Copyright ⓒ2017 KSII

Practical Text Mining for Trend Analysis: Ontology to visualization in Aerospace Technology

Yoosin Kim[1] Yeonjin Ju[2], SeongGwan Hong[2], Seung Ryul Jeong[2]
[1] Big Data Analytics Department, University of Seoul, Seoul, Korea
[2]Business IT Graduate School, Kookmin University, Seoul, South Korea
yoosin25@uos.ac.kr, juy3625@kookmin.ac.kr, play_w7@kookmin.ac.kr, srjeong@kookmin.ac.kr
*Corresponding author: Seung Ryul Jeong

Received July 4, 2017; revised August 2, 2017; accepted August 14, 2017;
Published August 31, 2017

Abstract

Advances in science and technology are driving us to the better life but also forcing us to make more investment at the same time. Therefore, the government has provided the investment to carry on the promising futuristic technology successfully. Indeed, a lot of resources from the government have supported into the science and technology R&D projects for several decades. However, the performance of the public investments remains unclear in many ways, so thus it is required that planning and evaluation about the new investment should be on data driven decision with fact based evidence. In this regard, the government wanted to know the trend and issue of the science and technology with evidences, and has accumulated an amount of database about the science and technology such as research papers, patents, project reports, and R&D information. Nowadays, the database is supporting to various activities such as planning policy, budget allocation, and investment evaluation for the science and technology but the information quality is not reached to the expectation because of limitations of text mining to drill out the information from the unstructured data like the reports and papers. To solve the problem, this study proposes a practical text mining methodology for the science and technology trend analysis, in case of aerospace technology, and conduct text mining methods such as ontology development, topic analysis, network analysis and their visualization.

Keywords: Text mining, Ontology, Trend Analysis, Aerospace Technology

"A preliminary version of this paper was presented at ICONI 2016, and was selected as an outstanding paper."

https://doi.org/10.3837/tiis.2017.08.022 **ISSN : 1976-7277**

1. Introduction

Advances in science and technology are driving us to the better life but also forcing us to make more investment at the same time. The government has supported the policy investment to carry on the promising futuristic technology successfully. Indeed, a lot of resources has supported for science and technology R&D for several decades. However, the performance of the public investments is not reaching to the expectation, so thus many institutes leading the public investment for the future science and technology (S&T) are seeking how the investment would be managed well and required that planning and evaluation about the new investment should be on data driven decision with fact based evidence. In this regard, the government wanted to know the trend and issue of the science and technology with evidences, and has accumulated an amount of database about the science and technology such as research papers, patents, white reports, and R&D information[1].

Nowadays, the database is supporting to various activities such as planning policy, budget allocation, and investment evaluation for the science and technology but the information quality is not reached to the expectation because of limitations of text mining to drill out the information from the unstructured data like the reports and papers. For instance, a web system of an institute accumulated over four million S&T policy database and services the technology trend and information of R&D investments. Despite of the service with the huge data, the institute felt keenly the necessity of improvement of the text mining quality to satisfy users' requirement. Many institutes and companies are using text mining technology but there are always various limitations such as the mining procedure, linguistic asset, analytics, etc. To support the difficulties, this study proposes a practical text mining methodology for the science and technology trend analysis.

After emerging big data, especially unstructured data, text mining is considered as a suitable methodology to analysis them. Text mining is expected to find the hidden pattern or relationship, to extract valuable information from overloading text data through combining various techniques with data mining, natural language processing, machine learning, and knowledge management[2]–[4]. Additionally, text mining with unstructured big data requires a linguistic resource organized well to extract more accurate information. This linguistic asset, called ontology, should be developed by experts having professional knowledge in a specific domain. Noy and McGuinness [5] explained "an ontology defines a common vocabulary for researchers who need to share information in a domain." In addition, they explained why an ontology was developed. There are sharing common comprehension for somewhat to analysis, reusing of domain knowledge, making assumptions on a specific area, and separating domain specific knowledge from the common.

Therefore, in this research, we proposes a practical text mining methodology for the science and technology trend analysis, and conducts mining tasks such as making an ontology, topic analysis, and network relationship analysis. Especially, we are more focusing the domain based ontology because it is a critical resource to lead the text classification and topic extraction. In case of ontology development, we try to make an aerospace technology ontology, one of biggest and hottest domains in S&T. After developing the ontology, we conduct to analysis the trend and issue of aerospace technology with topic analysis and network analysis.

2. Related works

2.1 Text Mining

After emerging big data analysis with unstructured data, text mining has been raising as a remarkable method to extract the information and the clue from unstructured text documents. According to Feldman and Sanger[4], "Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suit of analysis tools (p.1)." In this regarding, text mining is expected to find the hidden pattern or relationship, to extract valuable information from overloading text data. The valuable information is extracted and applied in various domains such as marketplace forecasting[6], stock price prediction[2], [7], consumer reputation[8]–[10], and trend analysis[11].

Additionally, text mining combines various techniques with data mining, natural language processing, machine learning, and knowledge management[2]–[4], [12]–[14]. They are executed in several steps like data collection, NLP, feature extraction, statistical analysis, and visualization, so that text mining drives various results such as information extraction, categorization, clustering, visualization and summarization through the steps[4], [12]. Text mining could expand its techniques and skills into brand new area for text analytics. Representatively, opinion mining and sentiment analysis is one of spin-off from text mining, and social network analysis within SNS is very matched with text mining as well[2], [15].

By the way, text mining with unstructured big data requires a linguistic resource organized well for more accurate information extraction[4], [16]. The linguistic asset, called ontology, taxonomy, or dictionary is very important in text mining, and then it should be developed very well. So thus the ontology is made and validated by experts having professional knowledge of a target domain. Indeed, K2Base, our research target, has dealt with various domains from basic science filed to high tech application sciences, so that the research team decided the aerospace technology as one of biggest and hot trendy test domains in S&T.

2.2 Ontology for Text Mining

In text mining with unstructured big data, a linguistic resource organized well is required to deal with text data in more accurate. The linguistic asset, called the domain ontology, which share common comprehension of the knowledge structure among people or software, should be developed by experts having professional knowledge in a specific domain[5]. Especially, the categorization methodology applies the linguistic sources such as a domain ontology and lexicon into the feature referencing and validation of terms[4]. In order to improve text mining service quality, the ontology should be developed in a right way and adjusted to match with the latest S&T trend. In this regard, we followed the ontology development 101 method of the Knowledge System Research Institute at Stanford University's which is a remarkable guide to develop an ontology[5], shown in **Fig. 1.**

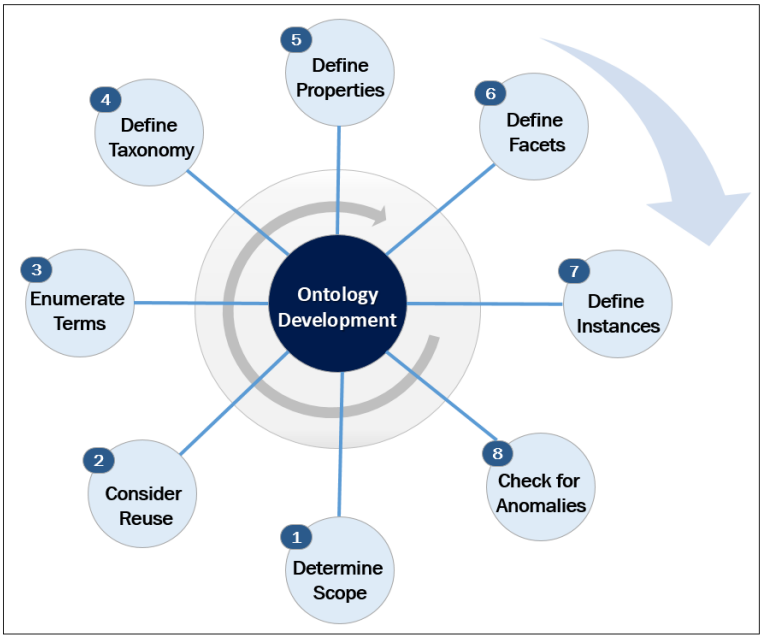

Fig. 1. Ontology development 101 method of Stanford Univ.

The ontology development 101 method suggested seven steps for developing an ontology. First step is determining the domain and scope of the ontology. The guide describes that the way to define the ontology scope is sketching a question list to be answered. Second step is considering reuse of existed ontologies. If we can discover and reuse an available one in the existed ontologies, it should be much helpful to develop the ontology. Third step is enumerating important terms in the ontology. It is useful to recognize a significant words list what the ontology state about or explain. Next fourth step is defining the classes and the class hierarchy. To make a class hierarchy, there are several approaches; top-down development process from general to specific classes, bottom-up from specific to general area, and combination process with top-down and bottom-up. Fifth step is defining the properties of classes. Since the classes cannot provide enough information to cover the competency questions of step one. Indeed, most of the remained words are likely to become properties of these classes. Sixth step is defining the facets of the slots. And last step is creating individual instances of classes in the hierarchy. After finish all procedure of ontology development, check for anomalies of the developed ontology.

3. Text Mining Methodology

In this study, we aimed to propose a practical text mining methodology for S& T trend analysis from data collection to visualization, and investigate it through a case study within an aerospace technology. As shown in **Fig. 2,** the proposed methodology is consisted with 4 phases; data collection, natural language processing, text analysis, and presentation. The methodology referred to the previous literature about text mining and opinion mining [2], [17], and is tailored as considering as practical text mining guide. Especially, we are more focusing the domain based ontology development because it is a critical resource to lead text analytics such as the topic extraction and text classification.

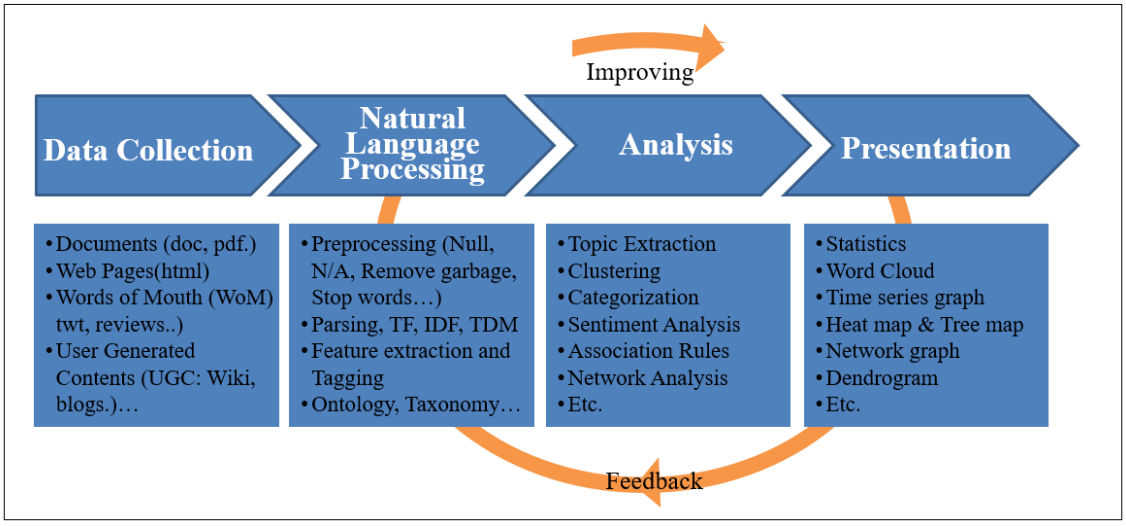

Fig. 2. Text mining methodology

3.1 Data Collection

The first step in the proposed text mining methodology is collecting source data from target sources. There are various text types of target sources such as working report, patent, research paper, online content, and social media data. Therefore the data collection method should be considered about data types and figures. Typically, four types are frequently used to gather the data; crawling robot, open API, database interface, and manual submission. The crawling robot, a kind of search engines in web portal site, explorers on Internet and gathers the web content which is targeted. Open API is easily to understand a protocol to access a database system and capture data. Especially, social network service such as Facebook and Twitter provide their API to access and collect their database. Another interface method is directly interfacing between deferent database management systems with their own interface protocol which they defined. Last, inconvenient but easy way is manual submission such key-in submission and data file uploading.

3.2 Natural Language Processing and Ontology Development

Whatever used in data collection, check the condition of gathered data, and qualify them for next step because the collected data might have too much garbage to analyze them without pre-processing. In this work, natural language processing for data cleaning is conducted as parsing text sentence, removing stop-words and disabled letters (i.e. html tags, punctuation, numbers, and emoticons), and convert countable data set such as term-document matrix or term list.

The qualified text data, after pre-processing, is ready to analyze right now, however, text mining requires the linguistic resources such as taxonomy, ontology, and sentiment dictionary, to obtain more accurate and efficient results. In previous research, the studies insisted that the domain-specific ontology has the same opinion and common comprehension of the knowledge structure among people or software in the domain, so thus it should be helpful to analyze text data including emotion, knowledge, opinion, thought, etc. In this regard, our proposed methodology defined the five steps for domain specific ontology development (**Fig. 3**). The ontology development process is redefined as tailoring ‘Ontology

Development 101'methodology and concludes 1) determine scope, 2) consider reuse, 3) extract terms form source data, 4) define taxonomy, 5) validate the preliminary ontology.

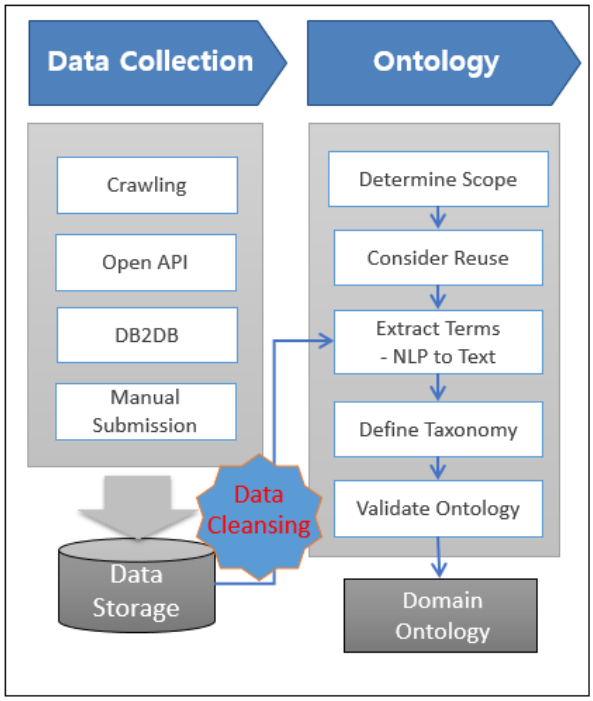

Fig. 3. Developing domain ontology after data collection

3.3 Text Analytics

The next phase is text analytics to mine hidden insight from the collected text data. There are various text analytics such as topic extraction, categorization, clustering, sentiment analysis, time series analysis, and so on. For instance, topic and buzz analysis are generally related with hot topic and issues on the market and society. Spam mail filtering is a method using text categorization analytics such as rule-based system or machine learning algorithms. Sentiment analysis tries to distinct the authors' mind, emotion, thought about products, services, people, and events. Sentiment analysis applies sentiment dictionary or machine learning approach to judge sentiment polarity, but the linguistic sources are very important in any way. And also statistics is applied in many way, to descript the characters of data set, compare figures in the same group or others, reveal relationship among variables and predict somewhat in future.

3.4 Presentation and Visualization

After text data analysis, various deliverables are generated as descriptions, tables, graphs, diagrams, and images. Whatever it conducted, the presentation in harmony with visual outputs is preferred to reader, who might be a client or a boss. That's why visualization in big-data analysis is emerging. Therefore, the last phase in the proposed methodology is about presentation of text mining results using visualization and others. There are simple one such as word cloud (tag cloud) to more complicating outputs such as time series graph, network analysis graph, and Dendrogram. Indeed, effective presentation is able to make text mining

results easier to understand. In this regard, the major purpose of the visualization is to make results simple, clear, and easy to comprehend its meaning and use it for decision-making.

4. Ontology and Trend Analysis

In this part, we tried trend analysis of Science and Technology with the proposed text mining methodology. Since the area of Science and Technology is too much broad, we targeted a specific domain, aerospace technology, which is one of biggest and hottest domains in Science and Technology. As following the proposed methodology, the text mining conducted data collection, domain specific ontology development, text data analysis, and visualization.

4.1 Data collection

To investigate the text mining methodology for trend and pattern analysis in aerospace technology, we collected data from the R&D data service system of the public institute[1]. The collected data is consist with research papers, patent documents, R&D project reports, and media news content (**Table 1**). Aerospace technology R&D project data has rich and various information categories, goals, keywords, effects, etc. Meanwhile, media news content and patent have poor information to investigate the trend of aerospace technology. Patent data, especially, has ten thousands aerospace technology patents but is seemed legal documents keeping away from the aerospace technology trend.

Table 1. Collected Data

Sources	No. of Doc	Contents
R&D Project	1,689	Year , title(detail), author, agency, category(Standard classification of Science and Technology), goals, keywords(Korean, English), contents, expectancy effects
Media News Content	1,048	Title, Year, Contents
Research article	910	Title(Korean, English), Author, Date, Categories & Codes(Classified Korea Research Foundation, KDC), pages, Keywords(Korean, English)
Patent	10,546	Date, title of invention(Korean, English), Abstract, applicant(Korean, English), Inventor(Korean, English)
Working Report	3,622	Date, Contents

4.2 An Ontology for Aerospace Technology

In this study, we developed the domain specific ontology for aerospace technology using the ontology development process in the proposed methodology; 1) determine scope, 2) consider reuse, 3) extract terms form source data, 4) define taxonomy, 5) validate the preliminary ontology.

Step 1. Determining scope

The research team decided to focus on aerospace technology domain as the scope of ontology. The ontology, the linguistic asset, is used for trend analysis of aerospace technology, therefore, it has common understanding of the knowledge structure among aerospace technicians. We thus referred to the aerospace technology categories of S&T Standard Classifications and roadmap for industrial space administration of Small and Medium Business Administration.

Step 2. Consider Reuse

According to the ontology development guide, it is considered if there are the existing linguistic resources, such as a dictionary, a white list, a synonym list, a stop-words list, and so on. If the linguistic assets were found, data analysts have to check them and decide whether reuse or not through term matching test. We tried to find the existing linguistic asset to be an ontology of aerospace technology, but we couldn't find out available resource.

Step 3. Extract terms by NLP

In this step, we set the collected data as two group, term extraction data and term validate data. Term extracting data is the data set to extract and choose the preliminary terms to be in the ontology, and research article and R&D project information were selected for the work. In project data, we conducted to split the terms, remove stop-words and disable terms, and remain meaningful words for aerospace technology. After extraction work, 3766 terms in R&D project data and 879 terms in research articles were left, and then 4608 terms except duplicated 37 words were remained finally.

Step 4. Define Taxonomy with domain experts

For attribute grant, we asked for review and selecting available terms to aerospace experts first. And then we set a term on the proper category and add the attributes through discussion with them. In this work, we and aerospace experts reviewed and applied categories and codes of S & T Standard Classifications, and Korea Research Foundation. As the result, the preliminary ontology in aerospace technology includes 4075 terms within eight categories.

4.5 Validate Ontology by term matching within documents

After determining the ontology, we conducted to test it through matching between ontology terms and the collected data. The matching rate is how many documents are recognized as the aerospace technology papers by the ontology terms. The terms separate the paper whether it is related with the aerospace technology and tag the features and detail categories at the paper. We used dataset in two way, extraction data set and texting data set. The matching rate in extraction data set is obviously higher as 87% and 78%, but the result in testing data set is relatively lower with 55% of news and working patent and 34% of patent (**Table. 2**). In fact, terms in the patent data are close to the legal terminology not to scientific content.

Table 2. Validating the Ontology

Data Set		Matching in	No. of Doc.	Ontology
Extract terms in	R&D Project	Title(detail), Goals, Contents	1689	1470 matching, 87.03%
	Research article	Title	910	706 matching, 77.58%

Test terms in	Media News Content	Title, Contents	1043	576 matching, 55.23%
	Patent	Title, Contents	10546	3592 matching, 34.06%
	Working Report	Contents	3622	2017 matching, 55.68%

4.3 Analysis and Visualization

In this phase, we tried practical text mining for aerospace technology trend analysis, and applied the domain-based ontology what we developed in front of phase. Analytics for text mining, indeed, is very wide and conducted with various methods, algorithms, and techniques. Applications of text mining are various in topic extractions, text classification, clustering, sentiment analysis, and predictions as well. However our research is more focusing on the domain specific ontology development and its practical use in text mining, therefore, this mining part is conducted into three representative analysis; topic analysis using word clouds, network analysis using association rules between terms, and emerging trend analysis on time series graph.

4.3.1 Topic analysis using word cloud graph

A word cloud (or tag cloud), is a visualization deliverables with word frequency in a text data set. Since the word cloud is a simple and powerful tool to see the issues and topic in the text data, so it has recently been popular to visualize the topic analysis in the text content. By the way, if the analysis is used without the linguistic sources, like the ontology, the result of it might be unclear by too much unavailable words. We conducted word cloud with the developed ontology and could see clearer and better graphs. A sample image of the word cloud is the analysis result of the research papers and shown the critical terms such as “space”, “aerospace”, “cosmology”, “satellite”, and “detection of lunar”, and so on (**Fig. 4**).

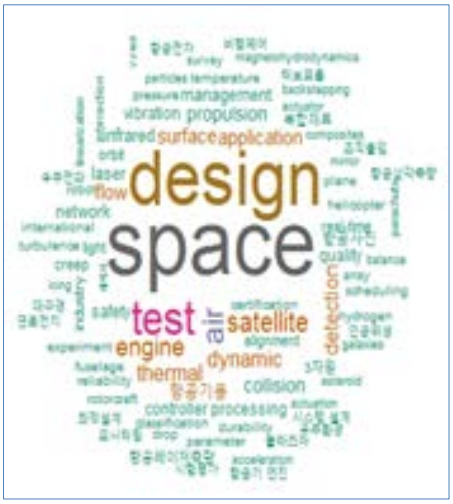

Fig. 4. A sample of topic word-cloud

4.3.2 Network graph

Social network analysis graph can visualize the relations among nodes which are terms in the ontology, the clusters which are close together, and the centrality, which is located in center of the group. To attempt the network analysis, first of all, node fairs are extracted using association rules analysis and then centrality of the nodes were calculated. In this work, we adopted the ontology to drop the clean result out. Following **Fig. 5** is the sample image of network analysis about research paper and thus it shows two groups clustering space and control.

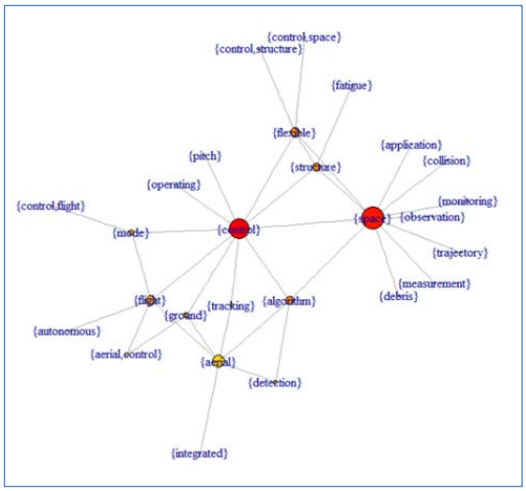

Fig. 5. A sample of network analysis graph

4.3.3 Trend analysis on time series graph

Time series analysis is another popular method to catch the emerging topics and item. We conducted time based trend analysis with aerospace technology keywords. In here, we reuses six keywords from the result of network analysis; cosmology, formation, galaxies, physics, properties and structure. The keywords' frequencies flow is investigated on a time graph as shown in **Fig. 6**. In the past, "structure" was presented frequently, especially in 2009, but it had decreased till 2012. By the way, all of the keywords' frequencies flow have increased after 2012 year.

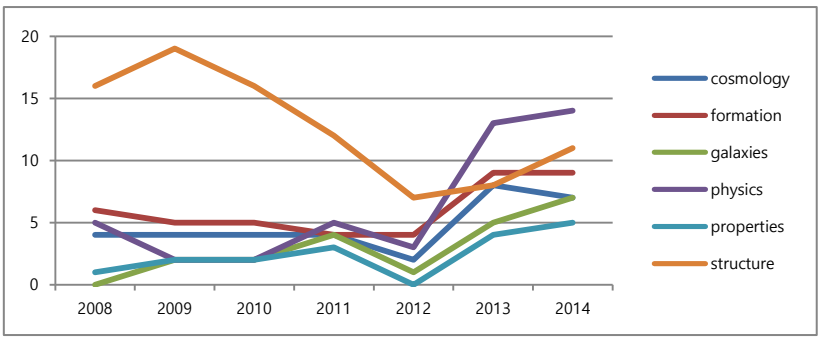

Fig. 6. Time series graph of project data

6. Conclusion

Advances in science and technology are driving us to the better life, so the government has invested science and technology R&D for several decades. The government have wanted to know how the investment is planned, executed, and managed within many institutes which conduct public investment. That is the reason several institutes and companies have accumulated huge S&T data and serviced the trend analysis which assists R&D investment activities such as appropriate and timely R&D planning, budget allocation and coordination, and performance evaluation. And as all of we know, most part of the accumulated data is unstructured text data such as documents, reports, and text records, hence text mining is considered to analyze them.

However the text mining system for S&T trend and issue analysis do not reach to the high performance to extract the information in unstructured data yet, because of limitations in text mining methodology, analytics, and especially linguistic resources. To support the difficulties, this study proposed a practical text mining methodology for the science and technology trend analysis and validated the methodology through the case study within an aerospace technology domain. Furthermore, we are more focusing the domain based ontology development because it is a critical resource to lead text analytics such as the topic extraction and text classification.

A proposed practical text mining methodology is consisted with four phases; data collection, natural language processing, text analysis, and presentation. The methodology referred to the literatures about text mining, opinion mining, and ontology development methods, thus it is tailored as a practical text mining guide. Moreover, we described the ontology development process in detail. And while we practiced and validated the methodology with a case of an aerospace technology, one of biggest and hottest domains in S&T, we could see the feasibility of the method. Indeed, we developed the aerospace technology ontology, conducted the trend and issue analysis with topic extraction, network relation, and time series analysis. In the results, the aerospace specific ontology with four thousands terms was delivered, it was used in text analytics like topic extractions, issue trend analysis, keyword network analysis and their visualizations.

As purposing in this study, we proposed the practical text mining methodology including the domain based ontology development, furthermore, showed the feasibility of the methodology and the ontology within the case of aerospace technology domain. We expect it would be a guide to not only aerospace technology but also other science and technology. The proposed methodology, indeed, was applied into upgrade the linguistic asset, ontology, and text mining techniques of an institute. Nevertheless, there are still many improvements in future. First of all, this research just studied on a domain, aerospace technology, so thus future studies should be expanded to the other domains such as medicals, computer technology, and so on. Another improvement is applying in various text analytics since this study was conducted just into three text analysis methods except the text classification, opinion mining, and predictions. Those text analytics would test the proposed text mining methodology and the ontology more, and should improve them.

References

[1] "K2BASE," *Korea Institute of Science & Technology Evaluation and Planning*. [Online]. Available: www.k2base.re.kr.

[2] Y. Kim, S. R. Jeong, and I. Ghani, "Text Opinion Mining to Analyze News for Stock Market Prediction," *Int. J. Adv. Soft Comput. Its Appl.*, vol. 6, no. 1, 2014. Article (CrossRef Link).

[3] W. He, S. Zha, and L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry," *Int. J. Inf. Manage.*, vol. 33, pp. 464–472, 2013. Article (CrossRef Link).

[4] Ronen Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2006. Article (CrossRef Link).

[5] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," *Stanford Knowl. Syst. Lab.*, p. 25, 2001. Article (CrossRef Link).

[6] Y. Kim, R. Dwivedi, J. Zhang, and S. R. Jeong, "Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5," *Online Inf. Rev.*, vol. 40, no. 1, pp. 42–61, 2016. Article (CrossRef Link).

[7] R. Schumaker and H. Chen, "A discrete stock price prediction engine based on financial news," *Computer (Long. Beach. Calif).*, no. January, pp. 51–56, 2010. Article (CrossRef Link).

[8] M. Chau and J. Xu, "Business Intelligence in Blogs: Understanding Consumer Interactions and Communities," *MIS Q.*, vol. 36, no. 4, pp. 1189–1216, 2012. Article (CrossRef Link).

[9] Z. Zhang, Q. Ye, R. Law, and Y. Li, "The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews," *Int. J. Hosp. Manag.*, vol. 29, no. 4, pp. 694–700, 2010. Article (CrossRef Link).

[10] Y. Kim, N. Kim, and S. R. Jeong, "Stock-index Invest Model Using News Big Data Opinion Mining," *Journal Intell. Inforamtion Syst.*, vol. 18, no. 2, pp. 143–156, 2012. Article (CrossRef Link).

[11] S. Kim, "Research Trends of the Credibility of Information in Social Q&A," *J. Korean Soc. Inf. Manag.*, vol. 29, no. 2, pp. 135–154, 2012. Article (CrossRef Link).

[12] G. Chakraborty, M. Pagolu, and S. Garla, *Text Mining and Analysis*. 2013. Article (CrossRef Link).

[13] J.-M. Lee and J.-Y. Rha, "Exploring Consumer Responses to the Cross-Border E-Commerce using Text Mining," *J. Consum. Stud.*, vol. 26, no. 5, pp. 93–124, 2015. Article (CrossRef Link).

[14] Y. Kim, D. Y. Kwon, and S. R. Jeong, "Comparing Machine Learning Classifiers for Movie WOM Opinion Mining," in *Proc. of KSII Trans. Internet Inf. Syst.*, vol. 9, no. 8, pp. 3178–3190, 2015. Article (CrossRef Link).

[15] B. Pang and L. Lee, *Opinion Mining and Sentiment Analysis*. 2008. Article (CrossRef Link).

[16] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, p. 168, 2004. Article (CrossRef Link).

[17] Y. Kim and S. R. Jeong, "Opinion-Mining Methodology for Social Media Analytics," in *Proc. of KSII Trans. Internet Inf. Syst.*, vol. 9, no. 1, pp. 391–406, 2015. Article (CrossRef Link).

Yoosin Kim is a Visiting Professor for Big-data Analytics Dept. in the University of Seoul. He received a Ph.D. with a research for Stock Index Prediction of News Big-data from Kookmin University in Seoul, Korea. He was a post-doctoral researcher in the University of Texas at Arlington, a data scientist at Accenture, and business analyst at SK. He has studied a fire-risk prediction model, Social Economic Index, a public service process mining methodology, and big data analytics.

Yeonjin Ju pursues a Master's degree at Kookmin University, Korea. Her research interests include text mining and business analytics. She has consulted for a number of organizations including Korea Institute of S&T Evaluation and Planning and Statistics Korea. Her current research topics include visualizations and big data analytics.

SeongGwan Hong pursues a Master's degree at Kookmin University, Korea. His research interests include text mining and big data analytics. He has consulted for a number of organizations including Ministry of the Safety and Statistics Korea. He has studied a Social Economic Index, and Korean beef price prediction by time series analysis.

Seung Ryul Jeong is a Professor in the Graduate School of Business IT at Kookmin University, Korea. He holds a B.A. in Economics from Sogang University, Korea, an M.S. in MIS from University of Wisconsin, and a Ph.D. in MIS from the University of South Carolina, U.S.A. Professor Jeong has published extensively in the information systems field, with over 60 publications in refereed journals like Journal of MIS, Communications of the ACM, Information and Management, Journal of Systems and Software, among others.