

Analysis of Research Trends in Regional Innovation Using Text Mining

Ju Seop Park*, Soon-Goo Hong*, Na Rang Kim* and Bo Ra Kang**

*Department of Management Information Systems, Dong-A University, Korea
E-mail: juseop60@naver.com, shong@dau.ac.kr, whitecoral@hanmail.net

**Department of Library and Information Science Kyungpook National University,
Korea

Abstract

To aid local governments in solving various regional innovation issues related to regional development, trend analyses should first be conducted. In this study, 579 abstracts published in academic journals between year 2003 and year 2015 were analyzed to examine the research trends of topics related to regional innovation through a keyword frequency analysis and a social network analysis, both of which are text mining techniques. As a result of these analyses, the most frequent keyword that appeared through the clustering of participating entities was regional innovation system during the Roh Moo-Hyun administration. During the Lee Myung-Bak administration, the most frequent keyword obtained through the participation of local residents was regional innovation focused on overall business development, which continued through to the Park Geun-Hye administration. This study suggests a big data analysis method to derive the core problems related to regional innovation and may trigger follow-up research. Furthermore, the results of this study can be used as basic data for local governments and administrative agencies to establish regional innovation policies.

Keywords: regional innovation, text mining, trend analysis, social network analysis, big data analysis

1. Introduction

A decline in birth rates and an aging society can cause critical problems in the long term, such as changes in the composition of the economically active population and an increase in costs [1]. A decreasing population, in particular, diminishes the labor force and the national production capacity. Therefore, many countries facing low population growth have been focusing on regional innovation, which can promote productivity and play a key role in the growth of knowledge-based economies [2].

An analysis on the trends of research topics related to regional innovation will allow us to understand the characteristics of regional innovation policies, determine which regional innovation policies are currently being introduced and established, and predict which improvements are required. Further, the results of this analysis will allow us to establish the paradigm and identity of regional innovation [14], and they can be used as basic data for the staff of local governments to establish policies. Therefore, we aimed to conduct a trend analysis on how regional innovation policies for regional development have been researched. In this study, we collected a total of 579 papers published between year 2003 and year 2015 from the Korea Citation Index (<https://www.kci.go.kr/>) and analyzed their abstracts using text mining techniques. The analysis periods were classified based on the following presidential terms: the Roh Moo-Hyun administration (2003.3.1–2008.2.28), the Lee Myung-Bak administration (2008.3.1–2013.2.28), and the

Park Geun-Hye administration (2013.3.1–2015.12.31). The analysis techniques used were a keyword frequency analysis and a social network analysis.

2. Literature Review

2.1 Text Mining

Text mining is a representative big data analysis technique that extracts valuable and meaningful information from atypical or semi-structured text using natural language processing. Text mining mainly processes atypical data such as text, e-mails, and HTML files. Text mining technology extracts meaningful information from a vast amount of text by grasping their context, identifying correlations between information, and classifying or clustering texts. The text mining process collects text and classifies it into parts of speech; then, it converts it into a form that is easy to analyze through preprocessing, which removes stop words and handles synonyms. After preprocessing, the text is analyzed using correlation, classification, prediction, clustering, and document summary. Previous studies on trend analyses using text mining techniques can be classified into three fields: business and technology, academic, and social [3].

First, in the business and technology field, Ahn et al. [4] examined the detailed technology trends of the Internet of Things by creating a document-term matrix and conducting K-means clustering with text mining techniques after organizing the patent data of the Internet of Things by company in order to analyze technical development trends using United States patent documents. Tseng et al. [5] analyzed patented technologies using text mining techniques such as text segmentation, summary extraction, feature selection, terminology linkage, cluster creation, subject identification, and information mapping.

Second, in the academic field, Ahn et al. [6] compared and visualized the results of extracting joint research networks and research topics through an analysis of the titles, authors, abstracts, and keywords of papers on management information systems between 1980 and 2015 to examine the research trends of information systems over time. Hung et al. [7] analyzed 119 journal articles and proceedings papers in the SCI/SSCI DB to examine mobile learning trends.

Third, in the social field, Chang-Yong et al. [8] classified news reports on North Korean trends published monthly by the Ministry of Unification between 2008 and 2015 in order to analyze the trends of North Korea. After establishing a keyword dictionary with negative and positive expressions in these news reports based on word frequency and the correlations between words and documents, they conducted a time-series analysis to find correlations between past cases of North Korean provocation. Mostafe [9] conducted text mining with over 3,000 random tweets to evaluate the emotions of customers regarding famous brands such as Nokia, T-Mobile, IBM, and DHL. Through this analysis, they introduced a technique to find positive and/or negative emotions.

2.2 Social Network Analysis

Social network analysis is a methodology used to analyze social structures and investigate the relationships between individuals, groups, and societies through networks in social studies [10]. This analysis method derives implications from these relationship patterns by focusing on relationships among the members of an organization and expresses a social environment based on these [11].

Lee [12] classified analysis techniques into six categories according to analysis level after creating a network. In the first, the overall characteristics of the network are analyzed at the network level. This is a macroscopic analysis using analysis indices such as network size and density. In the second, the characteristics of each node and the link

relationships among these nodes are analyzed at the node level. In the third, the characteristics of nodes in the network are analyzed, which includes analyses of the reciprocity, transitivity, and clustering coefficients of specific nodes. In the fourth, a centrality analysis is conducted, which examines the size of the influence of each node in the network. In the fifth, the sub-groups of the total network are distinguished by an analysis at the group level, and various characteristics of the sub-groups are analyzed. In the sixth, the characteristics of the network formed around specific egos are analyzed at the ego network level.

3. Research Methodology

3.1 Analysis Framework

The analysis framework showing the research process used in this study is illustrated in Figure 1.

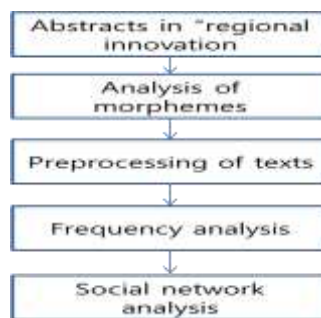


Figure 1. Analysis Framework

3.2 Analysis Outline

The abstracts of papers published in South Korea since 2003 were analyzed using data from the Korea Citation Index (<https://www.kci.go.kr/>). The following keywords related to regional innovation were selected through a review of related literature: regional development, resident participation, regional cooperation, regional governance, regional network, and local governance. With these keywords, 240 abstracts published during the Roh Moo-Hyun administration (2003.3.1–2008.2.28), 219 abstracts published during the Lee Myung-Bak administration (2008.3.1–2013.2.28), and 120 abstracts published during the Park Geun-Hye administration (2013.3.1–2015.12.31) were analyzed. After tagging each part of speech with a morpheme analyzer for all the abstracts, only nouns were used in the analysis. The KrKwic application was used to conduct a keyword frequency analysis [13], and the NetMiner 4 application was used to conduct network analysis and visualization. Before the analysis, only the words related to regional innovation were selected, and three regional innovation experts brainstormed 18 major keywords each for keyword selection to match the relative comparison level.

3.3 Frequency Analysis

Table 1 outlines the frequency analysis of the journal abstracts conducted based on South Korean government administration in three periods. Before the frequency analysis, keywords were selected to match the relative comparison level after excluding the related keywords, such as innovation, economy, and development.

During the Roh Moo-Hyun administration (the first period), the keyword that showed the highest frequency was cluster at 165 times, followed by strategy, system, technology, regime, government, business, and network. The high frequency of these keywords

implies active research on regional innovation systems. During the Lee Myung-Bak administration (the second period), the keyword that showed the highest frequency was project at 165 times, followed by business, support, resident, and strategy. The high frequency of project implies active research on regional development projects. During the Park Geun-Hye administration (the third period), the keyword that showed the highest frequency was project at 131 times, followed by resident, industry, development and government. This means that the research trends during this period were similar to those during the Lee Myung-Bak administration.

Table 1. Frequency Analysis of Regional Innovation Keywords

Divi	First Period 2003.3.1–2008.2.28			Second Period 2008.3.1–2013.2.28			Third Period 2013.3.1–2015.12.31		
	Keyword	Fre q.	Ratio	Keyword	Fre q.	Ratio	Keyword	Fre q.	Ratio
1	Cluster	165	8.3%	Project	165	8.1%	Project	131	9.3%
2	Strategy	156	7.9%	Business	154	7.6%	Resident	122	8.7%
3	System	142	7.2%	Support	149	7.4%	Industry	108	7.7%
4	Technology	141	7.1%	Resident	141	7.0%	Development	107	7.6%
5	Regime	141	7.1%	Government	125	6.2%	Government	92	6.5%
6	Government	133	6.7%	Strategy	124	6.1%	City	86	6.1%
7	Business	131	6.6%	Promotion	121	6.0%	Business	81	5.8%
8	Network	130	6.6%	City	106	5.2%	Growth	80	5.7%
9	Nation	118	6.0%	Culture	98	4.8%	Participation	77	5.5%
10	City	105	5.3%	Participation	97	4.8%	Culture	73	5.2%
11	University	94	4.7%	Structure	91	4.5%	Political scheme	67	4.8%
12	Culture	90	4.5%	Nation	90	4.4%	Resource	63	4.5%
13	Knowledge	87	4.4%	Technology	88	4.3%	Institution	59	4.2%
14	Project	80	4.0%	Cluster	82	4.3%	Role	55	3.9%
15	Entity	69	3.5%	Institution	81	4.0%	Center	55	3.9%
16	Environment	69	3.5%	University	78	3.9%	Environment	54	3.8%
17	Education	66	3.3%	Education	76	3.8%	Network	50	3.6%
18	Outcome	64	3.2%	Capital	72	3.6%	Community	48	3.4%

3.4 Network Analysis

The network relationships among the major keywords used in regional innovation research during the Roh Moo-Hyun administration (the first period) are shown in Figure 2. A high value of weight, which indicates the degree of linkage between nodes, means a high frequency of simultaneous appearance and a high correlation with the keyword. Correlation values were high for policy, industry, research, development, and cluster. This implies that the entities of the region, such as industrial entities and research institutes, had high correlations with regional innovation systems that pursued regional development through networking. The correlations among the sub-nodes in Table 2 show that technical innovation and corporate innovation have high correlations. This suggests that various studies supporting the regional technical innovation policies were conducted,

especially those on the technical innovations of local companies. Furthermore, industrial policy showed a high correlation among the sub-nodes, which implies that many studies were conducted on regional industrial clusters and policies.

Table 2. Descending Order by Weight during the First Period

node 1	node 2	weight
innovation	policy	1554
innovation	industry	1343
innovation	research	1123
innovation	development	917
innovation	cluster	872
policy	industry	871
innovation	implementation	804
innovation	technology	789
innovation	enterprise	718

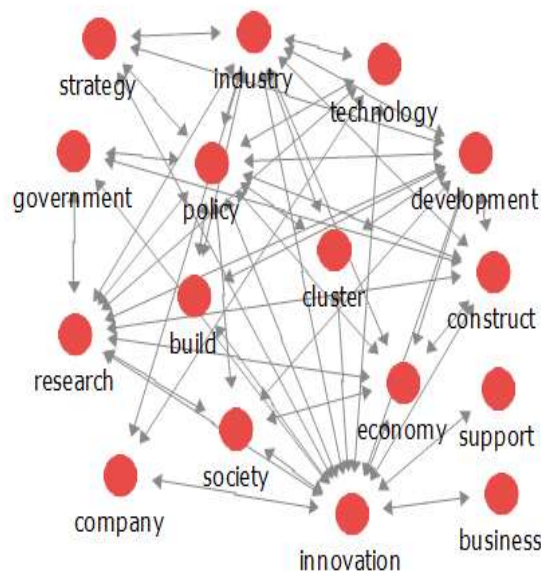


Figure 2. Relationships between Regional Innovation Networks During the First Period

The regional innovation network relationships among the keywords during the Lee Myung-Bak administration (the second period) are shown in Figure 3. Table 3 shows high correlations of economic development, policy development, industrial development, and social development. Whereas many studies on regional innovation systems were conducted during the Roh Moo-Hyun administration, many studies on the economic development of regions were conducted during the Lee Myung-Bak administration.

Among the sub-nodes in Table 3, corporate innovation showed a high correlation, which suggests that many studies were conducted on social enterprises and innovative corporations. Furthermore, industrial research also showed high correlations among all the sub-nodes. This implies that many studies were conducted on regional industrial clusters and policies during the Roh Moo-Hyun administration.

Table 3. Descending order by Weight during the Second Period

node 1	node 2	weight
development	economy	1123
development	policy	1006
innovation	research	863
development	industry	822
development	society	808
development	research	805
innovation	policy	729
research	industry	645
innovation	enterprise	637

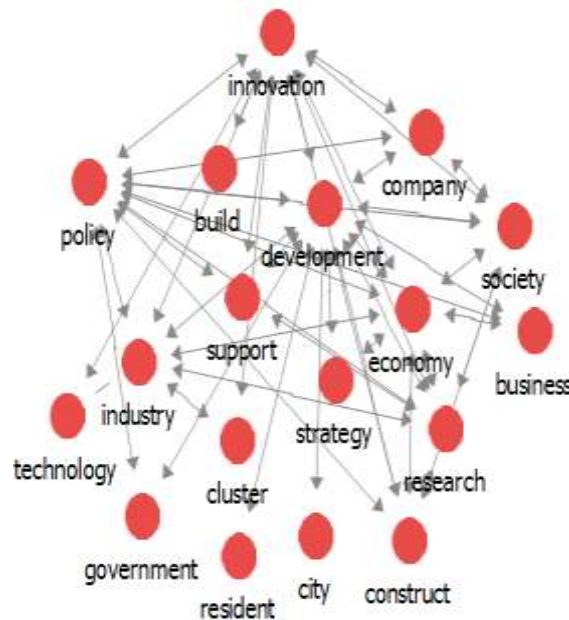


Figure 3. Relationships between Regional Innovation Networks during the Second Period

The regional innovation network relationships among the keywords during the Park Geun-Hye administration (third period) are shown in Figure 4 and Table 4. Policy development, project development, and resident development showed high correlations. This suggests that many studies were conducted on regional economic development

during the Lee Myung-Bak and Park Geun-Hye administrations. In Table 4, corporate innovation showed a high correlation among the sub-nodes, which suggests that many studies were conducted on regional industrial policies. However, studies on regional industrial clusters have been marginal since the Park Geun-Hye administration.

Table 4. Descending Order by Weight during the Third Period

node 1	node 2	weight
development	economy	1123
development	policy	1006
innovation	research	863
development	industry	822
development	society	808
development	research	805
innovation	policy	729
research	industry	645
innovation	enterprise	637

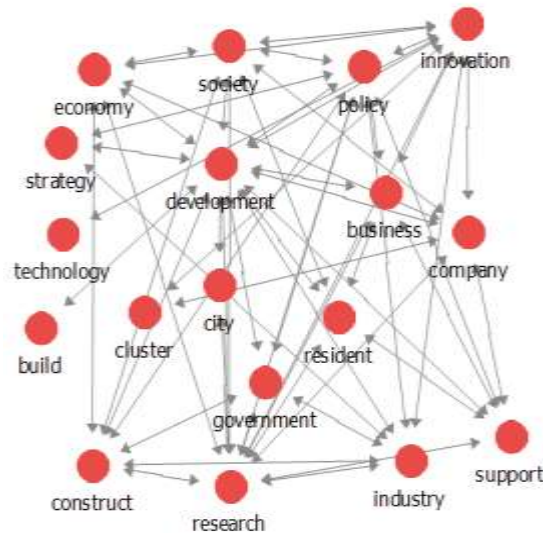


Figure 4. Relationships between Regional Innovation Networks during the Third Period

4. Conclusion

To solve regional innovation problems related to regional development, a trend analysis should first be conducted on how they have changed. Therefore, in this study, 579 paper abstracts published between 2003 and 2015 were analyzed using data from the Korea Citation Index.

Through the clustering of participating entities, the network analysis results revealed that the main subject of regional innovation research was on regional innovation systems

during the Roh Moo-Hyun administration, whereas it was determined to be on overall project development during the Lee Myung-Bak and Park Geun-Hye administrations through the participation of regional residents. Furthermore, the technical innovations of regional companies were main subjects of research during the Roh Moo-Hyun administration, whereas the regional innovations of social enterprise policies were main subjects of research during the Lee Myung-Bak and Park Geun-Hye administrations. In addition, regional industrial clusters and policies were main subjects of regional innovation research during the Roh Moo-Hyun and Lee Myung-Bak administrations, whereas regional innovations in regional industrial policies were main subjects of research during the Park Geun-Hye administration.

This study suggests a big data analysis method to derive core problems related to regional innovation and may trigger follow-up research. Furthermore, the results of this study can be used as basic data for local governments and administrative agencies to establish policies related to regional innovation. This study had a limitation in that it was not easy to determine operation definitions because not many studies have been conducted on regional innovation subjects. Thus, the major keywords were selected by brainstorming with regional innovation experts, but more objective methods are required to obtain more valid results.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A3A2046781).

References

- [1] T. Y. Seo and E. J. Lim, "The direction and corresponding strategies for the low-birth rate and old-aging era in Korean social studies", *Social Studies Education*, vol. 52, (2013), pp. 23-35.
- [2] <http://press.cnumedia.com/news/articleView.html?idxno=12855>.
- [3] J. S. Park and S. G. Hong, "Research of trends in Busan regional innovation policies using text mining", *The Korean Journal of Local Government Studies*, vol. 20, (2016), pp. 1-20.
- [4] S. J. Ahn, J. C. Kim, J. H. Lee, S. S. Park and D. S. Jang, "Business technology trend analysis using patent information", *Proceedings of KIIS Spring Conference*, vol. 25, (2015), pp. 13-14.
- [5] Y. H. Tseng, C. J. Lin and Y. I. Lin, "Text mining techniques for patent analysis", *Information Processing and Management*, vol. 43, (2007), pp. 1216-1247.
- [6] J. K. Ahn, S. D. Kim and H. W. Kim, "Exploring dynamics of information systems research trend using text mining approach", *The Korea Society of Management Information Systems*, vol. 18, (2016), pp. 73-96.
- [7] J. L. Hung and K. Zhang, "Examining mobile learning trends 2003-2008: A categorical meta-trend analysis using text mining techniques", *Journal of Computing in Higher Education*, vol. 24, (2012), pp. 1-17.
- [8] C. Y. Lee and H. S. Moon, "Study on analysis of North Korea's news trends associated with provocations using text mining", *Journal of National Defense Studies*, vol. 59, (2016), pp. 103-124.
- [9] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments", *Expert Systems with Applications*, vol. 40, (2013), pp. 4241-4251.
- [10] J. Scott, "Social network analysis", *SAGE Journals*, (2000), pp. 1-205.
- [11] K. Y. Gwak, "Social network analysis", Cheongam Publisher, (2014).
- [12] S. S. Lee, "Network analysis methods", Nonhyeong Publisher, vol. 39-41, (2012), pp. 177-178.
- [13] H. W. Park and L. Loet, "Understanding the KrKwic: A computer program for the analysis of Korean text", *Journal of the Korean Data Analysis Society*, vol. 6, (2004), pp. 1377-1387.
- [14] Y. C. Choi and S. J. Park, "Research trends in Korean public administration: application of network text analysis method", *Korean Public Administration Review*, vol. 1, (2011), pp. 123-139.

*Corresponding author: Soon-Goo Hong, professor
Department of Management Information Systems,
Dong-A University,
E-mail: shong@dau.ac.kr