

소셜 빅데이터를 활용한 건축안전 분야 증거기반정책 도입 방안

조영진 부연구위원, 김신성 연구원

요약

- 건축·도시 분야는 민원이 다발하는 대표적인 분야이며, 건축·도시를 둘러싼 정치·사회·경제 등 외부 환경에 대한 변화도 매우 빨라 증거기반정책 도입이 필수적
- 소셜 빅데이터 분석은 기존의 설문조사, 인터뷰조사를 통해서 수행되는 여론 모니터링을 대체할 수 있는 방법으로, 일련의 절차를 수립하면 실시간 여론 파악 가능
- 지난 3년간의 소셜 빅데이터를 수집·가공·분석하여 건축안전 분야의 여론 변화와 동향을 분석하였으며, 이를 통하여 건축안전 분야의 정책의제를 도출
- 분석 결과 건축물 화재안전에 대한 여론의 요구가 높고 내진설계에 대한 대책이 시급하며 건축물 구조 안전에 관한 국민의 불안감 해소가 필요한 것으로 나타남

정책제안

- 건축안전 분야의 소셜 빅데이터 분석을 통해 도출한 정책의제는 화재안전·내진설계·구조안전이며, 이는 객관적·과학적인 증거기반의 설명이 가능
- 빅데이터를 활용한 증거기반정책 개발은 기존 의견기반정책 개발과 비교할 때 장단점이 있어, 이를 고려하여 정책 주제에 따라 추진하는 것이 바람직함
- 건축·도시 분야의 빅데이터 활용은 증거기반정책뿐만 아니라 스마트시티, 스마트빌딩 등에 대한 요소 기술이기 때문에 관련 분야의 지속적 연구와 정책적 지원 필요

1 증거기반 건축 · 도시 정책 수립의 필요성과 빅데이터 활용 가능성

■ 건축 · 도시 분야에서 증거기반정책(Evidence-Based Policy) 도입의 필요성

- 우리 사회가 정보화 사회로 진입함에 따라 정책도 통념, 관행, 경험, 직관에 의존했던 의견기반정책에서 객관적 데이터에 기반을 둔 증거기반정책으로 변화 필요
 - 의견기반정책은 정책 개발 실무자와 정책 결정자의 능력에 따라 정책의 품질이 결정되며, 따라서 실무자와 결정자의 전문성이 높을수록 유용한 수단임
 - 건축 · 도시 분야는 전문 분야가 광범위하여 정책 개발에 분야별 전문성이 요구되나 공무원 직무순환제를 운영하고 있는 우리나라에서 이를 기대하기 어려운 실정
- 증거기반정책은 국민의 관심이 높고, 다수 민원이 발생하며, 정책 수요가 빠르게 변하는 건축 · 도시 분야에 유효함
 - 증거기반정책은 정책 개발 과정에서 의사 결정이 과학적이고 객관적으로 이루어져 신뢰도가 높은 정책 개발이 가능하고, 정책에 대한 정량적 평가와 환류가 용이
 - 건축 · 도시 분야는 민원이 다발하는 대표적인 분야이며, 건축 · 도시를 둘러싼 정치 · 사회 · 경제 등 외부환경에 대한 변화도 매우 빨라 증거기반정책 도입이 필요

■ 증거기반정책의 정량 데이터로서 소셜 빅데이터¹⁾의 활용 가능성

- 전통적으로 여론 모니터링을 위하여 설문조사, 전문가 인터뷰 등을 활용하였으나, 증거기반정책을 위해서는 보다 객관적 · 과학적인 방법론이 필요
 - 기존 의견기반정책에서는 설문조사, 전문가 인터뷰 등을 활용하여 여론을 모니터링하고 정책을 발굴하였으나 설문조사는 샘플링, 전문가 인터뷰는 객관성에 문제가 대두
 - 소셜 데이터는 블로그, 트위터, 페이스북 등 1인 미디어로 여론을 내포하고 있어, 이를 종합한 소셜 빅데이터 분석을 통하여 객관적 · 과학적 여론 모니터링 가능

■ 소셜 빅데이터 분석을 통한 여론 모니터링과 정책의제 도출

- 소셜 빅데이터 분석은 실시간 조사가 가능하여, 기존 여론조사, 전문가 인터뷰에 비하여 빠르고 경제적으로 여론 모니터링 및 정책의제 도출이 가능
 - 소셜 빅데이터는 사람들 간의 인적 네트워크를 기반으로 발생하는 텍스트, 이미지 등의 데이터를 말하며, 크게 SNS · 블로그 · 커뮤니티 빅데이터로 구분할 수 있음
 - 소셜 빅데이터 분석을 위해서는 데이터의 수집 · 가공 · 분석 · 시각화 등 일련의 과정을 거치며, 각각의 과정에 따라 알고리즘을 설계하고 프로그램을 개발하고 분석을 수행

1) 빅데이터란 기존 데이터에 비해 규모가 크고, 복잡 다양하며, 생성 · 처리 속도가 빨라 새로운 관리 및 분석 방법이 요구되는 데이터를 지칭

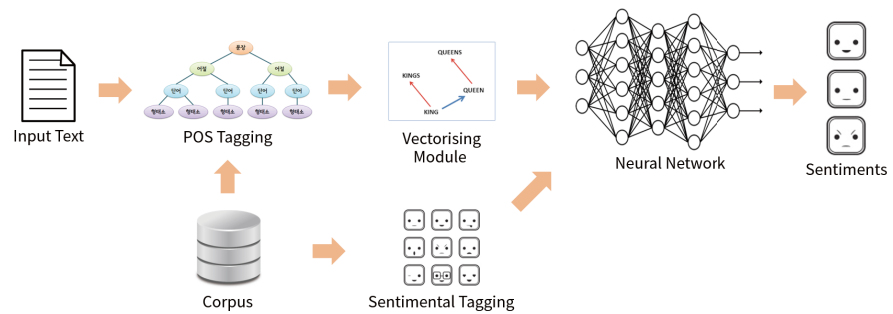
2 소셜 빅데이터를 활용한 건축안전 분야 여론 모니터링 수행

■ 분석 개요

- 국민 관심이 높은 건축안전 분야를 대상으로 소셜 빅데이터를 활용한 여론 모니터링 수행
 - (분석 대상) 뉴스(네이버), 블로그(네이버, 다음 티스토리), 트위터
 - (분석 기간) 2014년 4월 1일 ~ 2017년 4월 30일
 - (분석 방법) 키워드 빈도 분석, 연관 키워드 분석, 시계열 추이 분석, 감성 분석
 - (검색 키워드²⁾) ‘안전사고’, ‘지진’, ‘붕괴’, ‘지반침하’, ‘침수’, ‘화재’

■ 데이터 수집 및 분석

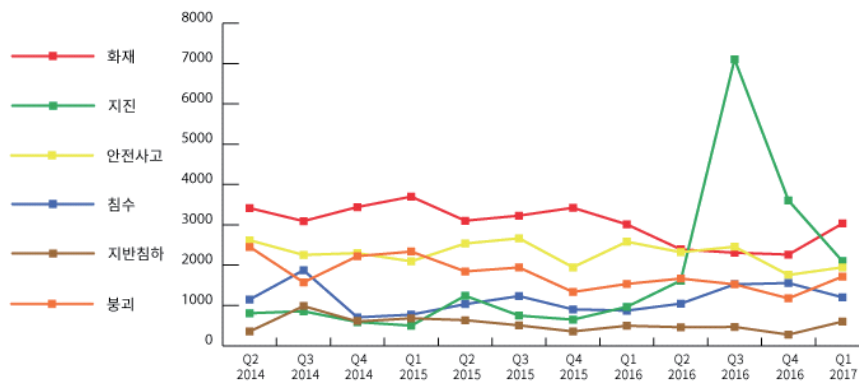
- 비용 효율적인 데이터 수집을 위하여 기존에 개발된 크롤링 소프트웨어 활용
 - 소셜 빅데이터 수집은 별도의 웹크롤러를 개발하는 것도 가능하나, 소셜 빅데이터의 경우 대표적인 빅데이터로 기 개발된 유무상의 플랫폼을 활용하는 것이 보다 경제적으로 기존 업체(주)아크릴의 웹정보 수집 시스템 활용
 - 자바 기반 시스템으로 개발된 웹크롤러를 통해 복수의 채널에 있는 데이터 저장
 - 자동 비주기적 크롤링 기술 적용(로봇방문자 차단 회피) 및 분산 처리 기술 적용
- 문서에 대한 키워드 분석의 대표적인 기술인 텍스트마이닝 기법을 적용하여 분석
 - 빈도 분석, 연관 키워드 분석, 시계열 추이 분석 등의 키워드 관련 분석은 통계 기반(TF-IDF)으로 키워드를 추출한 후 벡터화하여 클러스터링 후 분석
 - 감성 분석을 위해 분석 대상이 되는 텍스트 정보의 형태소를 분석한 후 이를 벡터화하고, 감성사전(문장에 긍정/중립/부정으로 태깅)과 텍스트 정보를 연결하여 분석
 - (주)아크릴의 분석 소프트웨어 ‘ZOOOMA Social’ 활용



감성 분석의 과정

2) 검색 키워드는 2017년도 국토교통부 업무계획을 분석하여 건축안전과 관련된 키워드를 추출함

3 건축안전 분야 주제별 여론 모니터링 결과



건축물 안전 관련 키워드에 대한 분기별 버즈량

■ ‘안전사고’ 관련 여론 모니터링 분석 결과

- 건축물 안전사고에 대한 버즈량³⁾은 총 27,476건(월평균 743건)이며, 2014년 5월과 10월, 2015년 7월, 2016년 7월에 급증
 - 2014년 5월 세월호 참사, 2014년 10월 성남 판교 환풍구 추락사고, 2015년과 2016년 7월 장마철 안전사고, 2016년 7월 지진 등이 영향을 준 것으로 분석
- 연관 키워드는 ‘안전(41.0%)’, ‘예방(30.8%)’, ‘관리(19.0%)’, ‘시설(18.8%)’, ‘발생(18.5%)’, ‘점검(17.1%)’, ‘시설물(11.9%)’, ‘사고(10.0%)’ 등
- 감성 분석 결과 긍정 16.75%, 중립 72.83%, 부정 10.42%로 중립을 제외하고는 긍정 감성이 부정 감성보다 조금 높게 나타남
- 안전사고는 건축물에서뿐 아니라 생활환경 전반에서 발생하는 경우도 많아 전반적으로 높은 버즈량을 유지하며, 특정 이슈 발생 시 버즈량이 급증
- 연관 키워드에서 예방과 관리에 대해 중요하게 생각하고 있으며, 감성 분석 결과 사고 이후의 해결 과정에 대해서도 여론의 관심이 높음을 확인

■ ‘지진’ 관련 여론 모니터링 분석 결과

- 지진에 대한 버즈량은 20,677건(월평균 559건)이며, 2016년 4월과 7월에 버즈량이 증가하였고 2016년 9월과 12월 사이 폭발적으로 증가

3) 버즈량은 온라인에서 해당 주제어(키워드)가 언급된 횟수를 말함

– 2016년 4월(944건)에 일본과 에콰도르 지진, 2016년 7월(1,166건) 울산 지진, 2016년 9월(5,565건) 경주 지진 발생이 원인

- 연관 키워드는 ‘발생(16.6%)’, ‘규모(9.4%)’, ‘원전(8.4%)’, ‘피해(7.9%)’, ‘안전(7.8%)’, ‘내진(7.4%)’, ‘건설(7.4%)’, ‘경주(7.1%)’, ‘설계(6.9%)’, ‘지역(6.9%)’ 순
- 감성 분석 결과 긍정 10.34%, 중립 71.32%, 부정 18.34%로 나타났으며, 중립 감성을 제외하면 부정 감성이 긍정 감성의 약 2배
- 국내에서 큰 규모의 지진이 발생하면서 지진에 대한 버즈량이 폭증하고, 감성 분석 결과 부정적인 감성이 상대적으로 높게 나타나 지진으로 인한 국민의 불안감이 고조되는 것으로 해석됨
- 건축물 내진설계 관련 키워드가 상위 연관 키워드로 랭크되어 있어 국민의 지진 불안감을 낮추기 위한 정책적인 전략으로 건축물 내진설계 관련 대책이 필요함을 확인할 수 있음

■ ‘붕괴’ 관련 여론 모니터링 분석 결과

- 건축물 붕괴에 대한 전체 버즈량은 21,323건(월평균 576건)이며, 2014년 5월과 10월, 2015년 1월과 3월 사이 버즈량 급증 구간이 나타남
 - 2014년 5월(1,307건) 북한의 고층 아파트 붕괴, 2014년 10월(948건) 판교 붕괴 사고와 고속도로 터널 부실 시공 이슈, 2015년 1월과 3월 사이(2,336건) 엠월드 붕괴 위험, 광주 아파트 옹벽 붕괴, 사당 종합 체육관 천장 붕괴, 용인고속도로 현장 붕괴 등이 연이어 발생하며 버즈량 급증
- 연관 키워드는 ‘사고(14.2%)’, ‘건설(10.5%)’, ‘발생(8.3%)’, ‘현장(7.3%)’, ‘북한(6.9%)’, ‘공사(6.3%)’, ‘안전(5.9%)’, ‘아파트(5.8%)’, ‘위험(5.7%)’, ‘건물(4.7%)’ 순
- 붕괴에 대한 긍·부정 여론은 긍정 10.04%, 중립 67.18%, 부정 22.79%로 나타났으며, 부정 감성이 긍정 감성의 2배 이상으로 높은 비율을 차지
- ‘건설’, ‘발생’, ‘현장’ 등 상위 연관 키워드에서 건설 현장의 붕괴 사고에 대한 여론의 관심이 나타나며, ‘안전’, ‘위험’, ‘피해’, ‘정부’, ‘우려’, ‘원인’ 등 연관 키워드가 지속적으로 등장하여 붕괴사고 원인 규명과 예방책에 대한 관심 또한 높아 대책이 필요할 것으로 분석됨

■ ‘지반침하’ 관련 여론 모니터링 분석 결과

- 지반침하에 대한 전체 버즈량은 6,420건(월평균 174건)으로 나타났으며 시기에 따라 버즈량 차이가 비교적 크게 나타나는데, 지하수에 의한 지반침하 현상인 경우가 많아 해빙기나 비가 많이 오는 시기에 관심도가 높아지는 양상이 나타남
- 연관 키워드는 ‘공사(14.0%)’, ‘싱크홀(12.0%)’, ‘안전(11.6%)’, ‘하다(9.8%)’, ‘현상(8.4%)’, ‘조사(8.4%)’, ‘착수(8.4%)’, ‘건물(7.9%)’, ‘점점(7.8%)’, ‘사고(7.4%)’ 순
- 감성 분석 결과 긍정 8.72%, 중립 72.35%, 부정 18.93%로 중립 감성의 비율이 높았으며, 중립 감성을 제외하고는 부정 감성이 높게 나타남
- ‘시설’, ‘건물’, ‘건축물’ 등이 꾸준히 연관 키워드로 등장하여 지반침하 시 시설물 · 건축물에 미치는 영향에 대한 여론의 관심이 높은 것으로 판단되며, 지반침하에 대응한 건축물 안전관리 방안 대책이 필요할 것으로 분석됨

■ ‘침수’ 관련 여론 모니터링 분석 결과

- 침수에 대한 전체 버즈량은 13,851건이며, 월평균 374건 기준으로 증감 폭이 큰 것으로 나타났으며, 매년 강우량이 많아지는 7월 전후로 버즈량 급증
- 연관 키워드는 ‘피해(27.5%)’, ‘지역(19.8%)’, ‘시설(16.4%)’, ‘사업(8.9%)’, ‘발생(8.8%)’, ‘설치(7.8%)’, ‘예방(7.2%)’, ‘상습(7.1%)’, ‘물(7.0%)’, ‘주택(6.3%)’ 순
- 감성 분석 결과 긍정 12.55%, 중립 75.01%, 부정 12.45%로 중립 감성을 제외한 긍정 · 부정 감성이 거의 비슷하게 나타남
- 침수 예방이나 피해와 관련된 다양한 연관 키워드가 등장하며, 상습 침수 지역에 대한 피해와 예방에 대한 여론이 많음
- 최근 빗물펌프장 등 우수 관리를 위한 키워드가 새롭게 등장하는 것으로 나타나, 우수 관리 정책에 대한 여론의 관심 증가를 확인할 수 있음

■ ‘화재’ 관련 여론 모니터링 분석 결과

- 화재에 대한 전체 버즈량은 36,403건(월평균 984건)으로, 여섯 개의 주제 가운데 버즈량이 가장 많은 것으로 나타남
- 화재와 관련된 연관 키워드는 ‘설비(16.1%)’, ‘발생(8.6%)’, ‘안전(6.7%)’, ‘사고(4.9%)’, ‘건설(4.9%)’, ‘건물(4.0%)’, ‘현장(3.3%)’, ‘위생(3.0%)’ 순

- 감성 분석 결과 긍정 7.94%, 중립 81.34%, 부정 10.71%로, 중립 감성을 제외하면 부정 감성이 조금 더 높은 것으로 나타남
- ‘설비’, ‘구조’, ‘탐지’, ‘경보’, ‘설치’ 등 연관 키워드를 볼 때 건축물 화재 예방과 유지·관리 방안에 대한 여론의 관심이 높음을 확인

4 건축안전 관련 정책의제 발굴과 시사점

■ 소셜 빅데이터를 활용한 여론 모니터링 결과에 기초한 정책의제 발굴

- 건축안전과 관련하여 여론의 관심이 가장 높은 주제는 건축물 화재안전으로, 건축안전 정책 수립 시 세부 정책의 우선순위 설정과 배분에 상대적 고려 필요
 - ‘화재’가 조사 기간 내 전체 버즈 중 약 29%를 차지하며 가장 높은 버즈량을 기록하고 있어 건축안전 분야 중 여론의 관심이 가장 높은 것으로 나타남
 - ‘화재’의 변이계수는 0.48 수준으로 ‘안전사고’와 더불어 여론의 관심이 지속적이며, 이에 정책도 단기적 처방보다는 중장기적 비전을 갖추어 추진할 필요가 있음
- ‘지진’은 여론의 관심이 가장 크게 변화한 주제로 건축물의 내진설계에 대한 국민적 관심이 높아 해당 정책 마련 필요
 - 월별 평균 버즈량 중 가장 높게 나타난 키워드는 2016년 9월 경주 지진의 영향으로 5,565건의 버즈량을 기록한 ‘지진’으로 전월에 비해 15배가 증가하였고, 주요 연관 키워드에 ‘내진(7.4%)’, ‘설계(6.9%)’가 포함되어 내진설계에 대한 국민적 관심 확인
- ‘붕괴’, ‘지반침하’, ‘지진’은 모두 건축물의 구조안전과 직접적인 관련이 있는 분야로, 감성 분석 결과 상대적으로 부정 감성이 높아 이에 대한 대책 마련 필요

건축안전 분야 키워드별 분석 종합

키워드	버즈량 분석				감성 분석		
	합계	월평균	표준편차	변이계수*	긍정	중립	부정
안전사고	27,476	743	299	0.40	16.75%	72.83%	10.42%
지진	20,677	559	1,898	3.40	10.34%	71.32%	18.34%
붕괴	21,323	576	398	0.69	10.04%	67.18%	22.79%
지반침하	6,420	174	187	1.08	8.72%	72.35%	18.93%
침수	13,851	374	349	0.93	12.55%	75.01%	12.45%
화재	36,403	984	475	0.48	7.94%	81.34%	10.71%

* 표준편차를 산술평균으로 나눈 값으로 여러 집단 간의 산포도를 비교

■ 여론 모니터링 수행의 의의 및 시사점

- 건축안전과 관련한 여섯 가지 주요 주제에 대해 3년간의 소셜 빅데이터를 분석하여 세 가지 건축안전 분야 정책의제를 도출
 - 정책의제 도출은 정책 수립 과정의 시작일 뿐이며, 이를 정책으로 개발하고 시행하기 위해서는 대안 개발, 대안 채택 등 여러 추가적인 과정이 필요
 - 이번 여론 모니터링 분석은 증거기반 건축·도시 정책을 위하여 소셜 빅데이터의 활용 가능성을 제시한 것으로, 건축안전 분야의 세부 정책 마련은 후속 연구 필요
- 빅데이터를 활용한 증거기반정책 개발은 기존의 의견기반정책 개발과 비교할 때 장단점이 있으며, 이를 고려하여 정책의 주제에 따라 추진할 필요가 있음
 - 증거기반정책 개발 시에는 의사 결정이 과학적이고 객관적으로 이루어져 신뢰도가 높은 정책을 개발할 수 있으며, 정책의제와 대안 개발이 정량적으로 이루어진 만큼 해당 정책에 대한 평가와 환류가 용이
 - 한편 데이터의 수집·가공·분석에 상대적으로 많은 시간이 소요되고, 분석 과정에 대한 설계가 정교하지 못하면 결과가 나오지 않는 경우도 있어 준비 과정에 상당한 시간이 필요
- 건축·도시 분야 빅데이터 활용은 증거기반정책의 개발뿐만 아니라 제4차 산업혁명 시대의 스마트시티, 스마트빌딩에 대한 직접적인 요소 기술이며, 관련 분야의 지속적인 연구와 정책적 지원이 필요
 - 인공지능과 초연결로 대표되는 제4차 산업혁명이 현실화되고 있는 지금, 국민 여론, 사회적 변화 및 위험 요인 등을 데이터를 통해 정량적으로 파악하여 정책에 활용하려는 노력이 세계적으로 확산되는 추세
 - 건축 분야의 빅데이터 분석 및 활용 기술은 제4차 산업혁명 시대 스마트시티와 스마트빌딩의 핵심적인 요소 기술로 관련 분야에 대한 지속적인 연구와 개발이 필요

조영진 부연구위원 (044-417-9692, yjcho@auri.re.kr)

김신성 연구원 (044-417-9835, sskim@auri.re.kr)

