텍스트 마이닝을 이용한 국내 건설기술정책 트렌드 분석

Trends Analysis on Construction Technology Policy using Text-mining

홍심희* 유정호**
Hong, Sim-Hee, Yu, Jung-Ho

키워드 : 건설, 트렌드분석, 워드 클라우드, 텍스트 마이닝
Keywords : Construction, Trend Analysis, Word Cloud, Text Mining

1. 서론

1.1 연구의 배경 및 목적

2016년 다보스포럼 이후로 4차 산업혁명에 따른 여러 기술은 많은 산업분야에 적용되고 있다. 4차 산업혁명은 개인과 사물에 네트워크를 부여하여 초연결성(Hyper-Connected)과 초지능화(Hyper-Intelligent)라는 트렌드를 만들어 냈으며, 다양한 산업구조 변화에 영향을 미쳤다. 또한, 4차 산업혁명은 폭발적인 데이터 생산에 기여하여 매년 대규모의 빅 데이터(big data)를 만들어내고 있다. 빅 데이터는 4차 산업혁명 기술의 기초가 되는 정보 및 정보기술로, 산업에서의 트렌드를 분석하여 전략수립, 리스크 관리 등 다양한 효과를 만들어 낼 수 있다.

건설에서도 AI, IoT, 드론 등 4차 산업혁명에 따른 신기술 적용이 활발하게 진행되고 있으며, 기술적용에 대한 니즈(Needs)가 꾸준히 증가하고 있다. 이에 따라 건설 역시 다양한 트렌드 가운데 건설에서 중요도 높은 이슈를 선정하고 이에 따른 미래 전략을 수립할 필요성이 있다. 따라서 본 연구에서는 국토교통부의 건설기술진흥 기본계획을 바탕으로 지난 10년간의 국내 건설기술정책 주요 이슈를 살펴보고, 이를 통해 시사점을 제시하고자 한다.

1.2 연구의 범위 및 방법

본 연구에서는 지난 10년간의 국내 건설기술정책의 트렌드 변화를 살펴보기 위해 국토교통부의 제4차(2008-2012년), 제5차(2013-2017년), 제6차(2018년-2022년) 건설기술진흥 기본계획과 2008년부터 2019년까지의 국토교통부 업무보고를 분석하였다. 학회지, 협회지, 신문 등은 간행물을 출판한 기관의 관점이 반영되었을 가능성이 높기 때문에 분석에서 제외하였다.

본 연구에서는 국내 건설기술정책 주요 이슈를 분석하기 위해, 파이썬 라이브러리 중 자연어처리가 가능한 KoNLPy 라이브러리로 텍스트 마이닝을 수행하였다. 건설기술진흥 기본계획의 기간에 따라 분석결과를 구분하였으며, 각각의 문서 별로 빈도분석을 실시하여 주요한 키워드를 도출하였다.

2. 현황고찰

2.1 텍스트 마이닝(Text Mining)

텍스트 마이닝(Text Mining)이란 텍스트를 분석하여 고품질의 정보를 도출하는 프로세스로, 일반적으로 텍스트 분류, 개념추출, 단어빈도분석, 패턴추출, 예측분석 등을 위해 주로 사용된다(Wikipedia, 2019). 텍스트 마이닝은 네트워크 분석(Network Analysis), 토픽분석(Topic Analysis) 등과 결합되어 사용되며, 결과를 표현하기 위해 다양한 시각화 기법이 적용된다.

다양한 시각화 기법 중 워드 클라우드 다이어그램이란 빈도에 따라 단어의 크기를 다르게 표현하는 시각화 기법으로, 소셜 네트워크 분석(Social Network Analysis, SNA) 등에 주로 활용된다(서민구 외1인, 2019). 워드 클라우드 다이어그램을 활용할 경우, 빈도를 기반으로 한 직관적인 트렌드 파악이 가능하며, 문서의 핵심 키워드와 핵심 개념을 직관적으로 파악할 수 있다.

* 일반회원, 광운대학교 대학원 건축공학과 석사과정, hshgl9786@kw.ar.kr
** 종신회원, 광운대학교 대학원 건축공학과 교수, 공학박사(교신저자), myazure@kw.ac.kr

2.2 기존문헌 고찰

텍스트 마이닝과 관련하여 건설 분야에서 수행된 기존연구는 표1과 같다. 먼저, 서민구 외 1인(2019)은 국내외 BIM 연구 동향을 파악하기 위해 텍스트 마이닝 수행하였으며, 관련된 결과를 워드 클라우드 다이어그램으로 표현하였다. Quian Chen et. al.(2018)은 건설 자동화 트렌드를 파악하기 위해 텍스트 마이닝을 적용하여 이를 통해 시사점을 도출하였다. Zhikun Ding et. al.(2018)은 텍스트 마이닝 기법을 활용하여 건물 에너지 분야의 연구 동향을 파악하였다. 김다솜 외 7인(2015)은 특정기술과 관련된 이슈를 도출하기 위해, 임시영 외 3인(2014)은 공간정보 분야의 연구동향을 파악하기 위해 텍스트 마이닝을 활용하였다. 마지막으로, 정철우외 1인(2012)은 텍스트 마이닝과 수학적 알고리즘을 활용하여 건설 트렌드 분석을 수행하였다. 이와 같이 텍스트 마이닝은 특정 주제의 특징과 트렌드를 파악하기 위해 주로 활용되었다.

표 1. 건설 분야에서의 텍스트 마이닝 관련 기존문헌 고찰

저자(연도)	내용
서민구 외 1인 (2019)	국내외 BIM 연구 동향을 파악하기 위해 텍스트 마이닝을 수행하였으며, 이에 따른 결과를 위드 클라우드 다이어그램으로 제시함.
Qian Chen et. al. (2018)	건설 자동화와 관련된 트렌트를 파악하기 위해 텍스트 마이닝을 활용하여 관련 패턴을 도출하였음.
Zhikun Ding et. al. (2018)	건물 에너지 절약과 관련된 연구 동향을 분석하기 위해 텍스트 마이닝을 활용하여 관련 결과를 도출하였음.
김다솜 외 7인 (2015)	특정 과학기술과 관련된 이슈를 도출하기 위한 텍스트 분석 기법 및 결과를 제시함.
임시영 외 3인 (2014)	국가 연구개발 보고서 및 논문에서 추출된 키워드를 분석하여, 공간정보 분야의 연구 동향을 파악하였음.
정철우 외 1인 (2012)	텍스트 마이닝과 TF-DI지표를 활용하여, 건설 트렌트 분석을 수행하였음.

3. 텍스트 마이닝을 활용한 국내 건설기술정책 트렌드 분석

3.1 텍스트 마이닝을 활용한 트렌드 분석 프로세스

본 연구는 다음과 같은 텍스트 마이닝 프로세스를 적용하여 키워드의 빈도분석을 수행하였다. 먼저, 파이썬 라이브러리 중 KoNLPy를 이용하여 수집한 pdf문서들의 데이터 전처리를 수행하였다. 데이터 전처리 과정에서 건설, 기술, 관리, 추진, 기준, 개발, 강화, 산업, 사업, 체계, 구축, 계획, 성격, 일정, 활용, 방안, 과제, 제도 등과 같이 건설과 관련된 일반적인 키워드는 제외하였다. 두 번째, 각 문서별로 빈도분석을 실시하여 빈도가 높은 키워드를 1위에서 50위까지 도출하였다. 세 번째, 1위에서 150위까지의 키워드 빈도를 워드 클라우드 다이어그램으로 표현하였다. 네 번째, 제4차, 제5차, 제6차 건설기술진흥 기본계획에서 도출된 결과를 비교하였으며, 업무보고와 비교하여 시기별 시사점을 도출하였다.

3.2 워드 클라우드 다이어그램을 활용한 국내 건설기술정책 트렌드 분석 결과

워드 클라우드 다이어그램을 활용한 시각화 결과는 다음과 같다. 먼저 해당문서들을 분석한 결과, 안전, 평가, 인력, 환경, 설계, 해외 등이 공통적으로 빈도 높은 키워드로 등장함을 확인할 수 있었다. 이에 따라, 안전, 인력, 환경 등의 이슈가 지속적

으로 논의되고 있음을 확인할 수 있었으며, 각 차별로 빈도의 순위는 다르지만 해외진출에 대한 이슈 역시 지속적으로 논의되고 있음을 확인할 수 있었다.

두 번째, 제 4차 건설기술진흥 기본계획(2008년-2012년)에서는 안전, 기관, 평가, 인력, 환경, 설계, 공사, 정보, 시설, 정책 등이 빈도 높은 키워드로 등장하였다. 2008년부터 2012년까지의 정책수행에 따른 업무보고를 분석하면, 환경과 관련하여 2011년도까지 4대강이라는 키워드가 등장하였으며, 2012년에는 세종특별자치시 출범과 더불어 세종시 이전이 본격화 되면서 세종시라는 키워드가 등장하는 것이 확인되었다. 이에 따라, 2008년부터 2012년까지는 안전, 인력양성, 해외진출, 환경 등과 같은 일반적인 건설이슈와 더불어 환경개선 정책 등이 수행되었음을 확인할 수 있었다.

그림 1. 제 4차 건설기술진흥 기본계획의 워드 클라우드 다이어그램

다음으로 제 5차 건설기술진흥 기본계획(2013년-2017년)에는 안전, 시설, 정보, 설계, 인력, 정책, 해외, 평가, 확대, 역량 등의 키워드 빈도가 높은 것으로 확인되었다. 또한, 2013년부터 공공기관의 녹색건축물 인증취득이 의무화되면서 녹색이라는 키워드가 등장하였다. 2013년부터 2017년까지의 업무보고를 분석하면 2017년에는 포항 지진으로 인해 지진이라는 키워드가 등장함을 확인할 수 있었다. 이에 따라, 2013년부터 2017년까지는 안전, 인력양성, 해외진출, 환경 등과 같은 일반적인 건설이슈와 더불어 환경정책들이 수립되었으며, 재난안전에 대한 이슈가 부각되었음을 확인할 수 있었다.

그림 2. 제 5차 건설기술진흥 기본계획의 워드 클라우드 다이어그램

마지막으로 제 6차 건설기술진흥 기본계획(2018년-2022년)에는 해외, 정보, 역량, 기술자, 안전, 평가 기업, 설계, 시장 등의 키워드 빈도가 높은 것으로 확인되었다. 또한, 스마트, 산업혁명, 혁신 등의 키워드가 새롭게 등장하였다. 국토교통부의 업무보고를 분석해보면, 제5차 진흥계획이 종료되는 시기인 2017년부터 스마트라는 키워드가 등장하였으며, 제6차 진흥계획이 시작되는 시기인 2018년에는 스마트, 첨단 등과 같은 키워드의 빈도가 증가하였음을 확인할 수 있었다. 이에 따라, 2018년부터 2022년까지는 안전, 인력양성, 해외진출, 환경 등과 같은 일반적인 건설이슈와 더불어 첨단기술 적용을 위한 기술자의 역량에 대한 이슈가 강화되었으며, 4차 산업혁명 기술 적용을 통한 부가가치 창출에 대한 이슈가 논의됨을 확인할 수 있었다.

그림 3. 제 6차 건설기술진흥 기본계획의 워드 클라우드 다이어그램

4. 결론

본 연구에서는 국토교통부의 건설기술진흥 기본계획을 바탕으로 지난 10년간의 건설 이슈를 도출하고 이를 통한 시사점을 제시하였다. 지난 10년간 건설은 주요한 정책과 맞물려 발전되어왔다. 2008년부터 2017년까지는 주로 안전, 인력양성, 해외진출, 환경 등과 같은 이슈와 맞물려 건설 발전이 이루어졌다. 2016년에 4차 산업혁명과 관련된 이슈가 본격적으로 논의되면서, 2018년부터 수립된 제 6차 건설기술기본진흥계획에는 스마트, 첨단, 4차 산업혁명 등과 같은 이슈가 본격적으로 등장하였다. 이에 따라, 건설에서도 4차 산업혁명기술 적용을 위한 세부전략들이 수립되고 있으며, 주요한 4차 산업혁명 기술과 건설간의 관계를 분석하여 이를 통해 전략을 수립할 필요가 있다.

본 연구는 정부에서 발생한 문서에 한정하여 건설트렌드를 분석하였음을 한계로 갖는다. 이에 따라, 향후연구에서는 보다 넓은 관점의 자료를 수집할 필요가 있으며, 건설 전반의 트렌드가 아닌 현재 다양한 연구가 진행되고 있는 4차 산업혁명 기술을 선정하고 이에 따른 트렌드를 제시할 필요가 있다. 또한, 수집한 자료를 정량적으로 비교할 수 있는 수학적 알고리즘을 구축할 필요가 있다.

감사의 글

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017R1A2B4012228).

참고문헌

김다솜, William Xiu Shun Wong, 임명수, 류신, 김남규, 박준형, 길우영, 윤한술 (2015). “텍스트 분석을 활용한 과학기술 이슈 여론 분석 방법론” 한국IT서비스학회, 14(3), pp. 33-48.

서민구, 이웅균 (2019). “워드클라우드를 이용한 국내외 BIM 연구동향 분석” 한국건축시공학회지, 19(1), pp. 95-103.

임시영, 이미숙, 진기호, 신동빈 (2014). “텍스트마이닝 기술을 이용한 공간정보 분야의 연구 동향에 관한 고찰 -국가 연구개발 사업 보고서 및 논문을 중심으로-” 한국공간정보학회지, 22(4), pp. 11-20.

정철우, 김재준 (2012). “텍스트마이닝을 활용한 건설분야 트랜드 분석” 한국디지털건축인테리어학회 논문집, 12(2), pp. 53-60.

Qian Chen, Borja Garcia de Soto, Bryan T. Adey (2018). “Construction Automation: Research areas, industry concerns and suggestions for advancement”, Automation in Construction, 94, pp. 22-38.

Zhikun Ding, Zongjie Li, Cheng Fan (2018). “Building energy savings: Analysis of research trends based on text mining”, Automation in Construction 96, pp. 398-410.

Wikipedia (2019). <https://en.wikipedia.org/wiki/Text_mining>