SUPPORTING AIR TRANSPORT POLICIES USING BIG DATA ANALYTICS: A DESCRIPTIVE APPROACH BASED EMERGING TREND ANALYSIS

Hyun-jung Kim[a], Nam-ok Jo[b], Kyung-shik Shin[c], Jin-seo Park[d], Ga-ram Sim[e] and Je-chul Kim[f]*

ABSTRACT

Qualitative research methods based on literature review or expert judgement have been used to find core issues, analyze emerging trends and discover promising areas for the future. Deriving results from large amounts of information under this approach is both costly and time consuming. Besides, there is a risk that the results may be influenced by the subjective opinion of experts. In order to make up for such weaknesses, the analysis paradigm for choosing future emerging trend is undergoing a shift toward implementing qualitative research methods along with quantitative research methods like text mining in a mutually complementary manner. The change used to implement recent studies is being witnessed in various areas such as the steel industry, the information and communications technology industry, the construction industry in architectural engineering and so on. This study focused on retrieving aviation-related core issues and the promising areas for the future from research papers pertaining to overall aviation areas through text mining method, which is one of the big data analysis techniques. This study has limitations in that its analysis for retrieving the aviation-related core issues and promising fields was restricted to research papers containing the keyword "aviation." However, it has significance in that it prepared a quantitative analysis model for continuously monitoring the derived core issues and emerging trends regarding the promising areas for the future in the aviation industry through the application of a big data-based descriptive approach.

KEYWORDS

Aviation; Big Data Analytics; Text Mining; Topic Analysis; Trend Analysis; Descriptive Approach

Corresponding author: Jin-seo Park, Tel: 82-44-211-3074, E-mail: park727@koti.re.kr
[a]Seoul School of Integrated Sciences & Technologies (aSSIST),
46, Ewhayeodae 2-gil, Seodaemun-gu, Seoul 120-808, Republic of Korea
[b,c]School of Business, EwhaWomans University
11-1 Daehyun-dong, Seodaemun-gu, Seoul, 120-750, Republic of Korea
[d,e,f] The Korea Transport Institute
370 Sicheong-daero, Sejong-si, 339-007, Republic of Korea.

1. INTRODUCTION

Recently, there has been a surge of interest in finding core issues and analyzing emerging trends for the future. This represents efforts to devise national strategies and policies based on the selection of promising areas that can create economic and social added value. The existing studies, including those dedicated to the discovery of future promising fields, have mostly been dependent on qualitative research methods such as literature review, expert judgement and the Delphi method. This approach involves the gaining and processing large amounts of information and the analyzing and reasoning results from them. Therefore, it takes a lot of time and effort to implement as a research methods for emerging trend analysis. Besides, there is a risk that the subjective opinions of experts might affect the results.

Efforts have been made to make up for the weaknesses of the conventional qualitative analysis approach designed to select key promising areas through discovery of future core issues and emerging trend analysis in various areas of academic research: the information and communications technology industry (Chung and Lee, 2012), the construction industry (Jeong and Kim, 2012; Korea Institute of science and technology Evaluation and Planning, 2010), the steel industry (Min et al., 2014), the public sector (Korea Agency for Infrastructure Technology Advancement, 2013; Korea Institute of S&T Evaluation and Planning, 2014). There needs to be a paradigm shift in toward implementing qualitative research methods along with quantitative research methods like text mining in a mutually complementary manner. The change is to ensure objective and practical emerging trend analysis results based on large amounts of data. However, even such studies have had shortcoming related to their dependence on simple keywords for analysis, which makes it difficult to derive meaning from data. Besides, no study has been carried out so far to develop core issues and analyze emerging trends in special domains like the aviation industry. Based on the issues, it identified aviation-related research trends and selected the promising areas for the future.

In this study, unstructured text data are quantitatively analyzed through text mining, which is a big data analytics technique. In this manner, the promising future areas for the air transport industry are selected based on objective data from aviation-related research papers. In order to compensate for the difficulties in grasping the meaning of single words in emerging trend analysis at keyword levels, this study will adopt topic analysis, which is a technique used to find out general themes latent in text document sets. The analysis will

lead to the extraction of topics, which represent keyword sets, thereby discovering core issues and conducting emerging trend analysis.

Research on core issue retrieval and emerging trend analysis for the aviation industry based on big data analysis is still in its incipient stages. So, the analysis targets for this study are restricted to data from aviation-related research papers. In the future, the scope is slated to expand to cover relevant domestic or international news articles and bidding information as well, thus increasing the reliability of analysis results.

On the basis of the topic analysis results, core issues for the aviation industry will be determined. Then, emerging trend analysis for the issues will be implemented by year in order to identify the changes they undergo in time series. Through these procedures, this study aims to prepare a system for developing key promising areas for the future aviation industry as well as for ensuring rapid response. Additionally, the promising areas selected based on the aforementioned results and the analysis of pertinent policy research reports will be compared with the areas in which the actual government investments are made. The results from this comparative analysis are expected to make useful reference materials for future policy development and budget establishment.

2. APPROACHES FOR EMERGING TREND ANALYSIS

The trend analysis is to understand the current phenomenon and further prospect and forecast the future emerging trends. Emerging trend is defined as "a topic area that is growing in interest and utility over time," (Kontostathis et al., 2004). Landford (1972) categorized three approaches for forecasting emerging technologies as intuitive approach, exploratory approach, and normative approach. Intuitive approach is to utilize expert's knowledge for forecasting the future about specific technologies. Experts prospect the future technologies based on specialized knowledge of expert or information previously provided to experts. This approach includes Delphi method, brain storming, cross-impact analysis, analogy, gap analysis, and monitoring. Exploratory approach is a technique for forecasting continual shifts in the future through trends represented during the period from the past to the present in the condition that is not directed to specific social demand. This approach assumes properties of technology or multiple determinants affecting technology development is changed by the time-series pattern, and analyzed using techniques such as trend extrapolation, growth curve, substitution curve, correlation, regression, and technometrics. Normative approach is to regulate the future technology shifts or the

demand for technology development, and then suggest optimal technological plan to satisfy the future demand. This approach contains relevance tree, scenario, morphology, mission-flow diagram, and simulations.

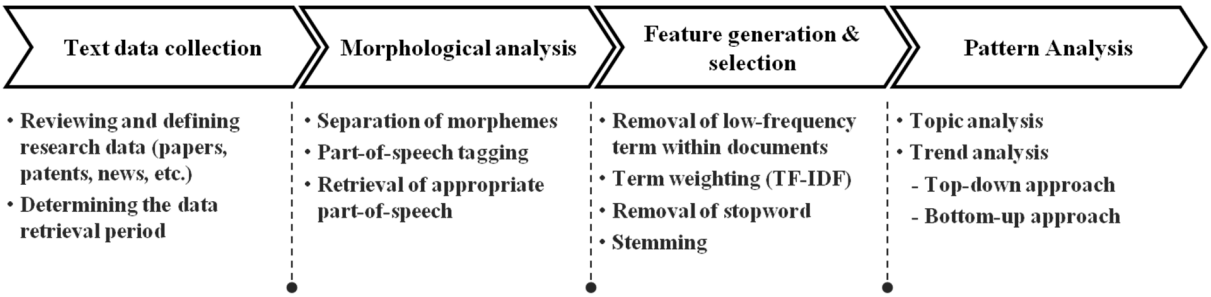
The existing studies have generally adopted qualitative research methods such as literature review, expert judgment such as the Delphi method among the various research methods for the selection of emerging issues based on trend analysis. This approach is contained to intuitive approach according to Landford (1972)'s study. Deriving results from large amounts of information under this approach is both time-consuming and costly, and it has weakness that may reflect subjective judgment. In this paper, we propose descriptive approach for analyzing changes of core issues and prospecting the promising areas in the future through trends analysis during the period from the past to the present. This approach mainly uses text mining techniques as an automatic data-driven approach. There has been no study that analyzes general trends using quantitative research methods such as text mining technique for the purpose of identifying core issues and prospecting the future promising areas of the aviation industry. This study is the first attempt to apply descriptive approach in the trend analysis for the aviation industry. The proposed approach extracts the latent themes of the aviation industry within document sets by using topic analysis, and analyzes changes of topics by year.

3. TEXT MINING

Text mining refers to a technique designed to obtain meaningful knowledge by deriving hidden patterns or relations from a large amount of unstructured text data comprised of natural language. This technique is based on the NLP (Natural Language Processing) technology that can understand languages spoken by humans. Text mining is considered as a part of the partial area of data mining, it differs from general data mining. Even though text mining is to search for meaning within unstructured text data, data mining is to a technique designed to discover patterns in structured data. While most of studies on knowledge discovery in data have focused on structure data, it is required for analyzing unstructured text data in that most of the available data is text data from various data sources according to the rapid of growth in the data available on the web (Feldman and Dagan, 1995). Text mining is used to extract meaningful information from vast amounts of text data, identify the association with other information such as structured data, and find out the categories corresponding to the text. It outperforms a simple information retrieval in terms of the scope of results that can be generated. The stages of performing text mining in

this study are illustrated in Figure 1.

Figure 1: Stages of Text Mining



3.1 Text Data Collection

The first stage is the process of collecting research data after selecting the sources. It is dedicated to collecting vast amounts of text data from a variety of sources such as office documents, email, news, blogs, and postings on social media. Data fit for analysis purpose should be selected.

3.2 Morphological Analysis

Morphological analysis is a stage when contents corresponding to terms, phrases and clauses are transformed into data forms suitable for language analysis processing. It is implemented through morpheme separation and part-of-speech tagging. Part-of-speech fit for research purpose should be extracted.

3.3 Feature Generation and Selection

After implementing morphological analysis, meaningful terms should be selected in order to make it possible to discover patterns hidden in text and analyze their trends. Term filtering involves the following procedures: removing low-frequency terms within documents, handling of stopwords, stemming, and assigning weights by term. First of all, terms are removed in case the number of documents containing them falls short of the minimum n. There are no established rules concerning this, so the minimum number should be determined through experiments. Terms difficult to understand, such as the definite article, as well as terms not used in the domain are processed as stopwords. In addition, terms with the same stem are processed as a single term in order to enhance the efficiency of text processing.

In order to store the processed text data as semantic information, term-specific weights are calculated by considering TF-IDF (Term Frequency-Inverse Document Frequency) which is widely used in the field of information retrieval, rather than just using term frequency (Salton & McGill, 1983). TF-IDF represents a value that can help determine the importance of a particular term in various document sets. TF is a value that reflects how often a particular term appears within a document. In general, the value increases proportionally to the importance of a term in a document. However, the frequent use of the term in a set of documents indicates that it is common. For this reason, not only term frequency but IDF (Inverse Document Frequency), which represents the reciprocal of DF (Document Frequency), is taken into consideration. IDF reflects how commonly a specific term appears within a set of documents. It is calculated by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient. The TF-IDF value is gained by multiplying the TF and IDF values, as shown in the following formula:

$$TF\text{-}IDF = TF \times log(N/DF)$$

Where

TF = Frequency of a terms within a document

N = Total number of documents

DF = Number of documents containing the term

IDF = Reciprocal of DF

Through the task of text preprocessing such as morphological analysis and feature generation and selection, an unstructured document collection is converted into a structured term-document matrix.

3.4 Pattern Analysis

In the last stage, information is reproduced through document classification or clustering based on the finally selected semantic information. Once a set of unstructured text documents is transformed into a structured, analyzable form, documents are clustered. Clustering is carried out through text clustering or topic analysis, grouping documents in accordance with similar characteristics. Text clustering and topic analysis are techniques used to discover clusters or topics hidden in a text document set. They involve clustering of documents in accordance with similarity based on the association of terms. Association

between terms is calculated by co-occurrence frequency within a set of documents. Text clustering is conducted by using such schemes as the EM (Expectation-Maximization) algorithm and the HAC (Hierarchical Agglomerative Clustering) method. Topic analysis was first based on a technique called LSA (Latent Semantic Analysis) suggested by Deerwester et al. (1990). Later, Hofmann (1999) proposed PLSA (Probabilistic Latent Semantic Analysis) by introducing a probabilistic concept into LSA. Lately, LDA (Latent Dirichlet Allocation), a technique proposed by Blei et al. (2003), is being used widely in various areas.

Conventional text clustering is based on the assumption that individual documents correspond to one theme. Thus, it has limitation that it is difficult to derive overall themes from large amounts of text documents. In contrast, topic analysis is based on the assumption that an individual document can contain complex themes dealing with various topics. A cluster or topic is represented as a set of multiple keywords. The task on naming of each cluster or topic should be determined directly by the researcher.

4. MODEL DEVELOPMENT

4.1 Research Data

Various information sources that can be used for generating core issues and analyzing emerging trends for the aviation industry include academic papers, policy research reports, patents and news articles. This study focuses on collecting aviation-related data from the academic research papers in Korea. For the trend analysis in this study, a total of 4,104 academic papers and policy research reports containing the keyword "aviation" were selected from among those published in Korea since 2000. Analysis focused on the period from 2000 and to September 2014, during which the nation laid the groundwork for take-off of its aviation industry beginning with the opening of the Incheon International Airport. Also taken into consideration was the fact that domestic aviation research started in a full-fledged manner in 2000. Detailed contents of the target data are presented in Table 1.

Table 1: Research Data

Source	Retrieval period	Frequency
NDSL academic papers	2000.1 ~ 2014.9	2,780
KISS academic papers		1,214
PRISM research reports		110
Total		4,104

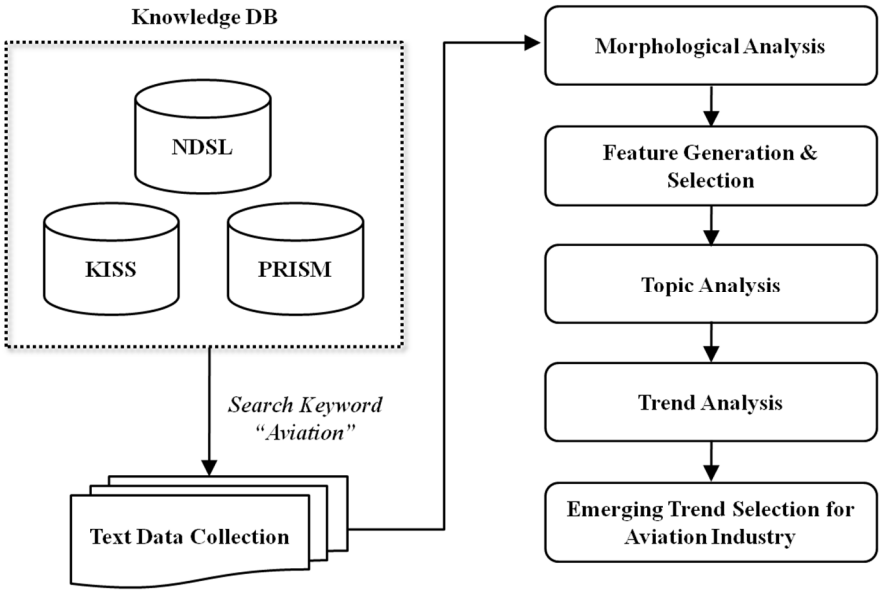
NDSL (National Digital Science Library) is a database that provides data on academic papers, patents, reports, trends, and factual information. As NDSL is focused on the science and technology fields, this study additionally used KISS (Korean Studies Information Service System), which is oriented toward social sciences. PRISM (Policy Research Information Service & Management) is a system designed to effectively manage the policy research tasks implemented by the central government agencies and share the policy research reports.

NDSL search led to the retrieval of 2,780 aviation-related papers, which had been published in 415 journals over a period of about 10 years. Large portions of the papers were found from the following journals: Journal of The Korean Society for Aeronautical & Space Sciences, Journal of the Korean Society for Aeronautical and Flight Operation, Journal of Surveying Geodesy Photogrammetry and Cartography, Aerospace Engineering and Technology, and The Korean Journal of Air & Space Law and Policy. Through KISS, 1,214 papers were discovered from 195 journals, which included prominent ones such as The Journal of the Korea Navigation Institute, Journal of the Aviation Management Society of Korea, Tourism Research, Journal of Tourism Management Research, Korean Journal of Hotel Administration, and International Journal of Tourism and Hospitality Research. Because NDSL and KISS databases were found to have the same document, only one of them was taken in order to avoid duplication. From PRISM, a total of 110 research reports were found in relation to aviation policies.

4.2 Experimental Design

Research on core issue generation and emerging trend analysis for the aviation industry is still in its early stages. Pilot experiment is conducted to build a framework based on descriptive approach for emerging trend analysis and examine the feasibility for the aviation domain application. Topic analysis, one of the text mining techniques, is employed as a way to retrieve the key issues affecting the aviation industry from academic research paper. The proposed research model and the analysis scope of this study are shown in Figure 2. Data-based model development based on descriptive approach for emerging trend analysis in the aviation industry is implemented through main two steps: text document collection and text mining based trend analysis.

Figure 2: Text Mining Based Trend Analysis



Text Data Collection

Various information sources that can be used for generating core issues and analyzing emerging trends for the aviation industry include academic papers, policy research reports, patents news articles, and bidding information. However, this study uses academic research papers and policy research reports. News articles data have copyright issue, and there also exist restrictions stemming from the difficulty of standardizing data crawled from a diversity of sources on the Web. Academic research paper data are relatively applicable to collect and analyze because they are composed of structured formats and items such as author, year of publication, abstract and keywords. In the case of policy research reports, data collection efforts faced limitations because of some organizations' practice of not disclosing the texts and due to the use of different items for mutually related information. In some cases, titles provided in summarized information by theme were used for data collection.

Research papers containing the term "aviation" in text data collection steps are retrieved through search of titles, abstracts and keywords. In general, the search keyword is selected from words specifically used in the pertinent domain, or chosen by experts. The keyword should be picked based on the relevance between research data and the purpose of the emerging trend analysis. This study is aimed at analyzing the overall aspects of emerging trends in the aviation sector, so the general term "aviation" was chosen as the search keyword. It will be necessary to prepare more sophisticated keywords when conducting

emerging trend analysis in the future targeting detailed areas within the aviation sector. Additionally, policy research reports are collected in order to compare the major issues dealt with in academic papers and policy research projects. The research reports are also analyzed to check whether proper investment is made in the core areas for the aviation industry.

Next is the process of clearly defining the scope of research data among a total of 4,104 academic papers and policy reports containing the keyword "aviation" selected for the emerging trend analysis in this study. Retrieved from search of NDSL, KISS, and PRISM databases are systematically organized and classified by year of publication, author, title, and organization in charge.

Text Mining-Based Trend Analysis

Based on the retrieved data, text mining, which is one of the big data analysis technique used to analyze the data in text form, is implemented through analyzing and processing technologies for the unstructured text data. Analysis targets are the theme of aviation-related academic papers and policy research reports. First of all, text preprocessing work, including morpheme analysis and feature selection, is implemented. This is followed by topic analysis to discover themes hidden in text document sets as well as trend analysis designed for time series review of the discovered topics. In topic analysis, documents are divided into groups based on similarity of themes. The similarity is determined by words' co-occurrence frequency within a document set. A topic is described as a set of one or more keywords, and topic analysis is based on the assumption that individual documents can handle not only single but multiple themes. Thus, topic analysis can be a useful tool for extracting general themes from large quantities of documents.

A topic extracted from a research paper can be understood as a set of keywords. The frequency of a particular topic's appearance within a document set reflects the level of interest in the concept the topic represents. Estimating the level of such interest makes it possible to analyze the trends regarding a particular concept. Eventually, it makes it possible to generate the core promising areas in the future for the aviation industry.

There are two approaches that can be used for trend analysis designed for time series review of the discovered topics. The first one is the top-down approach. In this method, the final topic is picked first, being followed by analysis of trends by year. After extracting a

topic from the entire data, frequency-based analysis is conducted by year based on the same topic. It can determine whether a topic is on the increase or decrease in terms of frequency. The second method is the bottom-up approach, which is to derive topics by era and analyze the trends. It makes it easy to grasp the topics that can be differentiated by era. However, it is difficult to examine the trends of topics taking place from the past up to the present. For this reason, this study adopts the top-down approach**.** By using this method, it implements trend analysis for the aviation industry, focusing on the identification of trends by year regarding the same topics.

This paper analyzes each topic's trends by using time series graphs showing yearly changes in the retrieved keywords and the discovered topics. The core issues and areas are generated through reflection of experts' opinions in addition to the keywords and the topics derived through topic analysis. This process is followed by the mapping of topics for each area. Finally, the topic trends are illustrated by using time series graphs by area. Through this process, topics showing growth trends are picked as the core promising areas for the aviation industry.

5. RESULTS AND ANALYSIS

A total of 261 keywords were retrieved as semantic information to be used for topic analysis from a total of 3,994 aviation-related academic research papers published between 2000 and 2014. The frequency and TF-IDF values of the retrieved keywords are exemplified in Table 2. We used TF-IDF values to weight terms for semantic information retrieval in the topic analysis.

Table 2: Frequency and TF-IDF of Derived Keyword Examples

No.	Keyword	Frequency	TF-IDF
1	Hub	76	7.534
2	Threat	82	7.317
3	Lidar	134	7.260
4	Job satisfaction	126	7.260
5	Disaster	86	7.260
6	Aviation safety	87	7.179
7	Airline service	74	7.179
8	Prevention	75	7.153
9	Cell	179	6.983
10	Low cost airline	160	6.960

On the basis of topic analysis results as well as experts' opinions in the aviation domain, the emerging topics were broadly classified into the following issue categories: aviation policy/air transport industry, airport, safety/security, and environment/technology. The topics and keywords corresponding to the core issue categories are presented in Table 3. Topic analysis led to the development of 23 topics, each of which was represented as a set of five keywords.

Table 3: Emerging Topics Derived for the Aviation Industry

Issue Categories	Topics	Keywords
1. Aviation policy/ Air transport industry	1.1 Aviation safety policy	Aviation, Safety, Accident, Aviation safety, Operation
	1.2 Airfare (low-cost carriers)	Carrier, Cost, Low cost airline, Operation, Value
	1.3 Distribution channels	Travel, Agency, Airline, Distribution, Channel
	1.4 Job satisfaction	Job, Satisfaction, Job satisfaction, Employee, Commitment
	1.5 Aviation agreements (baggage liability, etc.)	Law, Liability, Convention, State, Damage
	1.6 Flight attendant training/management	Flight, Attendant, Flight attendant, Commitment, Training
2. Airport	2.1 Airport service appraisal	Service, Quality, Passenger, Service quality, Satisfaction
	2.2 Aviation logistics	Airport, Cargo, Passenger, Facility, Logistics
	2.3 Airport hub strategy	Airport, Facility, Passenger, Security, Hub
	2.4 Noise control measures	Noise, Level, Airport, Measurement, Vibration
3. Safety/ Security	3.1 Air traffic control (collision prevention)	Control, Traffic, Controller, Response, Demand
	3.2 Air accident prevention	Accident, Passenger, Damage, Risk, Liability
4. Environment/ Technology	4.1 Eco-friendly high-efficiency fuel	Fuel, Cell, Power, Energy, Density
	4.2 Aircraft wing/shape design optimization	Aircraft, Operation, Wing, Landing, Stability
	4.3 Radar	Radar, Vehicle, Antenna, Traffic, Performance
	4.4 Sensor (error prevention)	Sensor, Error, Accuracy, Camera, Measurement
	4.5 Spatial resolution enhancement	Image, Camera, Resolution, Feature, Photo
	4.6 Unmanned aircraft	Vehicle, Path, Unmanned aerial vehicle,

		Aerial, Flight
	4.7 Lidar	Building, Lidar data, Lidar, Surface, Height
	4.8 Engine	Engine, Flow, Fuel, Performance, Temperature
	4.9 Composite materials	Material, Property, Composite, Temperature, Strength
	4.10 Digital map	Map, Digital map, Accuracy, Photo, Road
	4.11 Aerial photo	Land, Photo, Aerial photograph, Management, Construction

The derived topics, which can be understood as keyword sets, were mapped onto the four core issue areas. There are six topics in the aviation/air transport industry category, four in the airport category, two in the safety/security category, and 11 in the environment/technology category. Among the derived topics, the following 10 were selected as the highest ranked ones in terms of frequency: aircraft wing/shape design optimization, air traffic control (collision prevention), radar, flight attendant training and management, sensor (error prevention), airport service appraisal, spatial resolution enhancement (geographical features), aviation logistics, aviation safety policy, air accident prevention, unmanned aircraft, radar, engine, composite materials, and airfare (low-cost carriers). These can be considered to be the core issues being studied in the aviation sector. The selected topics and frequency of each core issue category are summarized in Table 4.

Table 4: Core Emerging Topics of the Aviation Industry

Rank	Categories	Topics	Frequency
1	Environment/ Technology	Aircraft wing/shape design optimization	614
2	Safety/Security	Air traffic control (collision prevention)	466
3	Environment/ Technology	Radar	441
4	Aviation policy/ Air transport industry	Flight attendant training/management	436
5	Environment/ Technology	Sensor (error prevention)	425
6	Airport	Airport service appraisal	423
7	Environment/ Technology	Spatial resolution enhancement (geographic features)	389

8	Airport	Aviation logistics	388
9	Aviation policy/ Air transport industry	Aviation safety policy	387
10	Safety/Security	Air accident prevention	370

We additionally retrieved the keywords and core issue categories to discover the investment areas from the aviation-related policy research reports. Then, they were compared with the areas in which the actual government investments were being made. The 10 keywords derived from the policy research reports are safety management, noise control measures, space, small craft, aviation accidents, aviation demand, aircraft certification, service appraisal, air traffic control, and unmanned aircraft. The level of correspondence was 60% between the academic research areas and government investment fields as shown in Table 5. It is determined that the non-matched areas such as small aircraft, aviation demand, and aircraft certification will be studied in the academic field for the future. The results from this comparative analysis are also expected to make useful reference materials for future policy development and budget establishment.

Table 5: Correspondence between Core Emerging Topics and Issues in Aviation Policy Research

Rank	**Categories**	**Keywords**	**Frequency**	**Correspondence**
1	Safety/Security	Safety management	13	○
2	Airport	Noise control measures	6	○
3	Aviation policy/ Air transport industry	Space	5	
4	Environment/ Technology	Small aircraft	4	
5	Safety/Security	Aviation accidents	4	○
6	Aviation policy/ Air transport industry	Aviation demand	4	
7	Safety/Security	Aircraft certification	3	

8	Airport	Service appraisal	3	○
9	Safety/Security	Air traffic control	3	○
10	Environment/ Technology	Unmanned aircraft	3	○

Trend analysis was conducted to identify the changes by era in the degree of interest in the finally chosen topics as well as to determine the promising fields for the future aviation industry. The data for 2014 were excluded from the trend analysis because they were available only for the first nine months. Finally, the topics corresponding to the core issues were analyzed through the top-down approach. First of all, topics are extracted from the entire data. Then, analysis is conducted to find out yearly changes in the degree of interest for the same topics. It represents a scheme to determine topics on the increase or decrease in terms of the level of interest. Frequency of a particular topic within a document set reflects the degree of interest in the concept represented by the topic. The annual degree of interest for an individual topic is calculated in the following manner:

Degree of interest = Frequency of a pertinent topic / Total number of documents

For trend analysis, the period from 2000 through 2013 was divided into three phases by reflecting relevant changes in the aviation industry. Then, time series patterns were analyzed for each topic. Phase 1 (2000~2003) was named the "period of stable growth and preparation for take-off," while Phase 2 (2004~2007) was called the "period of take-off and rapid change." Phase 3 (2008~present) was referred to as the "period of stability and second take-off."

Phase 1 saw the acceleration of the signing of aviation agreements as well as the introduction of international standards for aviation policies. In addition, efforts to secure aviation safety were intensified through the implementation of the confidential aviation incident reporting system, the aviation safety inspector system, the air operator certificate system, and the maintenance organization approval system. Following the 2001 opening of the Incheon International Airport, various standards were established, thus laying the groundwork for take-off the Korean aviation industry.

Phase 2 was a period when the hub strategy for Incheon International Airport was

implemented along with its second-phase construction project. An aviation safety management system was established to enhance the level of safety. In particular, low-cost carriers began to appear amid efforts to provide air services suitable for domestic situations. Their market participation was related to decreases in air demand caused by an economic slowdown and progress in the development of alternative modes of transport. Korea's first low-cost carrier was Hansung Airlines (presently T'way Air), which launched its flight services in August 2005. During the period of Phase 3, the 3rd-phase construction of Incheon International Airport began in a move to strengthen its competitiveness. The project is to be completed by the end of 2017. At present, two full-service carriers - Korean Air and Asiana Airlines - and five low-cost carriers are in operation in Korea. Other major developments included the preparation of autonomous air safety reporting regulations and the further easing of market entry restrictions for air transport operators. This was a period when the nation secured a basis for diversifying air transport businesses such as the operation of small aircraft.

The topics found to be on upward trends were picked as the core promising areas for the aviation industry. As demonstrated in Figures 3 to 6, researches pertaining to aviation were shown to be actively implemented during Phase 3, which were referred to as the stability period and the take-off period, respectively. The number in the following bar graphs was matched by the selected emerging topic in Table 3.

Figure 3: Trend Analysis for the Issue of Aviation Policy/Air Transport Industry

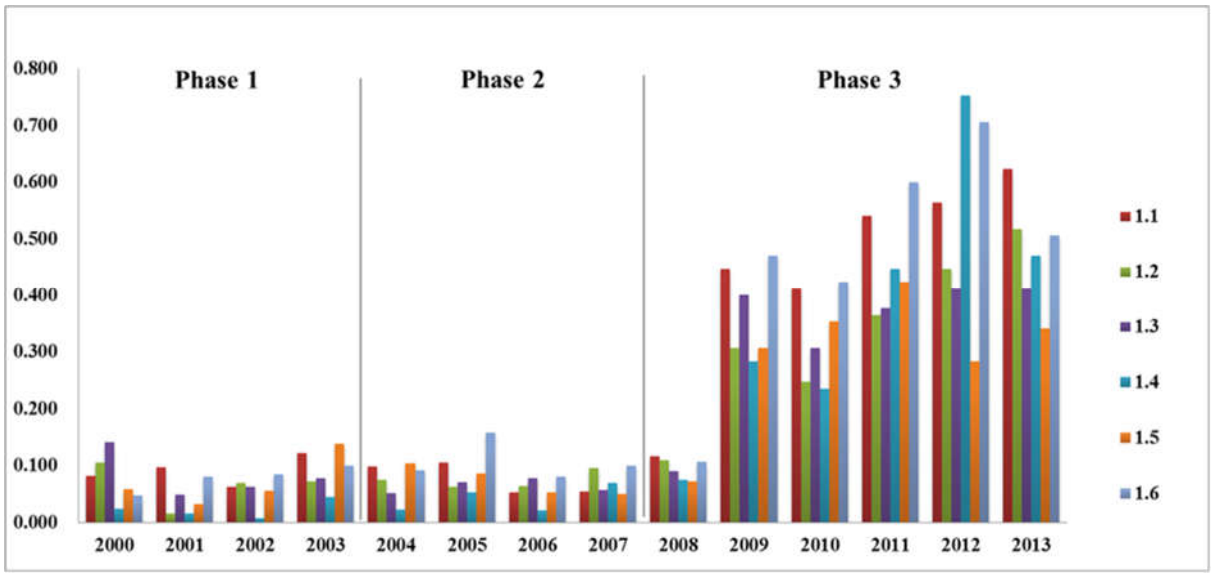


Figure 4: Trend Analysis for the Issue of Airport

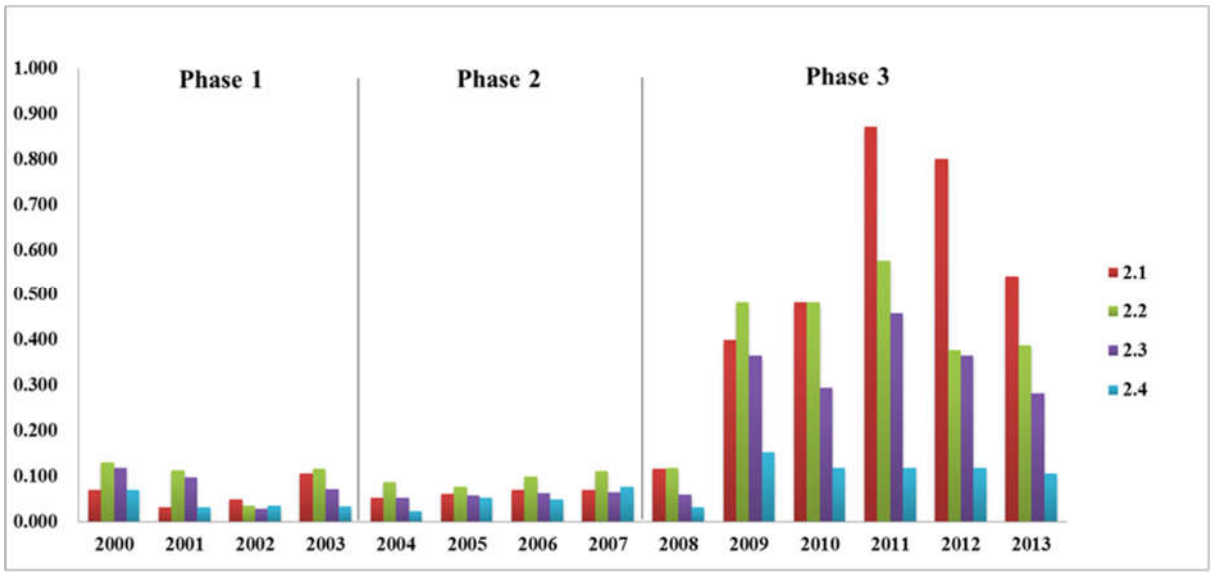
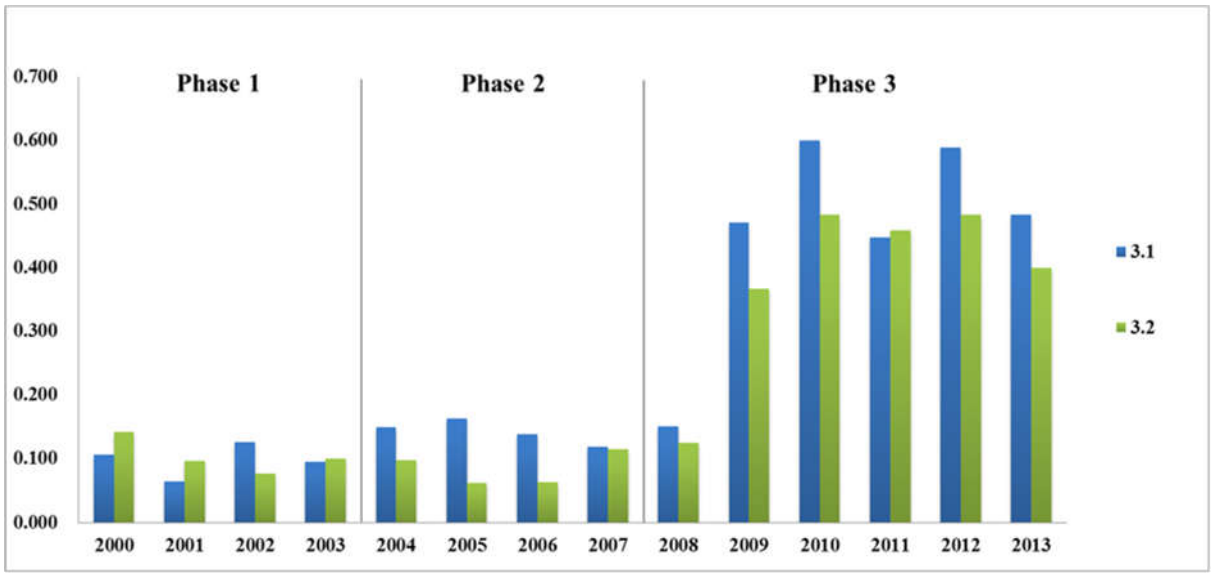


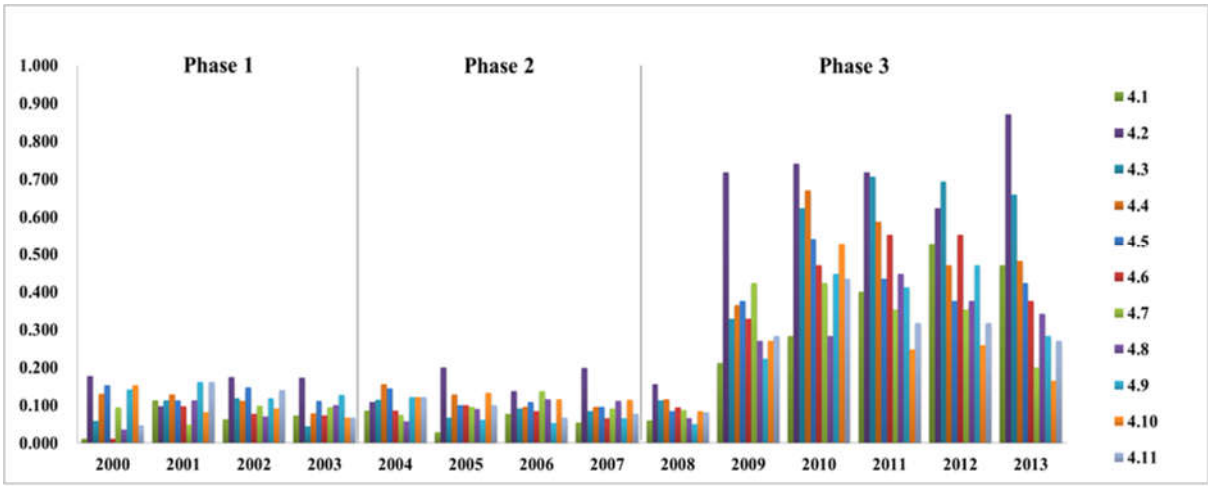
Figure 5: Trend Analysis for the Issue of Safety/Security



We categorized topics as three types through time-series pattern analysis: upward trends, downward trends, and types without upward or downward trends on the degree of interest for a topic. Types without specific patterns indicated that that it was consistently considered as important topics without trends or downward trends on the degree of interest for a topic. Topics with high degree of interest were selected as the future promising areas in the aviation industry by reflecting the number of three or more consecutive upward trends during the entire research period and the number of upward trends during phase 3,

considering that aviation-related studies were shown to be actively conducted in recent years.

Figure 6: Trend Analysis for the Issue of Environment/Technology



Topics such as aviation safety policy, airfare (low cost carriers), and eco-friendly high-efficiently fuel were found to be on upward trends. First, aviation safety policy, a topic that belongs to the issue category titled aviation policy/air transport industry, has drawn an increasing level of attention since the nation's aviation industry entered a period of stability and second take-off, as manifested in Figure 7. This was a period when international standards on aviation safety policy were introduced into the nation. In May 2008, Korea underwent a safety audit administered by the International Civil Aviation Organization (ICAO). Through the audit, Korea was certified to have implemented 98.89% of the relevant international aviation safety standards. Aviation accidents occur at lower frequency than those of other modes of transport such as trains, buses and ships. Once they occur, however, they cause large numbers of human casualties and serious damage to aircraft. Besides, aviation users become distrustful of the airlines involved. For these reasons, aviation safety policy is expected to keep attracting a high level of attention. There is a need to continuously improve the aviation safety system and to prevent accidents, flight delays and cancellations.

Second, airfare is a topic that belongs to the issue category of aviation policy/air transport industry. Particularly with regard to low-cost carriers, the airfare issue kept attracting a high level of interest beginning in 2009, as demonstrated in <Figure 8>. Low-cost carriers entered the aviation market as the nation was seeking a new air service system suitable for domestic conditions, following a drop in aviation demand caused by an economic slowdown and the development of alternative modes of transport. In August 2005, Hansung Airlines

(predecessor of T'way Air) began operations as the nation's first low-cost carrier. Four more budget carriers launched their commercial flights over the next several years: Jeju Air in June 2006, Jin Air in July 2008, Air Busan in October 2008, and Eastar Jet in January 2009. According to statistics compiled by the Ministry of Land, Infrastructure and Transport, the airline domestic market shares in the first quarter of 2014 reached 29.4% for Korean Air and 23.1% for Asiana Airlines. The share for low-cost carriers were 13.2% for Jeju Air, 11.9% for Air Busan, 7.9% for Eastar Jet, 7.3% for T'way Air, and 7.2% for Jin Air. Domestic low-cost carriers increased their flights and opened new routes. These activities led to a rise in the number of people using the budget airlines. As a result, the market shares of Korean Air and Asiana Airlines went down. The low-cost and full-service carriers are expected to compete fiercely over the market shares. Amid such a competitive atmosphere, the airfare issue will likely continue to attract a high level of attention.

Figure 7: Trend Analysis for Aviation Safety Policy

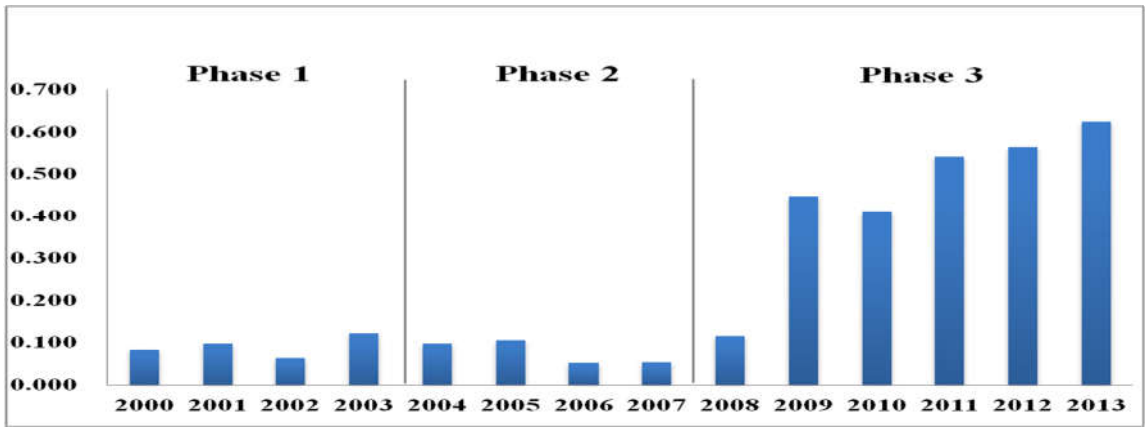
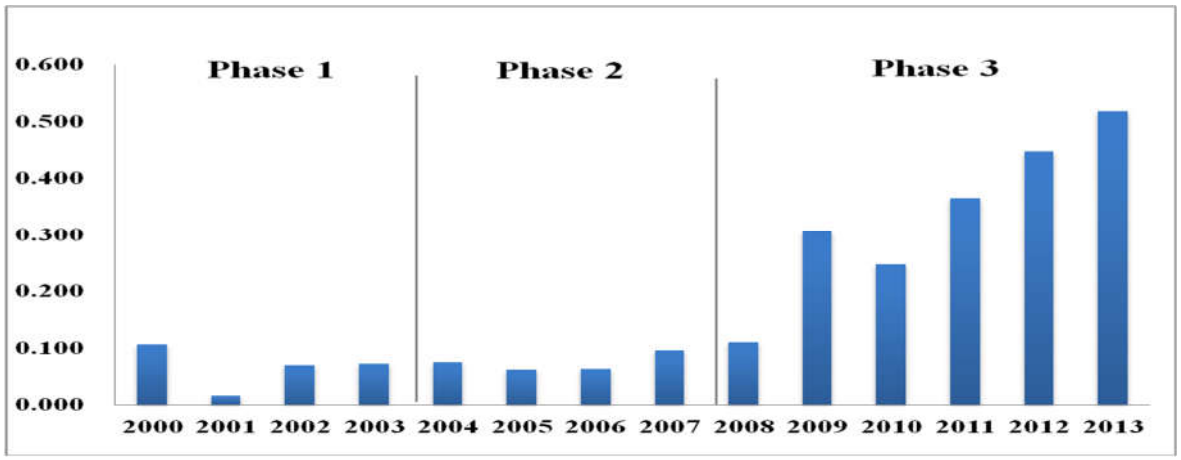
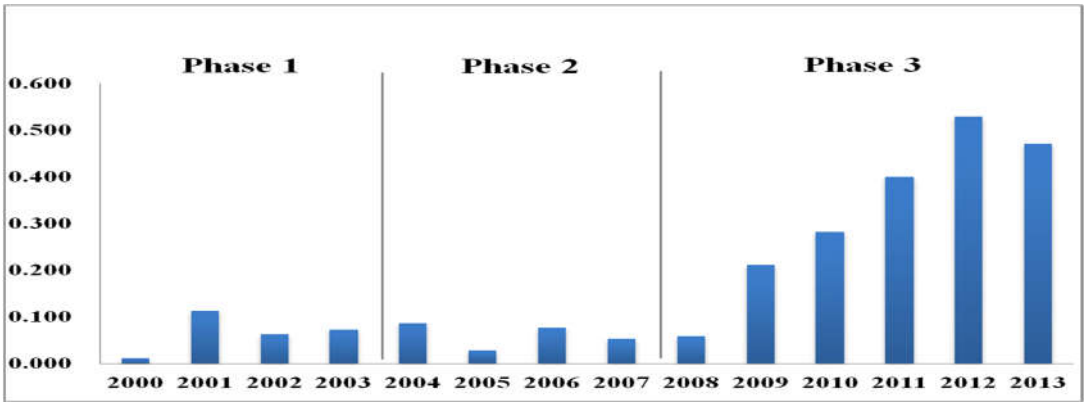


Figure 8: Trend Analysis for Airfare (Low-Cost Carriers)



Third, eco-friendly high-efficiency fuel is a topic that belongs to the environment/technology issue category. As shown in Figure 9, the level of interest in the topic has steadily increased since 2009, although it slightly went down in 2013. In relation to efforts to reduce greenhouse gas emissions from aircraft, the level of interest is particularly high for environment-friendly and high-efficiency fuels like hydrogen fuel cells, secondary batteries, and biofuels. In addition, a growing level of attention will likely be focused on matters related to greenization of the aviation industry, such as the use of composite materials, the use of fuel-efficient engines, and fuel conservation through shape design optimization.

Figure 9: Trend Analysis for Eco-Friendly High-Efficiency Fuel



Topics such as airport hub strategy, noise control measures, lidar, and aerial photos were found to be on downward trends. In contrast, the following are steadily regarded as important topics without specific patterns: circulation channel, job satisfaction, aviation agreement (baggage liability, carry-on baggage restrictions), flight attendant training and management, airport service appraisal, aviation logistics, air traffic control (collision prevention), air accident prevention, aircraft wing/shape design optimization, lidar, sensor (error prevention), spatial resolution enhancement (geographic features), unmanned aircraft, engine, composite materials, and digital map.

6. CONCLUSIONS

This study focused on deriving core issues for the air transport industry from aviation-related academic research papers by using a text mining method, a big data analysis technique, with a view to identifying the relevant trends and making predictions on promising areas for the aviation industry. It has limitations as its research for discovering the core issues and promising areas was restricted to academic research papers containing

the keyword "aviation." However, it has significance in that it has helped establish a quantitative research method for generating and steadily monitoring aviation-related core issues as well as for presenting directions of core promising areas in the future.

Research on extracting core issues and conducting emerging trend analysis for the aviation industry through the application of a big data-based descriptive approach is in its early stages. However, given the rapidly rising number of research papers, news and patents, it seems essential to prepare measures to cope with changes in relevant technologies and environments. In the future researches, efforts will be made to enhance the reliability of analysis results by selecting more detailed search keywords and expanding the scope of research sources to cover news articles, patents, and bidding information as well.

To increase the accuracy of the text mining based trend analysis results, in-depth studies will have to be carried out continuously on ensuring research data coding standardization and defining the pertinent standards as well as establishing an automation scheme for text representation. Such studies will help develop more objective analysis methods for discovering the core issues and promising areas pertaining to aviation, laying the groundwork for establishing mid- and long-term policies aimed at securing the competitiveness of the aviation industry. In addition, reflecting the opinion of domain experts will additionally considered for capturing the correct meaning of topics through a set of keywords.

REFERENCES

Blei, D. M, Ng, A. Y. and M. I. Jordan (2003), "Latent Dirichlet Allocations," *Journal of Machine Learning Research*, Vol.3, pp.993~1022.

Chung, J. H. and S. M. Lee(2012), "GSA-based future ICT technology prediction process," *ie Magazine*, Vol. 19, No. 3, pp.34~40.

Dalkey, N. C. and O. Helmer (1963), "An Experimental Application of the Delphi Method to the Use of Experts," *Management Science*, Vol.9, No.3, pp.458~467.

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., and R. A. Harshman(1990), "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science (JASIS)*, Vol. 41, No. 6, pp.391-407.

Feldman, R. and I. Dagan (1995), "Knowledge Discovery in Textual Databases (KDT)," *KDD*, Vol. 95, pp.112-117.

Hofmann, T. (1999), "Probabilistic latent semantic indexing," *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 50~57.

Jeong, C. W. and J. J. Kim (2012), "Analysis of trend in construction using text mining method," *Journal of the Korean Digital Architecture·Interior Association*, Vol. 12, No. 2, pp.53~60.

Jeong, D., Kim, J., Kim, G. -N., Heo, J. -U., On, B. -W., and M. Kang(2013), "A Proposal of a Keyword Extraction System for Detecting Social Issues," *Journal of Intelligence and Information Systems*, Vol. 19, No. 3, pp.109~122.

Kontostathis, A., Galitsky, L. M., Pottenger, W. M., Roy, S., & Phelps, D. J. (2004), A survey of emerging trend detection in textual data mining, *Survey of Text Mining*. Springer New York, pp.185~224.

Korea Agency for Infrastructure Technology Advancement (2013), "Technology Forecasting 2040; Land, Infrastructure, and Transport".

Korea Institute of science and technology Evaluation and Planning (2014), "KISTEP 10 future technologies for next 10 years".

Landford, H. W. (1972), Technological Forecasting Methodologies: A Synthesis, American Management Association, Inc.

Min, K. Y., Kim, H. T., and Y. G. Ji (2014), "A Pilot Study on Applying Text Mining Tools to Analyzing Steel Industry Trends: A Case Study of the Steel Industry for the Company "P"," *Journal of the Society for e-Business Studies*, Vol.19, No.3, pp.51~64.

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

* This work was translated into English from the original Korean study of Supporting Air Transport Policies Using Big Data Analysis published by the Korea Transport Institute in 2014.