A New Method of Key Word Extraction from Foresight Topics Based on Textmining and Complexity Network Analysis

Keun-Ha Chung[1], Cheol-Woo Jeong[2*]

Abstract

In this study, we present new methods for identifying keywords for foresight topics that utilize the internet, network analysis, and text mining techniques to draw objective and quantified information that support experts' qualitative opinions and evaluations in foresight. Furthermore, by applying this fabricated procedure, we have derived keywords to analyze priorities in architectural engineering.

Not much difference between qualitative methods of experts and quantitative methods such as text mining has been observed from comparison between technologies derived via qualitative method from "The 3rd Science and Technology Foresight" (control group) and technologies derived via quantitative method from the "S&T Vision for the Future towards 2040" (experimental group) in South Korea. Therefore, as a quantitative tool useful for drawing keywords for foresight, text mining can supplement quantitative analysis by experts. In addition, depending on the level and type of raw data, text mining can bring better results in deriving foresight keywords.

For this reason, research activities accommodating Internet search results and the development of text mining methods for analyzing current trends are in demand. Meanwhile, the question of whether to apply newly evolving methods to complex systems can only be answered through continuous research.

Keywords: S&T foresight, complexity, text mining, network analysis

1. Introduction

Forecasting the future of the nation's science and technology sector and making predictions about new technology have in recent days become important to the enhancement of national competitiveness. National strategies and policies are in fact established based on such foresight on the future. In particular, using foresight on the future to select and develop promising technologies has become a major issue, with the roles of government agencies and their research projects as a major theme. In addition, how governments and researchers should invest in and manage their limited resources and manpower as they conduct foresight projects has is also gaining much attention.

Such foresight methods are mostly based on qualitative opinions and subjective evaluation by

[1] *Korea Institute of S&T Evaluation and Planning, Technology Foresight Center, 12th floor, Dongwon Industries Building, Yangjae-dong, Seocho-gu, Seoul, Korea*

[2] *Korea Institute of Science and Technology, L8205 International Cooperation Center, 39-1 Hawolgok-dong, Seongbuk-gu, Seoul, Korea*

[*] Corresponding author. E-mail:smile@kist.re.kr

experts, while there are not enough studies on objective methods and attempts to adopt them.

While qualitative foresight by experts has been recognized as a very important factor, it is possible that they may contain biased opinions or assertions based on the tendency of the expert, political factors, and human relationships. Evaluation by experts is also prone to error and miscalculation due to the lack of objective data.

Therefore, in science and technology foresight, it is important to provide more specific and objective data and material so that experts can produce objective opinions and evaluations. The Japanese National Institute of Science and Technology Policy (NISTEP) uses a thesis map (2009) to connect foresight with national R&D, and many companies analyze patents through use of a patent map. Such studies are attempts to obtain quantitative foresight on the future via objective data and methods.

This study sought a way to deduce quantitative and objective data that can support experts in providing qualitative opinions and evaluations in future foresight by using data on the Internet instead of data from existing research paper or patents. In addition, instead of using thesis and patent maps to simply conducting searches on papers and patents and eliminating noise, this paper deduced keywords through text mining, making use of analysis of importance and frequency on the Internet, trend analysis of time and spacial information. The keywords were then listed in order of priority through use of the complex network method based on connectivity. This study presents the use of the three methods of source verification via the Internet, text mining, and complex network method as a model for supplementing existing quantitative methods that tend to display non-continuous change trends due to their inability to reflect specific situations or conditions, or that cannot reflect current trends of change.

In addition, this study verified this model by attempting a quantitative analysis of the construction sector of the ‘3rd Science and Technology Foresight Revision/Supplementation in South Korea’ (2008, hereinafter ‘3rd Science and Technology Foresight’) and deducing keywords required for foresight.

2. Examination of Existing Methods

The scenario method, which is the most widely used qualitative method for foresight (Kahn Herman, 1967) shows a wide range of glimpses into the future depending on the standards for classification. But because it cannot fully express the details required for decision-making, it has been used together with other methods. Other qualitative methods include SWOT, which analyzes interior/exterior environments, and the expert panel, which forecasts the future through knowledge and information obtained through discussion (Keenan, 2004), and the decision tree, which organizes the categorizes the types of possible futures into a structure resembling a tree’s branches (Sonquist & Morgan, 1964). Meanwhile, among quantitative methods, the Delphi method (RAND, 1949), which forecasts the future through agreement by experts, is currently the most widely used method, while the cross impact analysis (Gordon & Helmer, 1996) and real-time Delphi (Gordon & Pease, 2006) methods, which have improvements over the Delphi method, are also seeing use. Quantitative methods also include trend extrapolation, which analyzes trends using past data, and the analytic hierarchy process (AHP), which deduces priority through pairwise comparison (Saaty, 1970).

However, while qualitative methods such as the above can foresee the future even in the case of a lack of information, they can lead to biased results due to influence from a wide range of factors such as lack of information, tendencies of experts, political factors, human relationships (Hang-Sub Choi, 2007). Meanwhile, trend extrapolation, which is a linear method of analysis that applies mathematical and statistical methods on past data, lacks accuracy due to the addition and high number of new variables, and is particularly inaccurate in the case of specific situations and conditions (non-continuous change) (Se-Jun Lee et al., 2008). In addition, the AHP method can evaluate and prioritize variables necessary to decision making by matching them on a 1:1 basis, but cannot reflect current trends.

Quantitative foresight methods include trend extrapolation, which foresees the future based on past data, and the analytic hierarchy process (Saaty, 1970), which deduces priority through pairwise comparison. Quantitative methods that utilize statistics include the Delphi method (RAND, 1949) and cross impact analysis (Gordon & Helmer, 1996). These two methods foresee the future, which is filled with uncertainties, through agreements between experts. Methods for technology foresight that are currently the most widely used among governments and agencies are the Delphi and scenario methods.

The Delphi method was developed in 1948 by the American think tank RAND and is currently being used in various sectors including the military, R&D, and education sectors. The Delphi method is a structured communication technique in which experts rely on their instinct to reach consensus in the case a lack of data. In other words, the method collects data on the future via repeated feedback on surveys on anonymous experts. A variety of Delphi methods are currently being developed, as marked by the introduction of real-time Delphi (Gordon & Pease, 2006). Delphi is a statistical method in which experts are subject to surveys through which they subjectively forecast the point of realization of a specific technology, and the average of the forecasts becomes the theoretical point of realization. Most countries use this method to forecast the future. The scenario method is a forecasting method in which experts provide different scenarios for the future based on given variables. While this method is good for explaining long-term outcomes, it lacks the details that are required for decision making (Mats Lindgren & Hans Bandhold, 2002).

Therefore, this study, to enable a more quantitative analysis in selecting technologies subject to foresight, presents a quantitative models that uses Internet search engines, text mining and the complex network.

As an example for how search engines are utilized,

Table 1 Comparison of key foresight methods

Method	Characteristic	Strength	Weakness
Delphi survey	• Foresees the future by collecting views from experts on the given forecast assignment. • Analyzes statistics and is most widely used method.	• Method for foreseeing the future when data is lacking either slightly or profusely. • Utilizes the median figure to convert the column into figures when foreseeing (columns are converted into figures.) • The anonymity and independence of the method allows participants to freely express their views.	• Extensive time and effort is needed for selecting and sustaining experts. • A costly method requiring a minimum of 2 surveys and a duration of six months or more. • Disregards the interrelations between the forecast items.
Cross impact analysis	• Deduces numerical results similar with the Delphi survey, but complements the Delphi survey, which disregards the interrelations between the forecast items. • Prepares scenarios and mostly utilizes the cross analysis between scenarios.	• Structuralizes the interrelations between forecast items • Relatively easier to calculate and easier to grasp visually.	• A Delphi study must be conducted in advance.
Scenario building	• Involves the provision of the most plausible scenario of what may happen in the future. • A qualitative method of analysis that is used for mid and long-term future foresight.	• Offers an alternative future. • The method can be utilized diversely when combined with other forecast methods.	• Low accessibility to detailed technologies. • Inadequate to be used as data for making decisions and implementing actions.
Complex network model	• Utilizes the Internet and complex network analysis to deduce core technology. • Quantitative analysis method	• Reduces time and costs by not employing experts. • Expected to replace some of the experts' assessments.	• Effective when combined with other methods such as the Delphi survey or the scenario method.

the frequency of search words entered by U.S. citizens was used to analyze influenza development per region. The results matched the actual figures announced by the U.S. Center for Disease Control and Prevention. In fact, by using this method, the results on the flu outbreak were drawn approximately two weeks prior to the CDC announcement. (S.H. Lee & P.J. Kim etc. 2008)

Further, research on what factors are important for analyzing the inter-relations between different factors is dealt with in full scale in network analysis. This type of network theory first started in earnest with the graph theory in the field of mathematics in 1960. In social science, much research has been conducted starting in the 1930s as the social network theory. Fundamentally, in network theory, a network is connected by nodes and links, while in social sciences, nodes were treated as the agents and links as inter-relations. (Young-Soo Yoon, Seung-Byung Chae, 2005; Dong-Won Sohn, 2002; Yong-Hak Kim, 2004; Byung-Nam Kang, 2009) In social sciences, the concept of "center" in a network became an important factor following an experiment conducted by Alex Bavelas and Harold Leavitt of MIT in the U.S. in 1940.

Particularly, the agent at the center was seen to become more powerful within the group (Brass, 1984), more capable of personal innovation (Ibarra, 1993), exercise more influence over decision-makings (Friedkin, 1993), and be higher in personal performance. (Baldwin, Bedell & Johnson, 1997). Further, there were differences in promotion opportunities according to the position a person held within the corporate network (Burt, 1992), while a network that is placed in a structurally moderate position is capable of drawing out policy implications via strategically significant positions (Scott, 2000). Following the research on graph theory and the distribution function of connection numbers in random networks, Watts and Strogatz in 1998 proposed the 'small world network' model and measured the clustering coefficient through the power grid, the neural network of Caenorhabditis elegans and the actors' network. The links connecting the nodes in this model were explained as bringing the entire network closer together through shortcuts.

Based on these studies, network research began to grow at an explosive rate and also precipitated the announcement of the scale free network. In 1999, R. Albert, H. Jeong, A.L. Barabasi released a thesis on the physical connecting structure of the Internet, which is connected by hyperlinks. The results of this study showed that the distribution of connection numbers does not form a random network, but in fact follows the power law. This phenomenon was consequently discovered in other, various fields and was defined as a 'scale-free network.' Since the 2000s, network research has been applied to not only social science, but to a myriad of areas such as the metabolic reaction network, protein interaction network, thesis citations, stock market analysis, Internet structure analysis and contagious diseases studies.

3. Development of a New Future Foresight Method

The methods for deducing new keywords for future foresight reflect the trends that in turn reflect the time and space lacking in existing quantitative future foresight methods. In particular, utilizing Internet search engines capable of offering objective sources, text mining to draw out main keywords reflecting this trend, and complex network analysis to deduce priority based on the inter-relation of the main keywords are all being avidly studied in their respective areas. The end-results of each study, in particular, is being studied as a whole after being combined in a new mold called future foresight. In Table 2, existing research methods are compared with the new methods suggested in this paper. The following is the theoretical background of respective studies.

3.1 Complexity Network

To understand the complex and diverse natural and sociological phenomena in the world we live in, we have come in need of a new way of understanding various phenomena such as emergence or inter-relating phenomena. The concept that was deployed since the early 20th century to understand this complex world has been studied in various fields such as social and economic phenomena, and studies began in earnest

Table 2 Comparison of previous and current studies

Classification	Previous studies	Current studies	Characteristics
Resources	• The frequency or matching data of theses or patents	• Primarily uses the future foresight data of theses, patents and various experts to utilize the frequency of main keywords • Secondly uses frequency data using the Internet	• Securing the objectivity of data
Trends	• Reflects trends utilizing some of the recent frequency data, or uncapable of such reflection.	• Analyzes trends reflecting time and space by utilizing the text mining method	• Reflects the trends of time and space
Prioritization	• Nearly impossible when using quantitative methods	• Deduces prioritization through inter-relations	• A prioritization method that utilizes objective data

in 1984 when the Santa Fe Institute was established in New Mexico of the U.S. This complexity network comprises of innumerable natural phenomena that are connected by certain interactions, and thes interactions appear to be nonlinear and without any particular relations. Recently, there has been a self-similarity theories such as fractal, self organization and critical phenomena that are being suggested to partially explain these phenomena. These theories have been observed and proven through either macroscopic or microscopic approaches. Network theories were used to verify such complex phenomena, structure and inter-relations, in addition to determining the complex structure.

Out of the various methodologies such as agent-based-modelling, decision making model, game theory and system dynamics, this particular study aims to suggest a model for deducing the keywords for future foresight by utilizing networks. The theories related to networks have been developed in diverse fields. Network study started extensively in the field of social sciences in the 1930s, and to this day, social network analysis or social connection analysis is where the study is most widely applied. In social sciences, the study sought a scientific approach to human relationships and applied the values and significance of such relationships to the network theory as an essential theme. Particularly in network analysis, it is important to grasp the fundamental role of a node or the impact it has on a network based on the links between the nodes in the network. In general, a centrality[1] analysis is conducted to analyze the center of the network, and this analysis is the method for expressing how close the node is to the center.

3.2 Analysis Using Search Engines

The rapid growth of the Internet has spawned a huge amount of data online. According to studies, there were approximately 30 million websites in 1999, but this number has grown to 100 million in 2007 and is rising to this day at an exponential rate. We must now spend much time and effort to find real-time data or other important information on the fast growing Internet. For this reason, quantities of studies are being conducted on how to utilize the Internet data to deduce only the necessary information, and diverse methods such as text minding, search engines and search robots are being introduced.

In the area of networks, new methods utilizing IT (information technology) and the Internet have started to be introduced. Some studies utilize the vast data on the Internet to use methods such as data mining, text mining and search engines. For instance, at Google, studies on the usefulness of the Internet touched off after it was discovered in 2008 that the result of the exit polls for the 109th senate election were analogous to the analysis of Google searches.

This paper also seeks to deduce a virtual network by analyzing network links through diverse channels

[1] For further information on centrality analysis, refer to “G. Sabidussi, The centrality index of a graph. Psychometrika, 31(4) (1966) 581-603”

of information (web pages, theses, patents, blogs), and based on this network, suggest an analysis model. In particular, this study replaced the weight factor on the mutual link between the nodes with the end result of Google searches and analyzed this with the network. The Google search engine used for this analysis was realized through an algorithm called ‘page rank.’ It is also a search engine[2] created by a type of network algorithm expressed in terms of search results based on an analysis of the weight of hyper links connected to the web pages. The search engines created this way differs from previous search engines in that its search results are far more credible because arbitrary manipulation becomes difficult. Google has recently started to offer divers contents such as thesis search sites, patent search sites and the Google dictionary through its Open API (application programming interface) so that the sites can be used for various different purposes. Generally, an API was known as an interface for allowing an operating system or language to control certain functions, but in web 2.0, it has become expanded to an interface for enabling the usage of certain services on the web.

3.3 Text Mining Analysis

Most data can be generally categorized into structured data and unstructured data. Structured data is information turned into a data base by processing previous data to fit certain formats and conditions. Approximately 20 percent of this kind of information is comprised of information used to create, save and recycle data. Data mining refers to the extraction and processing of information in structured data, and this method is most commonly applied to creating data base systems and categorizing information.

Text mining is used to utilize the remaining 80 percent, which is unstructured data. This method is being utilized in various areas but more study is required. Text mining refers to the process of identifying not simply the keyword, but the context of the information the user is interested in out of the vast amount of data available. The explosive growth of information has called on the need for methods to automatically process large amounts of data, and methods are now being developed to discover covert patterns and searching for information pertaining to particular subjects.

Text mining is expected to yield methods for realizing technology unthinkable in the past and be applied to various fields. For instance, text mining may help track down criminals or terrorists by identifying past crimes that bear resemblance to recent crimes out of a vast crime log, or categorize the tens of millions of unstructured complaints posted on websites and weed out certain problems, or automatically identify an effective treatment for diabetes after sifting through prescriptions written out to innumerable patients.

Text mining is so far being applied to the Internet and the area of mining general data. Data mining methods using the Internet is being utilized in Internet search engines.

The general process of text mining is said to be diverse, but it generally goes through a 4-step procedure. Text mining follows the general procedure of unstructured data gathering to data treatment to data extraction to data analysis. This method deduces useful information via a mathematical model or algorithm at the data extraction stage. Regarding utilization, the results are used for search engines or for deducing other important keywords. Data can be extracted based on purpose, condition and environment to be used for text mining, and the data extraction method is one of the most important parts of text mining.

In particular, various mathematical algorithms and methods can be used for extracting data, among which the most simple yet powerful method is the TF-IDF(Term Frequency - Inverse Document Frequency) method. Spark (1972), under the premise that words simultaneously appearing in several documents are most widely used, supplied the method of IDF (Invert Document Frequency). Salton (1976), through the proposition that a frequently appearing word in a single document can represent the entire document proposed the method of calculating TF (Term Frequency). Wu & Salton (1981) reported on the term

[2] Refer to http://www.wikipedia.org

weight of these two methods. A closer look of TF-IDF shows that it is a method of finding the weight factor of the keyword to be used for information searches and text mining. It is in short, it is a statistical figure manifesting the significance of a word is in a certain document when there is a group of documents consisting of several documents.

$$TF\text{-}IDA = TF \times \frac{1}{DF}$$

TF : Frequency of certain words in a document

DF : Frequency of certain words in more than one document

IDF : DF's reciprocal Number

As seen above, TF-IDF[3] is a method utilizing frequency and has been verified over a long period of time. But more attention is needed to its high error rate stemming from its complex calculating methods and the methods and scope of data extraction. This makes the method difficult for analyzing current trends or situations, and consequently called for a new algorithm to supplement the current importance to better reflect current trends.

TF-DI (Term Frequency - Data Index) is a text mining method aimed at analyzing future trends. Created for a special purpose (trend analysis), it was developed to modify TF-IDF and supplement its weakness by analyzing the weight factor that gauges the importance of keywords per year.

The most important feature of TF-DI is that it deduces the most significant keywords in the document to analyze the frequency of the word based on the amount of data on the Internet. Further, it does not use frequency of a certain word between documents in a group of documents. Rather, it was built to use the weight factors per year to analyze trends. Consequently, it uses the principle idea of TF-IDF, which relies on the importance of frequency, but for effective trend analysis, the importance of the document is gauged by utilizing the Internet to analyze the weight factor per year. This type of analysis uses the current information available on the Internet and utilizes the weight factor. Its key strength is that it accurately reflects current data or information.

The most notable difference between TF-IDF and TF-DI is that firstly, TF-IDF's frequency analysis uses the frequency of documents within groups of documents, while in TF-DI, only the frequency of keywords in reports and theses considered as important and the recent search results on the Internet. This frequency is not limited to a certain group of documents, and can serve as a standard reflecting recent trends. Secondly, TF-IDF measures significance based on the number of documents containing certain words, but the TF-DI method use the concept of time as a variable to measure significance. With the implementation of time, it suggests a more useful way of analyzing the most up-to-date trends. Thirdly, the importance of TF-IDF widely fluctuates based on which group of documents were selected, but the TF-DI has a smaller error rate because it uses results based on the Internet. It is also capable of extracting the importance of different figures under diverse conditions to analyze trends.

The difference in these two methods show that while TF-IDF and TF-DI may appear to be similar in assessing significance, the inclusion of the time concept renders the two completely different in their method of deducing importance. The following Table 3 analyzes the comparative advantages of TF-IDF and TF-DI based on their respective strengths and weaknesses.

A closer look at the detailed algorithms of TF-DI shows that it was developed based on 2 propositions.

Propositions

1. Frequency decides the importance of keywords deduced from a document or process.

2. High frequency in terms of yearly analysis shows that it is an important keyword.

The first proposition indicates that the high frequency of a certain keyword is a reflection of its importance. Therefore, keywords with high exposure rates on the Internet was chosen based on how

[3] For further information, refer to "Salton G. and McGill, M. J. 1983, Introduction to modern information retrieval. McGraw-Hill"

Table 3 Analysis of strengths and weaknesses of TF-IDF and TF-DF

	TF-IDF	TF-DF	Comparative Analysis
Qualities	Capable of extracting words in certain documents to deduce prioritization of the word based on its relation to the document.	Trend analysis for future foresight and prioritization is possible because words are extracted from various documents and because the Internet is utilized.	TF-IDF can be used as a search engine algorithm, while TF-DF utilizes the results of the search engines
Usage	Deduces certain keywords, deduces prioritization of certain documents.	Deduces certain keywords, deduces prioritization of keywords to analyze future trends	Trend analysis approachableness TF-IDF < TF-DF
Accessibility to the date	The frequency of a word in a document is dependent on the type and group of document, and if they are not appropriate, the results can be quite different so too much time and costs are consumed.	As the document is not the criteria, there is no restriction on the type of group of documents. It also can assume certain situations or conditions to extract necessary keywords from various documents	Data collection approachableness TF-IDF < TF-DF
Usability based on the type of document	The frequency of a word in a document is heavily affected by the type of document, such as whether it is a thesis or patent	As the word frequency is deduced from the Internet, it is not affected by the type of document and can therefore collect various words from which ever data conforming to its purposes, such as from reports, theses and conference material.	Restriction of documents TF-IDF > TF-DF
Inclusion of the latest trends	Theses are used as target to reflect the most recent research trends for future foresight. bit it is difficult to identify the scope or number of recent theses, and expert assistance is mandatory.	As the Internet is used, it is easier to reflect recent trends, and can read trends based on identical conditions in various data sources such as theses, patents and blogs.	Inclusion of latest trends TF-IDF < TF-DF
Availability of yearly data	The TF-IDF Algorithm renders yearly analysis almost impossible because the documents need to be categorized according to time, which is impossible.	TF-DF is easier to analyze trends because it was created with the yearly frequency and weight factors in mind.	Analysis of yearly data TF-IDF < TF-DF
Margin of error	The bigger the number of words in a document, the lower the similarity of the document so that the bigger the document becomes, the lower the assessment.	The number of certain keywords do no have notable impact, and the more the words, the eaiser to analyze.	Margin of error TF-IDF > TF-DF
Boundary of target document	When dealing with documents for a wide-ranged topic. the number of words in the document increases, and even if documents deal in similar topics, the word composition may be completely different.	The year and respective analysis of each word is possible, and analysis of a combination of certain topics is possible.	Analysis of diverse keywords TF-IDF < TF-DF
Portability for developing IT system	When there are more documents, the calculation process become more complex and slower, and more time is needed for system development because more data bases and indices are used.	The structure is simple and enables quick calculation. Flexible system development is possible because it utilizes the Google search engine's Open API.	Usage and development approachableness TF-IDF < TF-DF

frequent they appear so that the results of Internet searches were selected as the frequency.

The second proposition predicts that the more recent the frequency of certain keywords, the more important they are, so that DF (data frequency) was selected based on yearly frequency and weight factors (weight rises when time becomes more recent).

Based on these two propositions, the TF-DI method multiplies the relative frequency of the keyword and the weight factor to produce the following formula.

$$(TF\text{-}ID)_i = \sum_{j=1}^{n} TF_j \times DI_j$$

$$TF_j = \frac{i - frequency}{(total\ frequency)_j}$$

$$DF_j = \frac{j}{n}$$

i : Classification of keywords

j : First occurrence in the starting year(j=1)

n : Yearly occurrence during the analysis puriod (j=1, 2, 3...)

4. Study Methods and Pilot Application

4.1 Study Method and Process

This study uses text mining and complexity network to deduce the necessary keyword for future foresight in the field of construction and analyze the priority for trend analysis. In particular, the study applied text mining and complexity network model to the construction field and 'S&T Vision for the Future Towards 2040 (2010)' of the 3rd Science and Technology Foresight to deduce main keywords. The deduced keywords were tested of their effectiveness by comparative analysis with the 'Construction and Transportation R&D Long-Term Plan (2008-2012)' announced by the government in 2007.

This study first conducted a two-step analysis to deduce the keywords for future foresight in the area of construction in the 3rd Science and Technology Foresight researched by expert Delphi method and the S&T Vision for the Future Towards 2040 using the scenario method.

In the first stage, the study used text mining to deduce keywords reflecting recent trends from various subjects of analysis. In the second stage, this paper analyzed the priority of the deduced keywords. Ultimately, the paper compared the 3rd Science and Technology Foresight that deduced long-term technology and the construction area technology deduced from the long-term S&T Vision for the Future Towards 2040 based on the scenario method with the 'Construction and Tansportation R&D Long-Term Plan (2008-2012).

For the first stage text mining analysis, the paper formed a new group of keywords centering on the keywords most frequently exposed in the S&T Vision for the Future Towards 2040 and the 3rd Science and Technology Foresight. Then, the frequency of these keywords were analyzed based on their Internet exposure. Next, TF-DI and Google's API was used for the analysis. This method used the extent of Internet exposure of the keywords to determine the frequency, then used the weight factors based on the most recent year to analyze the trend.

The second step analyzes the priority of the keywords deduced from text mining. In particular, the complexity network was used after assuming that there is a link between limited keywords. The analysis was also focused on the centrality network and to enable such an analysis, weight factors were used between nodes and links, while the links between nodes were minimized and optimized.

A closer look at the specific technologies used for these 2 steps show an attempt at structural analysis based on the TF-DI, which is a method of text mining, along with programs using Google's AJAX Search API, the network centrality of complexity networks and the path finder method.

4.2 The 3rd Science and Technology Foresight

To draw the keywords for foresight, this study selected 42 detailed technologies conforming to the 11 priority science technologies in the area of construction and transportation among the technologies in the "3rd

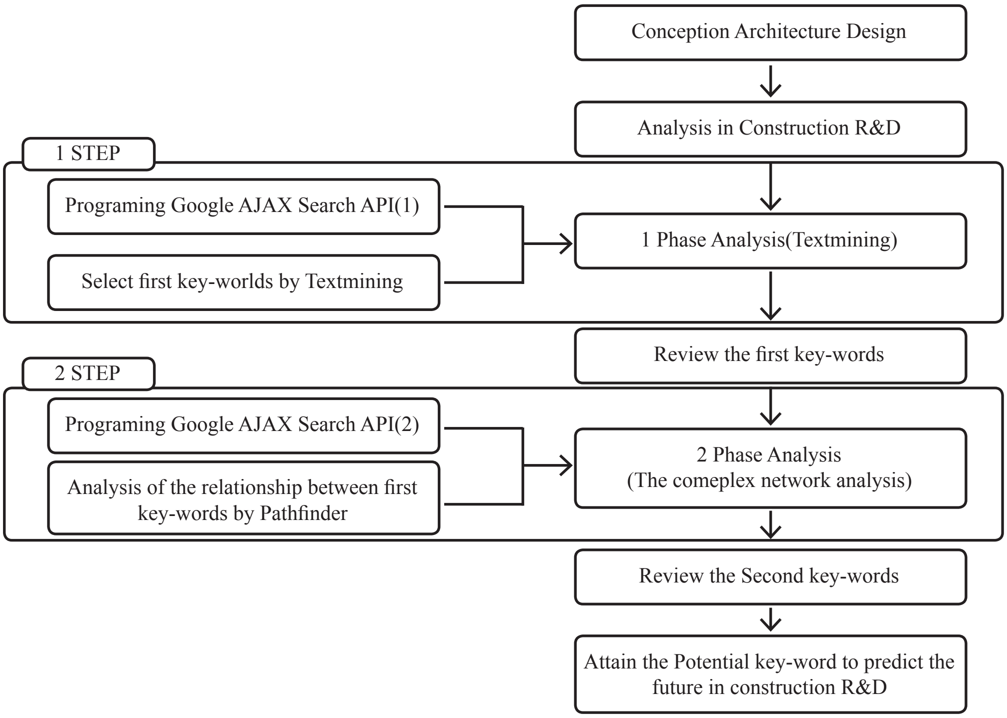
Conception Architecture Design
Analysis in Construction R&D
1 STEP
Programing Google AJAX Search API(1)
Select first key-worlds by Textmining
1 Phase Analysis(Textmining)
Review the first key-words
2 STEP
Programing Google AJAX Search API(2)
Analysis of the relationship between first key-words by Pathfinder
2 Phase Analysis
(The comeplex network analysis)
Review the Second key-words
Attain the Potential key-word to predict the future in construction R&D

Figure 1 Analysis process for future foresight

Science and Technology Foresight."

Primarily, this study drew 449 detailed technologies based on the 182 future strategy technologies drawn from the 3rd Science and Technology Foresight (761), Nationally Promoted Priority Technology Total Roadmap (90, 348 detailed technologies), Japan's 8th Technology Foresight (859), the 3rd Basic Plan Priority Technologies (273). EU Promising Technologies (40)[4], along with the Nationally Promoted Priority Technology Total Roadmap (90, 348 detailed technologies) and the new R&D business assignments from related government ministries (6 technologies, 25 detailed technologies), with focus on need-type assignments drawn from existing scientific research. These technologies maintain the form of mid-level technologies, rather than that of detailed technologies comprising the main project units. The technologies were analyzed and drawn according to similar level based on the above process.

Secondly, this study selected 100 priority science technologies and 411 detailed technologies out of 449 candidate technologies after considering the assessment of the following by experts in the fields of the industry, academia and research: The technological ripple effect over a span of 5 years; contribution to improving the quality of life; contribution to value-adding in the industry; contribution to national security and international society; urgency. Among the 100 priority science technologies, this study distinguished between priority strategic technologies (40) for securing strategies on a national level, and the strategic technologies (60) that are less urgent than priority strategic technologies but should be secured for pursuing mid and long-term strategies on a national level. During this process, 42 detailed technologies were selected based on 3 priority strategic technologies and 8 strategic technologies related to construction and transportation.

During the text mining analyzing phase, the analysis was conducted in the four stages of unstructured data gathering, data treatment, data extraction, and data analysis.

During the unstructured data gathering stage, this study selected 42 detailed technologies conforming

[4] For additional information on EU promising technologies, refer to "Emerging Science and Technology priorities in public research policies in the EU, the US and Japan, European Commission, 2006"

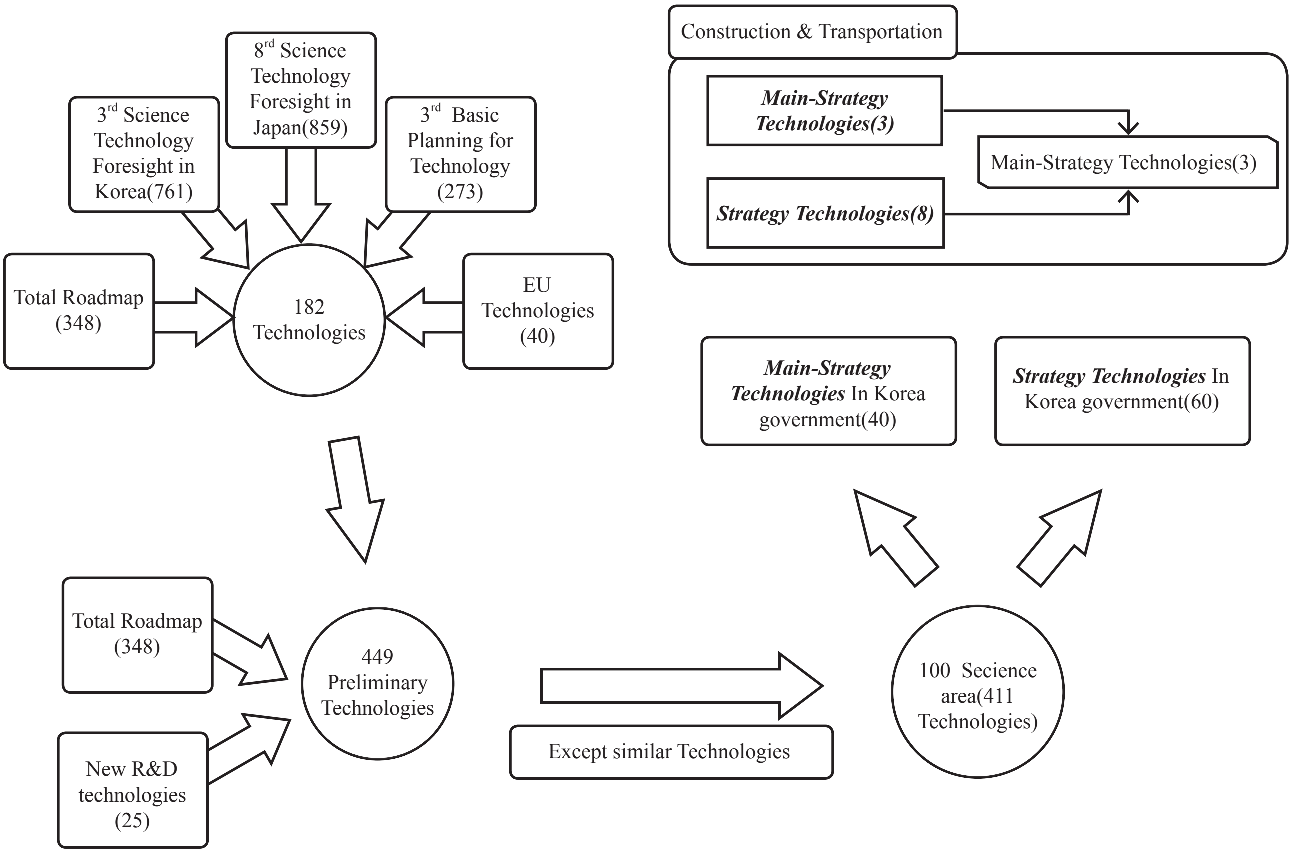
3rd Science Technology Foresight in Korea(761)
8rd Science Technology Foresight in Japan(859)
3rd Basic Planning for Technology (273)
Total Roadmap (348)
182 Technologies
EU Technologies (40)
Total Roadmap (348)
New R&D technologies (25)
449 Preliminary Technologies
Except similar Technologies
100 Secience area(411 Technologies)
Main-Strategy Technologies In Korea government(40)
Strategy Technologies In Korea government(60)
Construction & Transportation
Main-Strategy Technologies(3)
Strategy Technologies(8)
Main-Strategy Technologies(3)

Figure 2 Process of deducing keywords

to the 11 priority science technology in the field of construction and transportation in the 3rd Science and Technology Foresight.

The data processing stage is where the data source capable of extracting the information applying to respective technologies is processed based on the deduced subject data and technology list. In this stage, this study referred to the compendium on the 11 priority science technologies and the data (documents) collected on 42 detailed technologies to extract related keywords and put them into English.

In the data extraction stage, TF-DI and programs were applied to the model for forecasting future keywords to deduce the central keywords. Weight was given to end search results from Google's search engine that dated between January 1, 2000 to December 31, 2009 (10 years).

In the data analysis stage, existing English keywords grouped according to the categories of priority science technology and the newly deduced keywords were analyzed and regrouped under new names based on their newly defined correlations. 30 assignments were ultimately deduced.

■ Data treatment stage

This study referred to the compendium on the 11 priority science technologies and extracted keywords related to the future of the 42 detailed technologies.

In this stage, the data source capable of extracting the data applying to respective technologies was processed based on the deduced subject data and technology list. The extracted keywords were corrected and translated into English after referring to the collected data (documents).

Table 4 Key detailed technologies and English keywords

Area	Priority science technology	No.	Detailed Technology	English Keyword
Construction and transportation	Super tall building	1	Super tall building system and advanced material technology	Skyscraper Skyscraper structure Skyscraper material
		2	Swift construction and construction management technology	Skyscraper construction management
		3	Pilot Project planning and design technology	Skyscraper planning Skyscraper design
		4	Advanced environment and M&E element technology	Skyscraper environment
	Base technology for construction	5	Construction productivity, safety enhancement technology	Intelligent facility maintenance Facility safety management
		6	Energy/resource conserving construction technology	Green building
		7	Advanced construction material development technology	Advanced construction material
		8	Construction material quality certification/testing technology	Construction quality certification
		9	New construction production system and construction standardization technology	Construction Standard Construction production system
		10	Intelligent facility maintenance and management technology	-
		11	Deep underground space construction/utilization technology	Deep underground
	Super long-span bridge	12	Super long-span bridge design standards and methods	Long-span bridge Long-span bridge design
		13	Vibration/wind effect estimation and vibration reduction technology	Long-span bridge structure
		14	Advanced construction technology and special equipment development technology	Long-span bridge equipment Long-span bridge management
		15	Advanced material development technology	Long-span bridge advanced material
	Future urban planning	16	U-Eco City (environment-friendly city construction technology)	Green city
		17	Complex space development technology	Complex spatial
		18	Urban regeneration system technology	Urban regeneration
	Intelligent geographical information system	19	Land monitoring technology	Land Monitoring
		20	Spacial information-based infrastructure technology	Spatial information
		21	U-GIS based construction informatization technology	Ubiquitous GIS
	Future habitat	22	Intelligent education facility construction technology	Intelligent education facility
		23	Silver and environment-friendly housing environment technology	Silver housing Environment-friendly housing
	Future transportation system	24	ITS technology	Intelligent transportation system
		25	U-Transportation technology	Ubiquitous Transportation
		26	Future transportation safety enhancement technology	Transportation safety
		27	Integrated transportation and transfer system technology	Integration Transportation
	Marine and aviation safety and efficiency	28	Small aircraft certification technology	Small aircraft certification
		29	Air transport efficiency enhancement technology	Air transport efficiency
		30	Aviation safety technology	Aviation safety
		31	Marine safety technology	Marine safety
		32	Marine transportation management technology	Marine transportation management
	Next-generation high-speed railway	33	High-speed railway technology	High-speed rail
		34	High-speed magnetic levitation train technology	Magnetic levitation train
	Urban railway system	35	Advanced light rail and new transporation system technology	Light rail
		36	Urban magnetic levitation system technology	-
	Advanced logistics	37	Integrated intelligent container system technology	Intelligent logistics
		38	U-customs system technology	Ubiquitous logistics
		39	Intelligent transportation and logistics informatization technology	-
		40	Intelligent logistics center and high-efficiency equipment technology	Logistics automation
		41	Future transportation system technology	-
		42	Shipping logistics technology	Shipping logistics

■ Data extraction stage

Weight factors for the English keywords extracted in the data treatment stage was deduced based on TF-DF (Term Frequency - Data Frequency). Searching was done on Google's search engine, with search results drawn as weight factors. The period for search was set between January 1, 2000 and December 31, 2009 (10 years), and keywords were selected by TF-DF. Technologies that covered wide areas or that were representative were excluded, while detailed technologies from the same technology were screened based on priority (example: among super tall building facility and environment, super tall building design, and super tall building planning, the one with the highest priority was selected). In addition, technologies that displayed TF-DF results that were too low were excluded (example: ubiquitous logistics, spacial information-based infrastructure technology, environment-friendly housing, u-transportation technology, small aircraft).

Table 5 Detailed technologies extracted from TI and keywords

	Detailed technology	English keywords	TF-DF	Selection
1	Super tall building	Skyscraper	11290.114	X
2	Advanced light rail	Light rail	5206.1173	
3	Super tall building facility and environment	Skyscraper environment	3108.8757	
4	Super long-span bridge equipment	Long-span bridge equipment	2792.2539	
5	High-speed magnetic levitation train technology	Magnetic levitation train	2616.0213	
6	Future transportation safety enhancement technology	Transportation safety	2564.0572	
7	Super tall building design	Skyscraper design	2404.26	X
8	Aviation safety technology	Aviation Safety	2252.0875	
9	Facility safety	Facility safety management	2087.2799	
10	High-speed railway technology	High-speedrail	1901.2445	
11	Intelligent facility maintenance and management	Intelligent facility maintenance	1607.9018	
12	Land monitoring technology	Land monitoring	1538.7327	
13	Super long-span bridge design	Long-span bridge design	1482.1231	X
14	Marine transportation management technology	Marine transportation management	1462.6673	
15	Super long-span bridge material	Long-span bridge advanced material	1357.7789	X
16	Complex space development technology	Complex spatial	1197.1244	
17	Air transport efficiency enhancement technology	Air transport efficiency	1107.9809	
18	U-GIS based construction informatization technology	Ubiquitous GIS	1090.3	
19	Super tall building planning	Skyscraper planning	963.1944	X
20	Deep underground	Deep underground	796.05731	
21	Marine safety technology	Marine safety	795.93804	
22	Intelligent education facility construction technology	Intelligent education facility	692.78602	
23	ITS technology	Intelligent transportation system	673.7842	
24	Super long-span bridge	Long-span bridge	514.95725	X
25	Super long-span bridge structure	Long-span bridge structure	505.10814	X
26	Construction production system	Construction production system	454.95227	
27	Urban regeneration system development	Urban regeneration	454.27505	
28	Advance construction material	Advanced construction material	420.49649	
29	Construction quality certification	Construction quality certification	391.69061	
30	Shipping logistics technology	Shipping logistics	336.51971	
31	Construction standardization	Construction standard	195.91045	

Table 5 Detailed technologies extracted from TI and keywords (cont'd)

	Detailed technology	English keywords	TF-DF	Selection
32	Energy/resource conserving construction technology	Green building	186.51108	
33	Super long-span bridge construction management	Long-span bridge management	109.88248	X
34	silver housing	Silver housing	104.33369	
35	Eco-friendly urban construction technology	Green city	82.691987	
36	Logistics automation	Logistics automation	64.618014	
37	Intelligent logistics	Intelligent logistics	50.574406	
38	Integrated transportation and transfer system technology	Integration Transportation	37.161948	X
39	Ubiquitous logistics	Ubiquitous logistics	29.538344	X
40	Swift construction and construction management technology	Skyscraper construction management	25.117828	X
41	Spacial information-based infrastructure technology	Spatial information	14.310017	X
42	Environment-friendly housing	Environment-friendly housing	13.049058	X
43	Super tall building material	Skyscraper material	9.047507	X
44	Super tall building structure	Skyscraper structure	6.665581	X
45	U-Transportation technology	Ubiquitous transportation	2.5495492	X
46	Small aircraft certification technology	Small aircraft certification	1.3581987	X

■ Data analysis stage

In the data analysis stage, existing English keywords grouped according to scope of the science technology and newly deduced keywords were analyzed and regrouped under new names based on their newly defined correlations. Detailed technologies were put into a technology list and group names were modified.

Table 6 Finalized technology list and English keywords

Division	no.	Technology list	English keyword
Construction infrastructure	1	Intelligent facility maintenance and management	Intelligent facility maintenance
	2	Facility safety management	Facility safety management
	3	Energy/resource conserving construction technology	Green building
	4	Advanced construction material	Advanced construction material
	5	Construction material quality certification and testing technology	Construction quality certification
	6	Construction standardization	Construction Standard
	7	Construction production system	Construction production system
Large structures	8	Super tall building facility and environment	Skyscraper facility
	9	Deep underground	Deep underground
	10	Super long-span bridge equipment	Long-span bridge equipment
	11	Complex space development	Complex spatial
Future railway	12	High-speed magnetic levitation train technology	Magnetic levitation train
	13	High-speed railway technology	High-speed rail
	14	Advanced light rail	Light rail
Efficient land management	15	U-GIS	Ubiquitous GIS
	16	Land monitoring	Land monitoring
	17	Spacial information-based infrastructure	Spatial information

Table 6 Finalized technology list and English keywords (cont'd)

Division	no.	Technology list	English keyword
Aviation and marine	18	Air transportation efficiency enhancement	Air transport efficiency
	19	Aviation safety technology	Aviation safety
	20	Marine safety technology	Marine safety
	21	Marine transportation management technology	Marine transportation management
Future transportation	22	ITS technology	Intelligent transportation system
	23	Future transportation safety enhancement technology	Transportation safety
	24	Integrated transportation system	Integration Transportation
Future housing	25	Eco-friendly urban construction	Green city
	26	Urban regeneration system	Urban regeneration
	27	Silver housing environment	Silver housing
	28	Intelligent education facility	Intelligent education facility
Advanced logistics	29	Shipping logistics technology	Shipping logistics
	30	Logistics automation	Logistics automation

■ Network analysis stage

For network analysis of the construction sector, 30 component technologies were selected as nodes, and the relationship of each node was analyzed using Google AJAX Search API. Using pathfinder (PFNet), non-directional n(n-1)/2 links were simplified and optimized according to the priority of weight factors.

Preliminary steps were taken to display the 30 keywords drawn from the 3rd Science and Technology Foresight into a complex network. The Google search program was used to deduce weight factors for keywords based on relationships among one another, while pathfinder was used to minimize and optimize the deduced links. Priority in the optimized network was determined through degree centrality, betweenness centrality, and closeness centrality analysis.

Degree centrality analysis showed eco-friendly urban construction to have the highest level of priority, followed by advanced construction material, deep underground, high-speed railway technology, and future transportation safety enhancement technology. These technologies were found to act as hubs for other technologies and thus the most important technologies. Technology for large-scale eco-friendly housing environments and mass transportation means such as high-speed railways were found to be of importance.

In the betweenness centrality analysis, eco-friendly urban construction technology was found to have the highest level of priority, followed by future transportation safety enhancement technology, advanced light rail, and advanced construction material. These technologies were found to be of importance as they act as intermediaries between technologies. Eco-friendly urban construction technology and advanced construction material were found to act as hubs as well as intermediaries between other technologies.

The closeness centrality analysis was conducted to identify which technology was in located in the center of the entire network. Eco-friendly urban construction

technology and super long-span bridge equipment were found to have the strongest influence on the network.

Table 7 Degree centrality analysis of 3rd Science and Technology Foresight results

Priority	Degree centrality analysis	
	Future technology	Index
1	Eco-friendly urban construction technology	648.448276
2	Advanced construction material technology	629.310345
3	Deep underground technology	587.931034
4	High-speed railway technology	574.137931
5	Future transportation safety enhancement technology	541.724138

Table 8 Betweenness centrality analysis of 3rd Science and Technology Foresight results

Priority	Betweenness centrality analysis	
	Future technology	Index
1	Eco-friendly urban construction technology	0.535714
2	Advanced light rail	0.413793
3	Advanced construction material	0.400246
4	Future transportation safety enhancement technology	0.413793

Table 9 Closeness centrality analysis of 3rd Science and Technology Foresight results

Priority	Closeness centrality analysis	
	Future technology	Index
1	Eco-friendly urban construction technology	0.273585
1	Super long-span bridge equipment	0.273585

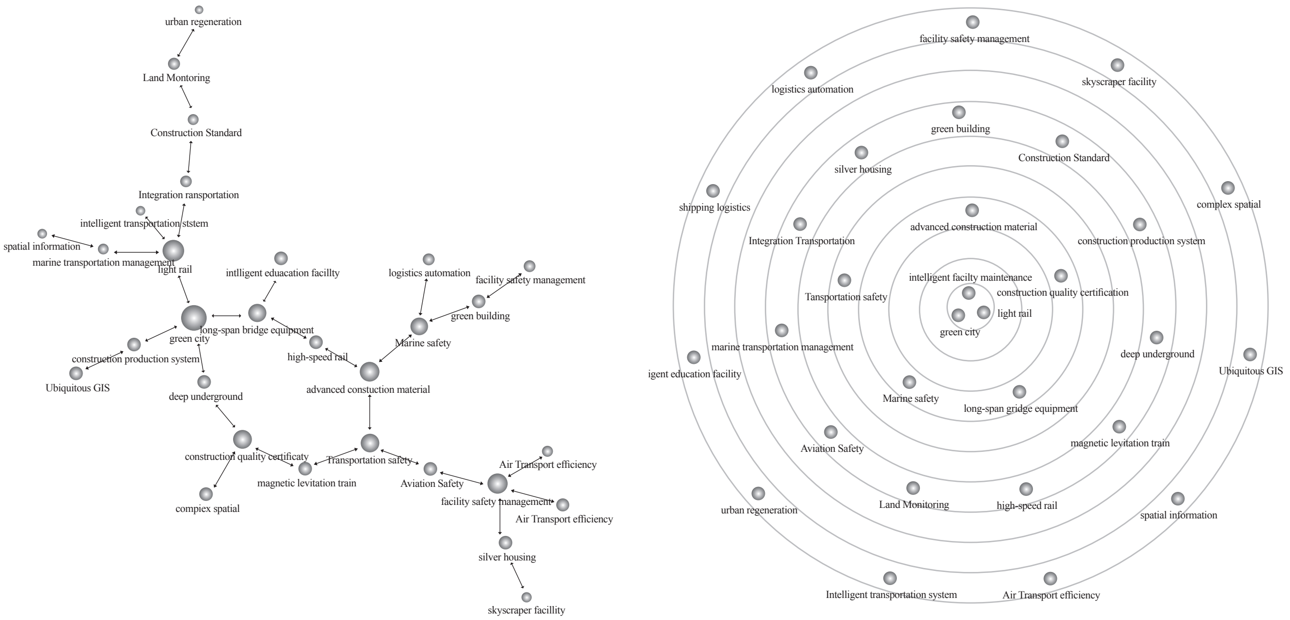

Figure 3 Network analysis of the construction sector

In the 3rd Science and Technology Foresight, environments that enhance the quality of life such as eco-friendly cities, deep underground spaces were found to have priority in addition to transportation means such as high-speed railways and advance light rail.

4.2 Result of pilot application

Analysis of the control group, the technologies derived from the 3rd Science and Technology Foresight and the experiment group, the technologies from the S&T Vision for the Future Towards 2040, showed that of the 29 technologies, with the exception of the 5 technologies of architectural structure transportation technology, space station construction, space transport craft, hydrogen vehicle, fuel cell transportation means, and customized public transportation, 24 technologies, or 82% of the technologies to be overlapping. Therefore, comparative analysis of technologies derived from 3rd Science and Technology Foresight through qualitative methods a technologies derived from the S&T Vision for the Future Towards 2040 through quantitative method (text mining) showed technologies drawn through the quantitative method to be not much different from those drawn by experts through qualitative methods. In addition, analysis of priority through the complex network showed environments

Table 10 Comparative analysis of the 3rd Science and Technology Foresight and the Construction and Transportation R&D Long-Term Plan (2007)

	3rd Science and Technology Foresight	Construction and Transportation R&D Long-Term Plan
Large buildings and urban environment	Eco-friendly urban construction	Super tall building
	Advanced construction material	Urban energy/ environment complex plant
	Deep underground	U-Eco City
High-speed and mass transportation means	Super long-span bridge equipment	Smart highway
	High-speed railway technology	High-speed rail system
	Advanced light rail	Light rail system
		Transportation and logistics system advancement
Other		Natural disaster-related technology

that enhance the quality of life such as eco-friendly cities, deep underground spaces and transportation means such as high-speed railways and advance light rail to have priority in the 3rd Science and Technology Foresight.

The construction sector, due to the characteristics of the human body and its structure, is less prone to changes toward nano and digital than other sectors. When it comes to housing, food, and clothes, humans maintain analog behavioral patterns and methods. Therefore, from the past until now, there has not been much change in the space and methods for housing. Skyscrapers or deep underground spaces are larger in scale, but in basic form they remain similar to what people have lived in the past.

However, control by a large, integrated system will be necessary in the future due to changes in shape and physical characteristics of construction material, city-level environment control, and convergence of transportation methods. In particular, as the rights and responsibilities, human dignity of each individual member of the society is expected to increase, the paradigm of the construction sector is expected to change from emphasis on the role on members to a society that satisfies the demands of individual members.

Therefore, in the future, skyscrapers and eco-cities will become environments that satisfy the needs of entire communities for anolog housing instead of spaces separated from the whole that satisfy the needs of individuals. Such paradigm change will require large-scale systems for individuals even if the change in population is not great. Logistical systems for handling mass volumes will also be required.

Public transportation systems such as rail systems, as they continue to develop toward means of quickly transporting large numbers of people, will see an increase in utility of personal space. For example, public transportation where people can enjoy their own space without interruption from others may be introduced.

The emergence of convergence technologies in energy and IT for easily controlling and operating such changes will enable control and operation of city-level systems, and this will enable the systems to grow in scale. Change in paradigm from such increase in scale and integration is also expected to become a keyword for change.

5. Conclusions

This study presents a quantitative foresight method that can supplement or partially replace qualitative methods carried out by experts, and verifies its validity by applying it to the construction sector. In particular, as a new quantitative method for extracting keywords for forecasting the future, this study proposes the use of Internet search engines to identify objective sources, text mining to extract keywords that reflect trends, and complex network analysis to prioritize the keywords.

In science and technology foresight, it is important to provide specific and objective data and material to experts so that their opinions and evaluations can be more objective. This paper, presents a method for adding objectivity to the variety sources such as research papers and patents. Utilization of the Internet in science and technology foresight can reflect current trends into foresight by making use of the complex and diverse information on the Internet. It is an area on which discussions on how to utilize data that reflect the recent trend of social networks such as Facebook and Twitter occur actively. It is also a quantitative method for forecasting the future on which many future studies are expected to be conducted.

Such convergence studies have led to the advancement of methods that have been studied in other areas, and have become the starting point for converging and newly applying those methods. In particular, text mining is expected to be useful as a tool for drawing keywords for forecasting the future, and is likely to see much development on its own.

Complex network analysis and Internet search engine analysis are also expected to see much individual research and development.

References

A. Quirin, O. Cordón, J. Santamaría, B. Vargas-Quesada, F. Moya-Anegón (2008), A new variant of the Pathfinder algorithm to generate large visual science maps in cubic time, Information Processing and Management, 44: 1611–1623.

A. Wagner and D. A. Fell (2001), The small world inside large metabolic networks, Proc. Roy. Soc. London Series B, 268: 1803-1810.

B. Bollobas (1981), Degree sequences of random graphs, Discrete Mathematics. 33(1): 1-19.

Byung-Nam Kang (2009), Science of Complexity Networks, Jipmundang. (in Korean)

Byung-Won Park (2007), Study on Improvement of Methodologies and Frameworks for S&T Foresight, KISTEP. (in Korean)

C. Tsallis and M. P. de Albuquerquer (2000), Are citations of scientific papers a case of nonextensivity?, The European Physical Journal B, 13: 777-780.

Cornelia Daheim (2007), Regional Foresight in Europe - 2 Examples: Duesseldorf and Linz, WFS Conference Minneapolis.

D. J. Watts and S. H. Strogatz (1998), Collective dynamics of ‘small-world’ networks, Nature, 393: 440-442.

Dong-Won Sohn (2002), Social Network Analysis, Gyungmunsa. (in Korean)

European Commission (2006), Emerging Science and Technology priorities in public research policies in the EU, the US and Japan.

G. Sabidussi (1966), The centrality index of a graph, Psychometrika, 31(4): 581-603.

G. Salton and M. J. McGill (1983), Introduction to modern information retrieval, McGraw-Hill.

H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi (2000), The large-scale organization of metabolic networks, Nature, 407: 651–655.

Hariolf Grupp and Harold A. Linstone (1999), National Technology Foresight Activities Around the Globe: Resurrection and New Paradigms, Technological Forecasting and Social Change, 60(1): 85-94.

Lee Myung-Hun Business School, http://www.emh.co.kr/xhtml/small_world_effect.html. (In Korean)

Jae-Yun Lee (2006), A Study on the Network Generation Methods for Examining the Intellectual Structure of Knowledge Domains, Korean Society for Library and Information Science, 40: 333-355.

James A. Dator (2002), Advancing Futures: Futures Studies in Higher Education, Praeger, Westport.

James T. Cushing (1998), Philosophical Concepts in Physics, Books Hill

Jang-Ho Park, A Critical Review on Genetic Determinism (Reductionist Perspective). (in Korean)

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, and Mark S. Smolinski & Larry Brilliant (2009), Detecting influenza epidemics using search engine query data, Nature 457: 1012-1014.

Fredrik Liljeros et al (2001), The web of human sexual contacts, Nature, 411: 907-908.

Martin Hilbert, Ian Miles, and Julia Othmer (2009), Foresight tools for participative policy-making in inter-governmental processes in developing countries: Lessons learned from the eLAC Policy Priorities Delphi, Technological Forecasting and Social Change 76(7): 880-896.

Mats Lindgren and Hans Bandhold (2002), Scenario planning: The link between future and strategy, Macmillan

M.E. J. Newman (2003), The Structure and Function of Complex Networks, SIAM Review, 45(2): 167-256.

M. Faloutsos, P. Faloutsos, and C. Faloutsos (1999), On Power-Law Relationships of the Internet Topology, Computer Communication Review, 29: 251-262.

Michael Marien (2002), Future Studies in the 21st Century: A Reality-Based View, Futures 34: 261-281.

Mikko Syrjänen, Yuko Ito, Eija Ahola (2009), Foresight for Our Future Society: Cooperative project between NISTEP(Japan) and Tekes(Finland), Tekes & NISTEP.

OECD (2001), Governance in the 21st Century: FUTURE

STUDIES.

P. Erdős and A. Réney (1960), On the Evolution of Random Graphs, Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 5: 17.

P. Erdős and A. Réney (1961), On the evolution of random graphs. II, Bull. Inst. Int. Stat. 38(4): 343-347.

Pirjo Ståhle (2007), Five Steps for Finland's Future, Tekes.

R. Albert, H. Jeong, and A.L. Barabasi (1999), Diameter of the World-Wide Web, Nature, 401.

R. Levin, D. Rubin, and J. Stinson (1986), Forecasting. in Quantitative approaches to management, NY, McGraw-Hill.

Rosario N. Mantegna and H. Eugene Stanley (1999), An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge.

R. W. Schvaneveldt, F. T. Durso, and D. W. Dearholt (1989), Network structures in proximity data, In G. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory, Vol. 24, Academic Press, New York.

S. A. kauffman, The origin of order : Self-organization and selection in evolution, Oxford University Press, New York, Oxford.

S. Redner (1998), How popular is your paper? An empirical study of the citation distribution, The European Physical Journal B, 4(2): 131-134.

Sang-Hoon Lee, Pan-Jun Kim, Yong-Yeol Ahn, and Hawoong Jeong (2008), Googling hidden interactions: Web search engine based weighted network construction, PACS.

Santa Fe Institute, http://www.santafe.edu/.

Schoemaker, J.H. Pau (1995), Scenario Planning: A Tool for Strategic Thinking, Sloan Management Review, Winter: 25-40.

Seung-Gu Ahn (2009), Analysis on the national R&D investment activities of Major countries in 2009, KISTEP. (in Korean)

Suk-Ho Sohn (2008), Evaluation of Korean Technology Foresight Program, KISTEP. (in Korean)

Wikipedia, http://www.wikipedia.org.

Yong-Hak Kim (2004), Social Network Theory, Parkyoungsa. (in Korean)

Young-Soo Yoon and Seung-Byung Chae (2005), Introduction to Complexity, Samsung Economic Research Institute. (in Korean)