

온라인 빅 데이터 분석 결과와 상수도 통계 비교를 통한 데이터 가치 추출

Data value extraction through comparison of online big data analysis results and water supply statistics

홍성진*, 유도근**

Sungjin Hong, Do Guen Yoo

요 지

4차 산업혁명의 도래로 사회기반시설물의 계획 및 운영관리에 있어 데이터 분석을 통한 가치 추출에 대한 관심은 매우 높은 상황이다. 데이터의 가용성과 접근성, 정부 지원 등을 평가하는 공공데이터 개방지수에서 한국은 1점 만점에 0.93점을 획득하여 경제협력개발기구 회원국 중 1위(2019년 기준)를 할 정도로 매우 높은 수준(평균 0.60점)이다. 그러나 공식적으로 발표 및 배포되는 사회기반시설물 관련 정보와 심도 있는 연구 분석이 필요한 정보는 접근이 여전히 제한적이라 할 수 있다. 특히 대표적인 사회기반시설물인 상수도시스템은 대부분 국가중요시설로 지정되어 있어 다양한 정보를 획득하고 분석하는데 제약이 존재하며, 관련 국가통계인 상수도통계에서는 누수 사고 등과 같은 비정상적 상황에 대한 사고지점, 원인 등과 같은 세부정보는 제공하고 있지 않다.

본 연구에서는 웹크롤링 및 빅데이터 분석기술을 활용하여 과거 일정기간 발생한 지자체의 상수도 누수사고 관련 뉴스를 전수조사하고 도출된 사고건수를 국가 공인 정보인 상수도통계자료와 비교·분석하였다. 독립적인 누수사고 기사를 추출하기 위해서 중복기사의 제거, 누수 관련 키워드 정립, 상수도분야 이외의 관련기사 제거 등의 절차가 필요하며, 이와 같은 기법은 R프로그래밍을 통해 구현되었다. 추가적으로 뉴스기사의 자연어 처리기반 정보추출기법을 통해 누수사고 건수 뿐만 아니라 사고발생일, 위치, 원인, 피해정도, 그리고 대상 관로의 크기 등을 획득하여 상수도 통계에서 제시하고 있는 정보보다 많은 가치를 추출하여 연계할 수 있는 방안을 제시하였다. 제시된 방법론을 국내 A광역시에 적용하여 누수사고 건수를 비교한 결과 상수도통계에서 제시하고 있는 누수발생건수와 유사한 규모의 사고건수를 뉴스기사분석을 통해 도출할 수 있었다. 제안된 방법론은 추가적인 정보의 추출이 가능하다는 점에서 향후 활용성이 높을 것으로 기대된다.

핵심용어 : 빅데이터, 가치추출, 상수도통계, 누수사고

감사의 글

본 결과물은 환경부의 재원으로 한국환경산업기술원의 지능형 도시수자원 관리사업의 지원을 받아 연구되었습니다(2019002950002). 이에 감사드립니다.

* 정회원 · 수원대학교 공과대학 토목공학과 석사과정 · E-mail : yd1163124@naver.com

** 정회원 · 수원대학교 공과대학 건설환경에너지공학부 조교수 · E-mail : dgyoo411@suwon.ac.kr