KOSDI ISSN : 1976-2208 (Print)
ISSN : 2671-5287 (On-line)
www.sodi.or.kr

Journal of the Society of Disaster Information
Vol. 18, No. 2, pp. 418-431, June 2022
https://doi.org/10.15683/kosdi.2022.6.30.418

Original Article

위성·지상정보 융합 재난 대응 기술 분야 유망기술 도출을 위한 연관 키워드 및 빅데이터 분석 기법

A New Scheme Exploiting the Related Keyword and Big Data Analysis for Predicting Promise Technology in the Field of Satellite·Terrestrial Information Convergence Disaster Response

이항원[1] · 김영억[2]*
Hangwon Lee[1], Youngok Kim[2]*

[1]Graduate Student, Electronic Engineering Department, Kwangwoon University, Seoul, Republic of Korea
[2]Professor, Electronic Engineering Department, Kwangwoon University, Seoul, Republic of Korea

***Corresponding author:** Youngok Kim, kimyoungok@kw.ac.kr

ABSTRACT

Purpose: We propose a new scheme for predicting promise technology and it improves the conventional scheme that misses important lists of patent because of insufficient search formula, and cannot reflect new trend of technology due to the unreleased period of patents. **Method:** In this paper, we propose a new search formula exploiting TF and TF-IDF with R programming as well as related keywords, and LDA topic modeling scheme is used for analyzing recently published papers in Satellite·Terrestrial Information Convergence Disaster Response. **Result:** By comparing both schemes with commercial DB, the proposed scheme can find more important patents, and can reflect new trend of technology, compared to the conventional scheme. **Conclusion:** The proposed scheme can be used to predict promise technologies in the field of Satellite·Terrestrial Information Convergence Disaster Response.

Keywords: Satellite, Terrestrial, Convergence, Big Data, Topic Modeling, Promising Tech

요 약

연구목적: 기존의 유망기술 도출 연구에서 대상 특허를 선별하는 과정에서 검색식이 충실하게 구성되지 않아 중요 특허가 분석 대상에서 누락되는 문제와 출원 미공개 기간에 변화하는 최신 기술 트랜드를 반영하지 못하는 문제들을 개선하기 위한 새로운 기법들을 제시한다. **연구방법:** 본 논문에서는 R 프로그래밍을 통한 TF와 TF-IDF 등의 연관 키워드 기반 검색식 구성 기법 및 LDA 토픽 모델링 기법을 제안하고, 위성·지상 정보 융합 재난 대응 기술을 중심으로 기존 방법과 비교 분석하여 제안된 기법의 유효성을 확인한다. **연구결과:** 상용 특허 DB 시스템을 활용한 제안된 방식 적용을 통해 대상 기술과 관련도가 가장 높은 주요(핵심)특허가 기존 방식(19건) 대비 17건 추가로 도출(36건, 89.5% 증가)되어 중요 특허 누락이 상당히 개선됨을 보였으며 Science ON에 등록된 최근 5년간 발표 논문을 대상으로 LDA 토픽 모델링을 수행하여 최신 기술 트랜드를 반영한 유망 기술 도출이 가능함을 보였다. **결론:** 본 논문을 통해 제시한 유망기술 도출 기법을 통해 위성·지상정보 융합 재난 대응 기술분야에서의 유망기술 도출을 통한 신규 연구개발에 유용하게 사용될 수 있는 방향성을 제공하였다.

핵심용어: 위성, 지상, 융합, 빅데이터, 토픽 모델링, 유망기술

Received | 8 June, 2022
Revised | 28 June, 2022
Accepted | 29 June, 2022

OPEN ACCESS

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted noncommercial use, distribution, and reproduction in anymedium, provided the original work is properly cited.

ⓒ Society of Disaster Information All rights reserved.

서 론

전 세계적으로 위성 보유 대수가 증가하고 우수기술이 발전함에 따라 환경·에너지·자원·식량 안보·재난 대응 등 다양한 사회문제 해결을 위해 위성정보 활용을 추진하고 있다. 위성 보유국 간 전략적 협력을 확대하는 등 각국의 위성 자원 활용 효율성 제고를 위한 다각적인 방안을 추진하는 가운데(Kim, et al., 2012), 한국의 경우 제3차 우주개발진흥 기본계획을 통해 재난·재해 등 국가 위기 대응 서비스, 해양·환경·농수산 등 공공 활용 서비스, 통신·항법 등 4차 산업혁명 기반 서비스, 한반도 정밀 감시 서비스 등 국가 위성의 활용도를 높이는 계획을 수립하여 Fig. 1과 같이 추진하고 있다(Ministry of Science and ICT, 2018).

Fig. 1. Examples of services using satellite information(Ministry of Science and ICT, 2018)

더불어, 국가 위성정보 활용체계 고도화를 통해 국민 편익 제고 및 위성 정보의 보급·활용을 촉진하기 위한 제2차 위성정보 활용 종합계획을 수립하였고(Cooperation of Related Ministries, 2018), 최근에는 연도별 시행계획을 수립하고 위성을 전략적으로 활용하기 위해 스마트한 3대 국가 위성 정보서비스 제공을 목표로, 신속·정확한 국가 안전·위기관리 서비스 제공, 다양한 국민 생활 밀착 서비스 확대, 미래선도 산업 기반 서비스 추진을 과제로 선정하였다(Cooperation of Related Ministries, 2021).

위성정보를 활용하는 산업 구조는 Fig. 2와 같이 정보를 송·수신하는 위성, 위성정보를 처리하여 사용 가능한 형태로 제공하는 공급자, 공급된 위성정보를 활용하거나 2차 가공하여 부가가치를 생산하는 사용자로 구성된다. 사용하는 위성의 종류에 따라 지구관측 위성, 통신위성, 항법위성으로 구분할 수 있고, 위성정보를 공급하는 사업자(민간, 공공)를 통해 정보를 최종 활용하는 사용자에게 전달되는 산업 구조를 갖는다. 사용자는 전달받는 정보의 종류에 따라 서로 다른 서비스에 활용하게 되는데 지구관측 위성의 경우 국방, 인프라, 환경 모니터링, 재해관리 등에 활용되고, 통신위성은 디지털 TV 등 영상서비스와 위성 통신에 대표적으로 활용된다. 항법위성은 위치기반 서비스(LBS) 및 항공·육상·해상의 교통 서비스에 활용된다(Cooperation of Related Ministries, 2018).

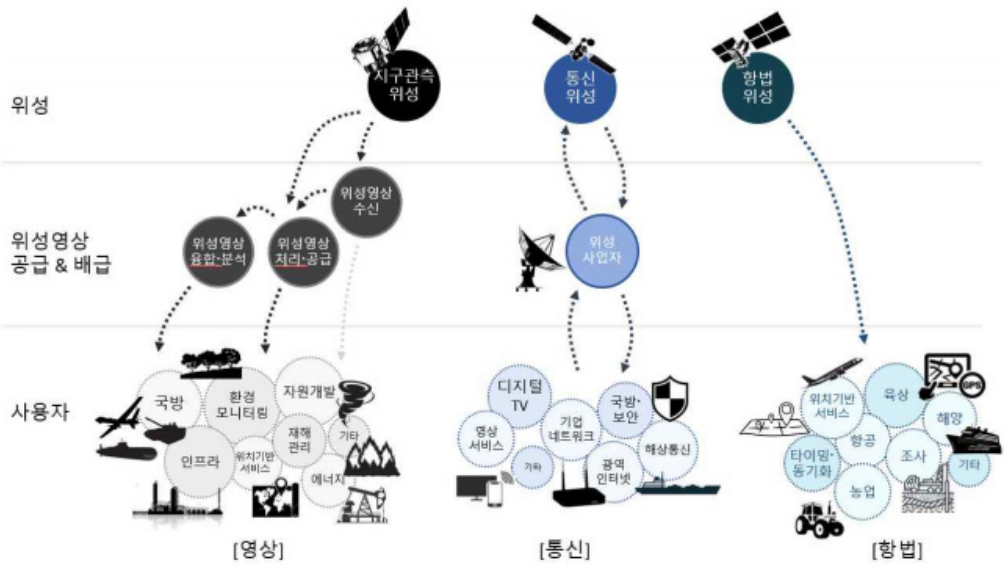


Fig. 2. Industrial structure of satellite information(Cooperation of Related Ministries, 2018)

최근에는 전 세계적으로 지구 온난화에 따른 기상이변이 발생하고 있고, 기상이변의 영향으로 발생하는 다양한 자연재해를 첨단기술을 이용하여 극복하고자 하는 연구가 다양하게 이루어지고 있다. 특히 인공위성 정보를 재난 대응에 활용하는 연구가 지속적으로 이루어지고 있어 이에 주목할 필요가 있다. 구체적으로는 다중 위성정보를 활용하여 적설 분포도를 작성하고, 이를 다양한 공간자료와의 GIS 공간분석을 통해 폭설 재난 시 적시에 의사결정을 할 수 있도록 지원한 연구(Kim et al., 2012), 인공위성 영상과 기계학습 알고리즘을 활용한 하천 재난 상황 인식에 관한 연구(Go et al., 2015), 위성 기반 재난관리 기술 및 GIS를 활용한 재난정보 분석 등 한반도 재난관리를 위한 다양한 연구(Kim et al., 2020), 기상위성을 활용한 겨울철 블랙아이스 발생 특성 분석(Park et al., 2017) 등의 연구가 위성정보를 활용한 재난대응 기술 개발을 위해 수행되었다.

그러나 보다 다양한 재난 환경에서 활용하기 위해서는 높은 해상도의 영상과 정확한 개체 식별이 가능한 수준의 정보가 제공되어야 하지만, 현재 국내에서 보유한 위성을 통해 제공 받을 수 있는 영상 정보는 이러한 요구에 미치지 못하는 상황이다. 이에 따라, 국가적 위기 극복 또는 국민을 위한 공공서비스로 활용할 수 있는 재난·안전 서비스 제공을 위해 위성과 지상 정보를 적절하게 융합할 필요가 제기되었으며, 광역 위성 영상과 지상 관측망을 이용한 강수 예측 모형 개발(Cho, 2010)와 같이 융합정보를 활용한 연구도 최근 소개되고 있다.

국가의 중장기 추진계획에 따라 위성의 수가 지속적으로 증가하고, 위성정보를 활용한 서비스의 중요성이 증대될 것으로 예상되는 가운데, 위성정보와 지상에서 확보할 수 있는 정보를 융합하여 재난·안전 분야에 활용할 수 있는 서비스 관련 유망 분야를 발굴하고, 선도적으로 R&D를 수행하여 유망기술을 확보해야 하는 중요한 시기로 판단된다. 그러므로 본 연구에서는 위성·지상 정보 융합 재난 대응 기술 분야의 특허를 조사·분석하여 특허가 선점되지 않은 기술 공백 영역을 도출하고, 논문 정보를 이용하여 최근 트랜드를 반영하는 방법으로 유망기술을 예측하는 새로운 기법을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 유망 기술 예측 방법 관련 연구 동향에 대해 소개하고, 3장에서는 유망 기술 도출을 위해 연관 키워드 도출을 통한 개선된 검색식 구성 기법과 최근 논문 정보를 활용한 빅데이터 분석 기법을 제안

한다. 4장에서는 위성·지상 정보 융합 재난 대응 기술을 중심으로 제안된 기법의 유효성 분석의 주요 내용을 제시하고, 5장은 본 논문의 결론을 기술한다.

기존의 유망 기술 예측 방법 관련 연구 동향 분석

일반적으로 기술 도출은 기술의 미래가 어떻게 변화할 것인가를 예측하거나, 가능한 미래를 인식하고 중요한 기술 발전 추세와 사회경제적 수요를 예측하는데 활용된다. 최근에는 단순한 기술 추세 전망을 넘어 국가나 기업의 장기적 생존을 보장할 수 있는 기술적 대안을 탐색하거나, 기술 발전 속도, 한계점, 유망기술분야 등을 전망함으로써 정부 정책, 기업의 미래 전략 등의 의사 결정자가 필요로 하는 정보를 제공하는 것도 포함된다.

글로벌 무한 경쟁 시대가 도래한 최근 들어 기술의 중요성이 커지면서 미래 유망기술을 선점하기 위한 노력이 전 세계적으로 이루어지고 있다. 이를 위해 해당 분야 전문가 그룹의 의견을 반영하는 델파이 기법과 AHP(Analytic Hierarchy Process) 분석 등을 활용해 왔는데, 이러한 정성적인 예측 방법은 유망기술 선정의 신뢰성을 담보하기 어려운 한계가 있어, 계량적인 통계 기법을 병행하는 방법으로 연구의 한계를 해결하려는 시도가 이루어지고 있다(Kwon, 2008; Kim et al., 2010).

정량적인 분석을 위해서는 정보가 일관성 있는 형식을 적용해야 하고, 쉽게 접근할 수 있도록 개방되어 있으면서 기술 내용을 상세하게 기재하여 기술을 예측하는 연구자가 이해하기 쉽도록 작성되어 있어야만 데이터의 누락 없이 신뢰성 있는 분석을 수행할 수 있다. 이러한 조건을 만족하는 빅데이터 정보는 논문과 특허 정도로 압축될 수 있는데, 특허의 경우 전 세계적으로 법령 및 각종 조약이 정한 바에 따라 통일된 형식으로 공개 및 개방되어 있고, 전문기관에 의해 체계적으로 DB화 되어 서비스가 제공되고 있어 접근이 용이한 장점을 가지고 있다. 또한, 특허제도가 기술 공개를 조건으로 청구범위로 정한 세부적인 기술에 대해 20년간 독점권을 부여하는 제도인 만큼, 특허는 제3자가 구현 가능하도록 세부적으로 기술내용을 적시하도록 의무화 되어 있다. 또한, 특허를 출원하고 등록을 유지하는데 상당한 비용이 소요되기 때문에 기업이 특허를 출원하고 등록을 유지한다는 것은, 출원한 특허가 시장에서 기업이 경쟁력을 확보하고 유지하기 위해 필요한 기술임을 간접적으로 보여주는 것으로 판단할 수 있어, 기업의 연구개발 방향성을 확인할 수 있는 중요한 지표로 활용되기도 한다.

유망기술을 도출하기 위한 연구로서 델파이와 AHP를 특허분석과 결합하여 미래기술을 제시한 연구(Lee et al., 2012), 특정 기술분야의 동향 및 미래기술 예측, 국가 정책 수립에 특허를 활용한 연구가 수행되었고(Jin et al., 2012), 특허 빅데이터 분석을 활용하여 관심 기술 및 산업분야의 미래기술을 예측하는 연구가 소개되었다(Jeon, 2011; Im et al., 2017; Choi et al. 2013; Yun et al., 2016; Lee, et al., 2021). 최근에는 다양한 빅데이터 분석 방법이 출현하면서 텍스트 마이닝 기반의 빅데이터 분석 Tool을 활용한 연구들 또한 발표되고 있다. 파이썬(Python)이나 R을 활용한 특허 빅데이터 분석 방법이 다양하게 시도되고 있으며(Zhang et al., 2013; Jang et al., 2013), TF-IDF와 네트워크 분석을 활용한 경쟁 우위 전략 수립(Yun et al., 2018)과 TF-IDF를 활용한 한글 자연어 처리 연구(Lee et al., 2019), 토픽 모델링을 활용한 디지털 헬스케어 관련 핵심기술 도출(Lee, 2019), 토픽모델링과 특허 통계분석을 활용한 라스트마일 물류 유망기술 도출 연구(Lee et al., 2021) 등이 이루어지고 있다. 이외에도 특정 기술 분야를 대상으로 유망기술을 분석한 연구에서 특허 정보를 활용한 유망기술 예측의 유효성을 확인하는 연구결과도 확인되었다(Kang, 2017; Garzaniti et al., 2021).

다만, 기존의 특허 정보를 활용한 유망기술 도출 연구는 대부분 특허 정보에 기존의 통계적 이론과 TOOL을 적용한 경우가 많았고, 대상 특허를 선별하는 과정에서 검색식이 충실하게 구성되지 않아 중요 특허가 분석 대상에서 누락되어 정확한

분석이 이루어지기 어려운 한계가 있었다. 또한, 전 세계가 특허 제도를 통해 특허 출원 1년 6개월 후에 문헌을 공개하도록 규정하고 있기 때문에, 미공개 기간에 변화하는 최신 기술 트랜드를 적절하게 반영하기 어렵다는 문제 또한 지속적으로 제기되고 있다.

본 논문에서는 이러한 문제들을 개선하기 위하여, 다양한 방법을 활용한 연관 키워드 도출을 통해 특허 누락이 최소화 될 수 있는 검색식 구성 방법을 제안하고, 또한 최근 5년간 발표된 유사 논문 정보를 활용하여 특허의 미공개 기간의 변화하는 기술 트랜드를 반영하는 방법을 제안하고자 한다. 이러한 연구를 보다 효율적으로 수행하기 위하여 텍스트마이닝 기반의 빅데이터 분석에 적합한 R 프로그래밍을 이용하고자 한다. 연관 키워드 도출을 위하여 TF(Term Frequency)와 TF-IDF (Inverse Document Frequency)를 적용하고, 유망기술 도출을 위해서는 LDA(Latent Dirichlet Allocation) 토픽 모델링을 수행하여 도출된 토픽 및 키워드를 기술 공백 영역과 매칭하는 방법으로 유망기술을 도출하고자 한다. Table 1은 유망 기술 도출을 위한 전체 프레임과 프로세스를 나타내고 있으며, 1단계의 연관 키워드를 활용한 검색식 구성 및 3단계의 최근 논문 정보 활용에 관해 2가지 기법을 제안하고자 한다. 1단계의 Proposed Method ①을 위해, 먼저 최근 5년간 발표된 관련 기술분야 논문을 검색한 후 제목과 요약을 대상으로 산출된 TF와 TF-IDF에 포함된 키워드 중 관련도가 높은 키워드를 반영함으로써, 최근 연구가 이루어지고 있는 분야에 대한 키워드 누락을 방지하고, 대상 기술과 관련도가 가장 높은 IPC(International Patent Classification) 분류의 최근 5년간 출원(공개, 등록)된 특허의 요약 및 청구범위를 대상으로 산출된 TF와 TF-IDF에 포함된 키워드 중 관련도가 높은 키워드를 함께 반영한다. 또한 3단계의 Proposed Method ②는 특허가 출원 1년 6개월 후에 공개되는 특징 때문에, 최신 기술변화를 충실하게 반영하기 어려운 한계를 극복하기 위한 것으로, 특허보다 앞서 발표되는 경향을 보이는 논문 정보를 활용하는 방법이다. 최근 5년간 발표된 논문을 대상으로 LDA 토픽 모델링을 실시하여 토픽과 키워드를 도출하고, 특허 분석을 통해 도출된 기술 공백 영역과 일치하는 Case를 최신 기술 트랜드가 반영된 유망기술로 예측한다. Table 2는 제안된 TF-IDF and LDA 토픽 모델링 에 대해 간략히 요약하였으며, 보다 구체적인 제안 내용은 다음 장에서 자세히 소개하고자 한다.

Table 1. Research frame and process

1단계	2단계	3단계「Proposed Method」②
• 기술분야 키워드 선정 • 특허조사를 위한 검색식 구성	• 유효특허를 대상으로 관련도가 높은 '주요(핵심)특허' 선별	• 최근 5년간 논문 정보 검색 • 필터링 → 유효 논문 선별
↓	↓	↓
「Proposed Method」①		
• TF, TF-IDF 활용 검색식 보완 • 논문 Keyword, 유사어 추천 등을 활용한 '연관 키워드' 도출	• '주요(핵심)특허' 심층분석을 통해, - O/S, T/S Matrix 분석	• 최근 5년간 발표된 유사 논문 대상 LDA 토픽 모델링 - R 프로그래밍 활용
↓	↓	↓
• 특허 검색 → 모집단 특허 • 노이즈 필터링 → 유효특허	• O / S, T / S 매트릭스를 활용한 '기술 공백 영역' 도출	• 기술 공백 영역과 LDA 토픽 모델링 결과 매칭을 통해 유망기술 도출

Table 2. TF-IDF and LDA topic modeling scheme

단계	제안된 방법 개요
「Proposed Method」 ① '연관 키워드'를 활용한 개선된 검색식 구성 (최근 5년간 발표된 논문, 최근 5년간 출원된 특허)	• TF(Term Frequency), TF-IDF(Inverse Document Frequency) - TF : 단어가 특정 텍스트(문서)에 나타난 횟수(빈도수) - DF : 단어가 사용된 문서의 수(문서 빈도) - $IDF = \log\frac{N}{DF}$, $TF-IDF = TF \times \log\frac{N}{DF}$
「Proposed Method」 ② 최근 기술 트랜드를 반영한 유망 기술 예측 (최근 5년 발표된 논문)	• LDA(Latent Dirichlet Allocation) Topic Modeling - 각 문서에 어떤 주제들의 존재를 나타내는 확률 모형이며, 문서 내의 숨겨진 주제를 찾아주는 알고리즘으로 주제별 문장 및 단어 확률을 파라미터를 통해서 도출하는 방식

유망 기술 도출을 위한 제안 기법

연관 키워드 도출을 통한 개선된 검색식 구성 기법

본 장에서는 특허 검색 과정에 누락되는 특허를 최소화하기 위하여 다양한 방법을 통해 연관 키워드를 도출하고, 이를 검색식에 반영하여 결과를 개선하는 방법을 제안하고자 하며, 제안 기법의 유효성을 분석하기 위해 기존에 동일한 주제를 대상으로 별도로 수행되었던 특허분석 결과와 비교 분석하는 방법으로 제안하고자 하는 연관 키워드 활용 방법의 유효성을 다음 장에서 분석하고자 한다.

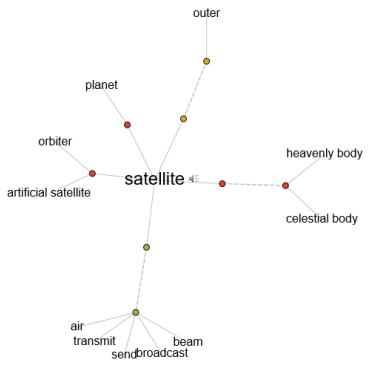
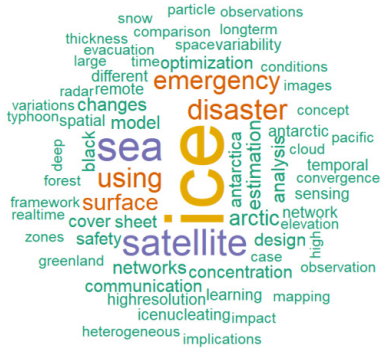
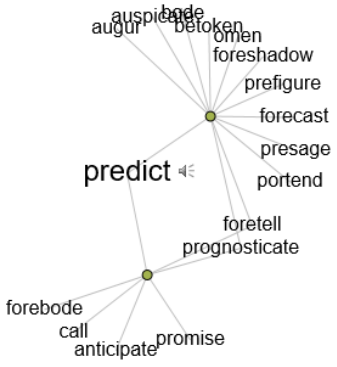
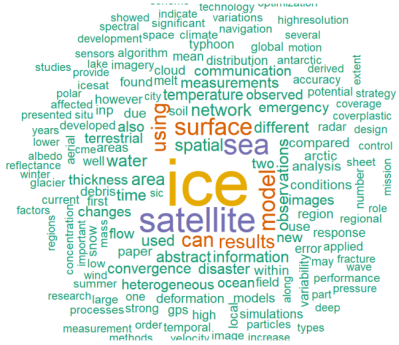
기존 방식으로 수행된 위성·지상 정보 융합 재난 대응 기술을 위한 특허 검색식은 [((정보) and (융합) or (재난) or (예측 or 대응))] or [((위성) and (지상) and (재난 or 안전) and (예측 or 대응))]의 다양한 유사어를 포함하여 작성되었으나, 여전히 특허 검색을 수행하는 주체의 기술에 대한 이해도와 몇몇 기술전문가의 기억력에 의존하게 되어 특허 검색에서 가장 중요한 부분인 검색식 구성에서 키워드가 누락될 가능성이 있다.

따라서 본 연구에서는 기존 방식을 포함하면서 연관 키워드를 활용하여 검색식을 확장하는 기법을 제안하고자 한다. 제안 기법에서는 연관 키워드를 도출하기 위하여 ① AI 키워드 추천, ② 유사어 제공 DB(visualthesaurus.com, wordsimilarity.com) 활용, ③ R 프로그래밍을 통한 TF & TF-IDF 결과 활용, ④ 기술분야 리딩 그룹의 유사 특허 별도 검색 방법을 적용하고자 한다. Table 3은 유사어 제공 DB(visualthesaurus.com, wordsimilarity.com) 및 R 프로그래밍을 통한 TF & TF-IDF 결과를 활용한 연관 키워드 도출에 대한 예시를 나타내고 있다.

논문 정보를 활용한 빅데이터 분석 기법

앞서 언급했던 것처럼 특허는 출원 1년 6개월 후에 공개되기 때문에 최신의 기술 변화를 반영하기 어려운 제도적 한계를 가지고 있다. 이러한 한계를 극복하기 위해서는 특허분석 결과에 최근의 기술적 변화를 반영할 수 있어야 하며, 특허 분석과 유사한 방법과 기준을 적용하기 위해서는 특허에 비해 조금 더 빠르게 기술적 변화를 감지할 수 있으면서 정량적인 분석을 적용할 수 있는 논문 정보를 활용하는 방법을 생각해 볼 수 있다. 논문은 발표와 동시에 공개되기 때문에 상대적으로 최근 정보를 반영할 수 있고, 유료 DB를 통해서만 확보할 수 있는 논문이 있는 반면, 한국과학기술정보연구원 (KISTI)에서 제공하는 Science ON(scienceon.kisti.re.kr)과 같이 무료로 활용할 수 있는 DB를 통해 부분적으로 기술 동향을 가늠해 볼 수 있다.

Table 3. Examples of related keyword derivation

visualthesaurus.com	wordsimilarity.com	TF & TF-IDF
	**Top 10 similar words or synonyms for satellite** terrestrial 0.723317 satellites 0.704552 gps 0.696245 gnss 0.683888 receivers 0.586278 geostationary 0.581637 receiver 0.580982 broadcast 0.576579 wifi 0.574702 wireless 0.572487	
	**Top 10 similar words or synonyms for disaster** disasters 0.744775 emergencies 0.646355 catastrophe 0.618946 accident 0.606509 crisis 0.598905 catastrophes 0.595486 earthquake 0.592666 outage 0.591425 outages 0.546964 emergency 0.546321	

LDA 토픽 모델링은 각 문서에 어떤 주제들의 존재를 나타내는 확률 모형으로, 문서 내의 숨겨진 주제를 찾아주는 알고리즘으로 주제별 문장 및 단어 확률을 파라미터를 통해서 도출하는 방식이다. 즉 단어로부터 전체 문서가 생성되어가는 과정을 확률 모형으로 모델링 한 것이다. LDA는 문서-단어 행렬을 입력데이터로 사용하고 문서에 포함되어 있는 단어들에 대한 디리클레 분포를 통해 토픽을 생성한다. 디리클레 분포를 기반으로 토픽별 단어의 분포와 문서별 토픽을 추정할 수 있다. 문서에 포함된 단어들의 분포를 확인할 수 있고, 이 정보를 바탕으로 문서와 주제, 주제와 단어의 베이지안 확률개념을 활용하여 토픽에 대한 문서, 단어의 확률값을 도출한다. 따라서 문서가 포함하고 있는 단어들을 보이지 않는 잠재된 변수를 통해 주제를 추론하는 것을 목적으로 하며, LDA 알고리즘을 통해 전체 문서의 주제들과 문서에 따른 주제 비율, 각 주제에 포함하는 단어들의 분포를 알 수 있다. LDA 토픽 모델링을 수행한 결과는 군집화된 특정되지 않은 토픽과 해당 토픽과 관련된 키워드의 구성으로 나타나게 된다.

본 논문에서는 해당 기술분야의 방향성을 판단하기 위한 목적으로 토픽 모델링을 제안하고자 하며, 의미 없이 사용되는 불용어(stop word)와 문장 부호 및 숫자 등을 제거하는 전처리 과정만 수행한 후에 적용하여 관련도가 높지 않은 일부 단어들도 포함되도록 하였다. 본 논문에서 제안하는 기법을 통해 특정 토픽에 할당된 키워드를 조합하여 대상 기술 분야의 의미 있는 방향성을 찾아낼 수 있고, 이를 기존의 특허 조사·분석을 통해 도출된 기술 공백 영역과 매칭하는 과정을 통해 유망기술을 예측할 수 있다. 구체적으로 본 논문에서는 최근 5년간 발표된 논문을 대상으로 LDA 토픽 모델링을 실시하여 토픽과 키워드를

도출하고, 특허 분석을 통해 도출된 기술 공백 영역과 일치하는 Case를 최신 기술 트랜드가 반영된 유망기술로 도출하는 기법을 제안하였다.

제안 기법의 유효성 분석

본 장에서는 위성·지상 정보 융합 재난 대응 기술을 중심으로 제안된 연관 키워드 도출을 통한 개선된 검색식 구성 기법과 논문 정보를 활용한 빅데이터 분석 기법의 유효성을 분석하고자 한다.

특허 검색 결과 및 유효·주요(핵심) 특허 비교 분석

기존 방법과 제안된 방법을 비교 분석하기 위하여 상용 특허 DB인 키워트(www.keywert.com) 시스템을 활용하여 검색하고 KIPRIS(www.kipris.or.kr)를 활용하여 보완하는 방법으로, 각 방식의 검색식을 특허 DB에 적용한 결과를 비교 분석하였으며 Table 4는 검색 결과를 비교하여 나타내고 있다. 또한, 노이즈 발생을 줄이기 위해 위성 활용 분야 IPC인 G08B와 영상처리 관련 IPC인 H04N, G06T로 한정하여 검색을 수행하였다.

Table 4. Patent search results reflecting search formula applying related keyword

구분	기존 방법	제안된 방법	
	검색건수	제안된 방법	추가된 연관 키워드
‘연관 키워드’ 도출	① 사전, 포털을 통한 유사어 검색 ② 특허 DB 활용 영문 ③ 키워드의 한글 표기법 확인 ④ 유사도 높은 특허 기술 용어 ⑤기술전문가 추천	⑥ AI 키워드 추천 ⑦ 유사어 DB ⑧ TF&TF-IDF ⑨ 유사특허 검색 합계	추정, 전망, estimation, SAR 등 14개 keyword crisis, tragedy, catastrophe 등 11개 keyword evacuation, cyclone, hybrid 등 27개 keyword city, highway, cctv 등 14개 keyword 66개 ‘연관 키워드’ 도출
검색 결과	744건 (KR) 320, (JP) 91, (US) 279, (EP) 54	1,633건 (KR) 727, (JP) 221, (US) 581, (EP) 104	

Table 4에서 보여진 바와 같이 기존 방식을 통해서는 총 744건의 모집단 특허가 검색되었고, 제안된 연관 키워드 방법을 활용한 특허 검색을 통해서는 총 1,633건의 모집단 특허가 검색되었다. 대상기술과 상이한 특허(노이즈)를 제거하는 필터링 과정을 통해 기존 방식의 경우 182건, 제안된 방식의 경우 562건의 유효특허가 선정되었다. 마지막으로, 유효특허를 전수 검토하는 방법으로 관련도가 높은 주요(핵심)특허 로서 기존 방식의 경우 19건, 제안된 방식의 경우 36건이 선정되었다. Table 5는 제안된 방법의 국가별 모집단과 유효 및 주요 특허에 대한분석 결과를 나타내고 있으며, Table 6은 기존 방식과 제안된 연관 키워드 방식의 특허 분석 결과를 비교하여 나타내고 있다. 기존 방법과 제안된 방법에 따른 검색 결과 및 유효·주요(핵심)특허를 비교하면, 검색식에 키워드가 누락됨에 따라 분석 결과에 영향을 미칠 수 있으며, 제안된 방법을 통해 분석결과가 매우 개선될 수 있음을 확인할 수 있다. 연관 키워드를 통해 17건의 주요(핵심)특허가 추가로 도출되어 유망기술 선정 과정에서 누락되는 특허 발생을 방지하였고, 이를 기반으로 이루어지는 O/S Matrix 분석 결과가 유망기술 예측과 직간접적으로 연결되어 있어 그 유효성을 확인할 수 있다. 다만, 주요(핵심)특허를 선별하는 과정에도 정성적인 관점이 반영되기 때문에, 연

관 키워드를 통해 추가로 도출되는 주요(유효)특허의 적정성을 정성적으로 확인하기 위해 기술전문가 중심으로 이를 검증하거나 연구개발(R&D) 과제 및 선행 연구결과와 비교분석하는 후속 연구가 이루어질 필요가 있다. 더불어 특허 빅데이터의 검색과 분석의 유효성 지표로 정확도(precision)와 재현율(recall)을 포함하여 결과를 제시하는 방법도 함께 검토되어야 할 부분이다.

Table 5. Result of selection of valid patent

특허청	모집단 특허(건)	유효특허(건)		주요(핵심) 특허	
		건수	비중(%)	건수	비중(%)
한국(KR)	727	269	47.9%	24	66.7%
유럽(EP)	104	17	3.0%	-	-
일본(JP)	221	91	16.2%	7	19.4%
미국(US)	581	185	32.9%	5	13.9%
합계	1,633	562	100%	36	100%

Table 6. Compare the impact of missing keywords

구분	기존 방법(A)	제안된 방법(B)	증감(B-A)
검색구간	최근 20년(2002.4.23.~2022.4.22.)		
모집단 특허	744건	1,633건	889건↑
유효특허	182건	562건	380건↑
핵심특허	19건	36건	17건↑

선별된 주요(핵심)특허를 기반으로 Table 7과 같이 목적(Object) / 해결방안(Solution) 매트릭스를 작성한 후 이를 활용하여 기술 공백 영역을 도출하였다. 가장 활발하게 출원이 이루어지고 있는 분야는 위성영상을 활용하여 접근이 어려운 지역의 재난을 예측하거나, 재난 예측의 정확성 제고를 목적으로 하는 분야인 것으로 분석되었다. 한편, 상대적으로 특허출원이 조사되지 않은 '이종 데이터 융합을 통한 실시간 예측·경보', '무인 비행체를 활용한 재난 예측 정확성 제고' 및 '피해 규모 산출 자동화' 등을 기술 공백 영역으로 도출되었다.

Table 8은 기술의 적용 분야(Target) 대비 수단(Solution) 매트릭스를 나타내고 있으며, 이를 통해 기술의 활용 분야 관점에서의 기술 공백 영역도 함께 확인하고자 하였다. 위성영상을 활용한 태풍, 지진, 화재, 홍수에 대한 예측 및 경보 관련 특허 출원은 활발하게 진행된 반면, '이종 데이터 융합'을 통한 재난의 예측이나 '무인 비행체'를 활용한 재난의 예측 등은 시급하게 특허 선점이 필요한 기술 공백 영역으로 도출되었다.

Table 7. Object / Solution matrix

목적/수단	복수(이종) 데이터 융합	위성영상 활용	GPS/GIS 정보 활용	무인비행체 및 항공사진 활용	위성통신(통신위성) 활용
접근이 어려운 지역 재난 예측/경보	-	[KR] 2018-0054169 [US] 13-800726 [JP] 2005-189979 [KR] 2014-0109792	[KR] 2020-0112287	[KR] 2018-0080961 [KR] 2016-0024134	-
재난 예측의 정확성 제고	[KR] 2018-0166857 [US] 11-815231 [KR] 2017-0173026 [JP] 2002-086659	[JP] 2020-151114 [KR] 2020-0035272 [KR] 2014-0093362 [JP] 2010-074047 [JP] 2002-254074 [KR] 2017-0057057	[JP] 2003-067376 [US] 10-291970 [KR] 2020-0017486	[기술 공백 영역]	-
재난 경보의 신속성 제고	[US] 10-017520	[KR] 2013-0151105	[US] 14-504691 [KR] 2014-0003334 [KR] 2019-0137721	[KR] 2016-0122516 [KR] 2020-0131097	[KR] 2005-0014562
실시간 경보/예측 시스템	[기술 공백 영역]	-	-	[KR] 2019-0037675 [KR] 2018-0138372	[KR] 2012-0110754
피해 규모 산출 자동화	[기술 공백 영역]	[JP] 2004-095493	-	[기술 공백 영역]	-
사각지대 해소 / 지속성 확보	[기술 공백 영역]	[기술 공백 영역]	[KR] 2019-0047760	[KR] 2019-0102764 [KR] 2016-0067649	[KR] 2019-0102858

Table 8. Target / Solution matrix

적용분야/수단	복수(이종) 데이터 융합	위성영상 활용	GPS/GIS 정보 활용	무인비행체 및 항공사진 활용	위성통신(통신위성) 활용
태풍	[기술 공백 영역]	[JP] 2020-151114 [KR] 2014-0109792	-	-	-
지진	[US] 10-017520	[JP] 2020-151114	-	-	[기술 공백 영역]
화재(산불)	[US] 11-815231	[KR] 2020-0035272 [KR] 2018-0054169 [KR] 2014-0093362 [KR] 2013-0151105	-	[기술 공백 영역]	-
홍수	-	[JP] 2005-189979 [KR] 2017-0057057	-	[기술 공백 영역]	[KR] 2012-0110754
싱크홀	-	[기술 공백 영역]	[KR] 2019-0047760	[KR] 2018-0080961	-
조난	-	-	[기술 공백 영역]	[KR] 2016-0122516	[기술 공백 영역]
산사태	-	[기술 공백 영역]	[JP] 2003-067376	[기술 공백 영역]	-
블랙아이스 (빙판)	[기술 공백 영역]	[기술 공백 영역]	[KR] 2020-0017486	[KR] 2020-0131097	-

논문 정보를 활용한 빅데이터 분석 기법의 유효성 결과

제안된 기법의 유효성을 분석하기 위해 Science ON에 등록된 최근 5년간 발표 논문을 대상으로 LDA 토픽 모델링을 실시

하였다. LDA 토픽 모델링은 R 프로그래밍의 'topicmodels' 패키지를 사용하여 수행하였다. 해당 알고리즘을 사용하여 LDA 토픽 모델링을 진행하게 되면 프로그램 내에서 자체적으로 최적의 파라미터 값이 자동으로 설정되어 분석이 수행된다. 본 연구에서는 불용어 처리 등 전처리를 통해 1차 정비된 데이터를 대상으로 LDA 토픽 모델링을 진행하였고, 이 과정에 결정이 필요한 군집 수(K)는 10으로 설정하여 분석을 진행하였다. Table 9는 LDA 토픽 모델링 결과를 나타내고 있으며, 각 Topic 별로 10개의 키워드가 도출된 것을 확인할 수 있다. Table 10은 특허 분석 결과와 논문 정보 분석 결과를 비교하여 유망 기술을 도출한 결과를 나타내고 있다. 첫 번째 컬럼은 특허 분석을 통해 도출된 기술 공백 영역 (A)를 나타내고 있으며, 두 번째 컬럼은 LDA 토픽 모델링을 통해 도출된 키워드를 조합하여 작성된 Topic (B)을 나타내고 있다. 세 번째 컬럼은 특허 분석 결과(A)와 논문 정보 분석 결과(B)를 'and' 조합을 통해 유망기술 여부를 판단한 결과를 나타내고 있다. 최근 5년간 발표된 논문에서 LDA 토픽 모델링을 통해 도출된 토픽을 매트릭스를 통해 도출된 기술 공백 영역과 매칭하면, ① 복수(이종) 데이터 융합을 통한 실시간 태풍 및 블랙아이스 대응 기술, ② 위성영상을 활용한 블랙아이스 대응 기술, ③ 위성과 무인 비행체를 활용한 재난 예측 정확성 제고 및 피해 규모 산출 기술, ④ 위성 통신을 활용한 조난 대응 기술을 유망기술로 예측할 수 있다.

Table 9. LDA Topic Modeling results for title of 98 papers in the last 5 years

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
satellite	ice	disaster	ice	based
ice	arctic	sea	using	realtime
learning	sea	heterogeneous	surface	data
deep	different	ice	estimation	analysis
antarctica	pacific	physical	sheet	prediction
spatial	ice nucleating	summer	disaster	typhoon
expansion	variability	black	satellite	safety
long term	high altitude	transfer	model	crime
sea	implications	based	safe	measurements
model	angles	data	debris	case
Topic 6	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
ice	ice	emergency	ice	ice
sea	disaster	satellite	remote	sea
changes	safety	networks	sensing	using
snow	effect	communication	satellite	estimation
surface	the	design	reflectance	high resolution
the	aircraft	evacuation	using	deformation
elevation	unmanned	constellation	concentration	design
satellite	network	centers	part	satellite
space	study	data	sheet	convergence
arctic	satellite	post disaster	optimization	model

Table 10. Prediction of promising technologies through matching of topic modeling results and white space of patent

기술 공백 영역 (A)		최근 5년 논문 Topic (B)	유망기술 판단 (A and B)
복수(이종) 데이터 융합	실시간 경보/예측	• realtime typhoon safety prediction based data measurements and analysis	유망기술
	피해 규모 산출	-	공백기술
	사각지대 해소	-	공백기술
	태풍, 블랙아이스	• realtime typhoon safety prediction based data measurements and analysis • black ice disaster based heterogeneous data	유망기술
위성영상 활용	사각지도 해소	-	공백기술
	싱크홀, 산사태, 블랙아이스	• black ice disaster based heterogeneous data • satellite remote sensing using ice reflectance	유망기술
GPS/GIS 정보활용	조난	-	공백기술
무인비행체 항공사진 활용	재난 예측 정확성	• high resolution estimation using satellite convergence model	유망기술
	피해 규모 산출	• disaster effect study using unmanned aircraft	유망기술
위성통신 활용	지진, 조난	• emergency evacuation using satellite network communication and design post disaster	유망기술

결 론

본 논문에서는 기존의 유망기술 도출 연구에 있어서 대상 특허를 선별하는 과정에서 검색식이 충실하게 구성되지 않아 중요 특허가 분석 대상에서 누락되는 문제와 출원 미공개 기간에 변화하는 최신 기술 트랜드를 반영하지 못하는 문제들을 개선하기 위한 새로운 기법들을 제시하였다.

먼저, 핵심 키워드가 특허 검색 과정에서 누락되는 것을 방지하기 위해 AI 활용 키워드 추천, 유사어 제공 DB 활용, 최근 논문의 Keyword, TF & TF-IDF 등을 활용하여 연관 키워드를 도출하여 검색식에 반영한 후 특허 검색을 실시하였고, 특허 출원 후 1년 6개월의 미공개 기간 때문에 가장 최근의 기술적 변화를 반영하기 어려운 특허 제도의 약점을 보완하기 위해, 최근 5년간 발표된 관련 분야 논문을 대상으로 LDA 토픽 모델링을 수행하여 최근 연구 동향을 확인할 수 있는 토픽과 키워드를 도출하였다. TF & TF-IDF와 LDA 토픽 모델링은 텍스트 마이닝 기반의 빅데이터 분석에 적합한 R 프로그래밍을 이용하여 수행하였다. 제안된 기법들을 통해 위성·지상 정보 융합 재난 대응 기술을 특허적 관점에서 조사 및 분석하였고, 향후 연구개발을 통해 특허 선점이 필요한 공백 기술 분야를 도출하였다. 또한 최근 5년간 발표된 논문 정보를 활용한 LDA 토픽 모델링 결과를 매칭하는 방법으로 위성·지상 정보 융합 재난 대응 기술 분야 유망기술을 도출하였다. 도출된 위성·지상 정보 융합 재난 대응 기술분야의 유망기술은 ① 복수(이종)데이터 융합을 통한 실시간 태풍 및 블랙아이스 대응 기술, ② 위성영상을 활용한 블랙아이스 대응 기술, ③ 위성과 무인 비행체를 활용한 재난 예측 정확성 제고 및 피해 규모 산출 기술, ④ 위성 통신을 활용한 조난 대응 기술 등이 예측되었다.

본 논문을 통해 제시된 유망기술의 적정성을 판단하기 위해 정부 중장기 정책 및 NTIS 연구개발 과제 동향을 검토한 결과, 제2차 위성정보 활용 종합계획의 "위성·레이더 등 다중정보 융합을 통한 예보지원 강화 및 위험기상 진단 정확도 향상", "천리안 2A/2B호 및 지상관측자료 융합을 통한 안개 및 황사 탐지 성능 향상"에서 본 논문에서 도출된 유망기술과 일치하는 정

책 방향을 발견할 수 있었고, 한국교통연구원에서 발표한 현황자료에서 적설 교통사고에 비해 블랙아이스 교통사고에서 사망자가 3.7배 많이 발생하는 유효한 통계분석 결과를 통해, 이에 대응하기 위한 연구개발이 시급하게 이루어질 필요가 있음을 확인할 수 있다. 또한, NTIS를 통해 연구개발 현황을 조사한 결과 블랙아이스(도로 살얼음) 대응 기술이 2017년 3개 과제를 시작으로 2020년 15개, 2021년 25개 과제로 급격하게 증가하는 추세를 보이고 있고, 연구비도 이에 비례하여 2017년 368백만원에서 2021년 7,750백만원으로 증가한 것에서, 블랙아이스 검지 및 예측을 통한 재난·안전 대응 기술에 대한 연구개발이 지속적으로 증가하고 있는 것을 확인할 수 있었다. 최근 발표된 439.5억원 규모의 위성정보 빅데이터 활용 지원 사업 관련 보도자료에서 '위성정보를 활용한 재난대응지원'을 포함하고 있어, 2022년 이후에도 위성정보를 활용한 보다 다양한 연구개발이 활발하게 이루어질 것으로 예측할 수 있다.

본 논문을 통해 제안하는 기법을 활용한 유망기술 예측 과정과 결과를 통해 해당 기술분야의 기술 동향과 핵심 키워드를 확인할 수 있고, 새로운 연구개발과 관련하여 유용하게 사용될 수 있는 방향성을 제공하였다. 연구개발 과제를 기획하는 연구자에게는 향후 연구개발을 위한 아이디어를 제공하고, 국가 R&D 정책 및 사업을 기획하는 정부 및 유관기관에게는 유망기술 분야 선정 또는 산업 지원 정책 방향 수립에 도움을 줄 수 있을 것으로 기대한다.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (NRF-2021R1F1A1049509). The present research has been conducted by the Excellent researcher support project of Kwangwoon University in 2022.

References

[1] Cho, S. (2010). Development of Precipitation Prediction Model using Wide Area Satellite Image and Ground Observation Network. Master's Thesis, Kyungpook National University.

[2] Choi, D., Kim, G., Park, S., Jang, D. (2013). "Prediction of AMOLED floating technology using quantitative analysis of patent keywords based on text mining." Korea Contents Association, 2013 Spring Comprehensive Conference, pp. 365-366.

[3] Cooperation of Related Ministries (2018). 2nd Comprehensive Plan for the Use of Satellite Information in Korea.

[4] Cooperation of Related Ministries (2021). 2nd Comprehensive Plan for Satellite Information Utilization in 2021.

[5] Garzaniti, N., Tekic, Z., Kukolj, D., Golkar, A. (2021). "Review of technology trends in new space missions using a patent analytics approach." Progress in Aerospace Sciences. Vol. 125, 100727.

[6] Go, B., Kim, H. (2015). "A study on the disaster recognition technology of the river using satellite images." Korea Multimedia Society, Vol. 19. No. 4, pp. 11-17.

[7] Im, H., Cho, S., Quan, J., Kim, J., Lee, C., Lee, K. (2017). "Prediction of future aviation security technology through patent big data cluster analysis." Journal of the Korean Society for Aviation Management, Vol. 15, No. 5, pp. 43-63.

[8] Jang, J., Zhang, J., Li, L., Kwon, H., Lee, C. (2013). "A study on the efficient extraction method of patent search key words using big data analysis tool R." 2013 Fall Korean Society of Safety Management Conference, pp. 387-391.

[9] Jeon, S. (2011). "Technology prediction of intelligent system using patent analysis." Korea Intelligent Systems Association, Vol. 21, No. 1, pp. 100-105.

[10] Jin, H., Lee, S. (2012). "Patent trend report about oriental medicine: The Korea institute of oriental medicine-oriented." Journal of the Korea Contents Association, Vol. 12, No. 6, pp. 223-229.

[11] Kang, J. (2017). Derive Promising Technologies for Cloud Computing using Patent Topic Modeling and Simultaneous Citation Analysis." Master's Thesis, Seoul National University of Science and Technology.

[12] Kim, S., Choi, J., Goo, S., Park, Y. (2012). "Heavy snowfall disaster response using multiple satellite imagery information." Journal of Korean Society for Geospatial Information Science, Vol. 20. No. 4, pp. 135-143.

[13] Kim, S., Goo, S., Park, Y. (2012). "Time-critical disaster response by cooperating with international charter." Journal of Korean Society for Geospatial Information Science, Vol. 20, No. 2. pp. 109-117.

[14] Kim, S., Noh, H., Lee, J., Kim, J., Kim, J. (2020). "Remote exploration and spatial information utilization technology for disaster management on the Korean Peninsula." Korean Journal of Remote Sensing, Vol. 36, No.5-4, pp. 1139-1151.

[15] Kim, W., Kim, D., Choi, Y. (2010). "A study on application limitation of AHP priority vector with expert measurement." Journal of the Korean Air Navigation Society, Vol. 18, No. 3, pp. 92-98.

[16] Kwon, T. (2008). "Critical issues in applying multi-criteria analysis into feasibility studies of public projects." Korean Society of Public Management, Vol. 22, No. 3, pp. 31-51.

[17] Lee, E. (2019). Derive Core Technologies related to Digital Healthcare using Patent Information from Domestic and abroad: Focusing on Text Mining. Doctor's Thesis, Kyunghee University.

[18] Lee, H., Kim, Y. (2021). "A study on white space search of wireless signal based passive tracking technology using enhanced search formula of patent analysis." Journal of The Korean Society of Disaster Information, Vol. 17, No. 4, pp. 802-816.

[19] Lee, H., Yu, J., Kwon, K., Lee, C. (2021). "Forecast of promising technology on last mile logistics by analyzing the patent information." Korean Society of Logistic Studies, Vol. 29, No. 2, pp. 43-56.

[20] Lee, J., Lee, M., Kim, J. (2019). "A study on Korean language processing using TF-IDF." Journal of Information Systems, Vol. 28, No. 3, pp. 105-121.

[21] Lee, T., Woo, B., Lee, J. (2012). "Selection of promising technologies based on the integrated application model of Delphi, AHP and patent analysis: Application to automotive parts industry." Korean Society of Corporate Management, Vol. 19, No. 6, pp. 283-303.

[22] Ministry of Science and ICT (2018). 3rd Basic Plan for the Promotion of Space Development.

[23] Park, G., Lee, S., Kim, E., Yun, B. (2017). "A case study on meteorological analysis of freezing rain and black ice formation on the load at winter." Journal of Environmental Science International, Vol. 26, No. 7, pp. 827-836.

[24] Yun, S., Han, K. (2018). "A study on patent data analysis and competitive advantage strategy using TF-IDF and network analysis." Journal of Digital Contents Society, Vol. 19, No. 3, pp. 529-535.

[25] Yun, Y., Kim, N., Choi, J., Park, J., An, D., Kim, Y. (2016). "Study on trend analysis of indoor evacuation path guidance system for disaster situation." Journal of The Korean Society of Disaster Information, Vol. 12, No. 2, pp. 130-135.

[26] Zhang, J., Jang, J., Kim, S., Lee, H., Lee, C. (2013). "A study on the efficient patent search process using big data analysis tool R." Journal of the Korea safety Management and Science. Vol. 15, No. 4. pp. 289-294.