

텍스트마이닝 기술을 이용한 공간정보 분야의 연구 동향에 관한 고찰

-국가연구개발사업 보고서 및 논문을 중심으로-

A Study on the Research Trends in the Area of Geospatial-Information Using Text-mining Technique Focused on National R&D Reports and Theses

임시영* · 이미숙** · 진기호*** · 신동빈****

Si Yeong Lim · Mi Sook Yi · Gi Ho Jin · Dong Bin Shin

요약 본 연구의 목적은 텍스트마이닝 기법을 활용하여 공간정보 분야의 연구동향을 파악하는 것이다. 이를 위하여 국가과학기술도서관에서 국가연구개발보고서와 논문을 추출하여 키워드에 대한 전처리를 수행한 후 분야별로 정리하였다. 정리된 키워드들을 통해 보고서 및 논문에서 키워드의 시기별 출현 빈도 및 변화를 살펴보고 이를 통해 공간정보 분야의 연구동향을 확인하였다. 분석결과 공간정보 분야에서는 시스템 관련 연구가 줄어드는 반면 활용 관련 연구가 늘어드는 추세가 있음을 확인하였다.

키워드 : 공간정보, 연구동향, 텍스트마이닝, 활용, 시스템, 데이터

Abstract This study aims to provide information about the research-trends in the area of Geospatial Information using text-mining methods. We derived the National R&D Reports and papers from NDSL(National Discovery for Science Leaders) site. And then we preprocessed their key-words and classified those in separable sectors. We investigated the appearance rates and changes of key-words for R&D reports and papers. As a result, we conformed that the researches concerning applications are increasing, while the researches dealing with systems are decreasing. Especially, with in the framework of the keyword, '3D-GIS', 'sensor' and 'service' xcept ITS are emerging. It could be helpful to investigate research items later.

Keywords : Geospatial Information, Research trends, Text-mining, Application, System, Data

1. 서론

공간정보산업은 창조경제의 기반으로 인식되는 산업이며 지난 시간 꾸준히 연구가 진행되어 온 분야이다. 기존 GIS(지리정보시스템)를 중심으로 발전되어 온 공간정보산업 분야는 최근 활용분야를 넓혀가고 있으며 타 기술들과의 융복합을 통해 더 많은 가치를 창출할 수 있을 것으로 기대된다[1,12]. 공간정보 분야는 기술에서 제도까지 아우르는 종합적 발전이 필요할 뿐 아니라 다양한 분야와 융복합을 통해 더욱 그 파급력을 확산시킬 필요가 있는 분야이다. 그러나 공간정보 분야에는 다양한 세부 전문가들이 많음에 비해 총괄적 관점의 전문가가 부족한 편이다. 융복합을

위해서는 관련 분야에 대한 이해가 중요한데, 이 점 역시 부족하다[10]. 따라서 종합적 관점에서 공간정보의 발전추이, 공간정보의 최신동향을 파악하고 이를 기반으로 기술개발 또는 산업발전을 위한 절차 수립에 노력할 필요가 있다.

특히 최근 빅데이터 기술이 발달됨에 따라, 이전에는 처리할 수 없었던 긴 시간에 걸친 방대한 양의 자료에 대한 분석이 가능해졌고, 이로 인해 학술적 동향을 분석하기에 적합한 기술 환경이 마련되었다. 따라서 본 연구에서는 빅데이터 기술 중 하나인 텍스트마이닝 기술을 활용하여 공간정보 관련 연구동향을 파악하고자 한다. 기존 데이터마이닝 기술이 수치형 및 범주형 자료에 대한 통계학, 데이터베이스 기술, 패턴인

* Si Yeong Lim, Principal Consultant, C&BIS. tome20@naver.com (Primary author)

** Mi Sook Yi, Professor, Dept. of Urban Information Engineering, Anyang University. mslee0414@anyang.ac.kr (Corresponding author)

*** Gi Ho Jin, Doctoral Student, Dept. of Environmental and Urban Engineering, Chungbuk National University. ghjin124@naver.com

**** Dong Bin Shin, Professor, Dept. of Urban Information Engineering, Anyang University. dbshin@anyang.ac.kr

식, 기계학습 등의 적용을 통해 자료를 예측, 분류, 군집화, 연관관계 규명 등을 수행하였던 것과 달리, 텍스트마이닝은 비정형 데이터 및 직관적 텍스트에 대해 어휘구조, 텍스트 전처리 등을 통해 의미를 도출하는 기술이다[4,7].

최근에 산업공학, 건설 등 다양한 분야에서 텍스트마이닝 기법을 활용한 연구가 활발히 진행되고 있다 [2]. Cho et al.[6]는 해외저널인 IIE Transaction에 게재된 논문들의 주제어들을 분석하여 산업공학 연구 분야 동향 파악 및 주제어간에 관계를 분석하였다. Cho et al.[5]는 국내 산업공학 관련 저널들에 게재된 논문들의 주제어들과 초록 내용을 바탕으로 산업공학의 주요 연구기법들의 동향을 파악하였다. Jeong et al.[8]은 인터넷 검색엔진을 이용하여 검색 결과 값을 주요 키워드의 빈도로 사용하여 건설 분야의 미래 트렌드를 분석하였다. Bae et al.[3]는 텍스트마이닝 기법을 이용하여 2000년 초반부터 최근까지 기후변화 관련 논문 중 식품관련 연구 성과물들에 대해 어떤 주제나 키워드가 연구되었는지를 알아보고, 이러한 주제나 키워드의 출현빈도 추세에 대하여 분석하였다.

한편, GIS 연구동향에 관한 연구도 진행된 바 있다. Sakong et al.[11]은 한국의 GIS관련 학회에서 발간한 1,222편의 논문을 대상으로 논문의 제목과 주요어(keyword)를 중심으로 GIS 관련 학술분야의 연구동향을 분석하였다. Kim[9]은 북한의 GIS 관련 연구 논문들을 조사하고 이를 주제별로 분류하여 남한의 GIS 연구동향과 비교하였다. 이러한 연구는 2006년까지 1,222편의 국내논문과 386편의 북한학술지 투고논문을 대상으로 주요어(keyword)를 중심으로 연구동향을 분석한데 비해 본 연구는 4000여개의 논문과 225개의 국가연구개발보고서에 대하여 키워드 및 요약문을 분석한 점에서 선행연구와 차별성이 있다.

본 연구에서는 공간정보 분야의 연구동향을 파악하기 위하여 국가연구개발보고서와 논문들을 분석하였다. 해당 자료에서 제시된 주요어(keyword)를 추출하고, 이에 대한 추세 및 비중을 중심으로 공간정보 분야의 연구동향을 확인하였다. 이를 통해 지난 시간동안 공간정보 분야의 연구에서 주요한 주제가 무엇이며, 국가연구개발보고서와 논문에서는 어떠한 차이점이 있는지를 점검하였다.

2. 연구 방법

2.1 연구 수행 절차

본 연구는 빅데이터를 활용한 데이터 분석 기법 중

비정형 데이터 분석의 하나인 텍스트마이닝 기법을 활용하여 공간정보 분야의 연구동향을 파악하는데 목적이 있다.

먼저 본 연구의 목적에 부합하도록 논문과 보고서를 검색할 수 있는 검색어를 선정하였다. 본 연구의 목적이 공간정보 연구동향을 확인하기 위함이었으므로 1차적으로 ‘공간정보’를 검색어로 하여 논문과 국가연구개발보고서를 수집하였다.

다음으로 수집된 논문에서 키워드를 추출하고, 추출된 키워드에 대해 연구목적에 부합하도록 전처리 과정을 거쳤다. 전처리 과정은 키워드를 조정하고 유사키워드를 정리한 후, 키워드를 특정 유형에 따라 분류하는 과정이다. 분류된 키워드를 중심으로 텍스트마이닝 기법을 사용하여 논문과 보고서의 요약 부분을 분석함으로써 공간정보 연구 분야의 동향을 추출하였다.

2.2 자료의 수집

본 연구에서 사용된 자료는 국가과학기술전자도서관(NDL)에서 ‘공간정보’를 검색어로 하였을 때 추출되는 논문과 국가연구개발보고서이다. 자료의 추출은 R 3.0.2 버전을 사용하였고, 자료의 추출 시기는 2013년 11월 5일이다.

추출 결과 논문은 1996년부터 2013년까지 총 4,000개, 국가연구개발보고서는 1994년부터 2013년까지 총 225개가 추출되었다. 논문의 경우 ‘제목’, ‘저널명’, ‘발행연도’, ‘키워드’를 추출하였으며, 국가연구개발보고서는 ‘제목’, ‘연구책임자’, ‘수행기관’, ‘발행연도’, ‘요약’을 추출하였다. 단, 해당 필드에 내용이 없는 자료의 경우는 삭제하여 분석 대상에서 제외하였다.

한 해에 나오는 자료의 양이 적으므로 분석 기간을 1996년(논문은 1994년)부터 1999년 1기, 2000년부터 2009년 2기, 2010년부터 2013년까지 3기로 구분하여 분석하였다.

2.3 키워드 추출 및 키워드 전처리

4,000개의 논문에서 저자가 제시한 키워드는 총

1) 국내에 GIS가 도입된 이후 지리정보와 공간정보라는 용어가 혼재되어 사용되다가 「국가지리정보체계의 구축 및 활용 등에 관한 법률」이 폐지되고 「국가공간정보에 관한 법률」이 2009년 2월 제정되고 국가공간정보정책 기본계획 및 시행계획 등이 수립·시행됨에 따라 ‘공간정보’라는 용어가 보다 보편적으로 사용되고 있다. 이에 본 연구에서는 최근의 연구동향을 파악하기 위하여 검색어를 ‘공간정보’로 설정하였다.

Table 1. Classifying the key-words into 5 sectors(system, data, application, methods, etc)

| Classification | N | Keyword |
|----------------|----|---|
| System | 23 | 3D GIS, Web GIS, Interoperability, Standardization, Component, XML, Database, GML, CORBA, Metadata, Data Stream, accessibility, Ontology, UML, Open Source, commonality, Spatial Index, maintenance, VRML, Mobile GIS, CBIR, SVG(Scalable Vector Graphics), Internet(WWW) |
| Data | 19 | LiDAR, aerial photograph, Satellite Image, DEM(Digital Elevation Model), Numerical Map, GPS(DGPS, RTK-GPS), RS(Remote Sensing), Digital MAP, land cover, Land registration Map, Change Detection, GCP(Ground Control Point), 3D Visualization, Ortho Imagery, Platform, Real-time, Sensor, Image Fusion |
| Application | 27 | LBS(Location-Based Service), U-City, Ubiquitous, landslide, ITS, BIM, mobile, land use, VMS, Moving Object, Climate Change, Renewable Energy, concrete pavement, Sevice, Augmented Reality, Telematics, Railway Station Sphere, Pedestrian, Personal Information Protection, UIS, smartphone, GNSS, Quality Assurance, Watershed management, SN(Sensor Network), Urban Restoration, Spatial Data Industry |
| Methods | 17 | Saptial Analysis, simulation, cluster analysis, AHP, NDVI, Kriging, genetic algorithm, regression analysis, Filtering, Index of Evaluation, Suitable Land Analysis, Finite Element Analysis, Factor Analysis, Structural Equation, GALIS, Contingent Valuation Model, Neurofuzzy |
| etc. | 9 | OGC, NSDI, NGIS, u-GIS, Google Earth, KOMPSAT-2, Infrastructure, KLIS, UPIS |

17,117개로 나타났다. 이 중 중복된 키워드를 제거한 결과 총 12,535개의 키워드가 나타났다. 여기서 언급된 중복은 동일한 표기로 이루어진 것을 의미한다. 12,535개 중 상위 30개의 키워드는 약 968개로 약 7.8%에 불과하였다.

그러나 위 키워드에는 예를 들어 GIS, 지리정보시스템 또는 Geographic Information System과 같이, 한글 표현 또는 약어 표현임에도 불구하고 표기방법에 따라 다른 키워드로 인식하는 문제가 있었다. 따라서 위 키워드를 직접적으로 분석에 사용할 수 없으므로 이를 위한 전처리 작업을 수행하였다. 공간정보 분야의 전문가 자문을 통해 약어, 한영, 띄어쓰기 등을 고려하여 12,535개의 키워드를 1차 정리하였다. 예를 들면, ‘LBS’, ‘위치기반서비스’, ‘location-based service’는 동일한 키워드로 처리하였다. 이 후 출현빈도가 매우 낮은 키워드들을 삭제하여 2차 정리하고, 총괄적 동향 점검을 위해 키워드들을 3차 집합화 하였다. 키워드들의 집합화는 시스템, 데이터, 활용분야, 기법, 기타로 구분하였고 해당되는 키워드들은 다음 Table 1과 같다.

3. 국가연구개발보고서 자료분석 결과

3.1 현황자료

국가연구개발보고서의 발간 현황을 살펴본 결과는 Figure 1과 같다. 먼저 시기별 발간 횟수를 보면 1999년에 급격히 상승하였고, 2010년에 또 한 번의 상승이

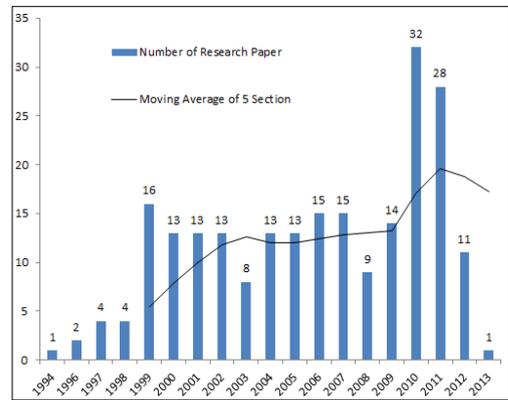


Figure 1. Number of published reports during the aimed period

있었다. 2013년의 경우, 11월에 수집된 데이터므로 발간되지 않은 보고서가 있을 것이라 예상되지만, 2012년의 경우는 11편으로 하강한 것을 알 수 있다.

다음으로 수행 주체에 따라 살펴보면 Figure 2와 같이 ‘연구원’이 39%, ‘학교’가 35%, ‘산업체’가 17%, ‘기타’가 9%로 나타났다. 좀 더 상세히 살펴보면 연구원의 경우 정보통신연구진흥원(22.1%), 한국정보화진흥원(14%), 한국전자통신연구원(12.8%), 한국지질자원연구원(10.5%), 한국건설기술연구원(9.3%) 등 상위 5개 기관이 과반 이상을 수행하였다. 학교의 경우, 연세대(11.4%), 부경대(6.3%), 경원대/경희대/세종대/충남대(5.1%) 등 상위 5개 학교가 38% 정도를 차지했다. 산업체의 경우 (주)우

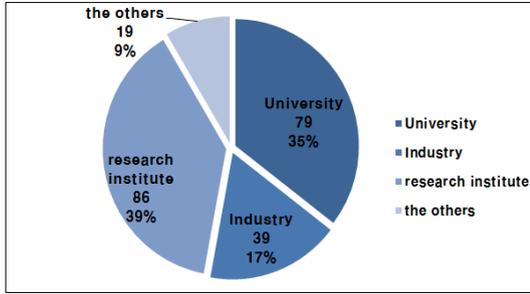


Figure 2. The share rate of report by performer

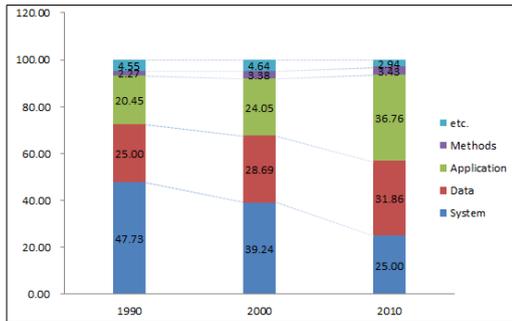


Figure 3. Variation of appearance rate reports' keywords in each sector

대갈스(12.8%), (주)비엔티솔루션/(주)공간정보기술/한국공간정보통신(주)/버추얼빌더스 등 상위 5개 업체가 49% 정도를 차지했다. 기타의 경우 한국지형공간정보학회(26%), 기업정보화지원센터(11%), 대한원격탐사학회(11%), 한국공간정보시스템학회(11%) 등 상위 4개 기관이 과반수 이상을 차지하는 것으로 나타났다.

3.2 주요 키워드 및 기간별 출현율

먼저 국가연구개발보고서에 출현한 주요 키워드와 그 점유율을 살펴보면, 데이터베이스가 전체 국가연구개발보고서 중 12.2% 출현했으며, 서비스(8.7%), 실시간(7.8%), 자연재해(6.8%), 3차원공간정보(5.8%), 표준화(5.8%)가 각각 그 뒤를 이었다.

기간별로 집합화된 키워드들의 출현율을 살펴보면 Figure 3과 같다. 1996년부터 1999년까지는 시스템의 비중이 높았으나 시스템의 비중은 점차 감소하고, 상대적으로 활용과 데이터 관련 키워드가 증가한 것을 확인할 수 있다. 이는 시스템에 대한 일차적 구축이 완료되었고, 보다 정확한 또는 보다 유용한 데이터에 대한 관심이 고조되고, 실제적으로 공간정보의 활용을 위한 수요가 발생하고 있다고 말할 수 있다.

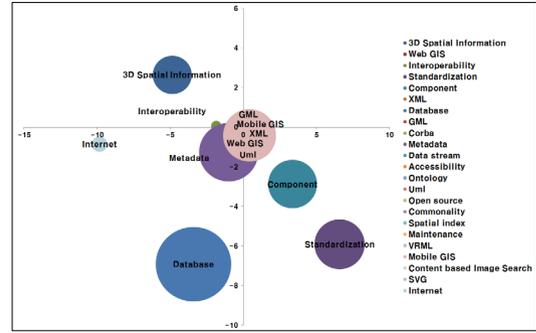


Figure 4. Comparison of keywords appearance rate in system sector during the aimed periods

3.3 시스템분야 상세 분석

시스템 분야의 상위 출현 키워드를 살펴보면 1기(1996년부터 1999년까지)에는 데이터베이스(18.2%), 인터넷(13.6%), 3차원 공간정보(9.1%), 상호운용성(2.3%), 표준화(2.3%), metadata(2.3%)로 나타났으며, 2기(2000년부터 2009년까지)에는 데이터베이스(14.8%), 표준화(6.9%), 3차원 공간정보(4.2%), 인터넷(3.8%), components(3.4%), metadata(1.3%)로 나타났으며, 3기(2010년 이후)에는 데이터베이스(7.8%), 3차원공간정보(6.9%), 표준화(2.9%), 인터넷(2.9%), 오픈소스(1.5%), 시멘틱 웹(1.0%)로 나타났다. 전반적으로 키워드들의 변화가 적은 편이며 키워드들의 출현율이 감소하고 있음을 확인할 수 있으며, 3차원 공간정보만이 2-3기 동안 증가하고 있다.

1-2기와 2-3기 동안의 증가율을 그래프화 하면 다음 Figure 4와 같다. 해당 그래프에서 1사분면에 위치한 키워드는 1-2기, 2-3기 모두 증가한 것이며, 2사분면에 위치한 키워드는 1-2기 증가, 2-3기 감소한 키워드를 의미한다. 3, 4사분면도 동일하게 해석 가능하다. 원의 크기는 각 기간 동안 출현율의 변화율을 의미하며 이는 1-2기의 증가율(감소율)과 2-3기의 증가율(감소율)의 비중에 대한 절대값을 의미하는데, 이 값이 큰 경우는 1-2기에 비해 2-3기의 변화가 상대적으로 크다는 의미이다. 즉 3사분면에 위치한 데이터베이스의 경우 크기도 매우 크므로, 1-2기, 2-3기 모두에 걸쳐 출현율이 급격하게 낮아지는 키워드라고 할 수 있다.

시스템분야에 대해서는 연구 수요가 작아지고 있으며 그 중 데이터베이스(2-3기의 출현율의 변화율이 -6.9로서 1-2기 출현율의 변화율 -3.4에 비해 2배 넘게 증가함)는 급격하게 출현율이 줄어드는 분야라고 할 수 있다.

이러한 현상은 국가연구개발사업에 의해 단순 지리 정보시스템을 구축해오던 시기가 지나고 3차원 공간 정보, 시맨틱 웹 등의 새로운 기술을 적용하는 시기로 전환하고 있음을 말한다. 또한 표준화, 메타데이터, 오픈소스 등의 키워드가 증가하고 있다는 점에서 공간 정보 분야의 시스템은 웹기반, 대용량화의 방향으로 진화하며 통합 또는 상호 운용성 확보라는 측면이 강조되고 있음을 확인할 수 있다. 단, 데이터베이스라는 키워드의 급격한 증가는 데이터베이스라는 키워드 보다 DBMS 등 다른 키워드로 대체하여 사용하는 경우가 증가하고 있기 때문일 수도 있다.

3.4 데이터분야 상세 분석

데이터 분야의 상위 출현 키워드를 살펴보면 1기에는 수치표고모형(9.1%), 실시간(9.1%), GPS (2.3%), 3차원 시각화 (2.3%), Platform(2.3%)로 출현했으며 2기에는 실시간(8.4%), 위성영상(5.5%), 원격탐사(4.2%), 수치표고모형(2.5%), Platform, LiDAR, 센서(각 1.69%)로 출현했으며 3기에는 실시간(6.9%), 위성영상(5.4%), 센서(5.4%), GPS(4.4%), 원격탐사, 정사영상(각 2.0%)로 나타났다.

Figure 5를 보면, 센서의 경우, 1-2기, 2-3기 모두에서 출현율이 높아지며 그 변화 역시 크므로(출현율의 변화율이 2.2) 지속적으로 연구 수요가 발생했다고 볼 수 있다. GPS의 경우, 1-2기에서는 약간 감소했으나, 2-3기에 크게 증가한 것으로 나타나므로 최근 다시 관심을 받는 주제로 볼 수 있다(출현율의 변화율이 2.5). 원격탐사의 경우 1-2기에는 출현율이 크게 증가했으나, 2-3기에는 출현율이 크게 감소하였으며 그 변화폭이 큰 편에 속하므로 일시적 유행에 따른 분야로 파악된다. 수치표고모형의 경우 조금씩 지속적으로 관심

이 떨어지고 있는 분야라고 볼 수 있으며 실시간의 경우 2-3기 동안 크게 감소하는 분야라고 볼 수 있다(출현율의 변화율이 2.4).

전반적으로 데이터 분야의 경우, 센서, GPS, 수치표고모형, 위성영상 등을 제외하면 1-2기, 2-3기에 걸쳐 출현율 변동이 적은 경향이 있다. 즉, 데이터 분야에서는 일반적으로 지금까지 꾸준히 연구되는 분야들이 있음을 의미한다.

이러한 현상은 공간정보의 특성으로부터 기인한다. 공간정보의 시발점은 다양한 형태의 데이터를 얼마나 정밀하게 수집할 수 있는가에 달려있다. 따라서 기존의 단순 측량을 넘어서 보다 정밀한 측정이 가능한 기술인 원격탐사와 LiDAR 등이 등장하였다. 또한 단순 수치지도를 넘어 항공사진, 위성사진을 활용하는 기술이 개발되었다. 그러나 센서 및 네트워크의 기술 발달에 따라 데이터 분야에서는 무엇보다도 ‘센서’와 ‘실시간’이라는 용어가 주요해졌다. 단, 기대되었던 ‘실시간’은 기술적 한계로 인해 주춤하는 경향을 보이나 ‘센서’는 꾸준히 발전하는 것으로 보아 공간정보 분야에서 데이터의 수집에 센서는 더 이상 뺄 수 없는 요소로 자리 잡았다고 볼 수 있다. 특이한 점은 GPS가 다시 부각되고 있는데, 이점은 GPS 관련 기술은 어느 정도 성숙하여 현시점에서 공간정보를 서비스하기에 적합한 기술로 활용되기 때문이라 생각된다.

3.5 활용분야 상세 분석

활용분야의 경우 1기에는 서비스(9.9%), 자연재해(4.6%), 지능형교통시스템(4.6%), 토지이용(2.3%), 2기에는 서비스(9.7%), 자연재해(5.9%), 모바일(3.0%), 지능형교통시스템(1.3%), 위치기반서비스, 유비쿼터스, 텔레매틱스(각 0.8%), 3기에는 자연재해(8.3%), 서비스(7.4%), 모바일, 스마트폰(각 4.4%), 유비쿼터스, 텔레매틱스, BIM(각 2.0%), 증강현실, 토지이용(각 1.5%)로 나타났다.

Figure 6을 보면 활용분야는 시스템, 데이터 분야와 다르게 1사분면에 많은 수의 키워드들이 위치하고 있는데, 이점은 활용에 대한 연구 수요가 전반적으로 증가했음을 의미한다. 자연재해의 경우 1-2기, 2-3기 동안 출현율이 증가하고 있으며 그 비중도 큰 편에 속하며(출현율의 변화율이 1.8), 모바일의 경우 1-2기, 2-3에 출현율이 증가하나 그 변화율은 크지 않으며(출현율의 변화율이 0.5), 서비스의 경우 1-2기에 비해 2-3기에는 출현율이 감소했다. 지능형교통시스템의 경우 1-2기에 크게 출현율이 감소했으며 2-3기 역시 출현율

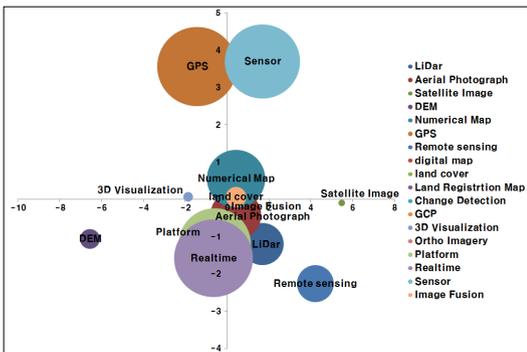


Figure 5. Comparison of keywords appearance rate in data sector during the aimed periods

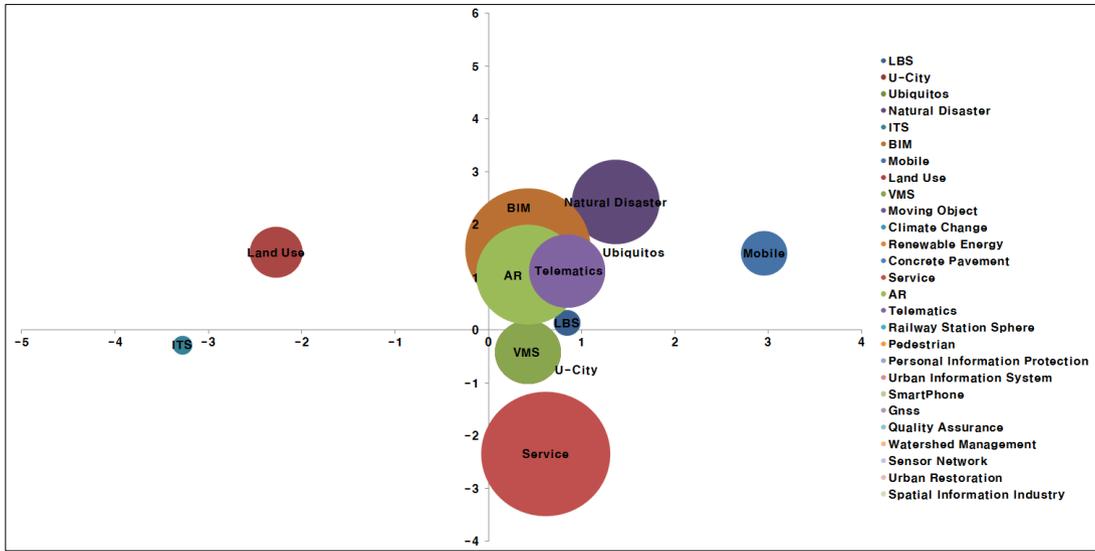


Figure 6. Comparison of keywords appearance rate in application sector during the aimed periods

이 감소하였으므로 어느 정도 활용에 대한 성과를 달성하여 추가적인 연구 수요가 없어지고 있음을 말하고 있다. 토지이용의 경우 1-2기에는 감소하였으나 2-3기에는 증가하고 있으므로 최근 다시 관심을 받는 주제라고 볼 수 있다.

위 현상의 원인은 공간정보의 활용이 서비스 제공에 초점을 맞추고 있으나 지능형교통시스템을 제외한다면 아직까지 서비스가 구체화된 분야를 찾기 어렵다. 이는 공간정보를 기반으로 서비스를 도출하는 일이 어렵기 때문으로 생각된다. 단, 새로운 분야로 자연재해 및 토지이용에 대한 관심이 커지고 있다는 점은 긍정적 현상이다. 또한 공간정보 서비스는 모바일 환경을 기반으로 제공되어야 함을 다수의 연구에서 인지하고 있다. 전반적으로 공간정보 활용에 관한 수요가 증가하고 있다는 점은 매우 바람직한 방향으로 고려되지만 아직까지 교통, 재난, 도시계획 분야에서만 활용되고 있다는 점은 공간정보 분야에서 타 분야와의 협업을 통해 보다 많은 융·복합연구가 수행되어야 할 것으로 판단된다.

4. 논문 자료분석 결과

4.1 현황자료

논문의 시기별 발표 횟수를 보면 Figure 7과 같이 2000년부터 급격히 증가하다가 2005년부터 2009년경에는 성장세가 둔화되었으며 2010년에 정점을 찍은

후 2011년, 2012년 감소하였다. 2013년 11월에 조사된 자료이므로 2013년 12월에 발간되는 논문집의 논문은 고려되지 않았으나 1년간의 비중을 생각하면 2013년도 역시 감소했다고 예측 가능하다.

발행처별로 살펴보면 대한토목학회(34.95%), 한국공간정보시스템학회²⁾(15.22%), 한국지형공간정보학회(15.16%), 지형공간정보학회(7.43%), 한국GIS학회(6.95%), 한국공간정보학회(6.22%) 한국측지학회³⁾(4.26%), 한국측량학회(4.06%), 개방형지리정보시스템학회⁴⁾(2.75%), 한국지리정보학회(1.65%), 대한원격탐사학회(1.35%) 순으로 나타났다.

4.2 주요 키워드 및 기간별 출현율

먼저 논문에 출현한 주요 키워드와 그 점유율을 살펴보면, 지능형교통시스템이 전체 논문 중 8.1% 출현했으며, 서비스(7.9%), 데이터베이스(5.4%), 수치표고모형(4.9%), 실시간(4.1%)이 각각 그 뒤를 이었다.

- 2) 한국공간정보시스템학회와 한국GIS학회가 통합을 하여 2010년 3월부터는 한국공간정보학회로 거듭나 학회명칭 및 학회지명이 변경되었다.
- 3) 한국측지학회는 2000년 4월에 한국측량학회로 명칭을 변경하였다.
- 4) 개방형지리정보시스템학회는 2005년부터 학회명을 한국공간정보시스템학회로 변경하고, 학술지명도 ‘개방형지리정보시스템학회 논문지’에서 ‘한국공간정보시스템학회 논문지’로 변경하여 발간하다가 2010년부터 한국공간정보학회로 통합되었다.

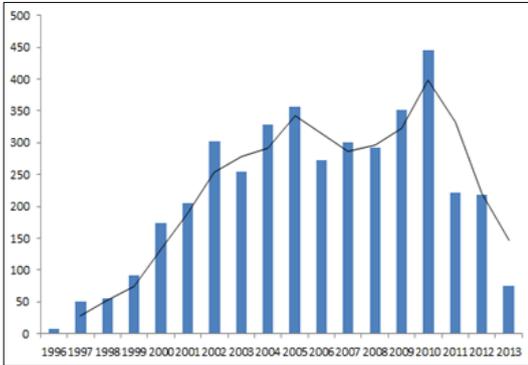


Figure 7. Number of published papers during the aimed period

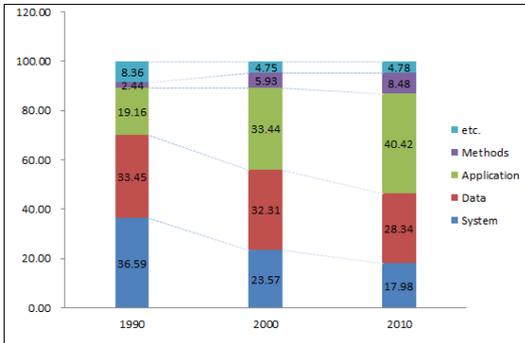


Figure 8. Variation of appearance rate of paper's keywords in each sector

기간별로 집합화된 키워드들의 출현율을 살펴보면 Figure 8과 같다. 1994년부터 1999년까지는 시스템, 데이터의 비중이 높았으나 시스템, 데이터의 비중은 점차 감소하고, 상대적으로 활용관련 키워드가 증가하였다. 또한 기법들에 대한 비중 역시 크게 증가하고 있다. 이는 곧, 시스템에 대한 연구는 데이터와 같이 연구가 진행되어 왔으며 최근에는 활용에 대한 연구가 40% 이상으로 증가하였으며 다양한 기법들의 접목이 이루어짐을 확인할 수 있다.

4.3 시스템분야 상세 분석

시스템 분야의 상위 출현 키워드를 살펴보면 1기(1994년부터 1999년까지)에는 데이터베이스(9.1%), 표준화(4.9%), 인터넷(4.2%), component(3.5%), maintenance(2.8%), 3차원공간정보, 상호운용성(각각 2.4%)이 나타났으며, 2기(2000년부터 2009년까지)에는 데이터베이스(5.9%), 3차원 공간정보(3.7%), component(2.4%), 인터넷(2.2%), 표준화(2.2%), maintenance(1.3%)가 나타났으며, 3기(2010년 이후)에는 3차원공간정보(4.2%), 데이터베이스(3.9%), 표준화(1.9%), maintenance(1.7%), component(1.3%), 인터넷(1.0%)이 나타났다.

Figure 9를 보면 대부분의 키워드 출현율의 변화는 작은 편이다. 그러나 온톨로지(출현율의 변화율이 9.9), xml(출현율의 변화율이 2.9) 등은 변화율이 큰 키워드이며, 온톨로지의 경우 1-2기 뿐 아니라 2-3기 동안에

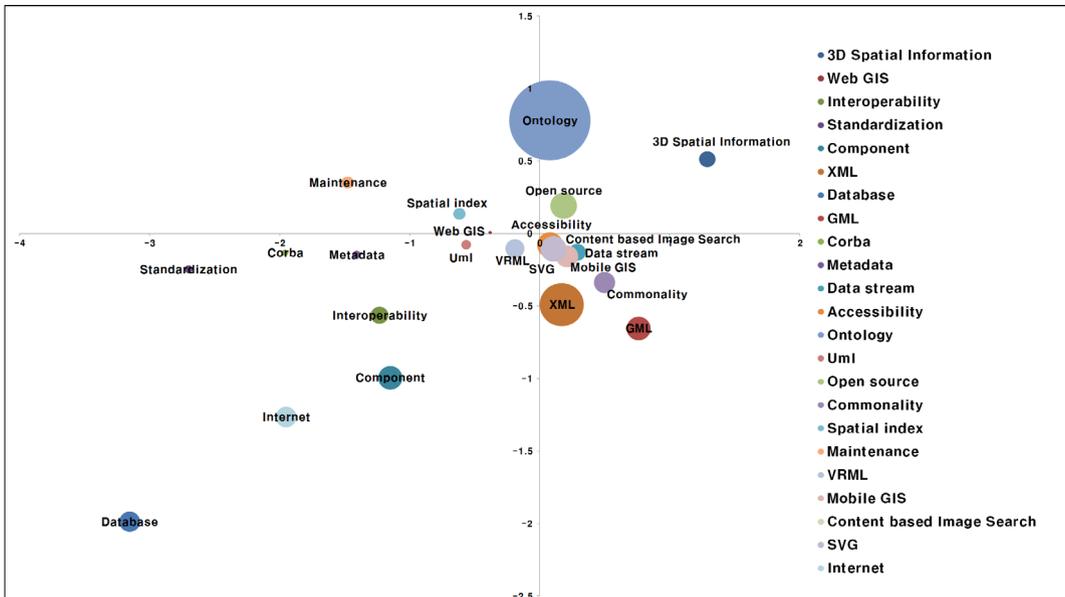


Figure 9. Comparison of keywords appearance rate in system sector during the aimed periods

연구수요가 증가하였으며 특히 2-3기에 크게 증가하였다. 3차원 공간정보의 경우 역시 1-2기뿐 아니라 2-3기 동안 꾸준히 연구수요가 증가하고 있는 키워드이다. 출현율의 상위권을 차지하고 있는 데이터베이스의 경우 1-2기와 2-3기 동안 꾸준히 감소하고 있는 것을 확인할 수 있다.

‘데이터베이스’, ‘3차원 공간정보’ 등에서 보듯이 국가 연구개발보고서와 유사한 형태를 보이지만 maintenance 등이 등장한다는 점과 온톨로지의 변화율이 크다는 점은 논문에서 보이는 차이점이다. 이러한 현상은 시스템분야에서는 재사용성 관점에서 기존의 성과물인 GIS에 대한 유지보수가 중요한 점으로 부각되고 있으며 온톨로지에 대한 보다 심도 깊은 학문적 연구가 진행되고 있기 때문에 파악된다. 국가연구개발보고서에서 변화율이 줄어든 3차원 공간정보가 논문에서 부각된다는 사실은 아직까지 3차원 공간정보는 실용화보다는 학술적 연구에서 더 많은 성과가 기대된다고 볼 수 있다.

4.4 데이터분야 상세 분석

데이터 분야의 상위 출현 키워드를 살펴보면 1기에는 수치표고모형(9.8%), 실시간(4.9%), 위성영상(4.2%), digital map(3.8%), GPS(3.1%)가 출현했으며 2기에는 수치표고모형(5.4%), 실시간(4.6%), GPS(4.4%), 위성영상(4.0%), 센서(2.6%)가 출현했으며 3기에는 센서(3.7%), 수치표고모형, Lidar data(각 3.2%), 실시간(3.0%), GPS, 위성영상(각 3.0%), land cover(1.6%)로

나타났다.

Figure 10을 보면, 수치표고모형의 경우 1-2기와 2-3기에 걸쳐 꾸준히 출현율이 감소하고 있으며 GPS의 경우 1-2기에 증가추세였으나 다시 2-3기에서는 감소추세로 변화하였다. 실시간의 경우는 1-2기, 2-3기에 감소하였으나 그 감소폭이 크게 나타났다. 센서와 LiDAR의 경우, 1-2기, 2-3기 모두 꾸준히 증가를 보이고 있다.

데이터 분야에서는 전반적으로 원점에 몰려 있는 현상을 볼 수 있는데, 이는 데이터와 관련된 키워드의 출현율은 시기와 상관없이 적당한 비율로 지속되어 왔음을 의미한다. 단 수치지형도의 경우 그 변화율이 매우 크게 나타나고 있다(출현율의 변화율이 41.7).

데이터 분야의 경우, 국가연구개발보고서와 유사한 형태를 보인다. 이점은 앞서 언급하였듯이 공간정보의 데이터적 특성에 따른 것으로 파악된다. 국가연구개발보고서와는 다르게 ‘LiDAR’의 변화율이 크게 나타난다는 점과 GPS에 대한 변화율이 감소한다는 점은 오히려 LiDAR에 대해서는 학술적 연구를 통한 기술개발이 필요한 반면 GPS에 대해서는 실용화할 수 있는 기술수준을 달성했다고 볼 수 있다.

4.5 활용분야 상세 분석

활용분야의 경우 1기에는 지능형교통시스템(8.7%), 서비스(4.9%), land use(1.7%), 자연재해, 모바일(각 1.0%)이 출현했고, 2기에는 지능형교통시스템(7.9%), 서비스(7.5%), 자연재해(3.0%), 모바일(3.0%), 유비쿼

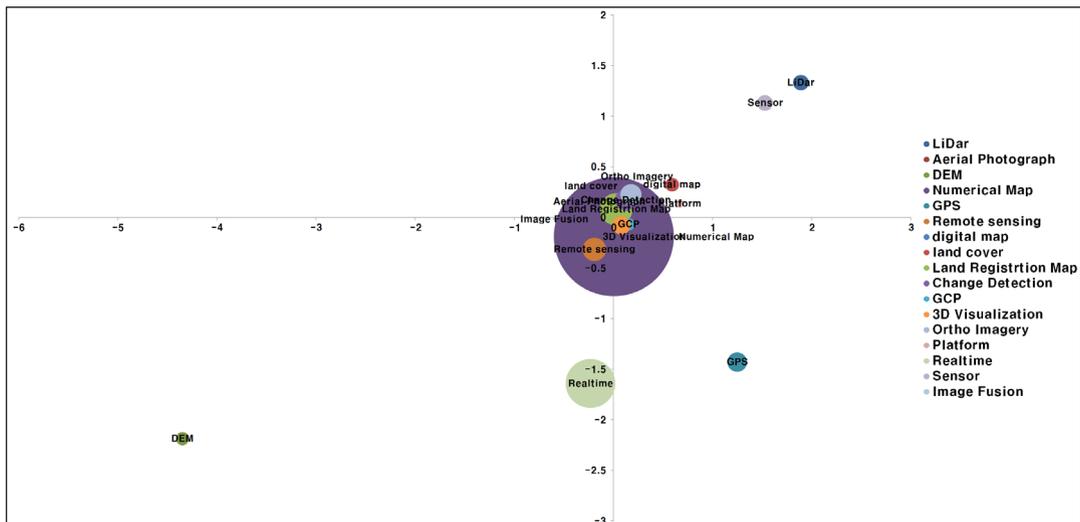


Figure 10. Comparison of keywords appearance rate in data sector during the aimed periods

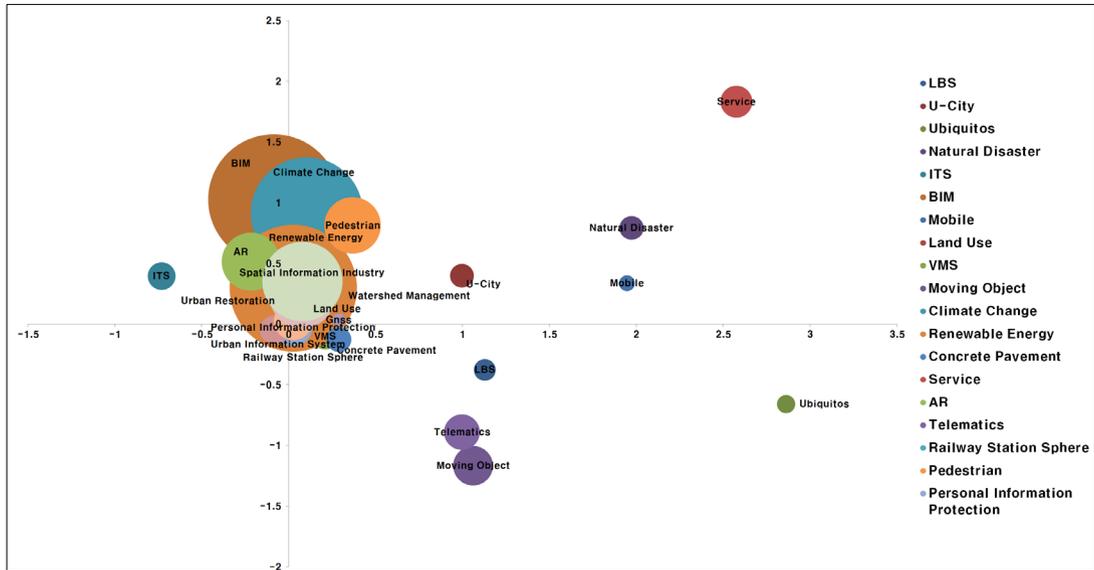


Figure 11. Comparison of keywords appearance rate in application sector during the aimed periods

터스(2.9%)가 출현했으며, 3기에는 서비스(9.3%), 지능형교통시스템(8.4%), 자연재해(3.8%), 모바일(3.3%), 유비쿼터스(2.2%)가 자주 출현했다.

Figure 11을 보면 활용분야는 3사분면에 위치한 키워드가 없으므로, 활용과 관련된 키워드들은 지속적으로 관심이 떨어지는 분야는 없이 다양하게 관심을 받고 있다고 할 수 있다. 서비스의 경우 1-2기, 2-3기 동안 지속적으로 높은 출현율을 보이고 있으며, 자연재해, 모바일, 유비쿼터스도시 등 역시 유사한 관심을 받고 있다. 위치기반서비스, telematics, moving object, 유비쿼터스 등은 2-3기에서 관심이 상대적으로 줄어들었으며 지능형교통시스템의 경우 1-2기에 관심이 줄었으나 다시 2-3기에 관심이 증가한 분야라고 볼 수 있다. 전반적으로 원점근처의 원들이 크게 나타나는데, 이는 출현율의 변화폭이 큰 키워드들이 지속적으로 나타난다고 볼 수 있다.

국가연구개발보고서와 유사하게 서비스에 대한 키워드들은 전반적으로 많은 관심을 받고 있다. 마찬가지로 상대적으로 지능형교통시스템에 대한 관심은 적어지는 현상을 발견할 수 있다. 더불어 자연재해에 대한 관심의 증가는 논문에서도 발견할 수 있는 주요한 사실이다. 논문에서 발견되는 특이한 점은 유비쿼터스 및 유비쿼터스 도시에 대한 키워드들의 상승세인데 이는 유비쿼터스와 공간정보의 뗄 수 없는 관계 때문이며, 결국 유비쿼터스도시는 공간정보를 기반으로 서비스를 제공해야하기 때문으로 파악된다.

논문에서는 국가연구개발보고서보다 더 많은 관련 주요어들이 등장하고 있으며 전반적으로 중심에 몰려 있는 것으로 보아 학술적으로 다양한 시도가 진행되고 있다고 긍정적으로 평가할 수 있다.

5. 결 론

본 연구에서는 국가과학기술전자도서관에서 4,000개의 논문과 225개의 국가연구개발보고서를 수집하여 공간정보 분야의 연구동향을 살펴보았다. 특히, 주요 키워드들을 분야별로 나누어 국가연구개발보고서와 논문의 분야별 동향을 살펴보았다. 먼저 국가연구개발보고서에 출현한 주요 키워드와 그 점유율을 살펴보면, 데이터베이스가 전체 국가연구개발보고서 중 12.2% 출현했으며, 서비스(8.7%), 실시간(7.8%), 자연재해(6.8%), 3차원공간정보(5.8%), 표준화(5.8%)가 각각 그 뒤를 이었다. 다음으로 논문에 출현한 주요 키워드와 그 점유율을 살펴보면, 지능형교통시스템이 전체 논문 중 8.1% 출현했으며, 서비스(7.9%), 데이터베이스(5.4%), 수치표고모형(4.9%), 실시간(4.1%)이 각각 그 뒤를 이었다.

본 연구의 분석 결과를 종합해 보면, 국가연구개발보고서와 논문의 경우 모두에 시스템 연구는 감소하는 추세이고, 활용 연구에 대한 비중이 크게 증가하고 있음을 확인할 수 있었다. 특히 주제어를 중심으로 살펴보면 3차원 공간정보, 센서, 서비스(ITS제외) 등이 많

이 부각되고 있다. 이러한 경향을 고려하는 것이 추후 관련 연구들의 방향 설정에 도움이 될 수 있을 것이다.

본 연구에서는 NDSL에서 제공하는 국가연구개발 보고서와 논문 자료를 분석하여 출현율이 높은 키워드, 변화율이 큰 키워드들을 찾아 공간정보 관련 연구 동향에 대해 알아보았다. 그러나 본 논문에서는 두 경우에 대한 단순 비교만을 수행하였는데, 실제로 논문은 새로운 이론을 탐구하는 쪽에 비중이 크고 국가연구개발보고서는 최신기술을 적용하는 경우가 많다. 따라서 논문이 새로운 이론이라면 해당 이론이 얼마나 국가연구개발보고서에서 확인이 되는지, 추가적으로 필요한 R&D 부문은 무엇인지, 논문과 국가연구개발보고서의 시간적 차이를 확인하여 현 시점에서 적합한 R&D 주제는 무엇인지를 확인하는 것도 향후 연구주제가 될 것이다.

References

[1] Ahn, J. W; Yi, M. S; Shin, D. B. 2013, Study for Spatial Big Data Concept and System Building, Journal of Korea Spatial Information Society, 21(5):43-51.

[2] Bae, J. H; Son, J. E; Song M. 2013, Analysis of Twitter for 2012 South Korea Presidential Election by Text Mining Techniques, Bibliographic Info: J Intell Inform Syst, 19(3):141-156.

[3] Bae, K. Y; Park, J. H; Kim, J. S; Lee, Y. S. 2013, Analysis of the abstracts of research articles in food related to climate change using a text-mining algorithm, Journal of the Korean Data & Information Science Society, 24(6):1429-1437.

[4] Chang, J. Y. 2013, A Study on Research Trends of Graph-Based Text Representations for Text Mining, Journal of the Korea Academia-Industrial cooperation Society, 12(8):3677-3688.

[5] Cho, G. H; Lim, S. Y; Hur, S. 2014, An Analysis of the Research Methodologies and Techniques in the Industrial Engineering Using Text Mining, Journal of the Koran Institute of Industrial Engineers, 40(1):52-59.

[6] Cho, S. G; Kim, S. B. 2012. Finding Meaningful Pattern of Key Words in IIE Transactions Using Text Mining, Journal of the Korean Institute of Industrial Egnineers, 38(1):67-73.

[7] Go, G. S; Jung, W. K; Shin, Y. G; Park, S. S;

Jang, D. S. 2011, A Study on Development of Patent Information Retrieval Using Textmining, Journal of the Korea Academia-Industrial cooperation Society, 12(8):3677-3688.

[8] Jeong, C. W; Kim, J. J. 2012, Analysis of trend in construction using textmining method, Journal of the Koran Digital Architecture-Interior Association, 12(2):53-60.

[9] Kim, C. H. 2007, A Review on GIS Research Trends In North Korea, Journal of the Korean Association of Geographic Information Studies, 10(4):189-197.

[10] Kim, K; Chun, J. Y; Shin, D. B; Lim, S. Y. 2011, Research Trends of U-City Theses in Korea and Oversea, Journal of the Korean Society for Geospatial Information System, 19(1):53-61.

[11] Sakong, H. S; Seo, K. H. 2007, A Review on GIS Research Trends using Content Analysis Method-Focus on the GIS Journals Published from 1993 to 2006, Journal of the Korean Association of Geographic Information Studies, 10(3):104-112.

[12] Yi, M. S; Ahn, J. W. 2013, Research on Principles to Transcribe Geographical Names in English for English Version Electronic Map Service, Journal of Korea Spatial Information Society, 21(5):53-61.

논문접수 : 2014.6.25

수정일 : 2014.8.18

심사완료 : 2014.8.18