



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Master's Thesis of Public Administration

**A Study on Policy Agenda Setting
using Social Big Data Analytics
- Focused on the Case of Seoul National University -**

소셜 빅데이터 분석을 활용한 정책의제설정
에 관한 연구

February 2019

**Graduate School of Public Administration
Seoul National University
Global Public Administration Major**

Kim, Geum Sun

**A Study on Policy Agenda Setting
using Social Big Data Analytics**
- Focused on the Case of Seoul National University -

Academic Advisor Park, Jeong Hun

Submitting a master's thesis of Public Administration
September 2018

Graduate School of Public Administration
Seoul National University
Global Public Administration Major

Kim, Geum Sun

Confirming the master's thesis written by
Kim, Geum Sun
December 2018

Chair Kim, Soon Eun (Seal)

Vice Chair Koo, Min Gyo (Seal)

Examiner Park, Jeong Hun (Seal)

Abstract

A Study on Policy Agenda Setting using Social Big Data Analytics

- Focused on the Case of Seoul National University -

Kim Geum Sun
Global Public Administration Major
The Graduate School of Public Administration
Seoul National University

Many emerging technologies, such as smart devices, internet news, and social media, have recently increased openness and interactions between public policy makers and consumers. Policy makers have come to actively use social media to gather public opinion for policy development; for this reason, it is becoming increasingly important to analyze Big Data from a variety of social media.

This paper explores the applicability of Big Data to the development of policy agendas in higher education field with recently developed programs and methodologies. Specifically, this study analyzes issues regarding colleges and universities revealed by newspaper articles and social media using text Big Data analytics for the 2015-2017 period. Based on this research, public opinion on major policies of the university were grasped macroscopically and emotional evaluation of policy image was investigated. Five issues were found from analysis results. First, press articles have emphasized not only education and research as core functions of the universities, but also publicity and integrity of its members. Second, there are no specific patterns in the media in terms of time. However, when a single event became a social issue, the number of social media

buzzes increased explosively. In particular, negative issues spreading through SNS tend to spread rapidly in the first week after the occurrence. Third, keyword structure of university policy through SNS and blog analysis showed different patterns. From blogs, texts were steadily extracted every year related to the original function of university. However, SNS keywords tend to be limited to specific issues. Finally, quantitative increase and negative image of social media on specific issues influenced policy making of the university.

This study explains that analyzing social media data using text Big Data analytics could serve as an effective way to identify social issues related to educational policies and guide advanced future policy directions. In addition to implications of these research results, limitations of this study include subjectivity intervention of emotional image computation, a list of dictionary morphemes, limitations of emotional language, and the lack of a detailed description of the analysis system.

Keywords: College and University Policy, Text Big Data Analytics, Newspapers and Social Media, Agenda Setting, Topic Modeling

Student number: 2016-23124

Table of Contents

Chapter 1. Introduction	1
1.1 Study Background	1
1.2 Scope of the Study	3
Chapter 2. Literature Review	4
2.1 Discussion of Theory and Precedent Study Review	4
2.1.1 Public Policymaking Process as a Political System	4
2.1.2 Big Data in Policy Making	5
2.1.3 Policy Agenda Setting using Big Data	7
2.2 Precedent Study Review	10
2.2.1 Precedent Study on Policy Formulation Using Big Data	10
2.2.2 Social Media Analysis Method in Precedent Studies	13
2.3 Education Environment and Big Data Analysis	13
2.3.1 Policies on the Higher Education of the Republic Korea	13
2.3.2 Policies of the Seoul National University	16
2.3.3 Implications for Policies in Social Media	19
Chapter 3. Research Design and Method	21
3.1 Research Question	21
3.2 Method	22
3.2.1 Research Model	22
3.2.2 Context of the Study	22
3.3 Research Procedure	23
3.3.1 Data Collection	23
3.3.2 Text Mining	25

3.3.3 Semantic Network Analysis	27
Chapter 4. Data Analysis and Findings	29
4.1 RQ1: Keyword Analysis	29
4.2 RQ2: Degree Centrality Analysis	39
Chapter 5. Conclusion	43
5.1 Implications	43
5.2 Future Study and Limitations	46
 Bibliography	 47
 Abstract in Korean	 51

List of Tables

Table 1. The Policy Process	4
Table 2. Key Findings of Precedent Studies	11
Table 3. Moon Jae-In Government's National Tasks in the Higher Education	15
Table 4. Keywords in '2017 Work Plan - SNU'	17
Table 5. Major Press Reports on Seoul National University	24
Table 6. Summary of Collected Data	24
Table 7. Four Steps of Framework of Social Network Data Analysis	27
Table 8. High Frequency Words in Newspaper	29
Table 9. High Frequency Words in SNS	30
Table 10. Top 10 Keywords in Newspaper by Frequency	32
Table 11. Policy Related News Articles and Keywords	34
Table 12. Keyword Frequency of Blogs in Naver and Daum	37
Table 13. Keyword Frequency in Facebook	37
Table 14. Keyword Analysis of SNU by Social Matrix (2017. 5.)	38

Table 15. Analysis of ‘Entrance examination’ by Social Matrix (2017. 5)	39
Table 16. Analysis of 'Siheung Campus' by Social Matrix (2017. 5)	39
Table 17. TF-IDF of Top Frequency Words in Newspapers	40
Table 18. Results of Degree of Centrality	41

List of Figures

Figure 1. A Model of the Political System	5
Figure 2. Data-based Scientific Policy Formulation Methodology	9
Figure 3. Research Model	22
Figure 4. Data Pre-processing	26
Figure 5. Yearly / monthly News Articles	31
Figure 6. Interest Trend of SNU by Google Trend	36
Figure 7. Degree of Centrality for the Top 20 Keywords.....	42
Figure 8. Network CONCOR Analysis of Top 20 Keywords	42

Chapter 1. Introduction

1.1. Study Background

Institutions of higher education are operating in an increasingly complex and competitive environment. They are under increasing pressure to respond to growing national, economic, political, and social changes in their focus to enhance academic reputation, increase student enrollment rates, and improve the quality of learning programs. In order to meet national and global needs, educational institutions have generally used surveys and public hearings to better understand the dynamics of people's opinions. However, due to the nature of public opinion that do changes rapidly or is not directly exposed to the public, it is becoming more difficult to address the recent complex social needs with the traditional survey methods. Moreover, these conventional survey methods are likely to face the problem of consistent distortion from self-reported data. However, now digital exposures, such as online press articles and social networking service (SNS), have provided an opportunity to systemically grasp public needs.

Policies are designed to address future problems. A policy is a relatively stable, purposive course of action followed by an actor or set of actors in dealing with a problem or matter of concern (Anderson, 2006). Policy making is a process of exploring and selecting a problem, and designing and implementing a policy to address the problem. The process of problem search and selection involves a future prediction process. Due to resource constraints within the organization, policy-makers need to choose and concentrate on policy decisions, so decisions based on large data analysis can be one of the solutions.

As society becomes increasingly informative, diverse types of data, such as text, images, audio, and video, are accumulating. The universalization of digital devices and the development of SNS and electronic newspapers have brought about quantitative and qualitative expansion of data. We are working to increase the potential value and impact of digital data. In the past, just piled up and discarded data were recognized as a valuable source today that could give new value as the analytical technology evolves.

Global IT companies, consulting firms, and governments are looking for ways to build digital data strategies. It is said that the success of the future business depends on how to use data. Governments and the public sector are also pursuing a variety of policies that utilize digital information. The US government monitored trends through Twitter's real-time analysis and used them in health policy. In Europe,

they have developed toolboxes, named 'WeGov', to use social media data analysis in various policy evaluation (Hwang et al., 2014). The Korean government also introduced the concept of government 3.0, and proposed various policy measures to support the policy formulation through the utilization of electronic data.

On the other hand, there are few cases for policy making that use Big Data in the education environment, despite growing interests in exploring the value of data in education institutions. The educational policy of the Republic of Korea is changing every time the government changes; furthermore, the citizens' satisfaction level with education is low (Kim & Baek, 2016).). Especially, high competition for college admissions and employment is the main factor that lowers the quality of life. As a result, a more systematic approach to higher education policy is needed. In addition, it is necessary to study a systematical approach and ways to utilize Big Data in the higher education field.

Seoul National University (SNU), which is the analysis case of this study, operates the SNS channel directly and collects Internet newspaper articles to provide information to members in the university. SNU seeks to strengthen the image of a university that is leading global education institution in accordance with the educational paradigm changes, such as liberal education and knowledge contribution. In addition, as a national university, it focuses on the utilization and communication of SNS from various perspectives in order to formulate a friendly public opinion of its research achievements, social contributions, management innovations, and multi-campus constructions.

Policy agenda-setting using Big Data analytics will be a way to make more systematic and objective process instead of traditional policy-making procedures in higher education. In addition, customized policies can be established based on demand analysis data, and administrative costs can be greatly reduced. Also, it is possible to access administrative innovation through Demand Response Monitoring instead of a one-way policy push on the part of suppliers.

This study will explore the applicability of Big Data to develop policy agendas in higher education field with the recently developed programs and methodologies. In particular, this study focuses on finding implications for policy makers to establish policy agenda by analyzing social media data related to SNU using Big Data analysis technique. This will be a way to make more systematic and objective process instead of traditional policy-making procedures in higher education. In addition, customized policies can be made based on demand analysis data, and administrative costs can be reduced.

1.2. Structure of Research

This paper is composed of 5 chapters entirely. Chapter 1 explains the meaning and direction of research on social Big Data analysis that is the subject of this study. It also included research background, purpose, significance, and structure. In Chapter 2, I discussed various theories related to this research, precedent study review on policy formation using social media analysis, and social media analysis methods. In Chapter 3, I presented the specific research questions to be discussed in this paper, and also described the concrete data collection and analysis of social data included in the Big Data analysis. In Chapter 4, I showed the research results focusing on the Big Data analysis of the research questions raised in this paper. The results of the research are presented in 2 sections, focusing on the analysis results of each research question. Finally, Chapter 5 summarizes the findings and discusses the conclusions. In addition, the limitations of this study are raised and future study is proposed.

Chapter 2. Literature Review

2.1. Discussion of Theory and Precedent Study Review

2.1.1. Public Policymaking Process as a Political System

Many researchers have developed various models, theories, approaches, concepts, and definitions for analyzing policy-making and the related component of decision-making. The term ‘public policy-making’ is often used in a very general way in which all aspects of political science and public administration can be lawfully included. Anderson, in his book *Public Policymaking* (2006), clearly defined the policy formation stages by applying a sequential approach focusing on problem formation, policy formulation and adoption, implementation, and evaluation. He discussed the theoretical approaches, which included political systems theory, group theory, elite theory, institutionalism, and rational-choice theory.

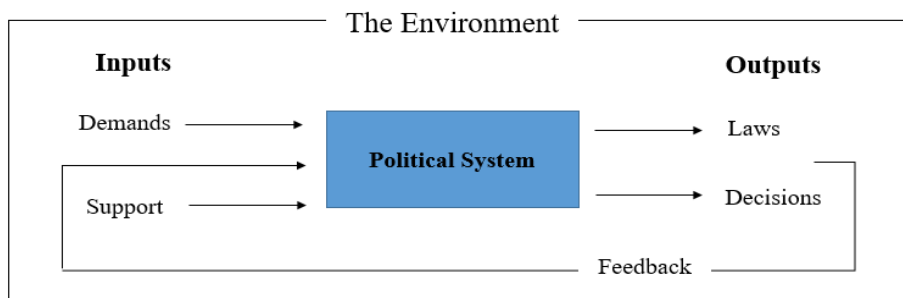
Table 1. The Policy Process

Policy Terminology	Definition	Common Sense
Stage1: Policy Agenda	Those problems, among many, that receive the serious attention of public officials	Getting the government to consider action on the problem
Stage 2: Policy Formulation	Development of pertinent and acceptable proposed courses of action for dealing with a public problem	What is proposed to be done about the problem
Stage 3: Policy Adoption	Development of support for a specific proposed so that a policy can be legitimized or authorized	Getting the government to accept a particular solution to the problem
Stage 4: Policy Implementation	Application of the policy by the government’s administrative machinery	Applying the government’s policy to the problem
Stage 5: Policy Evaluation	Efforts by the government to determine whether the policy was effective and why or why not	Did the policy work?

Source: Adapted from James E. Anderson, *Public Policymaking*, 2006, p. 4

However, since most of these approaches have not been specifically developed for the analysis of policy formation, decision-making in the public sector should be considered an inherently political process involving conflicts and struggles between conflicting interests, values, and desires on policy issues. Anderson described the policy formation process as a political system called a ‘black box’ because of the highly general and abstract nature of political influence. Inputs into the political system from the environment consist of demands and support. Demands are a claim to the actions that individuals and groups make to meet their interests and values. Support is provided when the public accepts decisions and actions made in the political system. Indeed, decisions are made and policies are developed within the black box. In Anderson’s book, he described the process of formalizing policies by taking the legislative process of the Family and Medical Leave Act (FMLA) as a case. The bill, in 1984, was triggered by the outrage of many women activists when the court first ruled that men were not allowed to take leave for childbirth. Since then, it has been legislated in 1993, through discussions and negotiations with interested members of Congress, the federal government, entrepreneurs, and the workers. Of course, there were many differences between the draft and the legislative bill in the scope and coverage of welfare. Stakeholder participation serves as a vehicle through which stakeholders become deeply involved in regulatory policymaking. Next, we discuss how stakeholder participation affects each step of policy making in the digital era.

Figure 1. A Model of the Political System



Source: Adapted from James E. Anderson, *Public Policymaking*, 2006, p. 19

2.1.2. Big Data in Policy Making

It has been argued that adopting and using information technology (IT) in an organization has an impact on organizational structure. There are few studies that have made an analysis that allows us to have a different perspective on the impact of IT in areas, such as policymaking in governments (Heintze & Bretschneider, 2000;

Höchtel et al., 2016). Heintze and Bretschneider (2000) attempted to explain the effect of IT on organizational change and organizational performance. They tried to prove their theories based on survey data of government organizations. The results showed that while the impact of IT on organizational structure is not significant, the introduction of IT has a direct effect on the performance of the organization and that the attitudes of organizational managers to IT also have a significant impact on organizational competence. It is argued that the organizations involved in this study are small, so it is difficult to generalize to all government agencies, but managers' attitudes to IT have an effect on decision-making.

The rapid development of information and communication technology has enabled various media based on networks, such as online articles and SNS, to function as important tools for increasing citizen participation in politics and administration. We think IT can be applied efficiently to describe policy decision-making in the digital age. 'The more quality and accurate information is available, the better the decisions will be' (Höchtel et al., 2016). In public administration, data can be considered as input to processes that seek new insights and can implement better regulations. Höchtel explained that government agencies can use Big Data or Big Data Analytics (BDA) to significantly speed up the decision-making process and improve the decision-making process. In Höchtel's article, the policy cycle process was divided into six steps: agenda-setting, policy discussion, policy formulation and policy acceptance, provision of means, implementation and continuous evaluation in the cycle. In the agenda-setting phase, Big Data plays an important role in finding the most pressing policy problem and setting priorities. In the two stages of policy formation and policy acceptance, since it is important to scrutinize before implementing them, the means of data collection can be used to examine the acceptance of specific policies among different groups of societies. Similar to the previous two stages, decisions on how to most efficiently provide the necessary personnel and financial means to implement the new policy, particularly in the budget and procurement processes, can be improved to allow detailed analysis of previous experiences. Implementation of policies can be affected by two aspects of Big Data: 1) the ability to pinpoint problem areas can be a way to implement different levels of policy enforcement. 2) the implementation of the new policy will immediately generate new data, which allow us to assess the effectiveness of these policies and improve future implementation processes by identifying problems. The policy evaluation step, which is the last stage of the policy cycle process, with advances in Big Data, enables the use of continuous evaluation of cases in public administration. Examples of continuous evaluation are the Automated Continuous Evaluation System of the US Army and the UK Government Program on

Performance Data of the UK. At the same time, however, these innovations should be applied not only to the introduction of new technologies but also to the bureaucratic environment with unique characteristics and organizational culture. Additionally, the introduction of BDA-based evidence-based policy decisions remains a challenge because all stakeholders must prepare the data in such a way as to examine the evidence in detail and draw conclusions based on facts and accessible figures so that the conclusions cannot be neglected even by stubborn politicians.

2.1.3. Policy Agenda Setting using Big Data

One of the main functions of mass media is to emphasize important issues in society, and the other is to make the problems into agenda by influencing the government through public opinion (McCombs & Shaw, 1972). In previous studies, media reporting attributes changed the perceptions of policy participants, so called mediating effects, which ultimately led to a higher impact on policy formulation than in policy outcomes (Jeong, 2009). Therefore, this study focuses on the first stage, agenda-setting, because the direction of the policy is determined in this process.

Policy agenda-setting as the first step of the policy cycle is generally done through social issues, public agendas, and governmental proposals (Cobb & Elder, 1971). Agenda-setting theory was formally developed by Max McCombs and Donald Shaw in a study on the 1968 American presidential election. In the 1968 ‘Chapel Hill study’, Combs and Shaw demonstrated a strong relation between the public's perceptions of the important election issue and the salience of issues in news content. McCombs and Shaw were able to determine the degree to which the media determines public opinion (McCombs & Shaw, 1972). According to their research, the evidence that mass media deeply change public attitudes in campaigns is not decisive, but there is much stronger evidence that voters learn from the immense quantity of information available during each campaign. Cobb and Elder also discussed that the role of mass media in defining the features of the agenda. These issues can be picked up by the media and therefore can be given enough attention to get on the public agenda. Once the issue has a presence on the public agenda, it becomes easier to maintain public interest and support, to create and influence policy decisions. The policy agenda is defined as a list of topics or problems that government officials or stakeholders pay attention to at a given point in time. This policy agenda-setting takes place through different processes depending on the characteristics of the political, economic, and social environment and the social problems themselves. It is more likely that the role of the general citizen will be underrepresented in the setting of these policies. Therefore, continuous role assignment and participation of citizens is an important issue of the modern policy

process. Agenda-setting theory represents ‘This ability of the news media to influence the salience of topics on the public’ (McCombs, 2002). That is, if a news item is covered frequently and prominently, the audience will regard the issue as more important.

In a world of developing digital and online medias, the dynamics of issue agendas are becoming more complicated. Big Data offers a critical test of extant and evolving theories of public opinion and public attention. Agenda-setting theory meets Big Data. Big Data is not a new concept; on the other hand, words that have been used only in special fields, such as astronomy, aerospace, and human genome information have begun to attract attention due to the development of IT (Gantz & Reinsel, 2011). Big Data is defined as a term for data sets that are large or complex that traditional data processing application software is inadequate to deal with them (Wikipedia). Big Data is used in a variety of ways, but there is no clearly agreed-upon definition. However, in consideration of diverse predefined meanings, Big Data is a vast amount of data that is difficult to handle with conventional storage methods as the IT environment changes and expands to include related technologies, platforms, and analytical techniques.

In 2001, META Group (now Gartner) defined data growth challenges and opportunities as being three-dimensional, such as increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources). Gartner, and now much of the industry, continue to use this ‘3Vs’ model for describing Big Data. As emphasized in recent applications, complexity, variability, and value have been added.

Major trends and issues related to Big Data include the following. First, when various communication functions and various sensors are installed in all functions of Internet environment and wearable sensors, the type and amount of data significantly increase. Next, Big Data related investments are increasing globally. In addition, companies try to provide customized services by analyzing customer information and services using Big Data. Insight Venture Partners, an IT investment firm, identified five Big Data issues to highlight in 2015: terrorism, open government data, customized medical care, digital education, and human rights.

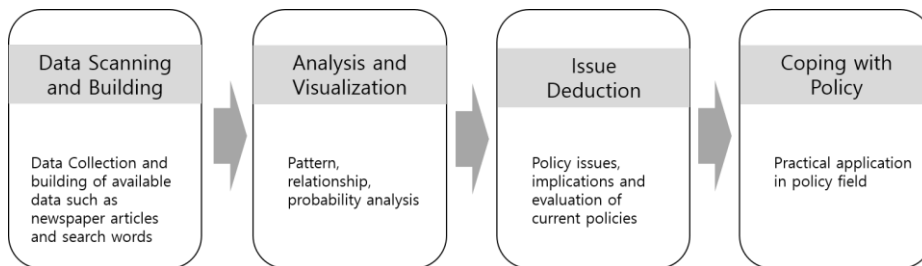
The government of South Korea recognizes the importance of data. Ministry of Science, ICT and Future (MSIP), The Ministry of the Interior (MOI), and other ministries are promoting various projects not only to create new business but also to utilize Big Data in various fields. In particular, since 2013, the government of the Republic of Korea has pursued ‘the Government 3.0’ policy to integrate public data and open data to public. Government 3.0 is a new paradigm for government operations to deliver customized public services and generate new jobs in a creative

manner by opening and sharing government-owned data to the public and encouraging communication and collaboration between government departments. Under this policy, the Government 3.0 service, which provides personalized services to the public by utilizing Big Data, reaches almost 1,800 cases proposed by each department and major public institutions (Yun, 2013).

As discussed above, the policy process consists of policy decision making, policy enforcement, policy evaluation, and policy feedback, including policy agenda-setting and policy alternative development. Organizations have generally used brainstorming, Delphi, and expert panels to develop their agendas. The policy agenda-setting through these methods has the potential to alienate the ‘lower voice’. In particular, it is more likely that the role of the general citizen will be underrepresented in the setting of these policies in order to establish agendas and develop alternatives. In addition, evidence-based policy-making is becoming increasingly important rather than experience or intuition. Evidence-based policy-making is attempting to improve policy reliability, minimize failure, and improve service quality by using objective and scientific evidence called data.

Big Data analysis at the policy formation stage is used for two purposes: to establish an accurate policy for data-driven future forecasting and to establish a practical policy that reflects public opinion. In this study, I will focus on the latter case, in which policy is formed by effectively reflecting public opinion through Big Data analysis.

Figure 2. Data-based Scientific Policy Formulation Methodology



2.2. Precedent Study Review

2.2.1. Precedent Study on Policy Formulation Using Big Data

Lim & Park (2015) experimentally analyzed Big Data for futures strategy in a local government, the case of Chungcheongnam-do province in the Republic of Korea. The analysis consists of the structural analysis of the evaluating for policy, analyzing public opinion using the newspaper which reported on regional issues and

SNS data as public opinion, and the visualization of the result. In addition, the government analyzed the characteristics, the relationship, and structure in SNS Big Data of each representative policy keyword of Chungcheongnam-do, such as ‘social economy’ policy and ‘3 innovations for agriculture’. From the results of the analysis, they found that different issues were highlighted according to the month or quarter of the period, and, geographically, Cheonan-city plays an important policy hub in Chungcheongnam-do. It also suggests implication that it should be made active in other cities besides Cheonan-city through promotion of those policies. However, a lack of opinion analysis, such as positive and negative analysis of policy, shows the limits of policy evaluation.

Kim & Baek (2016) analyzed issues on the college and university structural reform evaluation using text Big Data analytics. The topic of modeling, especially algorithm for Latent Dirichlet Allocation (LDA), was applied to extract and analyze the issues among 25 press releases of the Ministry of Education (MOE) and 625 articles in 10 major daily newspapers. The researchers examined the issues of university structural reform evaluation in previous research based on Big Data analysis. They examined the main concerns and issues in the college and university structural reform evaluation, how the interests and issues change over time, and whether other issues exist. In this study, however, there was a lack of data, such as lack of MOE documents and no Internet news, and further research would be needed to find the relationship between policies and deriving topics.

Kwon et al. (2014) analyzed bibliographic data of the papers in the field of education, media articles, and reports on future vision of education. This study has presented a new methodological approach to the development of agenda. Unlike the previous methods of brainstorming, Delphi and expert panels, this study showed that an agenda can be developed through analysis of research literature and media Big Data. From the perspective of information analysis in the past, the researchers analyzed the literature data of educational research and derived the agenda, including tutoring, academic achievement, and higher education finance. From the current information perspective, the researchers extracted areas such as curriculum, lifelong education, and school violence by analyzing media Big Data. Based on the future report, the agenda of creativity, convergence education, and IT were derived. It was shown that findings from analyzing media Big Data could be used in establishing education policy in the future using the derived agendas and related details. However, due to the broad scope of research, there was a limit to the extent to which comparative analysis could not be made between the agenda-derived areas and actual policies. In addition, from a qualitative point of view, there was insufficient reflection of expert opinion on education policy through panel, brainstorming or the

Delphi method in the course of research.

Many examples also exist of social media analysis as part of a marketing strategy. He, Zha et al. (2013) examined the social media sites of the three largest pizza chains and applied text mining to analyze unstructured text of the Facebook and Twitter sites. The results revealed that the three largest pizza chains are actively involves with social media and have committed substantial resources for their social media efforts. The data that they examined showed that they were committed to providing a delightful experience for their customers. As a result of the analysis, the social data showed positive and negative reactions instantly according to the service quality. In addition, they found that there were many more Facebook fans than Twitter followers, so the three pizza chains also offered more promotional and user engagement activities on Facebook than on Twitter. Social media competitive analysis allows a business to gain a possible business advantage by analyzing publicly available social media data. Many cases exist in which companies formulate social analysis as a policy as opposed to the public and education sectors. Future research will focus on finding innovative ways to turn businesses' social media fans from 'like' to 'buy'.

In the case of the Republic of Korea, Anti-Corruption and Civil Right Commission (2013) conducted a study analyzing civil complaints through text mining. The results showed the status of civil petitions by year, analysis of total civil petition trends, analysis of trends and characteristics of petitions according to ages, gender, region, civil petitions, and petitions. The purpose of this study was to extract features-specific keywords in cumulative data.

Kwon (2013) analyzed the characteristics of education Big Data and suggested ways to use it in education. In addition, research suggested how to utilize educational Big Data as individualized learning, content development, improvement of curriculum and education policy, and student guidance. In addition, global universities, such as Arizona State University, University of Florida, University of Maryland Baltimore County, and Towson University introduced cases of helping university students to select courses by using educational data. The main results of precedent studies are shown in table 2.

Table 2. Key Findings of Precedent Studies

<i>Author</i>	<i>Significance of Research</i>	<i>Limitation</i>
<i>Lim & Park (2015)</i>	A Tentative Approach for Regional Futures Strategy with Big Data - Through the Analysis using the Data of SNS and Newspaper	Need reaction analysis such as positive and negative analysis of policy

	<ul style="list-style-type: none"> · Analysis of news and SNS data · Evaluating for key policies · Time series and geographical analysis 	
<i>Kim & Baek (2016)</i>	<p>Analysis of Issues on the College and University Structural Reform Evaluation Using Text Big Data Analytics</p> <ul style="list-style-type: none"> · Analysis of MOE's press releases and articles in major newspapers · Focus on issues from previous research: (i) what are the main concerns issues in the college and university structural reform evaluation, (ii) how the interests and issues change over time, (iii) if there are other issues. 	<p>Lack of data</p> <p>No relationship between issues and policies</p>
<i>Kwon et al. (2014)</i>	<p>Developing agenda using education data: Focusing on Social Network Analysis</p> <ul style="list-style-type: none"> · Analysis of research literatures, media Big Data (news, blogs, and web documents), and internal future reports · Developing education agendas · Analysis of past, present and future demand perspective · Deriving agendas and related details 	<ul style="list-style-type: none"> · No relationship agendas and policies · Insufficient reflection of expert opinion (Need of panel, brainstorming or Delphi method)

Through the implication of the precedent studies, it was found that the most significant obstacle to utilization of social media analysis as a policy agenda is to establish the relationship between public opinion and policy accurately. In order to solve this problem, social media analysis should be conducted according to actual policies, and if many negative opinions exist, the collection of policy alternatives through additional analysis may be effective. In addition to quantitative or statistical methods called data analysis, it remains a challenging task to determine how the attributes of social media content affect policy formulation.

2.2.2. Social Media Analysis Method in Precedent Studies

Social media which was first used as a terms by Tina Sharkey in 1997 and is the extension of the concept of social relations into the Internet space, which means

interactive online tools and platforms that people use to share their thoughts and experiences. Typical examples of social media are blogs, social networks, instant messaging, and UCC, which share a variety of data formats, including text, images, video, and audio. Social media analysis is an emerging technology that attempts to extract meaningful information from unstructured textual data.

The unstructured data analysis model consists of three steps: collection of unstructured data, preprocessing of unstructured data, and time-series analysis of unstructured data. In the collecting phase of the unstructured data, APIs and free software are used to continuously collect and manage citizen discussions from a variety of social media, such as Twitter, Facebook, blogs, cafes, and online newspapers. In the preprocessing step of the unstructured data, keywords are extracted after the commercial data and duplicate messages are removed and refined from the public opinion data collected. In the time-series analysis of unstructured data, researcher can explore the meaning of public opinion and the patterns of change in forming subjects by analyzing time-series data, text mining, and social network analysis. Text mining is an extension of data mining to textual data (Romero & Ventura, 2010).

2.3. Education Environment and Big Data Analysis

2.3.1. Policies on the Higher Education of the Republic Korea

Though the South Korea government has implemented a series of measures to address the problems of dissatisfaction with the public education over the last decade, the problems still remain. The Moon Jae-in government, which was newly installed in May 2017, is speeding up ‘Educational Reform’ as the important national task. Educational Reform will be carried out by introducing a credit system at high schools, streamlining college admission procedures and enhancing fairness in those procedures (100 Policy Tasks Five-year Plan of the Moon Jae-in Administration 2017). Both the domestic and international environment surrounding universities is rapidly changing. Therefore, these tertiary education policies are most influenced by external factors. In Korea, the low birth rate has led to sharp drop in the number of school-age children and a decline in university enrollment rates. Also, the Fourth Industrial Revolution and the advent of an Intelligent Information society have increased uncertainty about the future society that could become a next generation growth engine or lose growth momentum (Shin, 2018).

South Korea’s birth rate was the lowest among 35 OECD countries whose

average was 1.68 in 2016¹. The Korea's total fertility rate reached a record low of 1.05 in 2017. The number of Korea's fertility rate was lower than 1.9 of France, 1.8 of the United States and the United Kingdom, and lower than 1.4 in Japan in 2016. Moreover, the South Korea's fertility rate is declining at the fastest rate among OECD, considering 2.18 in 1980 and 1.187 in 2013. The prolonged low fertility rate will demand various educational measures, such as relocation of school facilities and restructuring of universities due to the decline in school-age children. In addition, the low birth rate has a powerful ripple effect that can change the national social operating system through the transfer of the industrial manpower and the restructuring of the industrial structure due to the decrease of the economically active population. In particular, according to 'Basic Statistics of Education in 2018' issued by the Ministry of Education and the Korean Education Development Institute, college admission quota numbers are expected to reverse the number of high school graduates from 2021 due to low fertility, thus making it inevitable for university authorities to restructure their universities, such as the regulation on enrollment size and the consolidation of universities. MOE has assessed all universities to achieve the goal of reducing the freshmen student quotas, divides the evaluation results from 'A' to 'E' grades, and forced reductions and restructuring for 'D' and 'E' universities. Although the government basically had a social consensus on the necessity of restructuring the university, it confronted social resistance in the process of implementing restructuring measures. The way the government copes with the universal phenomenon of low fertility has been criticized because it focused more on the quantitative centering on the university restructuring rather than on the qualitative improvement of education (Shin, 2018).

The second factor is the arrival of the intelligence information society. The intelligence information society, commonly referred to as AI, IoT, Cloud, Big Data, and Wireless Network, has triggered structural changes in the industry that overwhelm existing production factors, such as labor and capital. The society led by intelligent information technology is called intelligent information society. In order to prepare for the emergence of the intelligence information society, which is referred to as the fourth industrial revolution, MOE has established 'The Mid to Long-term Education Policy Direction and Strategy in response to the Intelligence Information Society'. This strategy can be applied to 2030 by 1) education to maximize students' interest and aptitude, 2) education to develop thinking, problem solving ability and creativity, 3) customized education considering individual learning ability, and 4) education to contribute to social integration. Policies to cope with these goals include flexibility of university curriculum, development of virtual realistic contents,

¹ Data from <http://data.oecd.org> and <http://kostat.go.kr>

expansion of S/W education, and expansion of K-MOOC. However, it is being criticized that the government's reform measures are too fragmented compared to the targets, which will not be effective.

By pushing for university structural reforms, the government has been encouraging universities to adjust their functions in preparation for the upcoming social changes. The following table 3 shows university structural reforms of the Moon Jae-In government launched in 2017.

Table 3. Moon Jae-In Government's National Tasks in the Higher Education

<i>National Task</i>	<i>Goal</i>	<i>Detail</i>
49) <i>To strengthen educational services from early childhood to higher education</i>	- Provide fair opportunity regardless of economic conditions	- Reduce tuition and housing burden
50) <i>To innovate public education by revolutionizing classroom education</i>	- Enhancing fairness in college admission procedures	- Streamlining college admission procedures - Mid-term plan of college admission procedures
51) <i>To restore the ladder of hope in education</i>	- Strengthening education support for socially vulnerable class - Abolition of discrimination practices	- College admission Support for the socially vulnerable class - Career education for disabled students - Blind interview application in university admission
52) <i>To improve the quality of higher education and innovate lifelong learning and adolescents</i>	- Strengthen universities' publicity and competitiveness - Strengthen national vocational education responsibility	- Strengthening public service for higher education - Expansion of university autonomy - Improving the quality of college - Innovation of lifelong learning - Industry-academy cooperation

54) <i>To create a better educational environment to prepare for the future and make schools safe</i>	<ul style="list-style-type: none"> - Fostering creative and plural student for the Fourth Industrial Revolution 	<ul style="list-style-type: none"> - S/W education for digital literacy - Fostering leading teachers in digital education
76) <i>To strengthen fiscal decentralization for financial autonomy</i>	<ul style="list-style-type: none"> - Strengthen democratic accountability of education - Reorganization of Education Governance 	<ul style="list-style-type: none"> - Recovery of Education Democracy - National Education Committee - Reorganization of MOE - Communication with educational field

Source: Data from 'Education in Korea 2017' report of <http://www.moe.go.kr/>

2.3.2. Policies of the Seoul National University

SNU was incorporated in 2011, which means that SNU changed its legal status from a government organization to an incorporated university law. The incorporation of a public organization implies a set of measures to enhance the autonomy of the organization so that public organizations can exercise their own authority to make important decisions about the management of their organization. In particular, incorporation has been regarded as one of the core programs of new public management, an administrative reform movement that has been conducted in Europe and the United States for the past three decades (Rhodes, 1996; Farnham & Horton, 1996). However, the transition of SNU does not mean the complete separation of the university and the government. According to Article 32 of the Act on the Establishment and Operation of the SNU, SNU maintains its identity and accountability as a national university, and the government is supposed to provide financial support to SNU annually. SNU can establish policies with limited autonomy under government control.

The main characteristics of SNU's decision-making structure are horizontal rather than hierarchy. The board of directors, which is the highest decision-making body of SNU, consists of more than half of outside members. At the bottom of the board of directors, there is a multi-tiered organizational structure leading to headquarters, college, department, faculty, and others. Unlike the decisions made by the hierarchy and authority of the organizational structure in a typical large-scale organization in modern society, university decision-making is relatively horizontal, ranging from the board of directors and the president to individual professors. It is not a working hierarchical relationship. In other words, in a university organization with non-hierarchical characteristics, it tends to make it difficult to establish a policy

in one direction, while delegating management authority to the lower hierarchy to improve organizational management efficiency and effectiveness.

In accordance with the Act on the Establishment and Operation of SNU, SNU shall set the university performance goals for every four years and establish the annual university work plan reflecting the performance of the university. This legal regulation on the performance evaluation of national universities has an important meaning as an institutional device for examining accountability based on the autonomy of university administration. Table 4 below shows the main policies of SNU in 2017. The main policies of SNU are established through internal decision-making process and the contents and goals are determined through consultation with the Ministry of Education.

Table 4. Keywords in ‘2017 Work Plan - SNU’

<i>Domain</i>	<i>Key Word</i>
1. <i>Fostering Human Resources</i>	Leadership, Extracurricular activities, Liberal arts education, Academic Writing, Unification studies, Data-Science, Fostering Global Talents, Global education, Foreign language lecturing, Supporting master and doctor
2. <i>Supporting Students</i>	Customized scholarship program, Personal student support system, Healthcare, Students human rights
3. <i>Advanced Research</i>	Pioneer Project, World leading research, Convergence studies, Unification Studies
4. <i>Research Support System</i>	Research year policy, Advanced research infrastructure, Research support activities, Research ethics
5. <i>Career</i>	Industry-Academic Cooperation, Start-Ups Support
6. <i>Strengthen Public Education</i>	Admission selection, Accessibility of admission information, Education and undeserved area, Reducing gaps in basic academic ability
7. <i>Social Contribution</i>	Global volunteering, Future national policy agenda and research direction,
8. <i>Knowledge Sharing</i>	MOOK, Fellowship with a developing country
9. <i>Management innovation</i>	Internal regulations, Performance compensation system, Administrative capability enforcement, Statistical information system
10. <i>Financial Sustainability</i>	College development fund, Asset management, Integrated financial system, integrated procurement, Standardizing finance and accounting process, Open financial information
11. <i>Multi-Campus</i>	Campus development plan, Multi-campus, Smart-campus

Due to the environment and inherent processes, various phenomena of SNU's policies have characteristics as political processes. The policies of higher education institutions are formed through a political process in which various groups compete to gain greater influence in different interests, and the policy implementation cannot be completely free from such a political process (Anderson, 2006). Recently, as the communication of citizens and the formation of interest groups become easier, public opinion formation and pressure exercises are becoming more active, so policy participation is getting more active in order to attract favorable policies to oneself or their group. As a result, the debate and conflict over university policy decisions among these interest groups, such as internal colleges, student councils, labor unions and external civil groups, are intensifying (Chun, 2014).

Traditionally, students have been struggling with schools to engage more in university policy. The SNU student groups opposed the incorporation of the school in 2001, fearing privatization of the university would violate academic freedom and raise tuition fees. In 2013, students requested SNU to abolish the practice of paying the support fees that were included in the tuition fees and refund the support fees to students. Students who oppose the construction of Siheung campus have conflicted with the university authorities by occupying an administrative office building for 153 days in 2017. In addition, the Student Council has been actively engaged in expanding student participation in the presidential election system in 2018. Recently, students familiar with digital literacy have been actively advocating opinions on university policies through online communities. In response to these students' changing policy approaches, SNU's university authorities have also made institutional changes to make broader participatory decisions. SNU increased student participation in key decision-making bodies and expanded online policy commentary, such as online surveys and official SNS operations. According to the Regulations of the SNU President Recommendation Committee, which was revised in December 2017, 9.5% of students was guaranteed to participate in the Presidential Candidate Evaluation Team from the 2018 presidential election. In addition, the university conducts online surveys every year to gather student policy opinions, and investigates policy feedback by surveying students' satisfaction with each policy. Participatory decision-making that allows various actors of civil society to participate in the process of debate on public policy decisions in a variety of ways and to influence the public policy of the government is no longer a choice but a necessary step in policy agenda-setting (Irvin & Stansbury, 2004).

2.3.3. Implications for Policies in Social Media

The first step in the policy process, policy agenda-setting, is generally done through a series of processes, including social problems, social issues, public agendas, and government proposals (Cobb & Elder, 1971). The policy agenda can be defined as 'a list of subjects within the governmental agenda that makes active decision'. The 'governmental agenda' is a broader category that includes 'the list of subjects that are attracting attention (Cobb & Elder, 1971; Kingdon, 1984)'. In other words, it can be understood as a set of problems to solve social problems. However, not all social problems become policy agendas, only some of the problems that are socially noted and recognized by the public as an important agenda can enter into the policy process.

It is a common process for an organization to listen to the public opinion and to reflect it in policy formulation in the policy process. Government or public institutions allow citizen to participate through various methods, such as surveys, policy forums, policy councils, seminars, roundtables, and public debates, when setting agendas and determining specific policies. However, these direct citizen participation-based approach has the drawback that people's participation is rather showing format, the number of participants is limited, and the opinions of the interested parties are likely to be reflected in the selection of participants. In addition, since the preparation process is not simple and the costs are considerable, so the number of direct government-citizen meetings is rather small. In other words, traditional government-citizen communication methods are only appropriate to form the basis for a common understanding of key policy agendas, but they are less useful for locating policies that fit people and the local needs by timely understanding the changing public opinion trends. The agenda can be formed by the attributes of the issue, the issue maker, and the trigger device that activates the issue, and the role of the mass media in this process is important (McCombs, 2002; Lim, 2006; Lee et al., 2012).

According to social survey methods, a large sample size is more representative of the population. It extracts hundreds or thousands of samples from the population and estimates the characteristics of the population. However, unlike traditional social survey methods, Big Data Analysis can be based on millions of data and can be similar to the complete survey concept of "N = All", especially free from sampling errors. In addition, most of Big Data consists of digital traces automatically, in real time, that are left in the process of sending and receiving e-mails, ordering products, and sharing files. Therefore, Big Data differs from the existing social science data in that it is a material that reflects what has been made "naturally" rather than being

created by artificial intervention. However, Big Data and complete surveys are conceptually distinct and do not guarantee complete or systematic representation of the study subjects (Han, 2015). For example, what we find in research using Twitter resources is not about the entire population, but only about a specific category of 'Twitter users,' no matter how large the numbers are. Therefore, the researcher must pay attention to the question of who the data represents.

Since tertiary education policy is basically formed in interaction with interest groups in the present or future environment, there is a continuing flow of mutual checks or balances among the rationality, the equity, and the politics seeking satisfaction of collective interests throughout the decision-making process. In the process of selecting and spreading policies by the government in recent years, the emergence of new social media has a great influence on changing the process of spread and diffusion. The public can freely post the issue they think is important on the internet bulletin board, and some agendas are immediately issued to the netizen vote. Because of the mass and rapidity of the information delivery of the Internet and the social media network, issues that are constantly exposed in mass media are likely to become government agendas.

Chapter 3. Research Design and Method

3.1. Research Question

This study aims to explore the possibility of using Big Data analysis method of social media and internet news related to Seoul National University to understand characteristics of public opinion in SNU and set university policy agendas. To obtain valuable insights from a large amount of data, text mining has been used in many fields to identify future issues and establish long-term policies (Russell et al., 2014; Kim et al., 2014; Kwon et al., 2014; Lee & Kim, 2016). Therefore, it is meaningful to carry out Big Data analysis centered on newspaper articles and social media to determine university policy issues.

This study used text mining and semantic network analysis (SNA) tool to examine main policy issues of SNU that were discussed in Internet news and social media over the period of 2015-2017. Text mining refines words through stemming and natural language processing of common text data, extracts useful words from the text, and derives the frequency. Since text mining is not a simple keyword level, but a useful word level in the context of a sentence, it has features that enable meaningful interpretation in connection with context in unstructured textual data with complex and irregular forms (He et al., 2013). In addition, the importance of the word that conveys the meaning of the sentence can be grasped by the frequency of words used simultaneously in various documents (Kim & Baek, 2016). SNA makes it possible to grasp the relationship of each word from a structural point of view based on meaning of useful words in sentences. A significant relationship can be found between main words used in the sentence. Additionally, the use of the SNA tool provides the advantage of graphing relationships of key words for visualization of networks as well as for visual exploration and analysis.

Specifically, this study attempts to explore the direction of policy agenda-setting from a macro perspective by analyzing newspaper articles and social media about SNU using Big Data processing methods. This study attempts to answer the following research questions (RQ).

1.RQ1: What are the main issues related to SNU in news reports and social media (blogs and, Facebook)?

2.RQ2: What is the network centrality between core links related to SNU in news reports and social media?

This study analyzes major trends of policy issues related to SNU related in

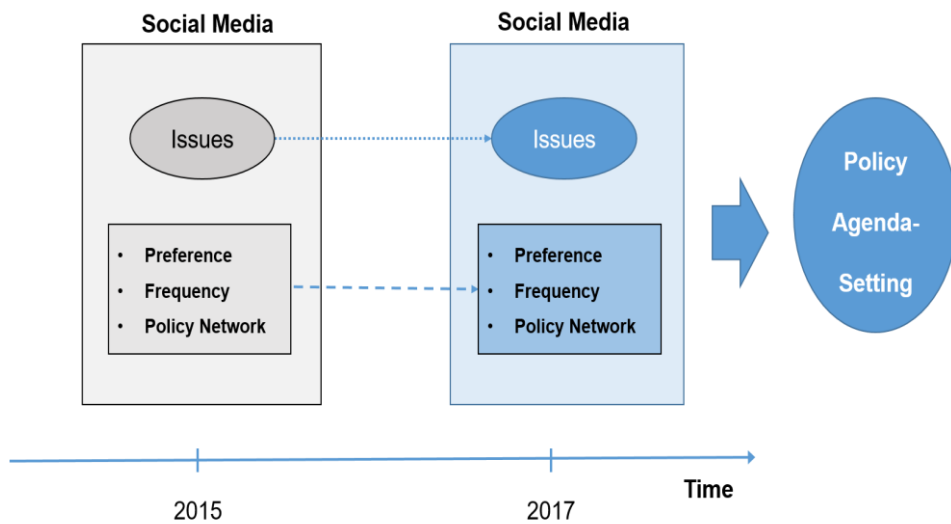
newspaper articles and social media and suggests future directions for university policy development through examining these research questions.

3.2. Method

3.2.1. Research Model

The process of policy formation suggests various stages of classification and modeling as discussed in the theoretical discussion above. On the other hand, it may be said that there is no optimal model that can explain it at a glance. This also means that diversified modern-day problems and policy prescriptions are complex. Therefore, this study aims to empirically investigate the influence of media on the efficacy of media in the context of the theory of Cobb & Elder (1971) and precedent studies (Lim & Park, 2015; Kim & Baek (2016).

Figure 3. Research Model



3.2.2 Context of this Study

Policy agenda-setting using public opinion also pays attention to higher education institutions. Today, the competitiveness of university education is an important factor that determines national competitiveness. Many countries around the world are making efforts to improve the competitiveness of their universities. Among these various efforts, it is a very important task for universities to offer education, research, and community services under the banner of autonomy and responsibility. In September 2011, SNU was transformed into a public institution

with rights and obligations for the operation of the university along with the announcement of ‘the Act on Establishment and Operation of SNU’. This can be seen as an opportunity for SNU to become a global leading university with a competitive edge. It became possible to establish an autonomous policy based on incorporation. Based on this autonomy, SNU should appropriately reflect needs of the public as a public institution. However, SNU has faced challenges of solving internal conflicts with students, professors, and employees and external conflicts with the government and the public in terms of admission policy, construction of multi campus, and restructuring method. Systematic analysis of public opinion could be an effective alternative to address these concerns. Thus, SNU was selected as a case study model in this study. By using these social media applications, not only public but also internal members can engage in activities such as suggesting ideas, customizing service, discussing public education and trading information with each other, giving praise and complaints, and providing feedback to the administrative body. Many colleges and universities employ public relations teams to monitor public opinion and run Facebook or Twitter to communicate with the public.

3.3. Research Procedure

3.3.1. Data Collection

In this study, Big Data analysis solutions 'BIGKINDS' and 'TEXTOM' systems were used. 'BIGKINDS' was used for data collection and analysis of newspaper articles, and 'TEXTOM' was used for data collection and analysis of SNS and blog data.

For the analysis of newspaper articles, data were collected using newspaper article search service (www.bigkinds.or.kr) provided by Korea Press Foundation. BIGKINDS, a search engine service, can search articles provided by various media such as nationwide daily newspapers, local daily newspapers, economic and business newspapers, and Internet newspapers published in Korea. In this study, eight major newspapers, five economic newspapers, and 25 local newspapers were collected. Major newspapers were Segye Ilbo, Kyunghyang Newspaper, Kookmin Ilbo, Hankyoreh, Seoul Shinmun, Hankook Ilbo, Munhwa Ilbo, and Tomorrow Newspaper. Economic newspapers were Daily Economy, Seoul Economy, Financial News, Korea Economy, and Herald Financial. Twenty-five local newspapers were collected, including Gyeonggi Ilbo, Gwangju Daily News and Chungbuk Ilbo. The article collection period was three years from 2015 to 2017.

Table 5. Major Press Reports on SNU

<i>Rank</i>	<i>Newspaper</i>	<i>Number of news articles</i>
<i>1</i>	Segye Ilbo	5,731
<i>2</i>	Kyunghyang Newspaper	5,314
<i>3</i>	Kookmin Ilbo	5,002
<i>4</i>	Hankyoreh Newspaper	4,842
<i>5</i>	Seoul Newspaper	4,611
<i>6</i>	Hankook Ilbo	3,992
<i>7</i>	Munhwa Ilbo	2,743
<i>8</i>	Naeil Newspaper	2,489
<i>9</i>	Others(30)	71,317
<i>Sum</i>		106,041

For the analysis of social media, data were collected using ‘TEXTOM’, a social media collection analysis site. Textom is a Big Data analysis solution using text mining technology. It collects and refines data in the web environment. It can consistently process matrix data. Textom supports data collection, refinement, and analysis. It also supports a variety of collection channels including Naver, Daum, Google, Baidu, Twitter, Facebook, and certain other sites. In this study, blog data from Naver and Daum, the largest portal sites in South Korea, and data from Facebook, an open social network service, were collected. A total of 1,814 texts were collected from blogs and 212 texts were collected from Facebook from 2015 to 2017.

All materials that helped describe and analyze the determined research problems were examined based on news and SNS analysis sites. Specifically, with keywords of 'Seoul National University' for gathering analysis data, 106,041 news reports and 2,026 social media data (Facebook and blogs) were obtained. The collection period of the data to be analyzed was from January 1, 2015 to December 31, 2017.

Table 6. Summary of Collected Data

	<i>News Data</i>	<i>Social Media Data</i>
<i>Data Source</i>	News list of 38 News Press	1) SNS (Facebook) 2) Blog (Naver and Daum)
<i>Term</i>	2015. 1. 1. ~ 2017. 12. 31.	2015. 1. 1. ~ 2017. 12. 31.
<i>Number</i>	106,041	2,026

3.3.2. Text Mining

Collected data were subjected to refinement procedures such as morphological analysis and data cleansing. These collected data were refined in order to remove particles, verb endings, and punctuation. These data were then refined again by the researcher after running the preliminary refinement process through Textom, a Korean language analysis program site. Then ‘word by word matrix’ data were extracted using Textom. The word by word matrix data is used to grasp the frequency between words co-occurring in a similar context. This allows a researcher to determine the relationship among words in the context. The matrix data between extracted key words were visualized using UCINET. Netdraw was used for network visualization. UCINET and Netdraw are software that can visualize the interwoven network of co-occurring words. They provide information on the connection structure of the quantified data. In addition, they show the relationship to centrality which describes the link strength among words and represent the role of a particular word in the overall network. Through this, it is possible to visually understand the semantic tendency of the SNU in the media and conduct a structural analysis on word roles. Clustering analysis was then performed through the Faction method to extract words with high interconnectivity. Faction cluster analysis is a method of grouping words with high interconnectivity and dividing one network into several subgroups. Cluster criteria are highly correlated. This means that a high degree of correlation exists between words in the subgroup, but no high correlation exists between the words outside the subgroup.

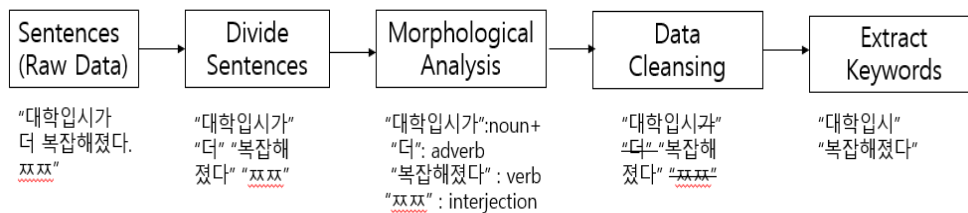
Data pre-processing was performed for the collected data from news and social media by the keyword of SNU, as shown in Figure 4. This was done to extract keywords from the refined data. Frequent occurrence of words in the analysis process means that these words are very important. Assuming that this premise is established, the order of importance of words can be determined by extracting key words from the text and performing frequency analysis. A group including surrounding terms around the search keyword is then determined. The term in the group is extracted. TF² (term frequency) and DF³ (document frequency, document number) are obtained. Frequency analysis includes frequency analysis (FA) which simply calculates how long a particular word appears, and term frequency-inverse document frequency (TF-IDF) which means the importance or weight of the word. The latter

² TF is the frequency of words that indicate how often a particular word appears in a range of documents.

³ DF is an index in which a particular word is used within a range of documents.

is obtained literally by multiplying TF by inverse document frequency⁴ (IDF). The word frequency is the number of times a specific word appears in a document. If the document is longer, the word frequency can also be larger regardless of its importance. Reverse document frequency refers to the value obtained by dividing the number of documents containing a specific word by the total number of documents followed by log transformation. TF-IDF increases when the word frequency value is larger and the value of the document containing the word is smaller. This allows one to filter out words that are common across all of them or remove words that are low in TF-IDF (Neto et al., 2000). In other words, if the word ‘CSAT’ does not appear frequently in common documents and its IDF value is high, it can be the key word in the document. However, in the collection of documents about CSAT, the word CSAT becomes a common cliché. Other words that are rarely used in each document would be assigned higher weights.

Figure 4. Data Pre-processing



Social media data are collected in a huge amount in an unstructured form. Thus a researcher need to decide on the analysis period and method. These collected data were analyzed by buzz quantity analysis, influencer analysis, related word analysis, and sensitivity analysis. Buzz monitoring was done to see how much a particular keyword was mentioned in social media. If a reference to a particular keyword was temporarily amplified, it could be inferred that something different than usual happened. Second, social media collection and analysis tools such as ‘TEXTOM’ use natural language processing⁵ technique to analyze the favorability and the related keywords mentioned at the same time of the message. Sensitivity analysis was done to understand whether the tendency mentioned through natural language processing of messages was positive or negative. Third, association keyword analysis was done to find another keyword that was mentioned at the same time as a specific keyword. Fourth, influencer analysis refers to those who exert influence on

⁴ IDF is the inverse of DF. The larger the DF, the lower the IDF.

⁵ Natural language processing is not a special programming language for computers, but a technology that enables computers to process natural language used in everyday conversation and spoken language.

social media. Klout, a social impact measurement index, was used to determine scores of social media users. Influence evaluation criteria were Twitter followers, frequency of updates, number of friends on Facebook, number of comments, likes, and so on. Using Clout, one can see how much influence someone has on social media and what kind of influencers they are. These four steps were used to analyze search terms that came into the analysis site.

Table 7. Four Steps of Framework of Social Network Data Analysis (Hwang, 2013)

<i>Framework of social data analysis</i>	<i>Contents</i>
<i>Buzz Monitoring</i>	How many references are to be in social media on a specific keyword.
<i>Influencer</i>	It means a person to exercise the influence of social media. To grant score by analyzing the participation of users of social media cloud client that provides the social influence measurement index and (Klout).
<i>Keywords associated issues</i>	It is intended to find another keyword mentioned at the same time as a particular keyword
<i>Sensitivity analysis</i>	It is a social media, by analyzing the related keywords that are referred to at the same time as the favorability message is on using natural language processing techniques, collected social analysis tools, social it is differentiated and search engine

3.3.3. Semantic Network Analysis

Semantic network analysis (SNA) based on social network analysis is an analytical method that focuses on the connection relation of words constituting a node. SNA can be used to investigate a specific social phenomenon by analyzing the structure of the network formed by the formation of one network by interactions between individuals (Kim & Chang, 2010). The network consists of nodes and links. A node represents a keyword and a link represents a connection frequency between keywords. Depending on the frequency and importance of the text, words to be analyzed are extracted from the text. These extracted words are visualized in a spatial structure to reveal the overall meaning of the text, or to analyze strengths and characteristics of nodes and connections in order to find the emphasized meaning. Once the frequency of the keyword is known, it can be visualized and network analyzed among keywords using UCINET, NetMiner, R, and so on.

This study used UCINET6 to identify the connection structure of word. The degree of the relationship was quantified by analyzing the ‘Degree Centrality’ and

‘Network Centralization’. Degree centrality refers to the degree of connectivity with other nodes. It is the number of lines connected to a word. A word with many connections is a word with a high degree centrality and a strong word. Network centralization refers to the degree to which an entire network is concentrated on one node in a network. In other words, the network becomes a highly concentrated network when an intensive connection is made to a small number of words. By contrast, the network becomes a low-concentrated network when various center points without concentrating on specific words exist (Kim & Chang, 2010).

NetDraw was used to visualize and clarify relationships between words. CONCOR (convergence of iterated correlations) was conducted to derive clusters formed by words with similarities. CONCOR analysis is a method used to identify nodes based on results of Pearson correlation analysis of simultaneous occurrence matrix of words. CONCOR is one of the most widely used structural equivalence analysis methods. Structural equivalence is a method used to find similarity between two nodes in a hierarchical structure (Lorrain & White, 1971).

Chapter 4. Data Analysis and Findings

4.1. RQ1: Keyword Analysis

Table 8 and 9 below show the 20 most frequently used words in newspaper articles, blogs, and Facebook. ‘Ph.D. program’, ‘Siheung’, ‘Student Council’, and ‘Press Conference’ were frequently used in both newspaper articles and SNS. ‘President’, ‘Ph.D. Program’, ‘Siheung (place)’, ‘Textbook’, ‘Chairman of the Board’, ‘Ministry of Education’, ‘Student Council’, ‘Law school’, ‘Oxy (company)’, and ‘Justice Department’ were frequently used in newspaper articles while ‘General admission’, ‘Recruitment’, ‘Entrance Examination’, ‘Graduation’, and ‘Shuttlebus’ were found to be frequently used in blogs and Facebook.

Table 8. High Frequency Words in Newspapers

No	Keyword	Frequency	No	Keyword	Frequency
1	President	63,348	11	Research Report	6,973
2	Ph.D. Program	19,829	12	Supreme Court	6,632
3	Siheung (place)	22,807	13	Humidifier Sanitizer	6,062
4	Textbook	14,904	14	Scholarship	4,524
5	Chairman of the Board	12,447	15	Examinee	4,282
6	Ministry of Education	11,390	16	Press Conference	4,049
7	Student Council	10,782	17	Outside Director	3,850
8	Law School	9,026	18	Sexual Harassment	3,792
9	Oxy (company)	8,377	19	Occupation of Administrative Building	3,008
10	Justice Department	7,162	20	Business Agreement	2,960

Table 9. High Frequency Words in SNS

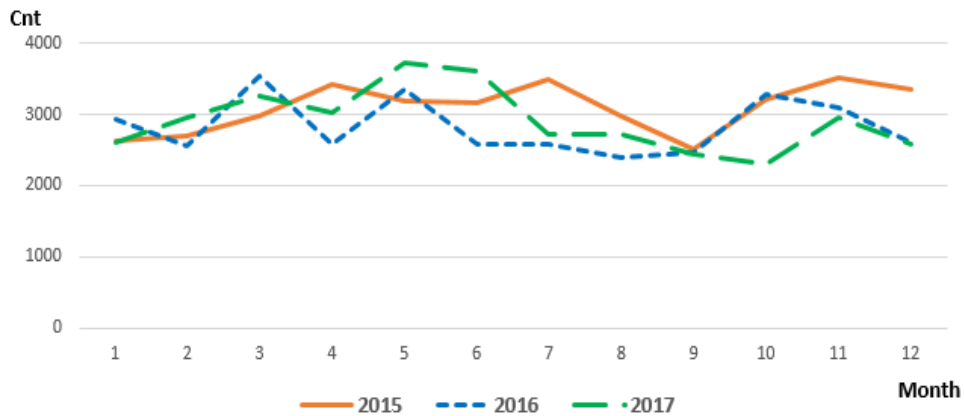
No	Blog		Facebook	
	Keyword	Frequency	Keyword	Frequency
1	Acceptance	141	Student Council	75
2	Preparation	101	Club	34
3	Laboratory	99	Festival	17
4	Occasional-Regular Admissions	96	Recruitment	15
5	In-Seoul	85	Entrance examination	14
6	Camp	62	Graduation album	7
7	SNU Institute for Global Social Responsibility	61	Course registration	7
8	General Admission	59	Shuttle bus	6
9	Building	52	The student body president	5
10	Location	41	Election	5
11	Recruitment	38	University Newspaper	5
12	Campus	38	Library	5
13	Admission	37	Big Data	4
14	Entrance examination	37	Machine learning	4
15	Announcement	37	Press conference	3
16	CSAT	33	Siheung campus	3
17	Ph.D. program	33	Graduation	3
18	Graduation	32	Negotiation	3
19	Waiting room	31	Recording	3
20	Shuttle bus	25	Attendance	3

Note: Words in bold are used simultaneously in blogs and Facebook

This study analyzed 106,041 news articles linked to Korea Press Foundation⁶ in order to identify which keywords were included. Total numbers of articles were 37,176 in 2015, 33,964 in 2016, 34,901 in 2017. The monthly average number was 2,990. There was no significant difference in the number of newspaper articles by year or by month.

⁶ Korea Press Foundation Supports digital news production and circulation. They support digitalization of news contents and the development of shared newspaper production systems.

Figure 5. Yearly / Monthly News Articles



Next, this study analyzed results of keywords using newspaper texts. This study analyzed the morpheme using a database of newspaper articles containing words ‘Seoul National University’. Keywords were then extracted from nouns. Basic association words of ‘university’, such as ‘student’, ‘professor’, and ‘education’ were excluded from keywords. Results revealed that events that became social issues showed overall high frequency among keywords. In particular, in 2015, moral hazard problems, such as professor’s scandals shocked the society. In 2016, a professor was arrested on charges of manipulating toxicology reports on Oxy products, and the press published many articles on the importance of authenticity of professor research. In addition, some university policies became social issues because of stakeholders’ objections. SNU Siheung Campus has promoted its business since June 2009 with a goal to establish a global start-up and industrial cluster. SNU held a board meeting in May 2016 and decided to conclude a business agreement for the Siheung Campus. The aim was to establish the Siheung International Campus in 2018. However, SNU has faced students’ objections that the Siheung Campus policy would lead to commercialization of the university. This conflict is not simply related to students. It is also related to the Siheung community as a whole. Therefore, according to the influence of diverse interest groups, a number of press articles, favorable or unfavorable, were reported. This issue in Siheung Campus is a representative example of the complexity of factors such as lack of communication between university authority and students, commercialization of universities, and interest of the local government and local people using academic background. Due to delays in policy decisions on new campus construction, there was continued media coverage from 2015 to 2017 which has had a negative impact on the organization's image. Total numbers of newspaper articles related to Siheung Campus were 462 and

463 in 2015 and 2016, respectively. It was increased to 671 in 2017 when university and students were in conflict. Keywords frequently used in Siheung Campus policy-related newspaper articles were analyzed in the order of ‘Siheung City’, ‘Enforcement Agreement’, ‘Baegot New City’, ‘Stay-in Strike’, ‘Council’, ‘President Sung Nak-in’, ‘Occupation of Main Building’, ‘University Headquarters’, ‘Halla Construction Company’, and ‘Press Conference’.

Table 10. Top 10 Keywords in Newspaper by Frequency

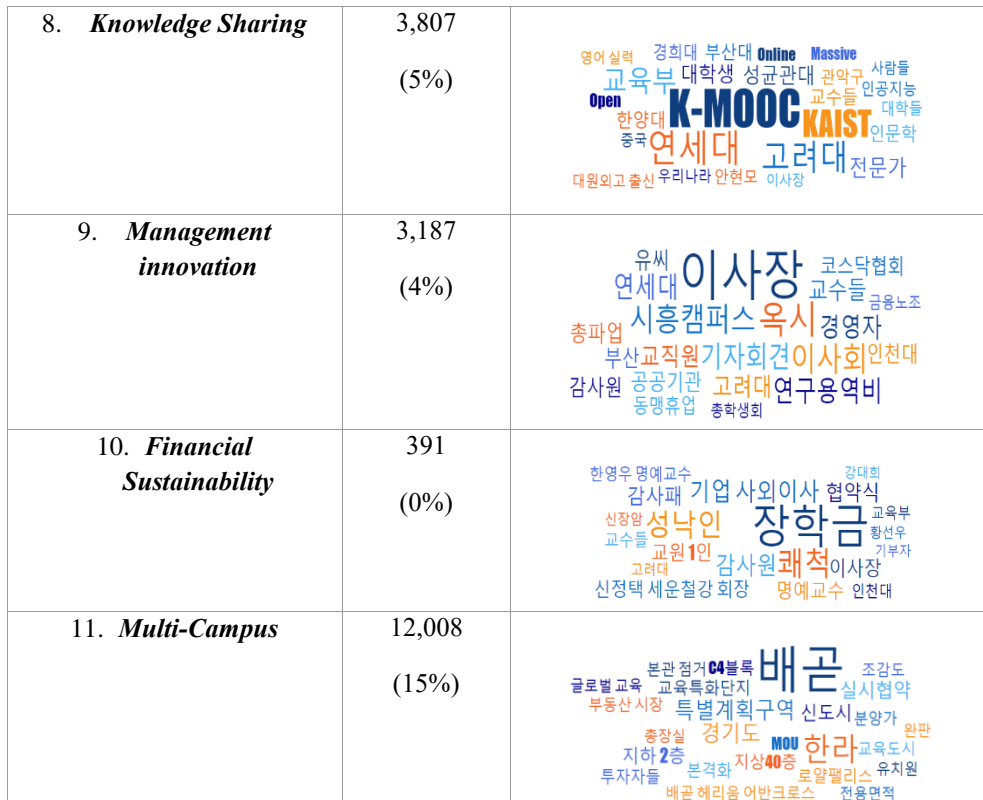
<i>Ran k</i>	<i>First half of 2015</i>	<i>Second half of 2015</i>	<i>First half of 2016</i>	<i>Second half of 2016</i>	<i>First half of 2017</i>	<i>Second half of 2017</i>
1	Sexual Harassment	Supreme Court	Oxy (company)	Siheung Campus (place)	Siheung (place)	Indefinite Suspension from School
2	Disciplinary committee	Kim InHye (person)	Research Report	Enforce-ment Agreement	Student Protest	Committee
3	Kang SeokJin (person)	Justice Department	Humidifier Sanitizer	Declaration of the State of Affairs	Kim MiKyung (person)	Council
4	Baegot (place)	History Textbook National-ization	Scholarship	Press Conference	Auto-Driving Car	Siheung (place)
5	Female Student	Outside Director	Law School	Park Geun-Hye (person)	Nullification of an Agreement	Press Conference
6	Siheung (place)	Graduate School	Search and Confiscation	Class Boycott	Chairman	Stay-in Strike
7	A Doctorate	Coming-out	Ministry of Education	Student Council	Physical Conflict	Graduate Student
8	Examinee	Law School	Arrest warrant	Occupation of Main Building	Graduate School	Severe Penalty
9	Cheating	Disciplinary Dismissal	Chairman of the Board	MOU	Ahn ChulSu (person)	Business Agreement
10	Professorship	Sexual Harassment	Graduate Student	SNU 70th Anniversary	Administ-rative Building	Artificial Intelligence

On the other hand, newspaper articles covering core functions of university such as education, admissions, research, and administration, were reported throughout the year. Total numbers of news articles searched by 'education', 'research', and 'student' keywords do not significantly differ by year. The total number of news items searched for by using 'education' keyword was 6,361 in 2015, 5,554 in 2016, and 5,705 in 2017. The total number of news articles searched for by using 'research' term was 6,827 in 2015, 7,369 in 2016, and 8,121 in 2017. The total number of news articles searched for by 'student' term was 4,931 in 2015, 4,683 in 2016, and 4,505 in 2017.

Next, this study examined the interest of the media and associative words on major policies of SNU. Total numbers of news articles related to major policies of SNU are shown in Table 11 below. In detail, the media coverage was focused on research and multi-campus policy the most. More than 10,000 news stories were searched. Key words such as 'Scientist', 'KAIST', 'IBS', 'Winner', 'Ministry of Science, ICT and Future Planning', and 'Commercialization' were found as associative words of SNU research policy. In relation to the policy of strengthening public education in SNU, which was also the keyword of high school education policy of Moon Jae-in administration, associative words such as 'Regular Recruitment', 'Examinee', 'Essay screening', and 'Admission Officer System' were searched as associative words. These associations imply that changes in the admissions system are closely linked to the publicness of university education. The current university admission system, which induces excessive private tutoring raises concerns that it cannot provide fair education opportunities socially vulnerable classes. News articles described that university admission policies should not only expand university autonomy, but also should enhance the public service of college education.

Table 11. Policy related news articles and keywords

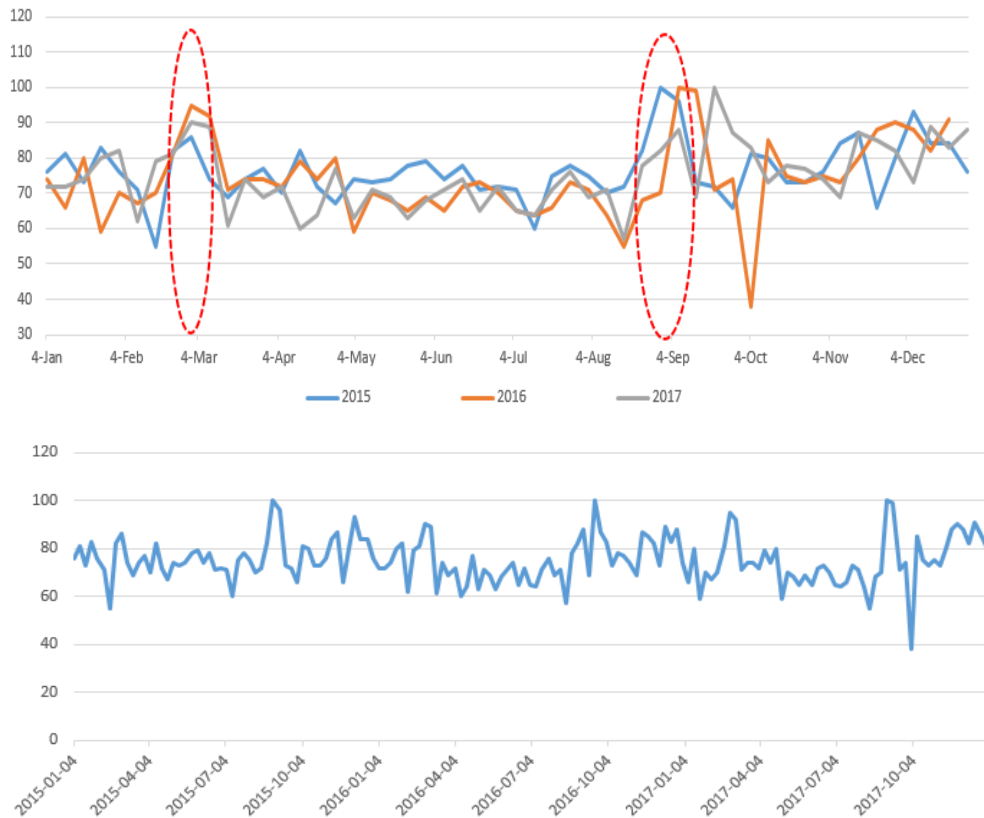
Policy	Frequency (Ratio)	Key Words
1. <i>Fostering Human Resources</i>	7,139 (9%)	우리나라 산학협력 성공관대 한양대 대한민국 장학금 교육부경쟁력 박사학위 연세대 고려대 MOU 과학기술 전문가 인문학 포스텍 KAIST 이공계 국방부 교육과정 전문연구요원제도 폐지 계획
2. <i>Supporting Students</i>	2,617 (3%)	실험실 자연계대학들 연세대 학부모 로스쿨 이화여대 수시모집수험생 장학금 배곧 부산대 지역균형선발전형 교육부성균관대 이사장 한라 경쟁률 생활비 기숙사 삼성전자 장학증서
3. <i>Advanced Research</i>	20,011 (25%)	사업화 이사장 KAIST 교수들 교육부 포항공대 미래창조과학부 산학협력 성공관대 연세대 과학자 SCI IBS 고려대 박사학위 전문가 중국 수상자 삼성 우리나라시상식 석좌교수
4. <i>Research Support System</i>	625 (1%)	미래창조과학부 과함자 RB코리아 DNA IBS 교수들 전문가 보고서 KBSI ICT 가습기 살균제 옥시 노벨상 뒷돈 교육부 RD KAIST 연세대 이사장 옥시레킷벤키저 삼성 생물학
5. <i>Career</i>	6,106 (8%)	대기업 포스텍 연세대 사업화 대학생 청년들 벤처기업고려대 이사장 MOU 일자리 KAIST 중소기업 배곧 한양대 스타트업 성균관대 세종시 이화여대 시흥
6. <i>Strengthen Public Education</i>	1,947 (2%)	자기소개서 일반고 비교과 대학 입시 고려대 수시모집 수시전형 대학들 수험생 연세대학부모들 학교생활기록부 학종논술전형 서강대 성균관대 모집인원 특목고 대학별 내신 성적 신입생
7. <i>Social Contribution</i>	959 (1%)	글로벌봉사단 충남대 삼성그룹 경기도 소비자 산학협동재단 봉사활동 CSR 임직원들이사장 삼성 고려대 한수원 행복수명지표 장학금 독스바겐 카카오게임 두산



Unlike media news articles, social media needs to be categorized and analyzed by media channel. Internet information portal search sites such as Google, Naver, and Daum are the most common and traditional channel used for searching information. Social media refers to online tools and platforms that people use to share their thoughts, opinions, experiences, and perspectives. They have characteristics of participation, disclosure, dialogue, community, and connections (You, 2010). Comparing the most widely used blogs and social networking features of social media, blogs are mainly aimed at information sharing. Furthermore, SNS has the purpose of relationship formation and entertainment, as well as information sharing.

Below is the annual trend of keyword search numbers from 2015 to 2017 for SNU. Searches in Google Trends were about twice the number of searches each year. The search volume increased in March and September, consistent with the beginning and end of the university semester. The difference in search volume from September to October seemed to change depending on whether there was a social issue such as a national audit. Search volume by year did not changed much. The average search volume per day was 76.31 in 2015, 74.85 in 2016, and 74.21 in 2017.

Figure 6. Interest Trend of SNU by Google Trend



This study analyzed the morpheme using social media texts (Blogs of Naver and Daum, Facebook) containing the word ‘Seoul National University’. Results of extracting words (keywords) around nouns were then obtained. Extracted results excluded basic association words of ‘university’, such as ‘student’, ‘professor’, and ‘education’. Tables 12 and 13 show keywords extracted mainly from social media by semester from 2015 to 2017. Unlike mass media, social media platforms generate a lot of text about students' personal interests as a way of expressing their opinions. On blogs, texts related to the main function of the university such as admission, student's affair, and campus life were extracted every year while keywords related to social issues have been extracted in Facebook. In other words, SNS keywords tended to be limited to specific issues while those in newspaper articles tended to be associated with various topics. As shown in Table 12, in blogs, 'Rolling Admission', 'Preparation' and 'Acceptance' related to the entrance examination system were extracted while 'A graduate', 'Reference list' related to the university's reputation, and academic affair(study) and part-time job(tutor) related to campus life have been extracted. Table 13 shows characteristics of SNS, in which public opinion on university issues is expressed as well as students' interest in school life.

'Administrative body', 'Student Council', and 'Siheung Campus', and 'Administrative Organization' extracted in 2016 and 2017 revealed conflicts between university authorities and the student council at the time on the construction of Siheung campus.

Table 12. Keyword Frequency of Blogs in Naver and Daum

<i>Rank</i>	<i>2015</i>	<i>2016</i>	<i>2017</i>
1	Rolling admission	Rolling admission	Acceptance
2	Acceptance	Acceptance	Professionalism
3	Reference list	Preparation	Laboratory
4	Study	Graduate(alumni)	Occasional-Regular Admissions
5	Part-time job	Camp	In-Seoul

Table 13. Keyword Frequency in Facebook

<i>Rank</i>	<i>2015</i>	<i>2016</i>	<i>2017</i>
1	Club	Club	Entrance examination
2	Festival	Student Council	Recruitment
3	Recruitment	Recruitment	Administrative body
4	Student council	Entrance examination	Student Council
5	Graduation album	Administrative body	Siheung Campus

Users are aware of blogs and SNS as relatively different types of media. Thus, when users access university-related information, they can make decisions on different paths depending on media channel. Lee & Hur (2012) have investigated media distinctiveness between twitter and other social media and described how information interpretation and responsiveness by internet users can be affected by the distinctive characteristics of twitter and blog media. In their study, characteristics of SNS including relationship formation patterns among users, channel diversity, immediateness of information communication, information flow within media, and media credibility were found to be more supportive than blogs. This indicates that blogs can be perceived as a suitable means of carrying out the essential role of institutional activities while SNS can be used as an appropriate means to bring about long-term institutional image improvement and a benevolent attitude through campaigns. Therefore, when obtaining ideas from policy agendas from social media, it is necessary to identify whether keywords of SNS are created by a particular interest group. Thus, the school need to look at other types of social media to balance

the majority of public views.

Sentimental analysis provides valuable clues to collect and analyze positive and negative public opinions and establish university policies. This is because such analysis method can identify the cause and frequency of positive and negative opinions. The media and the SNS can express subjective opinions about controversial social issues and events. Policy makers can analyze these arguments to complement strengths and weaknesses of policies. The following is an analysis of the relationship structure of words appearing in the SNS using network analysis method to see whether the public supports or opposes the policy. This study conducted a detailed analysis of associated keywords and sensitive keywords about university policies and specific keywords of 'entrance examination' and 'Siheung campus' regarding policy. In general, positive or neutral images were observed in university policies and entrance policy. On the other hand, some negative images about Siheung campus policy were highlighted without positive image. There was an opinion that there should be deep concern and improvement efforts. Based on results of sentimental analysis, the degree of public opinion can be quantified and the nature of the public opinion on the issue can be objectively grasped. In other words, when analyzing newspaper articles and SNS tweet volume about Siheung campus issues, if the negative coverage amount is 1,000 and the positive coverage amount is 800, the direction of public opinion on the issue can be judged to be negative and the degree of negativity can be quantified by a difference of 200.

Table 14. Keyword Analysis of SNU by Social Matrix (2017. 5.)

<i>Policy</i>	<i>Associative keyword</i>	<i>Sensitive Keyword</i>
Education	Korea, Professor, Administration, Elegant, Envious	Serious, Envious, Noble, Trustworthy
Research	Professor, Dissertation, Brazen, Plagiarism, Master	Brazen, Plagiarism, Good, New, Outstanding
Admission	Diversity, Special-purpose high school, College entrance type, Student, Economic power	Diverse, Indirect, worry, Deep, Passion
Administration	Abundance, Professor, Active, Career Opportunity	Well-known, Abundance, Unusual, Active
Campus	Student, Subway station, Conference, Long distance, Siheung campus	Long distance, State of Emergency, Small, Large, Extraordinary

Table 15. Analysis of ‘Entrance examination’ by Social Matrix (2017. 5)

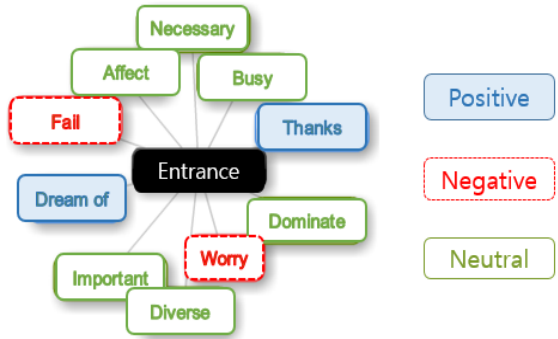

Rank	Associative keyword	
1	Student	 <p>The diagram for 'Entrance examination' shows a central black node 'Entrance' connected to several associative keywords: 'Necessary' (green), 'Affect' (green), 'Busy' (green), 'Thanks' (blue), 'Dominate' (green), 'Worry' (red), 'Diverse' (green), 'Important' (green), 'Dream of' (blue), and 'Fall' (red). To the right, three sentiment boxes are shown: 'Positive' (blue), 'Negative' (red), and 'Neutral' (green).</p>
2	Rolling admission	
3	Evaluation	
4	CSAT	
5	Region	

Table 16. Analysis of 'Siheung Campus' by Social Matrix (2017. 5)

Rank	Associative keyword	
1	Student	 <p>The diagram for 'Siheung Campus' shows a central black node 'Siheung' connected to several associative keywords: 'Worry' (red), 'Pitiable' (red), 'High' (green), 'Oppose' (red), 'Demand' (green), 'Different' (green), 'Crisis' (red), 'Punishment' (red), 'Violent' (red), and 'Urge' (green). To the right, three sentiment boxes are shown: 'Positive' (blue), 'Negative' (red), and 'Neutral' (green).</p>
2	Police	
3	Pitiable	
4	Protest	
5	Crisis	

4.2. RQ2: Degree Centrality Analysis

The second research question was to determine whether network centrality between core links was related to SNU in news reports and social media. Degree centrality assigns an importance weight based on the number of links held by each node. TF-IDF weighting model can be used to identify the importance of SNU-related words in newspaper articles and SNS. The frequency of a particular word in a newspaper article can be regarded as an important word in the document. However, the importance of this word is not high when the particular word has the same frequency as that in newspaper articles on other subjects. TF-IDF is one of text mining methods used to evaluate the importance of extracted words. It indicates the importance of words considering their frequencies of use in the document (Neto et al., 2000).

Table 17 shows importance of words extracted from contents of newspaper

articles. Higher frequency words tend to have larger TF-IDF values, such as words of 'president' and 'chairman of the board'. It can be explained that media coverage for the leader of the organization is important. Among top 10 keywords, 'Siheung' and 'Student Council' had relatively low values while 'Examinee' and 'Outside Director' among the bottom 10 keywords had relatively high values, indicating that there were a lot of media reports about 'Siheung'. However, these were limited to a single case. On the other hand, keywords with both high frequency and TF-IDF values such as 'Presidents' and 'Doctors' not only had a large number of reports, but also were widely used in various topics.

Table 17. TF-IDF of Top Frequency Words in Newspapers

<i>No</i>	<i>Keyword</i>	<i>TF-IDF</i>	<i>No</i>	<i>Keyword</i>	<i>TF-IDF</i>
1	President	95,457.37	11	Research report	19,604.50
2	Ph.D. program	34,416.74	12	Supreme Court	18,541.87
3	Siheung (place)	12,941.50	13	Humidifier Sanitizer	4,376.77
4	Textbook	34,863.60	14	Scholarship	16,702.56
5	Chairman of the board	72,697.26	15	Examinee	27,992.48
6	Ministry of Education	31,266.46	16	Press Conference	13,440.12
7	Student Council	11,108.43	17	Outside director	16,411.75
8	Law school	34,471.08	18	Sexual harassment	8,193.66
9	Oxy (company)	31,129.09	19	Occupation	10,738.81
10	Justice Department	24,031.08	20	agreement	13,391.84

It is considered prominent when one factor is distinguished in the network compared to other factors with unique characteristics. Among these factors, factors that are more related to other factors are considered to have higher degrees of centrality. Table 18 and Figure 7 show results of analyzing the degree of centrality for keywords in newspaper articles searched with 'SNU'. Degree of centrality measures whether a particular word in a semantic network is directly connected to other words. If a word has higher degree of centrality index value is, it has higher influence on the semantic network. Most degree of centrality index values ranged from 17 to 19. They were not significantly different from word frequency or TF-IDF rankings. 'President', 'Ph. D. program', 'Research Report', 'Supreme Court', and 'Press Conference' had the highest value of 19 while 'Oxy' had the lowest value of 7. If the degree of centrality is large, the size of nodes corresponding to the intersection of the network is large and nodes having a high frequency of coexistence are located relatively closely to each other. As shown in Figure 7, keywords related

to students such as ‘Occupation’, ‘Student Council’, ‘Examinee’, ‘Siheung’, and ‘Scholarship’ are located relatively close to each other. Key words related to government policy such as ‘President’, ‘Chairman of the Board’, ‘MOE’, ‘Law School’, and ‘Textbook’ have a strong degree of centrality because the node size is large.

This study grouped words with similar positions in the semantic network and performed CONCOR analysis in order to examine the relationship among groups. As shown Figure 8, results of CONCOR analysis are divided into three clusters: university, student, and education. More specifically, the university issue cluster included 'President', 'Press Conference', 'Business Agreement', and 'Supreme Court'. Student issue cluster included 'Siheung', 'Student Council', and 'Examinee' while education issue cluster included 'Scholarship', 'Law School' and 'Textbook'.

According to the degree centrality analysis of SNU-related newspaper articles, characteristics of the association between words were unclear because many words had the same value. However, visualization data by CONCOR analysis that classified similarity groups by repeatedly performing correlation analysis showed characteristics of the network. These characteristics of clusters identified in the CONCOR analysis can help set the direction and issues of university policy. Considering social phenomena contained in newspaper articles, there are different considerations in policy and research related to student, education, and university environment. In addition, social problems such as Humidifier Sanitizer of Oxy Company, students' school occupation caused by the Siheung campus, and outside director of professors are classified into separate clusters. Including more public opinion keywords in this analysis can help policy makers judge policy directions and issues.

Table 18. Results of Degree of Centrality

<i>No</i>	<i>Keyword</i>	<i>Degree</i>	<i>No</i>	<i>Keyword</i>	<i>Degree</i>
1	President	19	11	Research Report	19
2	Ph.D. Program	19	12	Supreme Court	19
3	Siheung	17	13	Humidifier Sanitizer	15
4	Textbook	18	14	Scholarship	18
5	Chairman of the board	18	15	Examinee	17
6	Ministry of Education	18	16	Press Conference	19
7	Student Council	17	17	Outside director	17
8	Law school	18	18	Sexual harassment	18
9	Oxy	7	19	Occupation of main building	16
10	Justice Department	18	20	Business agreement	19

Figure 7. Degree of Centrality for the Top 20 Keywords

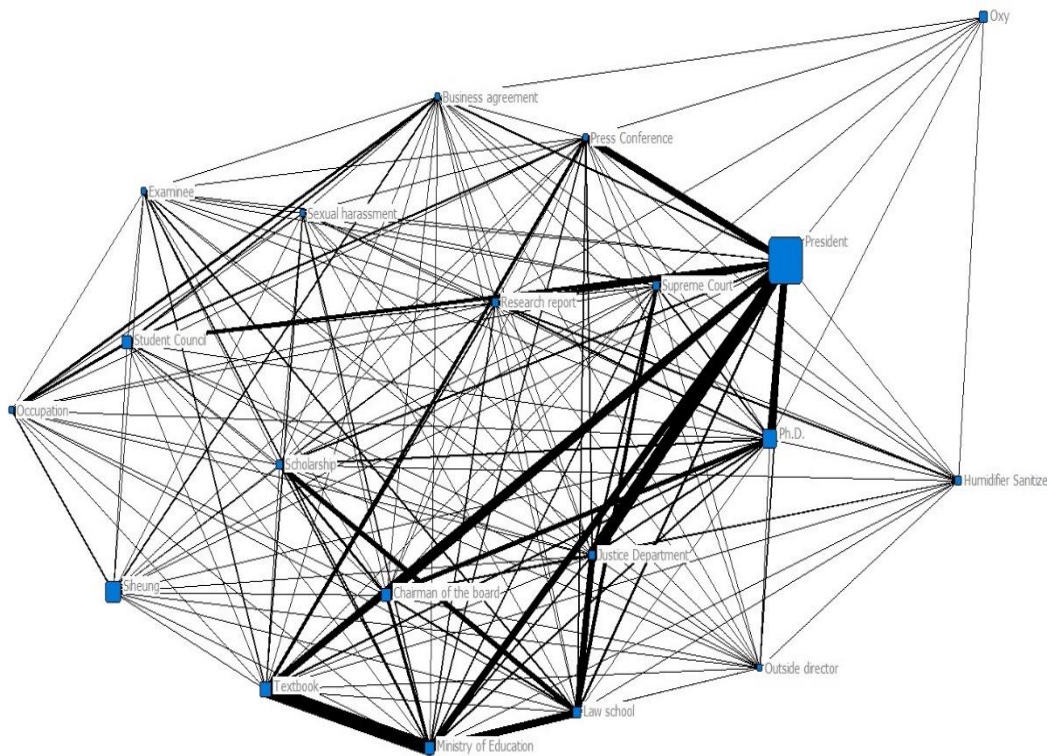
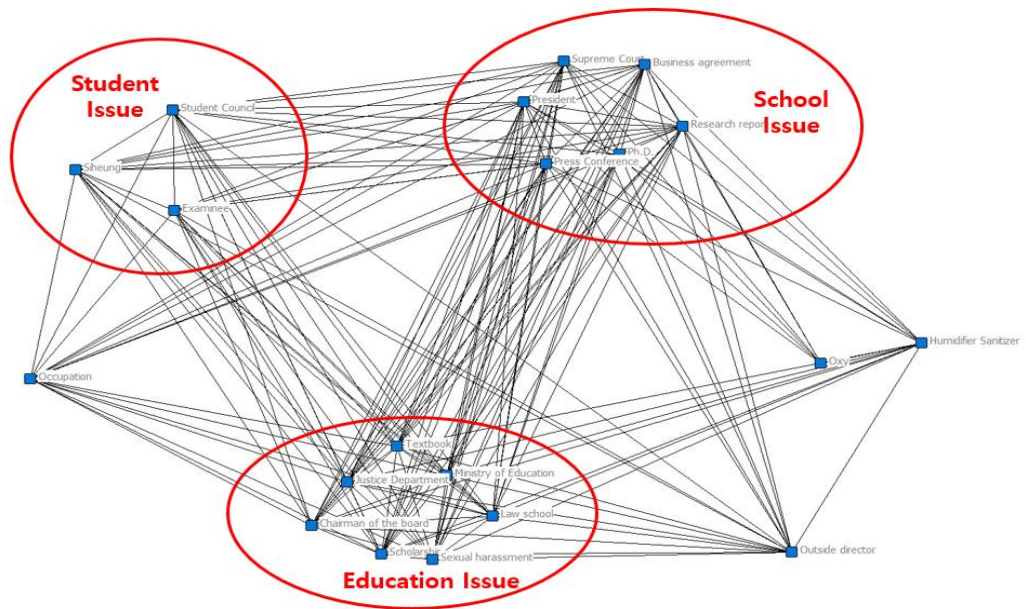


Figure 8. Network CONCOR Analysis of Top 20 Keywords



Chapter 5. Conclusion

5.1. Implications

The use of broad ideas in decision-making, with participation of broad actors enables them to reach high-level policy making and implementation. It is difficult to secure democratic legitimacy if only a few are involved in the policy making process (Cobb & Elder, 1971). The existing policy process consists of policy agenda-setting, policy formulation, policy adoption, policy implementation, and evaluation. The proliferation of social media has changed the policy environment, thus changing the policy setting process. As seen in the previous example, the fact that the policy agenda-setting is changing from organizational leadership such as government, political party, interest group, and university authority to private leadership is a theoretically important change.

Rapid development of information and communication technology has enabled various media based on networks such as online news articles and SNS to function as important tools to increase citizen participation in politics and administration. These societal and technological changes have led to interest in Big Data analysis to analyze diverse and large data in our society. Efforts have been made to launch new services or make improvements to existing policies so that responses of consumers or policy beneficiaries not only in private companies, but also in government agencies and public institutions, can be promptly analyzed by utilizing Big Data.

Since tertiary education policy is basically formed in interaction with interest groups in the present or future environment, there is a continuing flow of mutual checks or balances among rationality, equity, and politics seeking satisfaction of collective interests throughout the decision-making process. Policies of higher education institutions are formed through a political process in which various groups compete to gain greater influence with different interests. Policy implementation cannot be completely free from such political process. Recently, as communication of citizens and formation of interest groups become easier, public opinion formation and pressure exercises are becoming more active. Thus, policy participation is getting more active in order to attract favorable policies to oneself or their group. In this context, Big Data analysis of people's evaluation of major policies implemented at SNU seems to be meaningful in many ways.

This paper explored the applicability of Big Data to the development of the policy agenda-setting in higher education using recently developed programs and methodologies. In doing so, the study selects Seoul National University as a case of higher education institutions and conducts text mining and semantic network

analysis on the collected texts from internet news and social network services. This was based on analysis of 106,041 Internet newspaper articles from 38 media channels and 2,026 texts from blogs and Facebook for three years from 2015 to 2017 by searching with word of 'SNU'.

First, this paper examined news reports and main media of SNU and tried to figure out what issues were important in news reports and social media mentions. Results were used to deduce that press articles emphasized not only education and research as core functions of the university, but also focused on publicity and integrity of its members. In the keyword structure of the university policy through SNS and blog analysis, they showed different patterns. From the blog, texts were steadily extracted every year related to the original function of the university. However, in SNS, keywords related to social issues were extracted, although SNS keywords tended to be limited to specific issues as opposed to newspaper articles being associated with various topics. In Internet news analysis results, there were many keywords such as 'President', 'Ph.D. Program', 'Siheung (place)', 'Textbook', 'Chairman of the Board', 'Ministry of Education', 'Student Council', 'Law school', 'Oxy (company)', and 'Justice Department' about social issues related to SNU policies and its members. Results of SNS analysis showed that there were many keywords closely related to school life such as 'General admission', 'Recruitment', 'Entrance Examination', 'Graduation', and 'Shuttlebus'. As a result of analyzing media interest in 11 major policies of SNU's 2017 work plan, the media coverage was mostly focused on research and multi-campus policy. Since research policy is intertwined with various stakeholders and domains, there has been continuous media coverage throughout the year. On the other hand, multi-campus policy to build a new campus in Siheung City was centered on media coverage from 2016 to 2017 due to internal and external conflicts.

Second, the public has a relatively negative evaluation of the major policies promoted by SNU. It may be because the policy itself was wrong, but also because people with negative perspectives on the policy tend to actively post on their SNS. At this point, it may not be appropriate to evaluate public opinion of the policy from emotional analysis results as an absolute proportion of positive or negative. The purpose of sentimental analysis on public opinion is not to judge policy by absolute numbers, but to find policy alternatives, policy supply and beneficiary feedback from both positive and negative opinions. In addition, it can be deduced from results of this study that public opinion will deteriorate rapidly if policy makers do not respond to social issues at the right time. Due to delays in policy decisions on new campus construction, there was continued media coverage from 2015 to 2017 which had a negative impact on the organization's image.

Third, social media buzz can increase explosively when a single event becomes a social issue. Specifically, negative issues tend to spread rapidly in the first week through SNS after occurrence. In the period of heightened conflict, numbers of media reports and SNS mentions were increased. In addition, contents were negative. The quantitative growth and negative image of social media on specific issues has impacted on the university's policy decisions. Policy formation has progressed rapidly with the emergence of internal immoral events related to university members and the advent of external issues such as the nation's anti-graft law.

Next, there are no specific patterns in the media in terms of time. However, in SNS or blog, the amount of buzz increases at the beginning and end of the college semester. In blogs and Facebook, the number of buzzes can increase explosively when a single event becomes a social issue.

Finally, according to the CONCOR analysis, which repeatedly performs correlation analysis and finds an appropriate level of similarity groups, the characteristics of each network were clearly distinguished. The results of the analysis are divided into three clusters of university, student, and education, and each cluster could be grouped into associated keywords. One cluster can be viewed as one policy domain, and the keywords contained in it can be viewed as factors to consider when making policy agendas. In this study, since only 20 keywords were used in the CONCOR analysis, it was limited to find a wide range of keywords related to one policy domain. However, if this analysis is carried out including more keywords, this data will be helpful for policy makers to determine policy direction and problems.

The implication of this study is that institutional support and policy alternatives should be provided for areas of public interest. The meaning from a logical view is that analysis of news articles and social media may be an alternative method to establish policy agenda in view of difficulties in obtaining representation of opinion convergence in a college or university from survey which is a traditional polling method, similar to those of social media analysis. Also, macro trends in public opinion can be grasped quickly at low cost. Through this study, the possibility of data-based policy agenda-setting using data combining quantitative statistics and social media analysis was confirmed

The purpose of this study was to provide a new analytical method to supplement the limitation of information collection system for setting up existing policy agenda by quickly and effectively grasping actual contents of major issues related to a university in social big data. It has methodological significance.

Based on this research, a methodological possibility of forming bilateral symmetric relationship through interactions between university authorities and public by using social data analysis was confirmed. This implies a new paradigm of

policy agenda-setting and suggests that the role and value of policy formation will be added to social policy analytics in various public areas in the future, such as policy alternatives proposal, policy implementation monitoring, and policy impact evaluation feedback.

5.2. Future Study and Limitations

Future research will focus on finding innovative ways to selectively apply media analysis results according to policy area. This raises questions about whether institutional policies must be derived from the public's voice. In order to address complicated policy problems, it is necessary to establish policies based on accurate scientific evidence and objective data. In that sense, it is necessary to conduct a detailed study on which policy areas are better using social media Big Data analytics.

This study did not analyze individual characteristics. Instead, it analyzed data of the whole group of its members. Thus, an ecological fallacy could occur if applied to individuals (Song et al., 2014). In addition, since 'Integrity Awareness' is defined as the frequency of emotion words generated in online documents, it can differ from the operational definition of integrity awareness through existing researches. Analyzing the limited social Big Data for the 2015-2017 period might have limited the prediction of general recognition.

In social Big Data analytics, researchers' errors and bias can intervene in the lexical analysis stage because several manual tasks are used to designate words and phrases that may interfere with the validity of results (Kim et al., 2014). In detail, the preprocessing process started with collecting documents that were subjected to social Big Data analysis through a data mining process. A vocabulary dictionary was then designed for the purpose of analysis. Designing a vocabulary dictionary requires a very high level of rigor in that it can improve the accuracy and reliability of social Big Data results because it establishes a semantic and syntactic vocabulary system specific to research rather than general purpose. Big Data and complete surveys are conceptually distinct and do not guarantee complete or systematic representation of the study subjects. It only reflects about a specific category, therefore, the researcher must pay attention to the question of what the data represents.

Bibliography

- Bernanke, B. S., & Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3), 525-546.
- Carmack, R. C. (2016). *Acceptance of Artificially Intelligent Autonomous Self-Governing Technology (ALASGT): A Qualitative Case Study*. Northcentral University.
- Chun, Y.-H., (2014). Uncertainty of Autonomy Effects in Agentification: the Case of Seoul National University.
- Cobb, R. W., & Elder, C. D. (1971). The politics of agenda-building: An alternative perspective for modern democratic theory. *The Journal of Politics*, 33(4), 892-915.
- Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British journal of educational technology*, 46(5), 904-920.
- Farnham, D., & Horton, S. (Eds.). (1996). *Managing the new public services*. Macmillan International Higher Education.
- Frizzo-Barker, J., Chow-White, P. A., Mozafari, M., & Ha, D. (2016). An empirical study of the rise of big data in business scholarship. *International Journal of Information Management*, 36(3), 403-413.
- Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC view*, 1142(2011), 1-12.
- Grubb, W. N., Sweet, R., Gallagher, M., & Tuomi, O. (2009). OECD reviews of tertiary education: Korea.
- Han, S.-K. (2015). Doing Social Sciences in the Age of Big Data: Rethinking Analytical Strategy in the Changing Data Environment. *Korean Sociological Association*, 49(2), 161-192.
- Hwang, M.-H., et al. (2014). Spatiotemporal Analysis of Unstructured Big Data for Public Opinion Monitoring. KRIHS.
- Hwang, Y.-C. & C. KOH (2013). Analysis of Opinion Social Data on the SNS (Social Network Service) by Analyzing of Collective Damage Reply. *The Journal of Digital Policy & Management*, 11(5), 41-51.
- He, W., et al. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464-472.
- Irvin, R. A., & Stansbury, J. (2004). Citizen participation in decision making: is it worth the effort?. *Public administration review*, 64(1), 55-65.
- Jeong, U.-Y., et al. (2009). 'A Study of the Influence of the Press Reports on the

- Policy Making Process in Local Governments: Focused on Public Servants' Perceptions in Ulsan Metropolitan City.' *Korea Local Autonomy Institute*, 11(1), 207-232.
- KERIS (2014). 'Data System Upgrading and Analysis Study for Future Prediction of Educational Informatization Policy.'
- Kim, G.-H., et al. (2014). big data applications in the government sector. *Communications of the ACM*. 57(3), 78-85.
- Kim, J.-E. and S.-G. Baek (2016). Analysis of Issues on the College and University Structural Reform Evaluation Using Text Big Data Analytics. *Asian Journal of Education*. 17(3), 409-436.
- Kim, S.-C. and K.-H. Kim (2016). A Study on factors affecting the viewer rating of 'My Little Television': Focusing on SNS Big Data. *Journal of Digital Contents Society*. 17(1), 1-10.
- Kingdon, J. W., & Thurber, J. A. (1984). Agendas, alternatives, and public policies (Vol. 45, pp. 165-169). Boston: Little, Brown.
- Kum, H.-C., et al. (2015). Using big data for evidence based governance in child welfare. *Children and Youth Services Review*, 58, 127-136.
- Kwon, K.-S., et al. (2014). Developing agenda using education (Big) data: Focusing on Social Network Analysis. KERIS.
- Kim, S. H. & Chang, R. S. (2010). The Study on the Research Trend of Social Network Analysis and the its Applicability to Information Science. *Journal of the Korea Society for Information Management*, 27(4), 71-87.
- Kwon, Y.-O. (2013). Data Analytics in Education: Current and Future Directions. *Journal of Intelligence and Information Systems*, 19(2), 87-99.
- Lee J.-B., et al. (2012). A Comparative Study on Different Characteristics of Social Media and Product Information Processing and Evaluation. *Korea Association of Information Systems*, 21(1), 63-85.
- Lee, Y.-J. and D. Kim (2016). Methodological Implications of Employing Social Bigdata Analysis for Policy-Making: A Case of Social Media Buzz on the Startup Business. *Journal of the Korea society of IT services*, 15(1), 97-111.
- Lim, H.-J. and S.-H. Park (2015). A Tentative Approach for Regional Futures Strategy with Big Data - Through the Analysis using the Data of SNS and Newspaper. *Journal of the Korean Cadastre Information Association*, 17(1), 75-90.
- Lim, J. (2006). A CROSS-LAGGED ANALYSIS of agenda setting among online news media. *Journalism & Mass Communication Quarterly*, 83(2),

298(215).

- Lorrain, F., & White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, 1(1), 49-80.
- McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public opinion quarterly*, 36(2), 176-187.
- McCombs, M. (2002). The agenda-setting role of the mass media in the shaping of public opinion. In *Mass Media Economics 2002 Conference, London School of Economics*: <http://sticerd.lse.ac.uk/dps/extra/McCombs.pdf>.
- Min, G.-Y. and D.-H. Jeong (2013). Research on Assessment of Impact of Big Data Attributes to Disaster Response Decision-Making Process. *The Journal of Society for e-Business Studies*, 18(3), 17-43.
- Neto, J. L., Santos, A. D., Kaestner, C. A., Alexandre, N., & Santos, D. (2000). Document clustering and text summarization.
- Niemi, D., & Gitin, E. (2012). Using Big Data to Predict Student Dropouts: Technology Affordances for Research. *International Association for Development of the Information Society*.
- Park, J.-C., et al. (2015). Proposal of Smart era Online Learning Model with BigData. *Journal of the Korea Institute of Information and Communication Engineering*, 19(4), 991-1000.
- Park, J.-E. (2014). Future Strategy Center meaningful achievements and future direction. *Future Horizon*, (22), 15-18.
- Park, H.-J. (2014). A Study on Open Access and Use Plan to Public Data in Educational and Academic Sectors. Ministry of Education.
- Rhodes, R. A. W. (1996). The new governance: governing without government. *Political studies*, 44(4), 652-667.
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618.
- Russell Neuman, W., et al. (2014). The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data. *Journal of Communication*, 64(2), 193-214.
- Seo, Y.-S. (2016). Future government and bureaucracy due to the development of artificial intelligence technology. *The Korea Institute of Public Administration*, 17, 18-23.
- Shin, H.-S. (2018). An Analysis of Moon Jae-in Administration's Higher Education Policy and Its Tasks According to the Policy Environment Changes. *The Journal of Research in Education*, 31, 227-272.

- Song, Tae Min, et al. (2014). Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. *Yonsei medical journal*, 55(1), 254-263.
- Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *ICWSM*, 14, 505-514.
- You, H.-J. (2010). A Study on the Web PR Strategy using Social Media: The Case of Chungcheong Tourism. *The e-Business Studies*, 11(5), 97-116.
- Yun, M.-Y. (2013). Big Data use cases in Public Sector and Policy Activation Plan.

소셜 빅데이터 분석을 활용한 정책의제설정에 관한 연구

서울대학교 행정대학원
글로벌행정 전공
김금순

본 연구는 텍스트 데이터 분석 기법(Text Big Data Analytics)을 활용하여 서울대학교를 사례로 소셜 미디어에 나타난 주요 쟁점들을 분석한 것이다. 2015년부터 2017년 까지 3년 동안 언론 보도, 포털 블로그 및 SNS 등에 공개된 대중의 의견을 주제분석, 감성분석 기술로 분류하고 네트워크 분석과 연관분석 등의 기법을 적용하여 시각화 하였다. 우선 대학의 주요 정책에 대한 여론을 거시적으로 파악하고, 정책 형성과정에서 언론 보도와 소셜 미디어가 미친 영향을 사례 분석을 통해 조망하였다. 이는 대학이 빅 데이터를 활용하여 정책 의제 설정에 보완시스템으로써 활용될 수 있음을 보여준다.

분석 결과, 첫째, 언론 기사에서는 대학의 주요 기능인 교육, 연구 뿐 아니라 구성원의 도덕성까지 다양한 토픽이 나타났다. 또한 대학이 추진한 주요 정책들이 소셜 미디어 상에 많은 이슈가 도출되어 정책 참여자들이 정책 도입 또는 대안 탐색을 위해 쉽고 빠르게 활용할 수 있다는 점에서 본 연구는 탐색적이지만 정책적 및 분석방법론적 의의를 가진다고 할 수 있다. 둘째, 단일 사건이 사회적인 문제로 대두되면 소셜 미디어에서 급속히 전파되었으며, 특히 부정적 이미지는 SNS에서 통해 발생 첫 주 급속도로 전파되는 경향을 보인다. 셋째, 언론 기사에서는 다양한 토픽을, 블로그는 대학의 교육에 대한 보다 전문적인 콘텐츠를, 그리고 SNS에서는 특정 토픽에 집중하여 의견을 전파하는 경향이 있다. 이는 소셜 미디어 매체의 특성에 따라 다르게 정책에 반영해야 한다는 시사점을 제공한다.