be identified, long-term planning for technology development can be made by estimating the future value of these, let alone a short-term, urgent plan. This system can also be used to come up with the strategic position of a company in comparison with competitors and investigate the infringement of new technology on existing patents of other companies. It has a function to analyze the morphology of technology that belongs to a company, which leads to enabling investigating the competitiveness of competitors. *TechPioneer* can be regarded as a tool to manage large volumes of textual information which has been accumulated within companies. Since the growth of an internet industry and companies' thirst for knowledge, management departments have produced significant documents in great volume that need to be organized systemically and periodically (or in a real time). This system can provide up-to-date information to analysts by analyzing continuously produced documents on a regular basis.

However, this research has several limitations. While the framework of this system includes various techniques such as text mining, morphology analysis, network analysis and conjoint analysis, all analyses could not be completely conducted in one system. The outputs of network analysis and conjoint analysis are derived from Ucinet 6 (network analysis software) and SPSS 11.0 (statistical analysis software), providing input information for subsequent analyses in *TechPioneer*. Furthermore, the priority of alternatives is decided by just synthesizing scores of evaluators, totally depending on their expertise. However, the valuation of experts also needs to be supported by useful information such as a licensing fee and the impact of patents, enabling analysts to evaluate the value of patents on the basis of abundant sources for reliable assessment. Therefore, future research can be executed to overcome these limitations. Firstly, a complete system which integrates all necessary techniques needs to be developed in order to improve the applicability of this system. Secondly, a process that provides ample information for evaluating the value of patents can be implemented in the DMM. Various valuation techniques such as a benchmarking method, a monetary value method and a real option method can be applied to facilitate this process. Finally, in long-term planning, a method to prioritize alternatives needs to reflect the impact of technology breakthrough. Therefore, the schedule of long-term technology development can consider the future value of alternatives, which is estimated on the basis of the unexpected value of disturbing ideas than the valuation of existing patents.

Acknowledgements

This work was supported by the Korea Research Foundation Grant funded by Korea Government (MOEHRD, Basic Research Promotion Fund) (KRF-2006-214-D00192).

References

Arai, H. (2006). Japan's intellectual property strategy. *World Patent Information, 28*(4), 323–326.

Atkinson, J. (2002). Text mining: Principles and applications. *Revista Facultad de Ingenieria, U. de Tarapaca (Chile), 7*, 57–61.

Clifton, C. (2004). TopCat: Data mining for topic identification in a text corpus. *IEEE Transactions on Knowledge and Data Engineering, 16*(8), 949–964.

Cohen, W. W. & Hirsh, H. (1998). Joints that generalize: Text classification using WHIRL. In *Proceedings of the fourth international conference on knowledge discovery and data mining*, New York.

Cooper, A. C., & Schendel, D. (1976). Strategic responses to technological threats. *Business Horizons, 19*(1), 61–69.

Couger, J. D. (1995). *Creativity and innovation in information systems organizations*. New York: Boyd and Fraser Publishing Co.

Cutting, D. R., Karger, D. R., Pedersen, J. O., & Tukey, J. W. (1992). Scatter/gather: A cluster-based approach to browsing large document collections. In *Proceedings of ACM SIGIR international conference on research and development in information retrieval*. Denmark: Copenhagen.

Dhillon, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine Learning, 42*(1), 143–175.

Evans, J. R. (1990). *Creative thinking in the decision and management sciences*. Cincinnati, OH: South-Western Publishing Co.

Eysenck, H. J. (1995). *Genius: The natural history of creativity*. New York: Cambridge University Press.

Gatignon, H., Tushman, M. L., Smith, W., & Anderson, P. (2002). A structural approach to assessing innovation: construct development of innovation locus, type, and characteristics. *Management Science, 48*(9), 1103–1122.

Kobayashi, H. (2006). A systematic approach to eco-innovative product design based on life cycle planning. *Advanced Engineering Informatics, 20*, 113–125.

Kostoff, R. N., Boylan, R., & Simons, G. R. (2004). Disruptive technology roadmaps. *Technological Forecasting and Social Change, 71*, 141–159.

Lichtenthaler, E. (2004). Technological change and the technology intelligence process: A case study. *Journal of Engineering and Technology Management, 21*, 331–348.

Mani, I., & Maybury, M. T. (1999). *Advances in automatic text summarization*. Cambridge: The MIT Press.

Morrisa, S., DeYongb, C., Wua, Z., Salmanb, S., & Yemenub, D. (2002). DIVA: A visualization system for exploring document databases for technology forecasting. *Computers and Industrial Engineering, 43*, 841–862.

Porter, A. L., Kongthon, A., & Lu, J. C. (2002). Research profiling: Improving the literature review. *Scientometrics, 53*(3), 351–370.

Rohrer, R. M., Ebert, D. S., & Sibert, J. L. (1998). The shape of Shakespeare: Visualizing text using implicit surfaces. In: *Proceedings of the 1998 IEEE symposium on information visualization*. North Carolina.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*(1), 1–47.

Trippe, A. J. (2003). Patinformatics: Tasks to tools. *World Patent Information, 25*, 211–221.

Utterback, J. M., & Brown, J. W. (1972). Monitoring for technological opportunities. *Business Horizons, 10*, 5–15.

Van Raan, A. F. J., & van Leeuwen, T. N. (2002). Assessment of the scientific basis of interdisciplinary, applied research – application of bibliometric methods in nutrition and food research. *Research Policy, 31*(4), 611–632.

Watts, R. J., & Porter, A. L. (1997). Innovation forecasting. *Technological Forecasting and Social Change, 56*, 25–47.

Yang, J., & Liu, C. (2006). New product development: An innovation diffusion perspective. *The Journal of High Technology Management Research, 17*(1), 17–26.

Yoon, B. U., & Park, Y. T. (2004). A Systemic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technological Forecasting and Social Change, 72*, 145–160.

Zhu, D., & Porter, A. L. (2002). Automated extraction and visualization of information for technological intelligence and forecasting. *Technological Forecasting and Social Change, 69*, 495–506.