

Research Article

Research Trend Analysis of Geospatial Information in South Korea Using Text-Mining Technology

Kwan-Young Oh¹ and Moun-Jin Lee²

¹*Korea Aerospace Research Institute (KARI), Daejeon, Republic of Korea*

²*Korea Environment Institute (KEI), Sejong, Republic of Korea*

Correspondence should be addressed to Moun-Jin Lee; leemj@kei.re.kr

Received 26 April 2017; Accepted 2 July 2017; Published 17 August 2017

Academic Editor: Lei Zhang

Copyright © 2017 Kwan-Young Oh and Moun-Jin Lee. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of this study was to analyze geospatial information (GI) research trends using text-mining techniques. Data were collected from 869 papers found in the Korea Citation Index (KCI) database (DB). Keywords extracted from these papers were classified into 13 GI domains and 13 research domains. We conducted basic statistical analyses (e.g., frequency and time series analyses) and network analyses, using such measures as frequency, degree, closeness centrality, and betweenness centrality, on the resulting domains. We subdivided the most frequent GI domain for more detailed analysis. Such processes permit an analysis of the relationships between the Research Fields associated with each GI. Our time series analysis found that the Climate and Satellite Image domain frequencies continuously increased. Satellite Image, General-Purpose Map, and Natural Disaster Map in the GI domain and Climate and Natural Disaster in the Research Field domain appeared in the center of the GI-Research Field network. We subdivided the Satellite Image domain for detailed analysis. As a result, LANDSAT, KOMPSAT, and MODIS displayed high scores on the frequency, degree, closeness centrality, and betweenness centrality indices. These results will be useful in GI-based interdisciplinary research and the selection of new research themes.

1. Introduction

The term geospatial information (GI) was initially coined in Europe and is now used internationally [1]. GI is information that expresses geographical location and attributes in a form that computers can recognize [2–4]. GI can apply to all information on the Earth's surface, underground, and in the ocean and atmosphere [5, 6]. Representative examples of GI include satellite images and numerical maps. GI has maintained qualitative and quantitative growth with the development of Geographic Information Systems (GIS) [7]. Recently, GI has emerged as the core of the 4th industrial revolution and is expected to create more value by fusing with other technologies [8]. An understanding of GI and application fields is necessary for utilizing integrated GIs; therefore, it is important to analyze research trends that can confirm this evolutionary transition and changes in academic interest in GI.

As text-mining techniques have developed, it has become possible to analyze large amounts of text data, accumulated

over long periods, and a technical environment suitable for analyzing the academic trend has been created [9]. Text mining is a way of processing unstructured data and analyzing patterns that are latent in a text-based database [9, 10]. It provides a means to automatically extract natural language (character information) using a mechanical algorithm [11]. Recently, studies using text-mining techniques are increasingly used in various fields such as computer science, statistics, nutrition, and construction [12–14]. Hung and Zhang [15] analyzed the abstracts of Science Citation Index (SCI)/Social Science Citation Index (SSCI) theses published between 2003 and 2008, for research trends in the field of mobile learning. He et al. [16] analyzed Facebook and Twitter posts and analyzed buyer preferences for pizza chains. Lim et al. [17] analyzed trends in GI by analyzing the frequency and time series of keywords extracted from papers and reports published in South Korea. However, prior studies have been limited in terms of analytical methods. Since analysis is based on basic statistical analysis of

extracted keywords, only sections of research trends can be detected. These statistical methods do not show which topics are centrally located within the research flow or how the connection structure of each theme changes over time. To address this problem, network analysis methodologies such as cocitation analysis or cword analysis have been introduced [18, 19]. Cocitation analysis determines the characteristics of the document through an analysis of citation relationships between documents [20]. However, cocitation analysis has limitations in analyzing research trends using literature units [21]. Cword analysis is a method of detecting the relationships of keywords extracted from the literature and has advantages in research trend analysis. In particular, cword analysis is useful in that it enables the determination of the nature and strength of relationships between keywords. Recently, network analysis has been applied in text-mining research in various fields [22]. Kajikawa et al. [23] performed network analysis in the field of energy research to create a road map to sustainable energy. Additionally, Kajikawa and Takeda [24] derived promising results through an analysis of the network structure of studies of organic light-emitting diodes. Choi et al. [25] conducted a temporal analysis of the patent network to detect changing trends in technology. Through this type of keyword network analysis, it is possible to grasp large-scale trends in Research Fields.

In this research, GI research trends were examined using basic statistical methods and network analysis. The main target keywords of this trend analysis were GI and Research Field (GI-based). Papers relating to GI, over the past 20 years (1996–2015), were screened in the Korea Citation Index (KCI). Additionally, a set of domains (GI, Research Field) were extracted from the keywords presented in each paper. Basic statistical analyses and network analysis were conducted based on these extracted sets of domains. The results of these analyses allowed us to detect large-scale trends for GI, Research Field, and their interrelationships and to present new research themes that combine GI and related research. Managers and policy makers in the field of GI need to know researchers' interests and research priorities to allocate limited resources to GI fields appropriately. Thus, our research mainly aimed to demonstrate the status of GI research trends in South Korea and determine new directions for development.

2. Methods

In this study, basic statistical analysis and network analysis were performed using keywords from GI-related papers. As shown in Figure 1, the research procedure was divided into three stages: data collection, preprocessing, and analysis. First, GI-related papers were collected during the data collection stage and the keywords presented in the paper were extracted. Next, during the preprocessing stage, the keywords were categorized based on the classification scheme. Finally, in the analysis stage, basic statistical analysis and network analysis were performed.

2.1. Data Collection and Preprocessing. This study was limited to papers published in South Korea. GI-related papers were

collected via the KCI database (DB), which can search and collect all papers published in South Korea. We entered Geospatial Information into the KCI DB, with a collection period limited to the past 20 years (1996–2015). We excluded papers whose research objectives were system construction and services. A total of 869 papers were selected as research data. Preprocessing of the keywords was performed in two steps. In the first step, keywords relating to GI and Research Field were selected from the keywords listed in the collected papers. Keywords with similar meanings were changed to preselected keywords. For example, GIS and Geographic Information System were processed with the same keyword. In the second step, we classified the collected keywords according to a modified keyword classification system based on criteria presented by Lee et al. [26] and the Korea National Spatial Data Infrastructure (NDSI) Portal [27]. In particular, NDSI [27] provides a certified classification system for GI produced in South Korea. These classification systems resulted in 13 GI domains and 13 Research Field domains (Tables 1 and 2).

2.2. Analysis. We performed an analysis of research trends over the entire period (1996–2015) and also as divided into four periods: 1st term (1996–2000), 2nd term (2001–2005), 3rd term (2006–2010), and 4th term (2011–2015). The reason for dividing into the 4 periods is that the number of papers published in one year is small. We performed basic statistical analysis, including occurrence frequency and time series development of the GI and Research Field domains, and a network analysis focusing on the frequency of simultaneous domain appearances. Basic statistical analysis yielded the schematic flow of the GI and Research Field domains. Network analysis compared the relative importance of domains and visualized the connective structure of domains. Our network analysis included the calculation of four indices: frequency, degree, closeness centrality, and betweenness centrality. Frequency indicates the number of domains extracted from the papers. Degree is an index indicating how connected a particular node is to the surrounding nodes. Closeness centrality is the average of the number of specific trunk lines for one node to be connected to each node on the network. Finally, betweenness centrality is an index that measures the extent to which a specific node plays an intermediary role when constructing a network with other nodes. The scope of analysis was divided into major classification and subdivision classification analyses. Major classification analysis targeted the domains presented in Tables 1 and 2, and subdivision classification analysis targeted the keywords contained in each domain. We conducted all analyses using R and NodeXL [28], which are public software.

3. Results

3.1. Major Classification Analysis

3.1.1. The GI Domain. GI-related keywords extracted from the papers were divided into 13 GI domains, which are listed in Table 1. Figure 2 shows the frequency percentage and time series frequency of the GI domains over the entire study

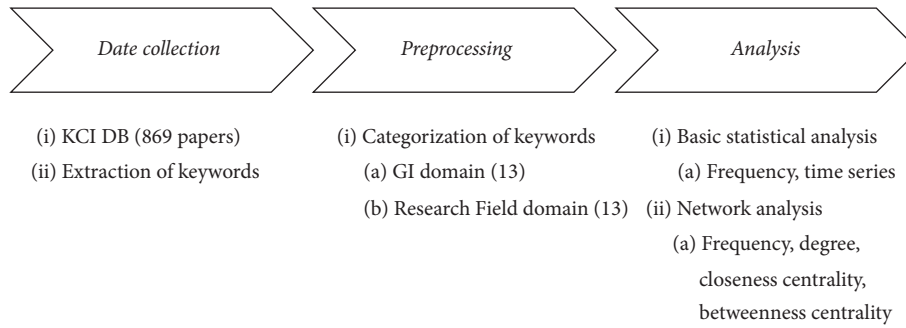


FIGURE 1: Research procedure.

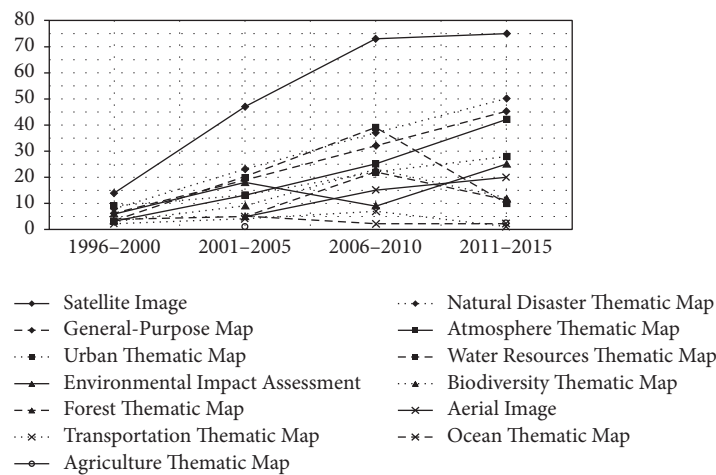
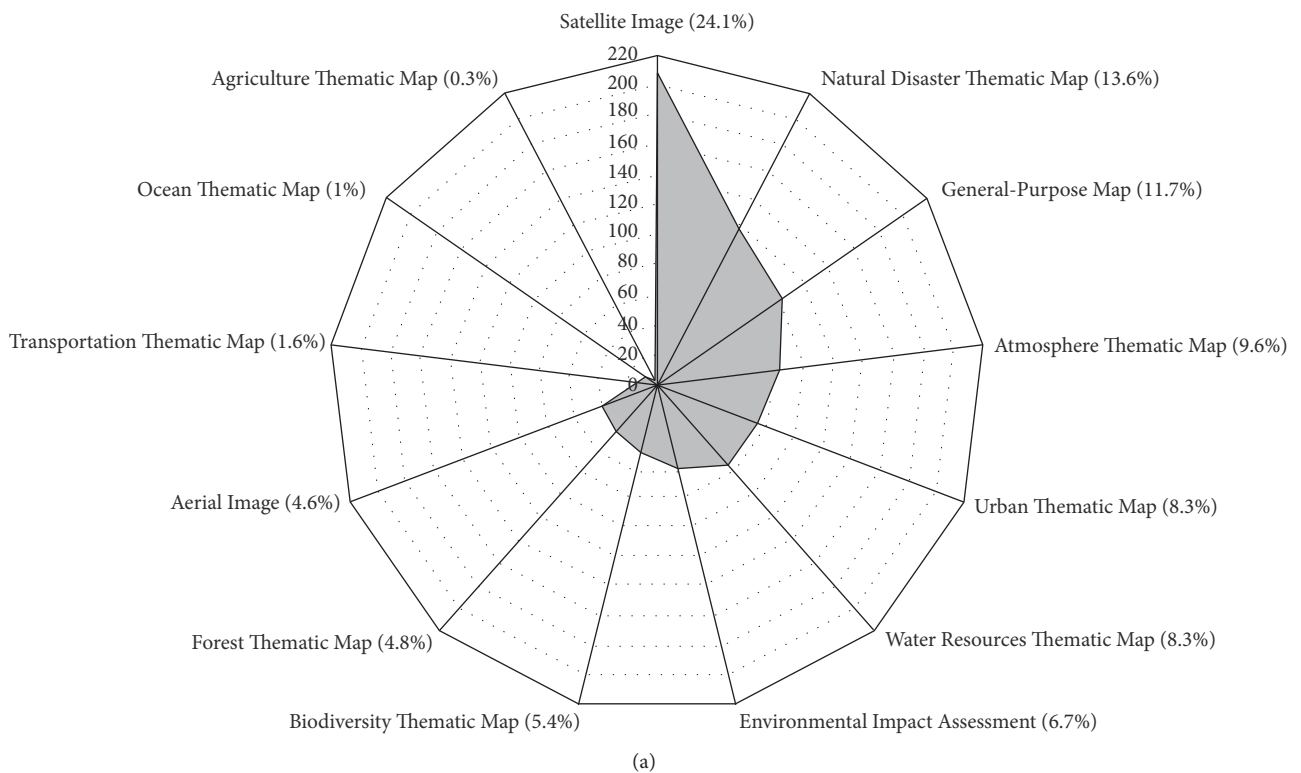


FIGURE 2: Frequency percentage (a) and time-dependent frequency change (b) of the GI domain.

TABLE 1: Geospatial Information (GI) domains.

Domain	Keyword examples
Satellite Image	KOMPSAT, GEOKOMPSAT, LANDSAT, MODIS, SPOT, <i>Cosmo-SkyMed</i> , ALOS, IRS
Aerial Image	Aircraft, Unmanned Aerial Vehicle (UAV), Drone, Lidar
General-Purpose Map	World map, Digital Elevation Model (DEM), Digital map, Electronic map
Environmental Impact Assessment	Land cover map, Land use map, Biotope map, Soil map
Biodiversity Thematic Map	Biodiversity census data, Ecological zoning map
Agriculture Thematic Map	Agricultural census data, Farm Program Atlas, Agricultural land use
Atmosphere Thematic Map	Climatic map, Air temperature, Amount of rainfall
Natural Disaster Thematic Map	Disaster census data, Landslide, Earthquake, Flood, Volcano
Forest Thematic Map	Forest classification, Wood age, Forest location
Ocean Thematic Map	Cadastral map, Water level (Tidal), Ocean-uses atlas, Sea surface temperature
Urban Thematic Map	Population density, Built-up area, Cultural facilities
Transportation Thematic Map	Rail, Road, Airport, Street
Water Resource Thematic Map	Water census data, Water supply facilities, Groundwater, Water resource

TABLE 2: Research Field domains.

Domain	Keyword examples
Natural Disaster	Landslide, Earthquake, Flood, Volcano, Forest fire
Agriculture	Agriculture ecosystem, Food science, Agricultural policy, Pest control
Forest	Forest cover loss, Forest ecology, Forest biomass waste, Forest canopy
Urban	Heat island, Urban growth, Surface temperature, Development density
Air Quality	Air pollution, Yellow dust (Asian Dust), Exhaust gas, Smog
Water Resource	Water pollution, Water quality, Base flow, Stream flow
Biodiversity	Species Diversity, Terrestrial species, Oceanic species, Habitat destruction
Climate	Greenhouse gas, Global warming, Carbon dioxide, Drought, Typhoon
Health	Carcinogen, Pathogen, Bacteria, Heat stroke, Respiratory disease
Noise	Noise pollution, Soundproofing, Hearing loss, Sleep disturbance
Ocean	Mudflat, Coast, Sea level, Red tide, Marine litter
Soil	Soil pollution, Land degradation, Pesticide, Herbicide
Waste	Hazardous waste, Wastewater, Bodily wastes, Recycling

period. GI domains with a frequency percentage equal to or greater than 10% were Satellite Image (24.1%), Natural Disaster Thematic Map (13.6%), and General-Purpose Map (11.7%) (Figure 2(a)). Among GI domains, the frequency percentage of Satellite Image was the highest, because Natural Disaster Thematic Map, General-purpose Map, and so forth are created by processing satellite images. The rate of occurrence of Natural Disaster Thematic Map was high because interest in natural disasters such as volcanic eruptions and earthquakes has increased in South Korea since the 1990s. In

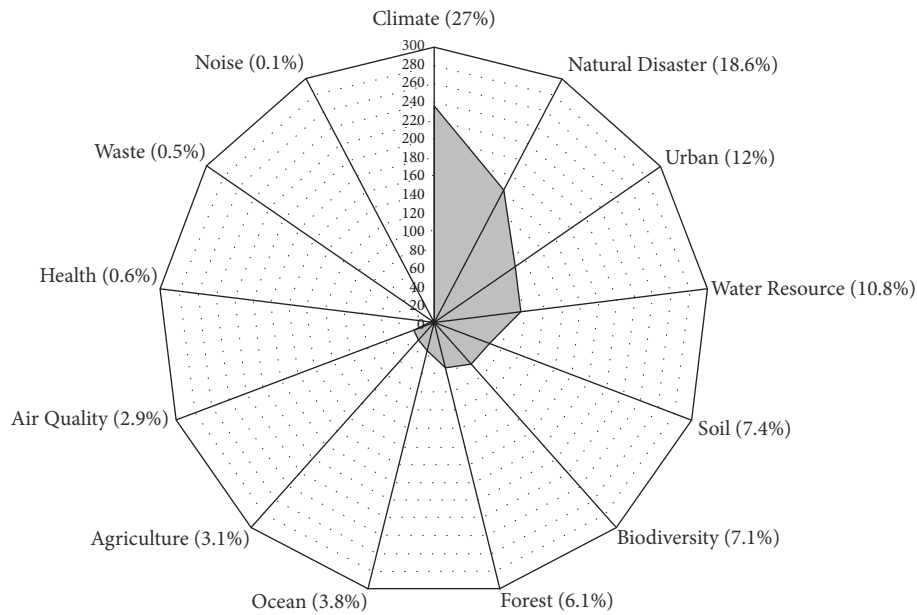
particular, there has been great concern about the volcanic explosion on Mt. Baekdu, located on the Korean Peninsula [29]. Additionally, the frequency percentage of the General-Purpose Map domain was high because it contains base maps for GIS-based spatial analysis tools such as digital maps and digital elevation models (DEMs), each of which is a separate domain (Table 1). The frequency percentage of the Water Resource Thematic Map, Biodiversity Thematic Map, and Forest Thematic Map domains gradually increased in periods 1–3; however, they sharply decreased in periods 3–4.

3.1.2. The Research Field Domain. Keywords related to Research Field were divided into 13 Research Field domains, as listed in Table 2. Figure 3 shows the frequency percentage and time series frequency of GI domains for the entire period. Research Field domains with a frequency percentage equal to or greater than 10% were Climate (27%), Natural Disaster (18.6%), Urban (12%), and Water Resource (10.8%) (Figure 3(a)). The high frequency percentage of the Climate domain was due to increases in damage caused by worldwide Climate change [30]. In the time series analysis, the rate of change of the frequency of domains, excluding Climate and Natural Disaster, was nearly constant (Figure 3(b)). The frequency of the Climate domain showed the sharpest rise in the 2nd and 3rd periods and showed a tendency to decrease in the 3rd period. In the Natural Disaster domain, the rate of change of frequency increased almost constantly over periods 1–4.

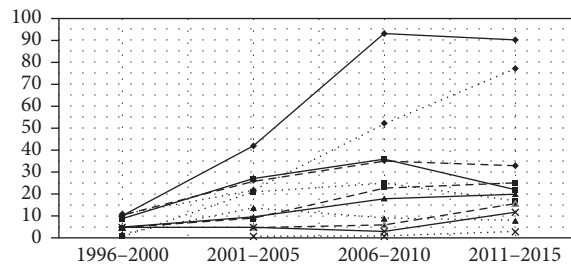
3.1.3. The GI-Research Field Network. Basic statistical analysis has limited application to the quantitative aspect of the frequency of domain. Through a network analysis of simultaneous domains (GI, Research Field), it is possible to compare temporal structural changes in the network and the relative importance of each domain in the network. Table 3 shows the number of links and nodes in the network over time. An increase in the number of links and nodes means that the structure of the network is becoming more complicated. The nodes indicate the domains classified based on the lists in Tables 1 and 2. Links indicate the relationships between the GI and Research Field domains; one link (GI-Research

TABLE 3: Number of links and nodes (1996–2015).

Division	1st (1996–2000)	2nd (2001–2005)	3rd (2006–2010)	4th (2011–2015)
Number of GI nodes (domain)	10	13	12	13
Number of Research Field nodes (domain)	9	11	13	11
Number of links (total)	33 (58)	57 (182)	74 (306)	71 (323)



(a)



(b)

FIGURE 3: Frequency percentage (a) and time-dependent frequency change (b) of the Research Field domain.

Field) was extracted per paper. In Table 3, the number of links is obtained by dividing redundant links in the network. The value displayed in parentheses is the total number of links that are not considered redundant, indicating the number of papers collected during each period. The number of links increased during periods 1–3 and decreased during the 4th period. However, since the number of links that were not duplicated has continuously increased over periods 1–4, we

cannot infer that the network scale was reduced during the 4th period.

Table 4 shows the results of network analysis by time period; for this analysis, we did not distinguish between the GI and Research Field domains, to observe their integrated importance in the GI-Research Field network. The maximum number of nodes (domains) was 26 (13 for each domain) (Tables 1 and 2). We display domains with the top 10 network

index (frequency, degree, closeness centrality, and betweenness centrality) values in Table 4. Italic cells in Table 4 represent the domains in which all network indices are within the top 10 in each period. These domains are relatively important within the GI-Research Field network. Therefore, it is reasonable to analyze the time series of the network around these domains. Two domains were in the top 10 for the entire period (periods 1–4): Satellite Image and Climate. Other domains repeatedly rose and fell in rank according to each period. We conclude that, irrespective of the periods, the research themes that receive steady attention are Satellite Image in the GI domain and Climate in the Research Field domain. These features are consistent with the results of our frequency analysis by period (Figures 2(b) and 3(b)). A high frequency indicates that the domain was frequently used as a research theme. The Satellite Image domain had the smallest variation in rank over time and was top ranking (1st to 3rd place) during periods 1–4. The Climate and Natural Disaster domains ranked higher than the Satellite Image domain during periods 3–4. Thus, the center of the GI-Research Field network moved from GI to Research Field. Such features were also observed in the analysis of degree, closeness centrality, and betweenness centrality indices in periods 1–4. Degree indicates how connected each node is to the surrounding nodes. At least three degree values ranked within the top 10 over the whole period. The domain with the highest degree was Climate, during the 4th period. In particular, the degree value of Climate continued to rise during all four periods (1st period: 5; 2nd and 3rd period: 10; 4th period: 11). This means that the degree value gradually increased with the use of various GI domains for Climate. The Environmental Impact Assessment Map domain ranked in 4th place (degree: 6) in the 1st period; however, it was far from 10th place during periods 2–3 and ranked 7th (degree: 8) in the 4th period. That is, the Environmental Impact Assessment Map domain may be GI that has recently regained attention. Closeness centrality is a measure of centrality in a network, calculated as the sum of the length of the shortest paths between the node and all other nodes in the network. Therefore, in order to assess the overall flow of a network, it is necessary to investigate those nodes with high closeness centrality values. The GI domains whose closeness centrality ranked within 10th place during the entire period were the Satellite Image and General-Purpose Map domains. For this reason, we conclude that these GI domains can be used universally, across all Research Fields. In the Research Field domain, Urban and Soil recently showed a tendency to be far from the center of the network. Urban and Soil ranked in 4th to 8th place during periods 1–3 but did not rank within 10th place in the 4th period. Water Resource was the domain with the largest fluctuation in ranking by period and showed a change of 2–10 places by period. Therefore, studies on water resources based on GI manifest repeated increases and decreases with no trend. The betweenness centrality is a measure of centrality in a network based on the shortest paths. The betweenness centrality for each node is the number of these shortest paths that pass through the node. Therefore, it is reasonable that the interdisciplinary research between nodes (Research Field domains) with different characteristics functions to mediate nodes with high betweenness centrality. In particular, the

Satellite Image, Climate, and General-Purpose Map domains had betweenness centrality values within the top 10 during the entire period. Therefore, it is effective to try to combine Satellite Image and GI within the Climate Research Field. The Urban Thematic Map domain ranked 1st place during the 3rd period but fell to 10th place in the 4th period. Figures 4–7 are a network structure diagram for periods 1–4. In Figures 4–7, red nodes are GI domains, and black nodes are Research Field domains. The size of the circles is an expression of the relative frequency of each node (domain). Dotted lines indicate nodes ranked within the top 10 in all network indices. The network structure diagram, by period, has a structurally simple form in the 1st period and the size of the nodes was relatively small. Over periods 2–4, the structure of the network changed to a more complicated form. In particular, the network structure during the 3rd period was the most complicated because it had the largest number of links throughout the entire period (Table 3). In Figures 6 and 7, the GI domains that were outside the dotted line during the 3rd period moved inside the dotted line during the 4th period. This means that the influence of some GI domains on the network increased and indicates that interdisciplinary research based on GIs will be feasible.

3.2. Subdivision Classification Analysis. This section presents detailed analysis results obtained by subdividing the GI domain with the highest frequency (Satellite Image). The Satellite Image domain was subdivided according to the satellite species. When the number of simultaneous occurrences of the subdivided Satellite Image and Research Field domains was less than 2, the set of domains was excluded from analysis. A basic statistical analysis was carried out separately for the entire period and for periods 1–4, and network analysis was conducted for the entire period.

3.2.1. The Satellite Image Domain. The Satellite Image domain was divided into 12 subspecialized domains based on the satellite species. Figure 8 shows the frequency over the entire study period and during each period. The frequency percentage was 10% or greater over the entire period for the satellites LANDSAT (50%), KOMPSAT (20.1%), and MODIS (10.9%) (Figure 8(a)); the frequency percentage of LANDSAT was 50%, which has a high relevance to the year when LANDSAT was launched. A total of eight LANDSAT satellites were launched by 2017 [31]. LANDSAT-1 was launched in 1972. The KOMPSAT and MODIS satellites were ranked in 2nd and 3rd place and were first launched in 1999 [32, 33]. In the time series analysis, the frequency of LANDSAT was found to be high overall (Figure 8(b)). The frequency of LANDSAT sharply increased during periods 1–2, and the rate of increase decreased during periods 2–3. During the 3rd and 4th periods, its frequency decreased sharply, possibly due to an increase in the use of KOMPSAT, which was developed in South Korea, because the papers that we collected for this study were published in South Korea [34]. A total of five KOMPSAT satellites were launched by 2017. KOMPSAT comprises four optical satellites (K-1, K-2, K-3, and K-3 A) and 1 Synthetic Aperture Radar (SAR) satellite (K-5). KOMPSAT satellites were launched between 1999 and 2015. The frequency of KOMPSAT increased steadily over periods 1–4

TABLE 4: Results of GI-Research Field network analysis (periods 1–4). Italic cells represent the domains in which all network indices are within the top 10 in each period.

Stage	Rank	Frequency	Degree	Closeness centrality	Betweenness centrality		
1 (1996~2000)	1	<i>Satellite Image</i>	14	<i>Satellite Image</i>	7	<i>Satellite Image</i> 0.032	<i>Satellite Image</i> 100.68
	2	<i>Natural Disaster</i>	11	<i>Natural Disaster</i>	6	<i>Natural Disaster</i> 0.029	<i>Environmental Impact Assessment Map</i> 72.92
	3	<i>Climate</i>	10	<i>Water Resource</i>	6	<i>Water Resource</i> 0.029	<i>Natural Disaster</i> 51.9
	4	<i>Urban</i>	10	<i>Environmental Impact Assessment Map</i>	6	<i>Environmental Impact Assessment Map</i> 0.029	<i>Climate</i> 48.29
	5	<i>Water Resource</i>	9	<i>Climate</i>	5	<i>Climate</i> 0.028	<i>Water Resource</i> 45.28
	6	<i>Urban Thematic Map</i>	9	<i>Urban</i>	5	<i>Urban</i> 0.028	<i>Urban</i> 44.4
	7	<i>Natural Disaster Thematic Map</i>	8	<i>General-Purpose Map</i>	4	<i>General-Purpose Map</i> 0.026	<i>Biodiversity</i> 35.63
	8	<i>Environmental Impact Assessment Map</i>	6	<i>Natural Disaster Thematic Map</i>	4	<i>Natural Disaster Thematic Map</i> 0.024	<i>General-Purpose Map</i> 23.85
	9	<i>General-Purpose Map</i>	6	<i>Biodiversity</i>	3	<i>Biodiversity</i> 0.023	<i>Natural Disaster Thematic Map</i> 15.24
	10	<i>Forest</i>	5	<i>Air Quality</i>	3	<i>Urban Thematic Map</i> 0.023	<i>Air Quality</i> 12.75
2 (2001~2005)	1	<i>Satellite Image</i>	47	<i>Satellite Image</i>	10	<i>Satellite Image</i> 0.027	<i>Satellite Image</i> 120.77
	2	<i>Climate</i>	42	<i>Climate</i>	10	<i>Climate</i> 0.026	<i>Climate</i> 105.91
	3	<i>Water Resource</i>	27	<i>General-Purpose Map</i>	8	<i>General-Purpose Map</i> 0.023	<i>Atmosphere Thematic Map</i> 69.36
	4	<i>Urban</i>	26	<i>Soil</i>	8	<i>Soil</i> 0.022	<i>Soil</i> 65.66
	5	<i>Natural Disaster Thematic Map</i>	23	<i>Natural Disaster</i>	7	<i>Natural Disaster</i> 0.022	<i>General-Purpose Map</i> 50.74
	6	<i>Natural Disaster</i>	22	<i>Urban</i>	7	<i>Natural Disaster Thematic Map</i> 0.022	<i>Agriculture</i> 50.23
	7	<i>Soil</i>	21	<i>Natural Disaster Thematic Map</i>	6	<i>Atmosphere Thematic Map</i> 0.021	<i>Natural Disaster</i> 45.67
	8	<i>Water Resource Thematic Map</i>	20	<i>Atmosphere Thematic Map</i>	6	<i>Urban</i> 0.021	<i>Natural Disaster Thematic Map</i> 35.93
	9	<i>General-Purpose Map</i>	19	<i>Water Resource</i>	5	<i>Agriculture</i> 0.020	<i>Urban</i> 33.84
	10	<i>Environmental Impact Assessment Map</i>	18	<i>Forest</i>	5	<i>Water Resource</i> 0.020	<i>Ocean</i> 17
3 (2006~2010)	1	<i>Climate</i>	93.0	<i>Climate</i>	10	<i>Climate</i> 0.025	<i>Urban Thematic Map</i> 109.99
	2	<i>Satellite Image</i>	73.0	<i>Satellite Image</i>	9	<i>Natural Disaster</i> 0.024	<i>General-Purpose Map</i> 63.73
	3	<i>Natural Disaster</i>	52.0	<i>Natural Disaster</i>	9	<i>Water Resource</i> 0.024	<i>Climate</i> 63.19
	4	<i>Water Resource Thematic Map</i>	39.0	<i>Water Resource</i>	9	<i>Satellite Image</i> 0.023	<i>Satellite Image</i> 54.21
	5	<i>Natural Disaster Thematic Map</i>	37.0	<i>General-Purpose Map</i>	9	<i>General-Purpose Map</i> 0.023	<i>Natural Disaster</i> 51.15
	6	<i>Water Resource</i>	36.0	<i>Urban</i>	8	<i>Urban</i> 0.023	<i>Water Resource</i> 47.83
	7	<i>Urban</i>	35.0	<i>Soil</i>	8	<i>Urban Thematic Map</i> 0.021	<i>Ocean</i> 47.38
	8	<i>General-Purpose Map</i>	32.0	<i>Urban Thematic Map</i>	8	<i>Aerial Image</i> 0.021	<i>Urban</i> 46.38
	9	<i>Soil</i>	25.0	<i>Water Resource Thematic Map</i>	7	<i>Soil</i> 0.021	<i>Aerial Image</i> 30.87
	10	<i>Atmosphere Thematic Map</i>	25.0	<i>Natural Disaster Thematic Map</i>	7	<i>Water Resource Thematic Map</i> 0.020	<i>Water Resource Thematic Map</i> 26.92

TABLE 4: Continued.

Stage	Rank	Frequency	Degree	Closeness centrality	Betweenness centrality				
4 (2011~2015)	1	<i>Climate</i>	90	<i>Climate</i>	11	<i>Climate</i>	0.027	<i>Climate</i>	91.07
	2	Natural Disaster	77	<i>Satellite Image</i>	10	<i>Satellite Image</i>	0.027	<i>Satellite Image</i>	64.6
	3	<i>Satellite Image</i>	75	Natural Disaster	8	Natural Disaster Thematic Map	0.024	Forest	59.87
	4	Natural Disaster Thematic Map	50	Natural Disaster Thematic Map	8	<i>Atmosphere Thematic Map</i>	0.024	Agriculture	52.88
	5	<i>General-Purpose Map</i>	45	<i>General-Purpose Map</i>	8	Natural Disaster	0.023	<i>Atmosphere Thematic Map</i>	44.9
	6	<i>Atmosphere Thematic Map</i>	42	<i>Atmosphere Thematic Map</i>	8	<i>General-Purpose Map</i>	0.023	<i>Environmental Impact Assessment Map</i>	38.63
	7	Urban	33	<i>Environmental Impact Assessment Map</i>	8	<i>Environmental Impact Assessment Map</i>	0.023	<i>General-Purpose Map</i>	33.8
	8	Urban Thematic Map	28	Water Resource	8	Water Resource	0.023	Natural Disaster Thematic Map	32.03
	9	<i>Environmental Impact Assessment Map</i>	25	Forest	8	Forest	0.023	Biodiversity	26.95
	10	Biodiversity	25	Biodiversity	7	Biodiversity	0.022	Urban Thematic Map	26.51

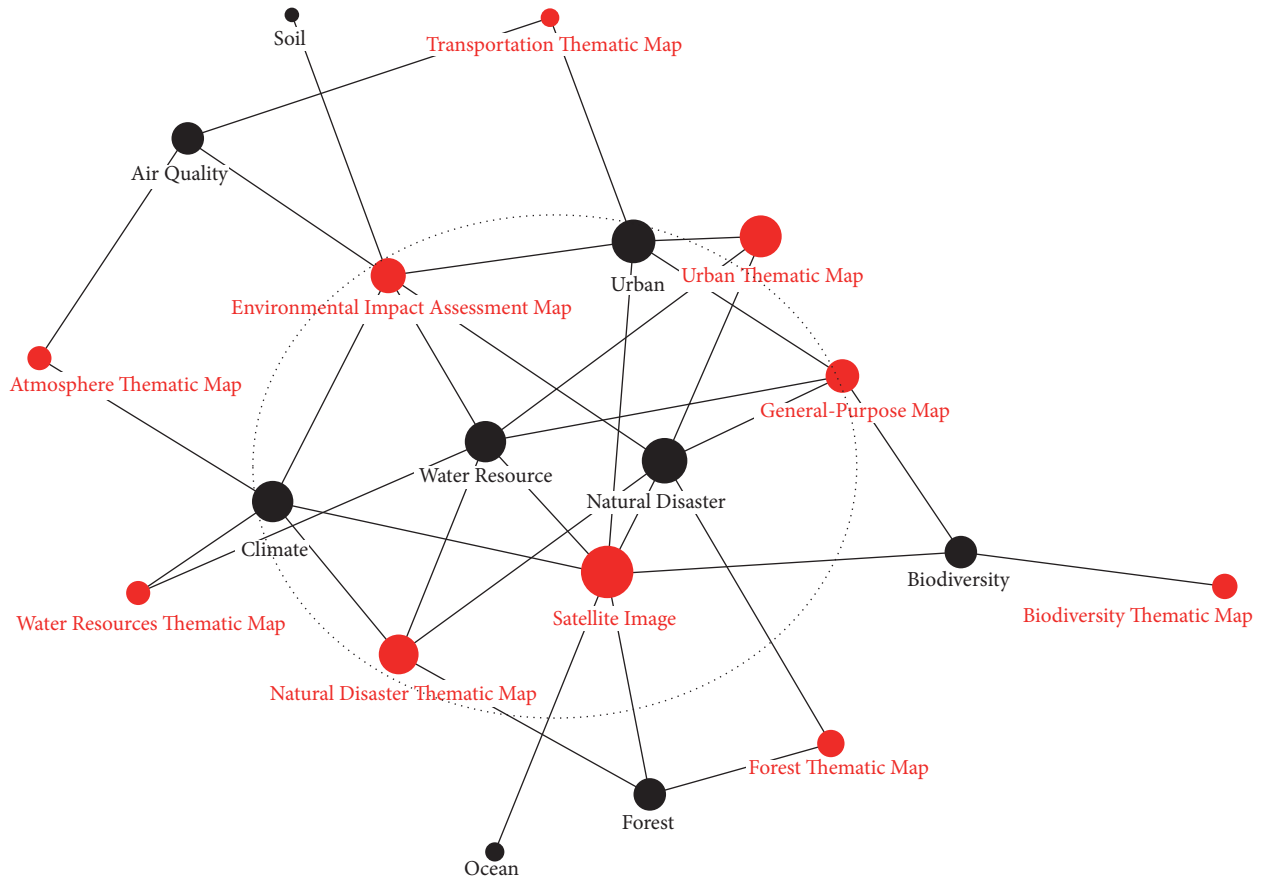


FIGURE 4: GI-Research Field network, 1st period (1996–2000). The red nodes are GI domains, and black nodes are Research Field domains.

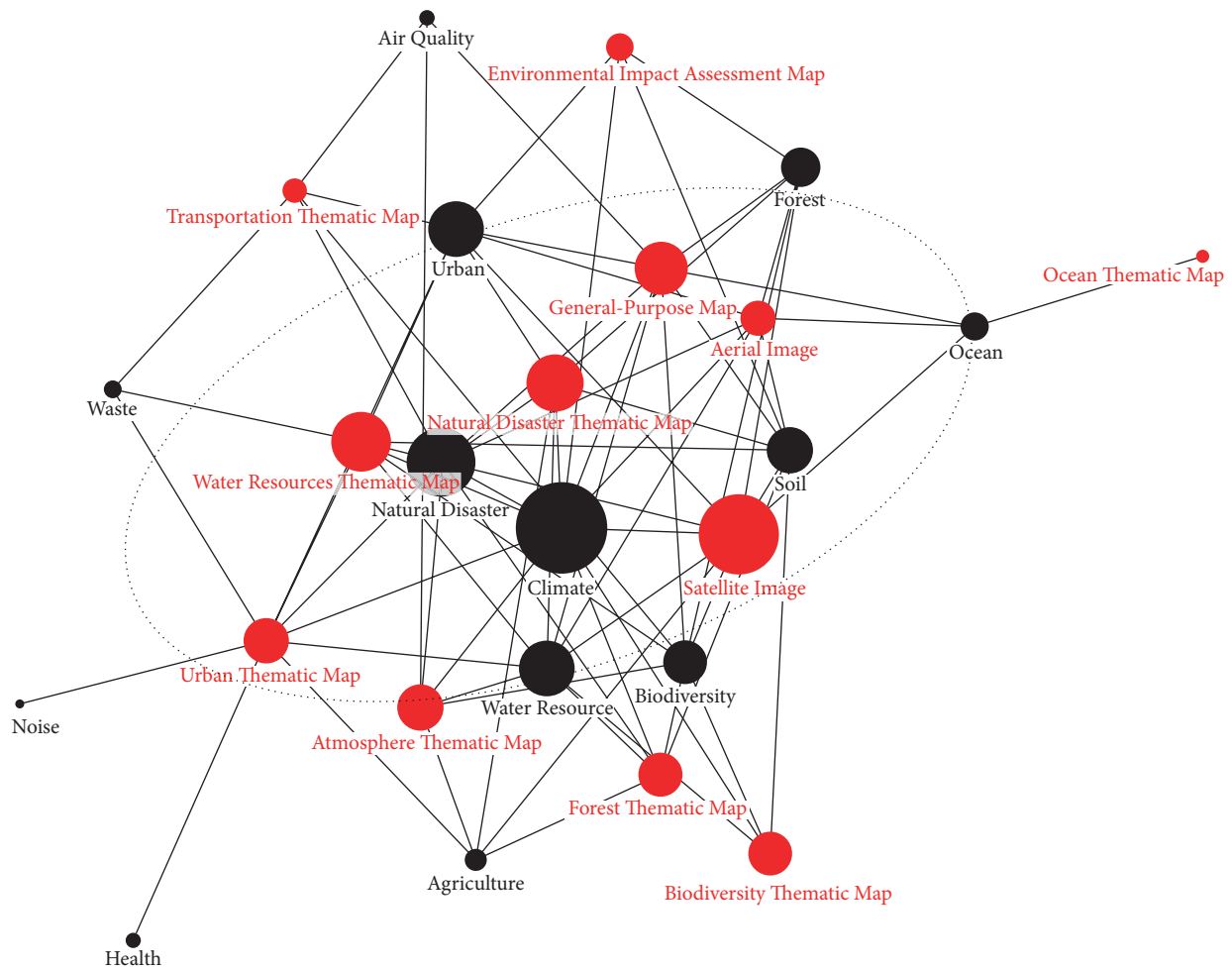


FIGURE 5: GI-Research Field network, 2nd period (2001–2005). The red nodes are GI domains, and black nodes are Research Field domains.

and was highest within the 4th period. When considering the collection period (1996–2015) of the papers and the launch year (2015) of K-3 and K-3A, we expect that the frequency of KOMPSAT will continue to increase in the future. The frequency of MODIS was lower than that of LANDSAT and KOMPSAT but continued to increase over periods 1–4. Next, the frequency of GEOKOMPSAT, which was launched in 2010 as the first Korean multifunction geostationary satellite, was high. GEOKOMPSAT-2A and GEOKOMPSAT-2 B will be launched from 2018 to 2019 [35]. Therefore, we expect the frequency of GEOKOMPSAT to continue to increase. Eight satellites, including SPOT, were used as Satellite Image domains, but not at high rates, due to the difficulty and cost of obtaining their data. The privatization of LANDSAT has had the effect of gradually decreasing its use as a keyword for the following reasons. First, the frequency of utilization of high-resolution optical satellites is increasing. The spatial resolution of LANDSAT-8 is 30 m (panchromatic band), whereas the spatial resolution of the KOMPSAT-2 is 1 m (panchromatic band). Medium-, low-, and high-resolution satellite images have advantages and disadvantages depending on the purpose of use. Therefore, it is difficult to

judge whether high-resolution satellite images (KOMPSAT) are replacing medium- or low-resolution satellite images (LANDSAT). However, it is reasonable to infer that Research Fields requiring high-resolution satellite imagery are increasing in number. Second, the demands for satellite images captured by various sensor types are increasing. Earth observation satellites include optical satellites (LANDSAT, K-2), SAR satellites (K-5), and high-spectral resolution satellites (MODIS, GEOKOMPSAT). The frequency of use of optical satellites (LANDSAT) was very high during periods 1–3. However, the frequency of the SAR satellite (K-5) and high-spectral resolution satellites (MODIS, GEOKOMPSAT) gradually increased during the 4th period.

3.2.2. Research Field Based on Satellite Image. Research Field based on Satellite Image was divided into nine domains. Figure 9 shows the frequency percentage for the entire study period and for each period. The domains whose frequency percentage was over 10% for the entire period were Climate (29.3%), Natural Disaster (16.7%), Forest (10.9%), and Water Resource (10.3%) (Figure 9(a)). Overall, Research Field domains with a wide spatial extent such as Climate

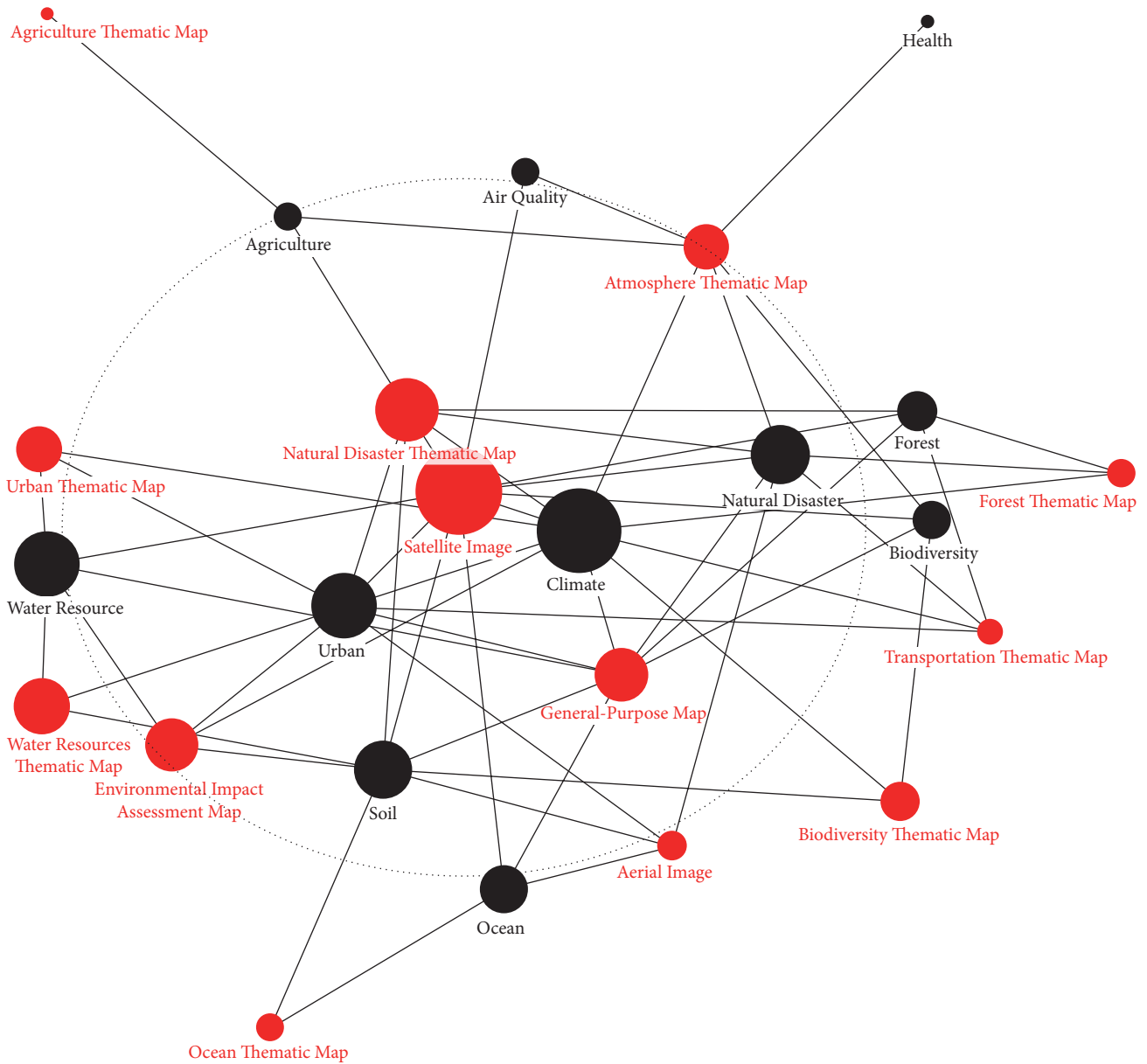


FIGURE 6: GI-Research Field network, 3rd period (2006–2010). The red nodes are GI domains, and black nodes are Research Field domains.

and Natural Disaster ranked highly, whereas Research Field domains with a narrow spatial range, such as Urban and Agriculture, were of low ranking. Figure 9(b) shows the time series frequency for periods 1–4. The frequency of the Climate domain was the highest over the entire period and rose continuously during periods 1–4. The frequency of Natural Disaster was also high and increased continuously during periods 1–4. Overall, the frequencies of the Climate, Natural Disaster, and Forest domains were high and continually increased. Conversely, the Urban, Water Resource, and Soil domains showed a tendency to decrease. These results may be due to differences in spatial extent and indicate the importance of time series change analysis, because Satellite Image is advantageous for analysis over a wide area and can detect time series changes in the region of interest.

3.2.3. The Satellite Image-Research Field Network. Table 5 and Figure 10 show the results of the Satellite Image-Research Field network analysis. We performed this analysis over the entire period because the structure of the entire network was relatively simple, such that time series analysis would hold no significant meaning. In the Satellite Image-Research Field network, 12 Satellite Image domains and nine Research Field domains were connected. The number of links connected by the Satellite Image-Research Field network was 34 in total. The top 10 domains for each index are shown in Table 5.

The results of a detailed analysis of each network index follow. The frequency of the Urban and Ocean domains was high, whereas the other network indices ranked below 10th place. These domains have relatively high frequencies but very little interaction with various other domains. The

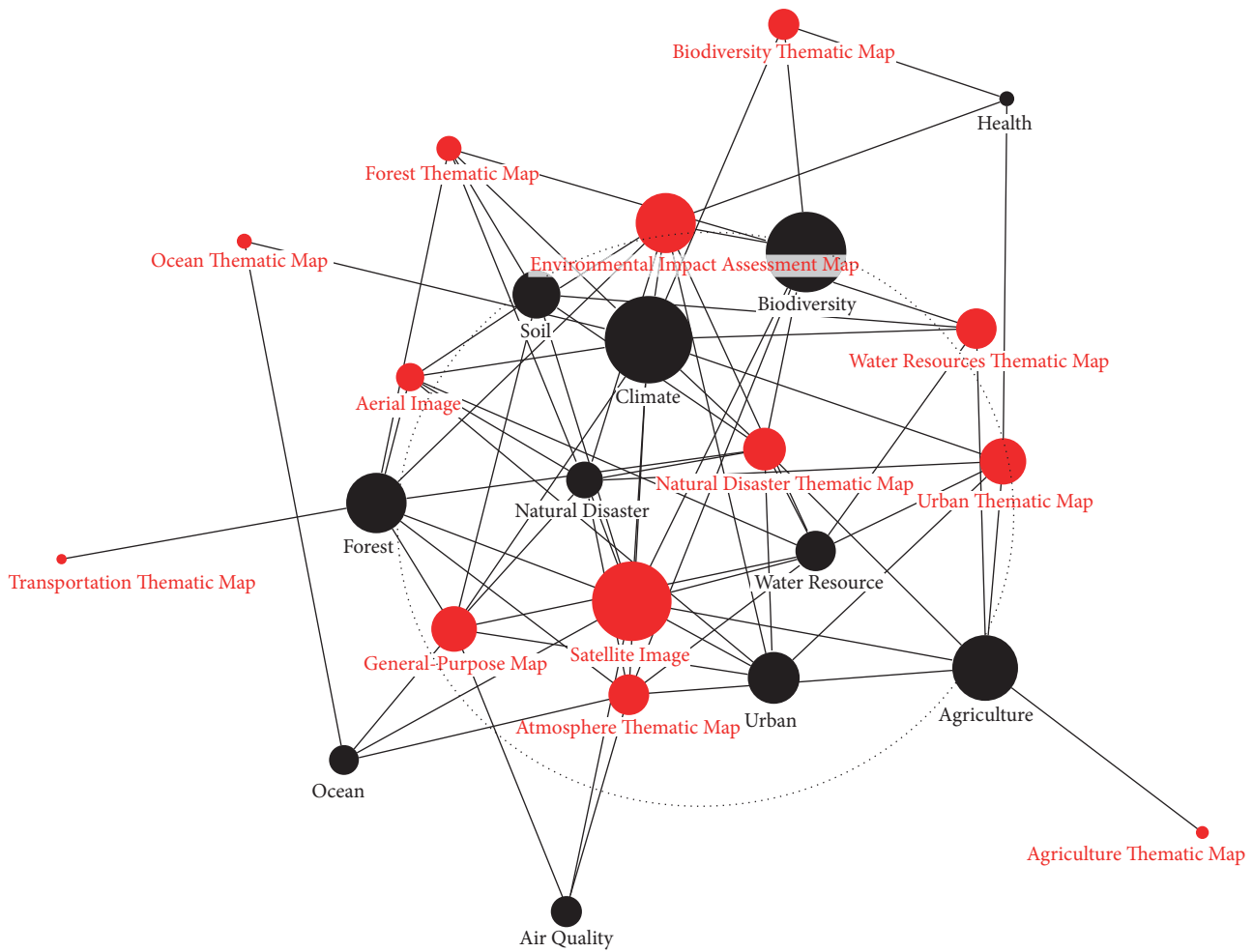


FIGURE 7: GI-Research Field network, 4th period (2011–2015). The red nodes are GI domains, and black nodes are Research Field domains.

TABLE 5: Results of network analysis (1996–2015). Italic cells represent the domains in which all network indices are within the top 10 in each period.

Rank	Frequency		Degree		Closeness centrality		Betweenness centrality	
1	<i>LANDSAT</i>	87	<i>LANDSAT</i>	9	<i>LANDSAT</i>	0.032	<i>LANDSAT</i>	140.804
2	<i>Climate</i>	51	<i>Natural Disaster</i>	8	<i>KOMPSAT</i>	0.029	<i>Natural Disaster</i>	136.670
3	<i>KOMPSAT</i>	35	<i>KOMPSAT</i>	7	<i>Natural Disaster</i>	0.028	<i>Climate</i>	86.810
4	<i>Natural Disaster</i>	29	<i>Climate</i>	6	<i>Climate</i>	0.025	<i>KOMPSAT</i>	72.887
5	<i>Forest</i>	19	<i>Forest</i>	5	<i>MODIS</i>	0.024	<i>Forest</i>	43.886
6	<i>MODIS</i>	19	<i>MODIS</i>	5	<i>Forest</i>	0.024	Agriculture	38.758
7	Water Resource	18	Soil	3	Soil	0.022	<i>MODIS</i>	30.487
8	Urban	17	Biodiversity	3	Biodiversity	0.022	<i>GEOKOMPSAT</i>	15.540
9	Ocean	12	Agriculture	3	Agriculture	0.022	Water Resource	8.645
10	<i>GEOKOMPSAT</i>	12	<i>GEOKOMPSAT</i>	3	<i>GEOKOMPSAT</i>	0.021	SPOT	4.032

frequencies of the Natural Disaster and Climate domains were 29 and 51, respectively, and the degrees were 8 and 6. In other words, in the field of Natural Disaster, we infer that many pilot studies are attempted, using various satellite images. To confirm the latest research trend using Satellite Image, an analysis of the Natural Disaster domain would be appropriate. Generally, closeness centrality will be high if the

degree is high; these were approximately 80% coincident in the domains within the top 10 in the present study. The domains with high closeness centrality were in the center of the network, such that they easily connect with the entire domain. Scholars studying satellite images for the first time need to search for related papers with high closeness centrality values and investigate the overall flow of the Satellite

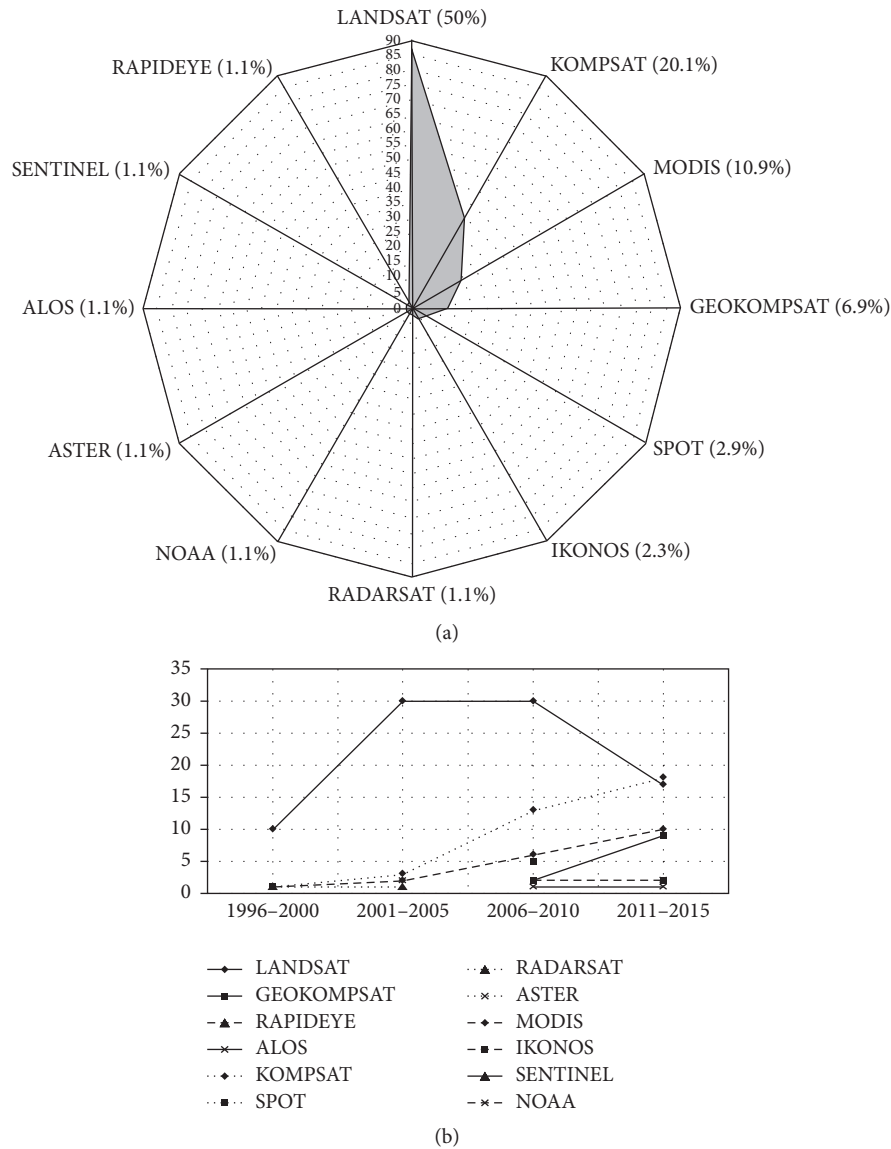


FIGURE 8: Frequency percentage (a) and time-dependent frequency change (b) of the Satellite Image domain.

Image application field. A domain with a high betweenness centrality value has connections with other research themes. Therefore, when trying to fuse different Research Fields, it is reasonable to choose a domain with high betweenness centrality. Figure 10 shows the Satellite Image-Research Field network. When examining the node inside the dotted circle, all Research Field domains are included; however, the Satellite Image domain comprised 41.6% of the total, indicating that the privatization of several satellites such as LANDSAT and KOMPSAT had a substantial effect. Satellite Image domains outside the dotted circle indicate pilot studies in some Research Fields.

4. Discussion and Conclusion

Recently, the GI field has grown in both quantity and quality. To increase the value of GI and apply it in various Research

Fields, it is important to establish research trends. In this study, we analyzed GI research trends using various network indices. We extracted domain pairs (GI, Research Field) from GI-related papers and performed frequency analysis, time series analysis, and centrality analysis on these pairs. We also conducted major classification and subdivision classification analyses. Subdivision classification analysis was performed by subdividing the representative GI domain, calculated from the major classification.

A total of 869 papers were collected from KCI DB, and one set of domains (GI, Research Field) was extracted from each paper. As a result of a frequency analysis of periods 1-4, only a few domains, such as Climate and Satellite Image, showed a sustained increase. As a result of the GI-Research Field network analysis, the Climate, Satellite Image, Natural Disaster, General-Purpose Map, and Natural Disaster Map domains had high-ranking values among all network

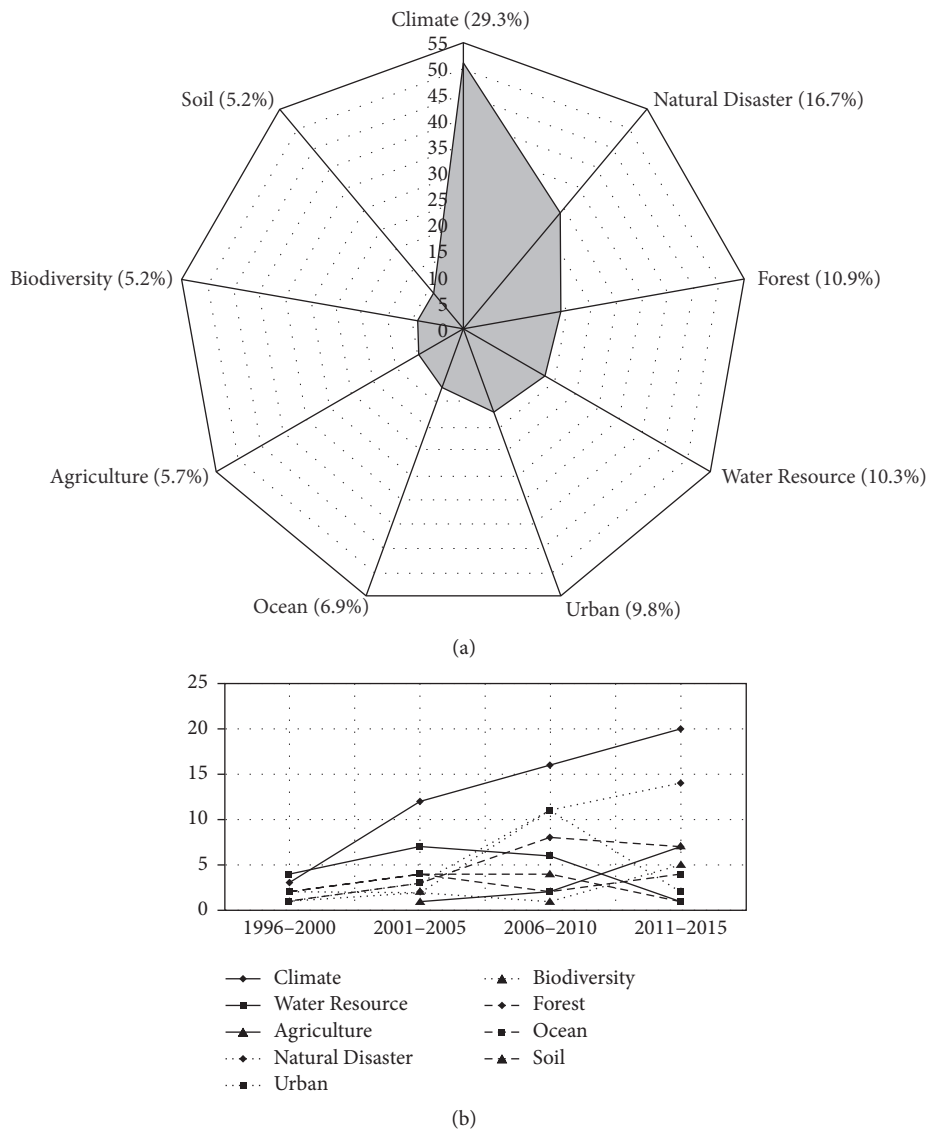


FIGURE 9: Frequency percentage (a) and time-dependent frequency change (b) of the Research Field domain, based on Satellite Image.

indices during the entire study period. An analysis of detailed research trends in the Satellite Image domain showed that the LANDSAT, KOMPSAT, MODIS, Natural Disaster, Climate, and Forest domains had high-ranking values among all network indices.

In the major classification analysis, we found that the Climate domain moved to the middle of the GI-Research Field network over time. The network indices of the Climate domain continued to increase throughout periods 1-4 and, in the 4th period, it had the highest ranking among all network indices. The high values of all the network indices indicate that its accessibility to other domains is at a peak. In other words, it is effective to focus on Climate in the interdisciplinary research mediated by GI. GI domains consistently occupying the top ranks over the entire period were Satellite Image and General-Purpose Map. The Satellite Image and General-Purpose Map domains are most common types

of data used in spatial analysis; therefore, this result was expected. However, to maintain continued growth in the value of the GI field, it is necessary to further strengthen the versatility of GI. Thus, we were encouraged to see that the Environmental Impact Assessment Map and Atmosphere Thematic Map domains moved from low to high rankings during periods 3-4.

Our subdivision classification analysis of Satellite Image showed that the privatization of several satellites on the network had a significant effect. In particular, LANDSAT displayed much higher values in all network indices. However, this phenomenon has been decreasing over time. In the time series frequency analysis, LANDSAT showed a sharp decline during periods 3-4, and KOMPSAT, MODIS, and GEOKOMPSAT showed a sustained increase over periods 1-4 (Figure 8(b)). To expand interdisciplinary research, it seems reasonable to try to center on the Satellite Image

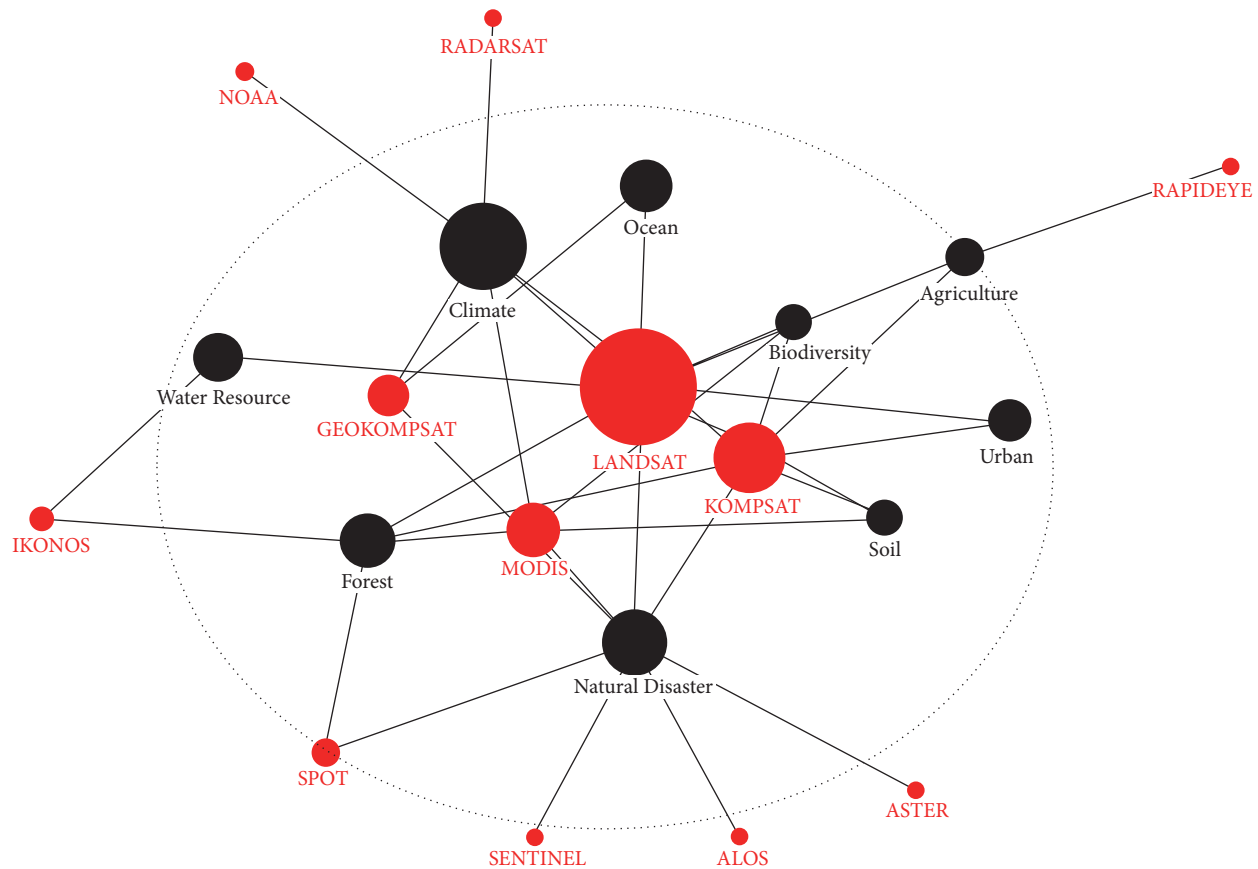


FIGURE 10: Satellite Image-Research Field network (1996–2015). The red nodes are Satellite Image domains, and black nodes are Research Field domains.

domain, which has a high betweenness centrality value. The convergence of Climate and Natural Disaster by mediating LANDSAT and KOMPSAT with high betweenness centrality would be an efficient method. It is necessary to derive new research topics through the combination of Satellite Image domains. It is reasonable that the consilience of Satellite Image domains shows an increased focus on Research Field domains with high betweenness centrality and closeness centrality values. For example, if researchers study natural disasters by combining KOMPSAT and ALOS, new research results may be derived.

In this research, we analyzed GI research trends using the text-mining method, with the following limitations. First, since this study examined only papers indexed by KCI, there is a limit to generalization of results to larger GI research trends. Second, we analyzed GI research trends using network indices (frequency, degree, closeness centrality, and betweenness centrality). However, for a more diverse analysis, more indicators should be analyzed. Third, we classified the keywords extracted from the papers into 26 domains to perform major classification analysis, resulting in a rather simplified analysis. Subsequent studies should overcome these limits to produce results that can be more widely generalized among Research Field and GI domains.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was conducted at Korea Environment Institute (KEI) with support from the Technology Advancement Research Program (TARP) funded by the Ministry of Land, Infrastructure and Transport of the Korean government via Grant 16CTAP-C114629-01.

References

- [1] P. Folger, *Geospatial Information and Geographic Information Systems (GIS): Current Issues and Future Challenges*, DIANE, 2010.
- [2] S. Fotheringham and P. Rogerson, Eds., *Spatial Analysis and GIS*, CRC Press, 2013.
- [3] K.-Y. Oh, H.-S. Jung, and K.-J. Lee, "Comparison of image fusion methods to merge KOMPSAT-2 panchromatic and multispectral images," *Korean Journal of Remote Sensing*, vol. 28, no. 1, pp. 39–54, 2012.

- [4] S.-W. Lee, A.-R. Song, and N.-W. Park, "Environmental impact assessment of nuclear power plant accident using spatial information modeling: a case study of chernobyl," *Korean Journal of Remote Sensing*, vol. 28, no. 1, pp. 129–143, 2012.
- [5] S.-Y. Cha, U.-H. Pi, and C.-H. Park, "Mapping and estimating forest carbon absorption using time-series MODIS imagery in South Korea," *Korean Journal of Remote Sensing*, vol. 29, no. 5, pp. 517–525, 2013.
- [6] N.-W. Park, "Geostatistical downscaling of coarse scale remote sensing data and integration with precise observation data for generation of fine scale thematic information," *Korean Journal of Remote Sensing*, vol. 29, no. 1, pp. 69–79, 2013.
- [7] F. G. Bonham-Carter, *Geographic Information Systems for Geoscientists: Modelling with GIS*, vol. 13, Elsevier, 2014.
- [8] B. Jaap, V. D. Menno, D. Sander, E. David, M. Rene, and V. O. Erik, "The fourth industrial revolution," VINT Research Report 3, SOGETI, 2014.
- [9] R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, Cambridge, UK, 2006.
- [10] M. W. Berry, "Survey of text mining," *Computing Reviews*, vol. 45, no. 9, pp. 548–554, 2004.
- [11] M. A. van Driel, J. Bruggeman, G. Vriend, H. G. Brunner, and J. A. M. Leunissen, "A text-mining analysis of the human genome," *European Journal of Human Genetics*, vol. 14, no. 5, pp. 535–542, 2006.
- [12] J. Y. Lee, H. Kim, and P. J. Kim, "Domain analysis with text mining: analysis of digital library research trends using profiling methods," *Journal of Information Science*, vol. 36, no. 2, pp. 144–161, 2010.
- [13] M. Krallinger, F. Leitner, and A. Valencia, "Analysis of biological processes and diseases using text mining approaches," *Bioinformatics Methods in Clinical Research*, vol. 593, pp. 341–382, 2010.
- [14] D. Delen and M. D. Crossland, "Seeding the survey and analysis of research literature with text mining," *Expert Systems with Applications*, vol. 34, no. 3, pp. 1707–1720, 2008.
- [15] 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, no. 1, pp. 1–17, 2012.
- [16] W. He, S. Zha, and L. Li, "Social media competitive analysis and text mining: a case study in the pizza industry," *International Journal of Information Management*, vol. 33, no. 3, pp. 464–472, 2013.
- [17] S. Y. Lim, M. S. Yi, G. H. Jin, and D. B. Shin, "A study on the research trends in the area of geospatial-information using text-mining technique focused on national R&D reports and theses," *Journal of Korea Spatial Information Society*, vol. 22, no. 4, pp. 11–20, 2014.
- [18] M. Callon, J. P. Courtial, and F. Laville, "Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry," *Scientometrics*, vol. 22, no. 1, pp. 155–205, 1991.
- [19] Q. He, "Knowledge discovery through co-word analysis," *Library Trends*, vol. 48, no. 1, p. 133, 1999.
- [20] B. Yoon and Y. Park, "A text-mining-based patent network: analytical tool for high-technology trend," *Journal of High Technology Management Research*, vol. 15, no. 1, pp. 37–50, 2004.
- [21] M. Rokaya, E. Atlam, M. Fuketa, T. C. Dorji, and J.-I. Aoe, "Ranking of field association terms using co-word analysis," *Information Processing and Management*, vol. 44, no. 2, pp. 738–755, 2008.
- [22] G. A. Ronda-Pupo and L. A. Guerras-Martin, "Dynamics of the evolution of the strategy concept 1962–2008: A co-word analysis," *Strategic Management Journal*, vol. 33, no. 2, pp. 162–188, 2012.
- [23] Y. Kajikawa, J. Yoshikawa, Y. Takeda, and K. Matsushima, "Tracking emerging technologies in energy research: toward a roadmap for sustainable energy," *Technological Forecasting and Social Change*, vol. 75, no. 6, pp. 771–782, 2008.
- [24] Y. Kajikawa and Y. Takeda, "Citation network analysis of organic LEDs," *Technological Forecasting and Social Change*, vol. 76, no. 8, pp. 1115–1123, 2009.
- [25] S. Choi, J. Yoon, K. Kim, J. Y. Lee, and C.-H. Kim, "SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells," *Scientometrics*, vol. 88, no. 3, pp. 863–883, 2011.
- [26] M. S. Lee, C. H. Lee, and J. Y. Kim, *Big Data Analysis on Demands for Environmental Policies*, Korea Environmental Institute, 2014.
- [27] Korea National Spatial Data Infrastructure Portal, <http://www.nsd.go.kr>.
- [28] NodeXL, <https://www.microsoft.com/en-us/research/project/nodexl-network-overview-discovery-and-exploration-in-excel/>.
- [29] A. Paone and S. H. Yun, "Pyroclastic density current hazards at the baekdusan volcano, Korea: analyses of several scenarios from a small-case to the worst-case colossal eruption," in *Updates in Volcanology-From Volcano Modelling to Volcano Geology*, InTech, 2016.
- [30] Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2014: Impacts, Adaptation, and Vulnerability*, Cambridge University Press, Cambridge, UK, 2014.
- [31] D. P. Roy, M. A. Wulder, T. R. Loveland et al., "Landsat-8: science and product vision for terrestrial global change research," *Remote Sensing of Environment*, vol. 145, pp. 154–172, 2014.
- [32] L. Hoonyol and H. Hyangsun, "Evaluation of SSM/I and AMSR-E sea ice concentrations in the antarctic spring using KOMPSAT-1 EOC images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 1905–1912, 2008.
- [33] Terra (MODIS), <https://terra.nasa.gov/about/terra-instruments/modis>.
- [34] EESA Earth Online (KOMPSAT), <https://earth.esa.int/web/guest/missions/3rd-party-missions/current-missions/kompsat-2>.
- [35] J. Kim, M. Kim, M. Choi et al., "Monitoring atmospheric composition by GEO-KOMPSAT-1 and 2: GOCI, MI and GEMS," in *Proceedings of the 36th IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2016*, pp. 4084–4086, chn, July 2016.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

