

빅데이터를 활용한 건축·도시 미래정책 개발체계 연구

Policy Development Framework of Architecture and Urbanism using Big Data

조영진 Cho, Young Jin
유광흠 Yu, Kwang Heum
김신성 Kim, Shin Sung

(a u r

기본연구보고서 2017-2

빅데이터를 활용한 건축·도시 미래정책 개발체계 연구

Policy Development Framework of Architecture and Urbanism using Big Data

지은이	조영진, 유광흠, 김신성
펴낸곳	건축도시공간연구소
출판등록	제569-3850000251002008000005호
인쇄	2017년 10월 26일, 발행: 2017년 10월 31일
주소	세종특별자치시 절재로 194, 701호
전화	044-417-9600
팩스	044-417-9608

<http://www.auri.re.kr>

가격: 20,000원, ISBN:979-11-5659-137-5

이 연구보고서의 내용은 건축도시공간연구소의
자체 연구물로서 정부의 정책이나 견해와 다를 수 있습니다.

연구진

- ! 연구책임 조영진 부연구위원
 - ! 연구진 유광흠 선임연구위원
 김신성 연구원
 - ! 외부연구진 (주) 아크릴
 - ! 연구보조원 김성태, 송민경, 안의순
-

- ! 연구심의위원 이명식 동국대학교 건축공학부 교수
 김성아 성균관대학교 건축학과 교수
 남영우 국토교통부 건축정책과 과장
 김상호 건축도시공간연구소 기획조정실장
 오성훈 건축도시공간연구소 도시본부장

제1장 서론

최근 건축도시를 둘러싼 사회·경제·환경 여건이 급격하게 변화하고 있다. 이러한 여건 변화에 국민의 안전을 보장하고 삶의 질을 향상시키기 위해서는 다가올 미래에 선제적으로 대응할 수 있는 건축도시 정책 개발을 위한 새로운 방법론이 필요하다.

증거기반 정책은 국민적 관심이 높고, 다수의 민원이 발생하며, 정책 수요도 빠르게 변하는 건축도시 분야에 유효한 수단이다. 그러나 지금까지 관련 연구는 매우 부족한 실정이며 건축·도시 분야 정책의 품질을 높이기 위해서는 분야 특성을 반영한 증거기반 정책 수립 방법론에 대한 연구가 시급하다.

미국, 영국 등 세계 주요 선진국들은 빅데이터를 기술의 혁신이자 새로운 문제해결의 방식으로 판단하면서 미래지향적인 사고로 발전해 나가고 있다. 우리 정부도 정부3.0 기조에 맞추어 건축·도시 분야 개방데이터와 소셜 데이터 등 빅데이터 기반 융합 방향 모색 필요하며, 특히 정보화시대에 빠르게 변화하는 건축·도시 정책의 수요와 변화를 진단하고 예측할 수 있는 방법론 개발이 요구된다.

본 연구의 목적은 첫째 건축·도시 분야 빅데이터 활용 방안 모색하고, 둘째 빅데이터를 활용한 건축도시 분야 정책개발 방법론 도출하고 마지막으로 증거기반 건축도시 분야 정책 개발체계 제시하는 것이다. 이를 위하여 빅데이터 선행이론을 검토하고, 건축·도시 분야 빅데이터 현황 조사를 실시하고, 정책개발을 위한 자료별 특성연구를 수행하고, 빅데이터 분석을 위한 모델을 설계하고 실증하고, 최종적으로 정책개발을 위한 빅데이터 활용방안 제시하고자 한다.

제2장 빅데이터 관련 선행이론

빅데이터에 대한 정의는 2001년 Gartner에서 정의한 데이터 규모(Volume), 빠른 속도(Velocity), 다양성(Variety)의 3가지 개념(3Vs)이 가장 대표적이며, 기존 정의를 종합하면, 빅데이터란 기존의 데이터에 비해 규모가 크고, 복잡 다양하며, 생성·처리 속도가 빨라 새로운 관리 및 분석 방법이 요구되는 데이터의 속성적 개념과 함께, 데이터를 활용·분석하여 가치 있는 정보를 추출하는 데이터 가치의 개념까지 포함하는 정보화 기술로 정의할 수 있다.

빅데이터의 특성은 데이터 규모(Volume), 빠른 속도(Velocity), 다양성(Variety), 정확성(Veracity), 가치(Value) 등 5가지로 구분하여 설명할 수 있으며, 자료의 형태에 따라 정형, 비정형, 반정형 데이터로 구분할 수 있다. 정형 데이터는 일정한 항목(필드)에 저장한 데이터를 의미하며 건축물 대장 정보, 토지거래 정도 등과 같이 데이터베이스로 관리가 가능한 반면, 비정형 데이터는 항목이 정해져 있지 않은 데이터로 블로그, 소셜 데이터, 뉴미디어 등을 통해 생산된 데이터를 의미한다.

소셜 데이터는 사람들 간의 인적 네트워크를 기반으로 발생하는 텍스트, 이미지 등의 데이터를 말하며, 크게 SNS, 블로그, 커뮤니티 빅데이터로 구분할 수 있다. 로그 데이터는 사람들이 일상생활에서 사용하는 모든 기기와 도구들이 남기는 디지털 흔적을 지칭하며, 기기와 도구의 종류에 따라 매우 다양하다. 미디어 데이터는 온라인 뉴스가 일반화되면서 생산되는 신문사, 방송사, 포털 등의 뉴스 데이터로 보도한 기사 원문의 텍스트 데이터와 그와 관계된 데이터를 의미한다. 정형 데이터는 일정한 항목(필드)에 저장한 데이터를 의미하며, 빅데이터 등장 이전부터 전통적으로 다뤄져온 대부분의 데이터가 정형데이터라고 할 수 있다.

빅데이터 분석을 위해서는 데이터의 수집·가공·분석·시각화의 일련의 과정을 거친다. 이를 위해서는 각각의 과정에 따라 알고리즘을 설계하고 프로그램 개발을 수행하여야 한다. 그러나 최근 공공과 민간에서 데이터 수집 및 가공 기술을 확보하기 어려운 사용자들을 위하여 쉽게 빅데이터를 처리하고 분석할 수 있도록 빅데이터 인프라 서비스를 제공하고 있다. 민간 업체의 경우, 포털사이트, 통신사, 신용카드 빅데이터 센터 등에서 주도적인 역할을 하고 있으며, 공공기관은 여러 기관에서 빅데이터와 관련된 센터를 운영 중이다.

제3장 빅데이터 기반 정책개발 사례

일반적으로 정책과정이란 정책을 개발하고, 개발된 정책을 집행하고, 집행된 정책을 평가하는 일련의 단계를 지칭한다. 정책과정론(policy process theory)은 정책문제의 인지로부터 목표설정·대안의 분석·결정·합법화·집행·평가의 과정을 거쳐 정책이 종결되기까지의 과정을 다룬 이론이라고 할 수 있다. 현재 정책학에서 가장 대표적으로 활용하고 있는 정책과정 모형은 Anderson(1975)의 모형으로 총 5단계로 구성되어 있으며, 문제형성, 정책형성, 대안채택, 정책집행, 정책평가 단계이다.

국내 사례를 살펴보면 SKT의 스마트 인사이트는 소셜 데이터를 실시간으로 분석하여 여론 동향을 파악하고, 국민권익위는 기존의 민원데이터를 텍스트 마이닝을 통하여 문제점을 도출하는 등 정책과정 중 문제형성 단계에서 빅데이터 활용 가능성을 보여준다. 또한 서울시 심야버스, SKT T맵 등 사례에서는 로그데이터를 중심으로 정책대안을 만들고 또한 다양한 대안 중 최적의 대안을 탐색하는 정책형성과 대안채택 단계에서 정량적 의사결정을 위한 빅데이터 적용방안을 확인 할 수 있다. 경기도는 소셜 데이터 분석을 통하여 경기관광의 문제점과 대안을 포괄적으로 탐색하였고, 서울시는 휴대전화·신용카드 등 로그데이터를 활용하여 대안(관광코스)을 개발하여 최적의 대안을 찾았다. 이 사례들에서 확인할 수 있듯이 관광분야라는 동일한 주제라도 정책과정과 달성하려는 목적에 따라 활용 빅데이터가 달리 적용됨을 알 수 있다.

국외 사례에서는 자체적인 알고리즘을 개발하여 빅데이터를 분석 활용하는 사례가 많이 나타났다. 뉴욕시의 CompStat은 기존 범죄데이터를 종합하여 범죄발생지역을 사전에 예측하는 알고리즘을 개발하였고, 구글은 검색데이터 기반의 독감 예측 알고리즘, 월마트의 소셜 게놈, FDNY의 FireCast 등 대다수의 사례에서 기존의 빅데이터 분석방법론에 구축 데이터의 특성에 기초한 자체적인 알고리즘을 구축하여 운영하고 있다. FDNY의 사례에서 뉴욕시 MODA의 DataBridge는 정부가 보유하고 있는 공공빅데이터의 구축 및 활용에 큰 시사점을 제시하고 있다. 우리나라 정부에서도 데이터공개와 활용정책에 많은 비용과 시간을 투자하고 있다. 그 결과 많은 양의 데이터가 민간에 공개되고 있으나, 각 부처별·기관별 구축하고 있는 데이터가 매칭 되지 않아서 실제 활용에 한계가 있다. 이러한 공공 빅데이터의 활용을 높이기 위해서는 뉴욕의 DataBridge와 같은 서로 다른 데이터 간에 이를 연계할 수 있는 시스템이 반드시 필요하다.

제4장 빅데이터 기반 건축도시 정책개발 체계 수립

정책과정 중 정책개발의 범위를 정리하면 정책과정의 5단계 중 정책집행 이전 단계인 문제형성, 정책형성, 대안채택 3단계에 해당된다. 문제형성 단계에서는 특정 주제·이슈에 대한 대중의 의견과 여론을 파악하기 위한 여론 모니터링이 필수적이며, 정책형성 단계의 빅데이터 활용은 다양한 데이터를 결합 및 연계하여 정책 대안을 만드는 것에 초점이 맞춰져 있다. 대안채택 단계는 정책형성 과정에서 도출된 다양한 대안 중 최적의 대안을 찾는 것이다.

빅데이터 기반 건축도시 정책개발 체계는 앞에서 정리한 문제형성 단계, 정책형성 단계, 대안채택 3단계로 구성하였다. 각 단계별 빅데이터 활용 체계는 ‘질의개발’, ‘분석내용 결정’, ‘데이터 종류 및 특성 파악’, ‘분석방법 설계’, ‘데이터 수집 및 가공’, ‘분석 및 결과해석’의 6단계로 구성하였다. 빅데이터 분석체계는 목적(질의)에 따라 수집하는 데이터의 종류 등에 차이가 있는 것이며 기본적인 빅데이터 분석의 틀은 동일하다.

문제형성 단계의 질의는 해당 주제에 대한 일반 국민, 이해관계자에 대한 의견 청취 내용으로 구성되며, 기존 설문지 문항 개발과 유사한 과정으로 볼 수 있다. 정책형성 단계의 질의는 해당 문제에 대한 현황(실태)에 대한 질의이며, 대안채택 단계의 질의는 대안에 따른 경제적·사회적 효과를 검토할 수 있는 질의라 할 수 있다.

질의개발을 완료한 후 분석내용을 결정한다. 문제형성 단계의 경우 전반적인 국민 여론을 분석할 것인지, 전문가 집단의 의견을 파악할 것인지 등의 대상을 결정하여야 하며, 정책형성 단계의 경우 질의에 따라 대안 도출이 가능한 내용 범위를 설정한다. 대안채택 단계는 경제성, 효과성, 실현가능성 등에 대한 분석내용을 결정한다.

분석 내용을 결정한 후 해당 내용에 관한 어떠한 데이터가 있는지 파악 한다. 분석 내용에 따라 소셜 데이터, 로그 데이터, 정형 데이터 등 다양한 빅데이터를 활용할 수 있다. 분석 대상 데이터를 결정한 후에는 분석 방법을 설계한다. 가장 중요한 부분은 이기종의 데이터를 통합하여 분석하는 방법을 설계하는 것이다.

분석방법 설계에 따라 데이터를 수집한다. 데이터의 수집방법은 정형데이터와 비정형·반정형 데이터로 구분할 수 있다. 정형데이터는 파일로 수집하거나, 데이터베이스를 바로 연결하여 수집한다. 비정형데이터는 웹 크롤러 등을 이용하여 공개된 자료를 자동으로 수집한다. 최종적으로 수집한 데이터의 분석을 수행한다.

제5장 빅데이터 기반 건축도시 정책개발 체계 실증

최근 국민 관심도가 증가하고 있고, 국토교통부 주요 업무계획에 포함되어 있는 건축물 유지관리 정책을 빅데이터 기반 정책개발 체계 실증 대상으로 선정하였다.

문제형성단계의 검증을 위하여 건축물 유지관리 정책의 이슈와 수요를 도출하기 위해 주요 질문을 도출하고, 주요 질문을 바탕으로 유지관리 정책에 대한 일반 국민 여론 분석을 목적으로 키워드 빈도분석, 연관키워드 분석, 시계열 추이 분석, 감성 분석을 시행하였다. 데이터 수집 기간은 시간, 데이터의 질과 양을 고려하여 최종적으로 2014년 4월 1일부터 2017년 4월 30일로 결정하였으며, 유지관리 분야에 대한 사전 검색을 통해 국민적 관심도가 높고 정책적 문제해결이 시급하다고 판단되는 ‘안전’ 분야를 주요 분석 대상으로 하였다. 분석결과 전반적으로 가장 많은 버즈가 발생한 키워드는 ‘화재’로 여론의 관심이 가장 높은 것으로 나타나 정책의제를 ‘건축물의 화재안전’으로 선정하였다.

정책형성 단계에 검증을 위하여 건축물 화재 안전 관련 질의를 도출하고, 분석모델을 설계한 후 분석을 수행하였다. 분석에 활용한 데이터는 건축물 화재통계(2012년~2016년)를 중심으로 건축물 대장정보, 기상정보, 인구총조사 정보를 결합하여 활용하였다. 분석결과 건축물 화재는 주거공간 화재가 발생 건수가 가장 많고 인명 피해도 집중되어 있다는 점을 확인할 수 있었으며, 기상데이터와 결합한 결과 월평균 습도 50%~85% 구간에서 10% 습도 하락 시 화재발생이 약 19% 증가한다는 사실을 도출하였다. 인구총조사 연동한 결과 대부분의 법정동은 화재 발생이 매우 적고 극소수의 법정동에서 매우 많은 화재가 발생하는 거듭제곱 분포(power law)를 따르는 것을 알 수 있었다. 또한 머신러닝을 통한 건축물 대장정보와 결합한 연관 분석결과 집합건축물의 주건축물이면서 비주거용이고 용적률 산정 연면적 1,653m² 초과하는 건축물은 화재위험이 최대 11.9배로 나타나는 등 화재발생 집중 건축물 유형을 도출하였다.

실증 과정을 통하여 최종적으로 건축물의 유지관리 정책 중 화재안전에 관련하여 ‘비주거 집합건축물 화재안전 강화 및 소방 관련 법제도 개선’, ‘산업시설 대형 화재 예방’, ‘광역시·도간 유기적 소방 대응 체계’ 이상 3가지 정책대안을 개발할 수 있었다.

제6장 결론

본 연구에서 개발한 빅데이터 활용 정책 개발 체계는 증거기반 정책개발을 시행할 수 있는 방법론으로 객관적이고 과학적인 데이터 기반의 정책을 개발할 수 있음을 연구를 통하여 실증하였다. 이러한 증거기반 정책개발 체계는 의견기반 정책개발 체계와 비교할 때 정책의 결정 과정이 투명하고, 특정 전문가나 집단의 의견이 아닌 데이터 기초하여 정책의 의제와 대안 등을 개발할 수 있으며, 데이터를 통하여 정책을 개발한 만큼 평가 및 환류과정에도 유익하게 활용될 수 있다.

빅데이터의 활용을 높이기 위해서는 공공데이터의 국가기초구역 단위로 제공이 빠르게 현실화 되어야 하며, 공공데이터의 국가기초구역 단위 제공 완료 후 민간이 보유하고 있는 휴대전화, 신용카드 등의 데이터 역시 해당 단위로는 제공할 수 있도록 제도적 기반을 마련하고 이를 시행할 필요가 있다. 또한 빅데이터의 활용을 위해서는 한 곳에서 생산된 데이터를 해석하는 것도 중요하지만, 각기 다른 기관의 데이터를 연계 활용하여 새로운 시사점을 찾는 것이 보다 더 중요하다. 이를 위해서는 해당 데이터를 연계할 수 있는 연계플랫폼 개발이 시급하다.

본 연구는 정책의 과정 중 정책 개발에 해당하는 문제형성과 정책형성을 중심으로 수행되었으며, 추후 정책집행과 정책모니터링에 관한 빅데이터 활용 연구가 필요하다. 또한 본 연구를 토대로 건축·도시 분야의 주제별 빅데이터 연구를 통하여 구체적인 개발 체계 수립이 이루어 질 필요가 있다.

본 연구에서는 빅데이터를 활용한 건축·도시 분야 미래정책 개발체계를 개발하고 이를 실증하였다. 이를 통하여 건축·도시 분야의 정책 개발에 다양한 빅데이터를 활용할 수 있음을 확인할 수 있다. 빅데이터에 기초한 정책개발은 증거기반 정책 수립에 기반이라 할 수 있으며, 데이터를 통한 정량적 분석을 통한 정책개발은 정책의 신뢰성과 합리성을 높일 수 있을 것이라 기대된다.

주제어

빅데이터, 정책개발체계, 증거기반정책, 건축물 유지관리, 건축물 화재

차례

TABLE OF CONTENTS

제1장 서론

1. 연구의 배경 및 목적	1
2. 선행연구 현황 및 차별성	4
3. 연구내용 및 방법	6
1) 주요 연구내용	6
2) 연구추진방법	8
3) 연구의 범위	9

제2장 빅데이터 관련 선행이론

1. 빅데이터 개념 및 특성	11
1) 빅데이터 개념 및 정의	11
2) 빅데이터의 특성과 형태	13
3) 빅데이터 종류와 특징	15
2. 빅데이터 분석 방법론 고찰	19
1) 빅데이터 구축 및 활용 과정	19
2) 빅데이터 수집 및 가공 방법	20
3) 빅데이터 분석 방법	27

제3장 빅데이터 기반 정책개발 사례

1. 정책의 개념 및 과정	33
1) 정책의 개념	33
2) 정책의 과정	34
2. 국내 사례	37
1) SK텔레콤, 인터넷·SNS 여론분석 시스템 ‘스마트 인사이트’	37
2) 파리바게뜨, 기상과 제품별 판매량을 분석한 날씨 판매지수	38
3) 서울시 심야버스 노선수립 지원시스템	39
4) 경기도 소셜 미디어 분석을 통한 관광 트렌드 파악	41
5) SK텔레콤 T맵을 통한 실시간 교통 정보 서비스	42
6) 국민권익위원회 민원정보 분석시스템	43
7) 서울시 관광사업	45
8) 국내 사례 종합 및 시사점	46

3. 국외 사례	47
1) 뉴욕 범죄현황 통계분석 및 예측 시스템 컴프스탯(CompStat)	47
2) 구글 검색어 분석을 통한 독감 유행 예측 시스템 Google Flu Trends	48
3) 미국 영화사 SNS 분석을 통한 영화 흥행 예측	49
4) 암스테르담 스마트 시티 프로젝트 에너지 아틀라스(Energy Atlas)	50
5) 월마트, 소셜 데이터 분석을 통한 소비자 패턴 조사	52
6) 미국 웰포인트 빅데이터 기반 의료정보 제공 시스템	53
7) 브라질 리우데자네이루 재해예방 통합관리 지능형 운영센터	54
8) 뉴욕시 소방서(FDNY)의 RBIS & FireCast	55
9) 국외 사례 종합 및 시사점	57

제4장 빅데이터 기반 건축·도시 정책개발 체계 수립

1. 건축·도시 분야 정책 현황	59
1) 건축정책관실 2017년 업무계획 분석	60
2) 도시정책관실 2017년 업무계획 분석	63
3) 건축·도시 분야 2017년 정책 종합	68
2. 빅데이터 기반 건축·도시 정책개발 체계 수립	72
1) 문제형성 단계 - 정책의제 선정	72
2) 정책형성 단계 - 정책대안 개발	75
3) 대안채택 단계 - 최적 대안 선정	78
4) 빅데이터 기반 건축·도시 정책개발 체계	79

제5장 빅데이터 기반 건축·도시 정책개발 체계 실증

1. 대상 정책 선정	85
1) 대상 정책(건축물 유지관리 정책) 선정 배경	85
2) 건축물 유지 관리 정책의 범위	85
2. 문제형성 단계 실증	89
1) 질문 도출	89
2) 분석내용 설정	89
3) 데이터 종류 및 특성 파악	89
4) 분석 방법 설계	89
5) 데이터 수집 및 가공	94
6) 분석 및 결과해석	95

차례

TABLE OF CONTENTS

3. 정책 형성 단계	121
1) 문제 형성 단계에 따른 정책 형성 방향	121
2) 질의개발	121
3) 분석내용 결정	121
4) 데이터 종류 및 특성 파악	122
5) 분석 방법 설계	122
6) 분석 및 결과해석	125
4. 평가 및 환류	145
제6장 결론	
1. 연구의 결과	149
2. 건축·도시 빅데이터 활용을 위한 정책제언	152
3. 연구의 한계와 기대효과	155
참고문헌	157
영문초록	161

표차례 LIST OF TABLE

[표 1-1] 선행연구 현황 및 차별성	4
[표 2-1] 데이터 단위(바이트)	11
[표 2-2] 빅데이터 형태에 따른 유형 구분	14
[표 2-3] 기존의 데이터 관리 방식과 빅데이터 관리 방식의 차이점	15
[표 2-4] 데이터 수집 기술	21
[표 2-5] 데이터 가공 방법	22
[표 2-6] 데이터 저장 방식	23
[표 2-7] 민간부문 빅데이터 분석 서비스 기관	25
[표 2-8] 공공부문 빅데이터 분석 서비스 기관	26
[표 2-9] 빅데이터 분석 방법	27
[표 3-1] 국내 빅데이터 기반 정책개발 사례 종합	47
[표 3-2] 해외 빅데이터 기반 정책개발 사례 종합	58
[표 4-1] 건축분야 정책 내용 분석(2017년 건축정책관실 업무계획)	68
[표 4-2] 도시분야 정책 내용 분석(2017년 도시정책관실 업무계획)	70
[표 4-3] 문제형성 단계 빅데이터 활용 예시	75
[표 4-4] 정책형성 단계 빅데이터 활용 예시	77
[표 4-5] 대안채택 단계 빅데이터 활용 예시	78
[표 4-6] 빅데이터 기반 건축·도시 정책 개발 단계별 활용 데이터 및 활용 예시	80
[표 5-1] 건축분야 정책 내용 분석(2017년 건축정책관실 업무계획)	86
[표 5-2] 건축물 유지관리 정책 범위	86
[표 5-3] 건축물 유지관리 정책 분야와 제도 현황	88
[표 5-4] 건축물 유지관리 분야 검색 키워드 선정	90
[표 5-5] 분석 대상 선정 결과	95
[표 5-6] 전국 연도별 화재 발생, 인명 및 재산피해 통계	125
[표 5-7] 건축물 용도별 및 발화지점별 화재 발생 및 피해 통계	126
[표 5-8] 건축물 발화요인 상위 5개 (발생건수 기준)	129
[표 5-9] 건축물 발화열원 상위 5개 (발생건수 기준)	129
[표 5-10] 건축물 화재 발생건수 상위 시군구 통계	136
[표 5-11] 건축물 화재 발생건수 상위 법정동 통계	139
[표 5-12] 건축물 화재에 대한 연관규칙 및 분석 지표	142
[표 5-13] 화재 발생 비율이 높은 건축물 유형	143
[표 6-1] 공간빅데이터	154

그림차례 TABLE OF FIGURES

[그림 1-1] 정책 수립 단계와 데이터 활용	2
[그림 1-2] 빅데이터 분석을 활용한 범죄수사분석(911테러)	6
[그림 1-3] 서울시 심야버스노선 결정과정	7
[그림 1-4] 연구흐름도	9
[그림 2-1] 빅데이터 개념의 등장	12
[그림 2-2] 빅데이터의 특성	14
[그림 2-3] SNS 빅데이터 분석 사례	16
[그림 2-4] 휴대폰 빅데이터로 유동인구 분석 사례	17
[그림 2-5] 언론기사 빅데이터 분석 시스템 사례	18
[그림 2-6] 서울시 제20대 국회의원 건축도시정책 공약 현황	18
[그림 2-7] 빅데이터 구축 및 활용 과정	20
[그림 2-8] 웹크롤러를 이용한 비정형 데이터 수집 과정	21
[그림 2-9] HDFS의 데이터 쓰기 과정(좌)과 데이터 읽기 과정(우)	23
[그림 2-10] 분류체계 기반 여론 토픽 분석의 원리	28
[그림 2-11] 토픽모델링 기반 여론 내용 분석의 원리	29
[그림 2-12] 토픽모델링을 활용한 Science 논문의 내용의 경향 분석	30
[그림 2-13] 환경 키워드에 대한 빈도 분석 사례	30
[그림 2-14] 다양한 연관어 분석 및 시각화 사례	31
[그림 2-15] 감성 분석 예시	32
[그림 2-16] 네트워크 분석 사례	32
[그림 3-1] 스마트 인사이트 분석 예시	37
[그림 3-2] 파리바게뜨 계산대(POS)에 제공되는 날씨지수	38
[그림 3-3] 서울시 심야버스 노선 검증 및 통행량 산출 알고리즘	40
[그림 3-4] 빅데이터 기반 심야버스 노선수립 지원시스템 인터페이스	40
[그림 3-5] 경기도 소셜 미디어 분석 과정	41
[그림 3-6] 국내와 경기도 관광 연관 키워드 관계도	42
[그림 3-7] 7가지 길 안내 방법	43
[그림 3-8] 민원정보분석시스템 지역별 민원 지도	44
[그림 3-9] K-pop 스타(EXO) 팬을 위한 관광 서비스 지도	45
[그림 3-10] NYC Crime Map	48
[그림 3-11] Google Flu Trends 시계열 예측	49
[그림 3-12] 소셜네트워크 데이터량	50
[그림 3-13] Use of Gas and Electricity	51

그림차례 TABLE OF FIGURES

[그림 3-14] 소셜 게놈(Social Genome) 개념과 분석 사례	52
[그림 3-15] 웰포인트 의료정보 제공 체계	53
[그림 3-16] 라우데자네이루 지능형운영센터	55
[그림 3-17] Example of a high risk building according to FireCast	56
[그림 4-1] 방재계획 수요 도출 휴대폰 빅데이터 활용 사례	74
[그림 4-2] 서울시 및 대전시의 교통카드 빅데이터 활용 사례	76
[그림 4-3] 미국 샌프란시스코 범죄지도	77
[그림 4-4] 소셜빅데이터와 정형데이터를 통합한 분석 절차 및 방법 예시	82
[그림 4-5] 빅데이터 분석을 통한 정책 개발 과정	83
[그림 5-1] 텍스트 마이닝 과정	91
[그림 5-2] TF-IDF를 이용하여 추출한 키워드와 키워드 중요도 수치 예시	91
[그림 5-3] Skim-gram 모형 개념도	92
[그림 5-4] 키워드 백터화 예시	92
[그림 5-5] 키워드 클러스터링 결과 예시	93
[그림 5-6] 형태소 분석 결과 예시	93
[그림 5-7] 웹 정보 수집 시스템 구조	94
[그림 5-8] 웹 정보 수집 시스템 결과	94
[그림 5-9] '안전사고'에 대한 키워드 분석 결과	96
[그림 5-10] '안전사고'에 대한 시계열 빈도	97
[그림 5-11] '안전사고'에 대한 연관키워드	97
[그림 5-12] '안전사고'에 대한 데이터 수집 채널별 연관 키워드 순위	98
[그림 5-13] 버즈량이 급증한 2014년 10월의 게시글에서 나타난 연관 키워드	98
[그림 5-14] '안전사고'에 대한 감성 분석 결과	99
[그림 5-15] '지진'에 대한 키워드 분석 결과	100
[그림 5-16] '지진'에 대한 시계열 빈도	100
[그림 5-17] '지진'에 대한 연관키워드	101
[그림 5-18] '지진'에 대한 데이터 수집 채널별 연관 키워드 순위	101
[그림 5-19] 2016년 4월 연관 키워드	102
[그림 5-20] 2016년 7월 연관 키워드	102
[그림 5-21] 2016년 9~12월 연관 키워드	103
[그림 5-22] '지진'에 대한 감성 분석 결과	103
[그림 5-23] '붕괴'에 대한 키워드 분석 결과	104
[그림 5-24] '붕괴'에 대한 시계열 빈도	104

그림차례 TABLE OF FIGURES

[그림 5-25] ‘붕괴’에 대한 연관키워드	105
[그림 5-26] ‘지진’에 대한 데이터 수집 채널별 연관 키워드 순위	105
[그림 5-27] 2014년 5월 연관 키워드	106
[그림 5-28] 2014년 10월 연관 키워드	106
[그림 5-29] 2015년 1~2월 연관 키워드	107
[그림 5-30] 2015년 3월 연관 키워드	107
[그림 5-31] ‘붕괴’에 대한 감성 분석 결과	108
[그림 5-32] ‘지반침하’에 대한 키워드 분석 결과	108
[그림 5-33] ‘지반침하’에 대한 시계열 빈도	109
[그림 5-34] ‘지반침하’에 대한 연관키워드	109
[그림 5-35] 2014년 8월 연관 키워드	110
[그림 5-36] 2015년 2~4월 연관 키워드	110
[그림 5-37] 2016년 7월 연관 키워드	111
[그림 5-38] 2017년 3월 연관 키워드	111
[그림 5-39] ‘지반침하’에 대한 감성 분석 결과	112
[그림 5-40] ‘침수’에 대한 키워드 분석 결과	112
[그림 5-41] ‘침수’에 대한 시계열 빈도	113
[그림 5-42] ‘침수’에 대한 연관키워드	113
[그림 5-43] ‘침수’에 대한 데이터 수집 채널별 연관 키워드 순위	114
[그림 5-44] 2014년 7~8월 연관 키워드	114
[그림 5-45] 2015년 7월 연관 키워드	115
[그림 5-46] 2016년 7~10월 연관 키워드	115
[그림 5-47] ‘침수’에 대한 감성 분석 결과	116
[그림 5-48] ‘화재’에 대한 키워드 분석 결과	116
[그림 5-49] ‘화재’에 대한 시계열 빈도	117
[그림 5-50] ‘화재’에 대한 연관키워드	117
[그림 5-51] ‘화재’에 대한 데이터 수집 채널별 연관 키워드 순위	118
[그림 5-52] 2014년 5월 연관 키워드	118
[그림 5-53] 2015년 1월 연관 키워드	119
[그림 5-54] ‘화재’에 대한 감성 분석 결과	119
[그림 5-55] 기계학습 및 빅데이터 분석 프로그램 Weka	123
[그림 5-56] 타이타닉 생존 데이터에 대한 연관규칙(Association Rule) 분석 결과	124
[그림 5-57] 전국 건축물 화재 피해 발생 추세	126

그림차례

TABLE OF FIGURES

[그림 5-58] 건축물 용도별 발화지점별 화재 발생 통계	128
[그림 5-59] 건축물 발화요인 및 열원별 화재발생 통계	130
[그림 5-60] 건축물 용도별 발화요인별 화재 발생 통계	131
[그림 5-61] 월별 및 월평균 온·습도에 따른 화재 발생 통계	132
[그림 5-62] 월평균 온·습도에 따른 화재 발생 통계	133
[그림 5-63] 시도별 월별 건축물 화재 발생건수 통계	135
[그림 5-64] 시도별 화재 발생 및 피해와 인구·건축 밀도 상관분석	136
[그림 5-65] 시군구별 건축물 화재로 인한 인명피해	137
[그림 5-66] 시군구별 건축물 화재로 인한 재산피해	138
[그림 5-67] 법정동별 건축물 화재 발생	140
[그림 5-68] 법정동별 건축물 화재 발생	140
[그림 5-69] 법정동별 건축물 화재 발생	140
[그림 5-70] 빅데이터 기반 정책개발 체계 실증 과정	145
[그림 5-71] 빅데이터 정책개발 체계와 의견기반 정책개발 체계 비교	148
[그림 6-1] 공공데이터 포털	153
[그림 6-2] AlphaGo	156

제1장 서론

- 1. 연구의 배경 및 목적
 - 2. 선행연구 현황 및 차별성
 - 3. 연구범위 및 방법
-

1. 연구의 배경 및 목적

□ 건축도시 분야 정책 개발을 위한 새로운 방법론 필요

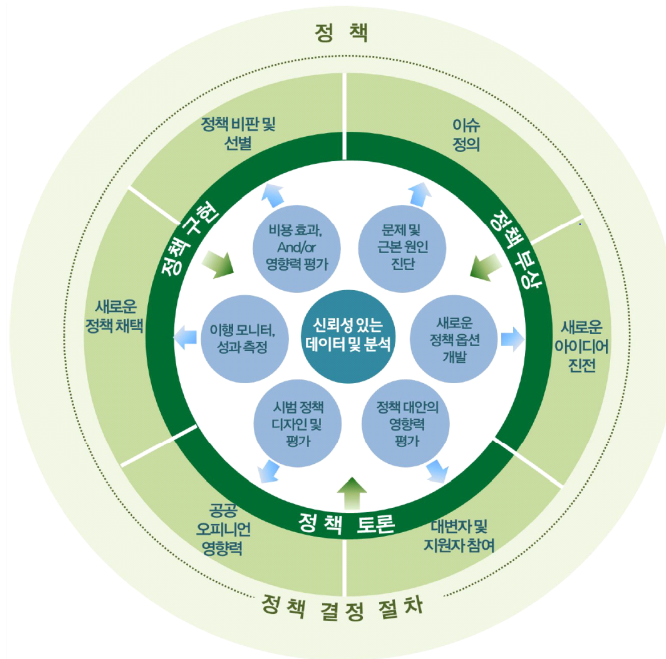
건축도시 분야 정책은 국민의 실생활에 미치는 영향이 크고, 관심도도 매우 높아 정책의 수립 및 집행에 앞서 국민의 인식에 대한 사전적인 고려가 필수적이다. 지금까지 건축도시 분야의 정책개발은 주로 정부주도로 이루어져 왔으며, 또한 국민의 여론 수렴도 전문가 자문이나 설문조사에 근거하고 있는 실정이다. 이러한 방식은 개인의 이해에 따라 의견과 결과가 달라질 수 있어 객관성 확보에 어려움이 있다.

최근 건축도시를 둘러싼 사회·경제·환경 여건이 급격하게 변화하고 있다. 따라서 이러한 여건 변화에 국민의 안전을 보장하고 삶의 질을 향상시키기 위해서는 다가올 미래에 선제적으로 대응할 수 있는 건축도시 정책 개발을 위한 새로운 방법론이 필요하다.

□ 건축도시 분야 정책 품질 제고를 위한 증거기반 정책(EBP)으로 전환 시급

인공지능과 초연결로 대표되는 제4차 산업혁명이 현실화되고 있는 지금 통념·관행·경험·직관에 의존했던 의견기반 정책(Opinion-Based Policy)이 증거기반 정책수립(EBP : Evidence-Based Policy-making)으로 점차 변화하고 있다. 또한 국민 여론파악, 사회적 변화 및 위험요인 등에 대하여 데이터를 통하여 정량적으로 파악하여 정책 입안에 활용하려는 노력이 확산되는 추세이며, 세계 각국의 정부는 정책의 신뢰성과 합리성 제고를 위해 증거기반 정책 수립에 노력을 기울이고 있다.

증거기반 정책은 국민적 관심이 높고, 다수의 민원이 발생하며, 정책 수요도 빠르게 변하는 건축·도시 분야에 유효한 수단이다. 그러나 지금까지 관련 연구는 매우 부족한 실정이며 건축·도시 분야 정책의 품질을 높이기 위해서는 분야 특성을 반영한 증거기반 정책 수립 방법론에 대한 연구가 시급하다.



[그림 1-1] 정책 수립 단계와 데이터 활용

출처 : Turner, Margery Austin (2013), *Evidence-Based Policymaking Requires A Portfolio Of Tools*, THE URBAN INSTITUTE

□ 빅데이터 시대, 미래예측 패러다임의 변화¹⁾

미래연구는 수리적이고 단편적인 예측(Forecasting)에서 종합적 전망에 근거한 대응 방안을 포함하는 미래전략(Strategic Foresight)으로 발전 중에 있다. 빅데이터 처리 기술이 발달함에 따라 텍스트 마이닝, 소셜네트워크 분석, 계량정보학 등 빅데이터 분석을 통한 통합적 미래연구가 주목받고 있다.

현재 주요 국가 정부와 산업계에서는 빅데이터를 각종 문제 해결 및 이슈 대응 뿐 아니라 미래전략과 전략적 의사결정의 중요한 도구로 활용 중이다. 또한 제4차 산업혁명 시대

1) 이영주(2015), 「빅데이터 시대, 미래전략의 새로운 접근법」, 한국정보화진흥원, 1p.

에 지금보다 훨씬 많은 분석 가능한 데이터가 발생할 것에 대비하여 유용한 정보를 선별하고, 활용하여 지식화 할 수 있는 역량을 키우기 위해 많은 노력을 기울이고 있다.

미래 전략은 미래 예측과 전망 뿐 아니라 전략적 대응방안을 모색하는 실천(Action) 지향적인 활동으로, 지혜를 탐구하고 축적하는 과정이다. 따라서 빅데이터 시대에는 증거와 객관적 합의에 의해 형성된 지식을 기반으로, 도덕이나 윤리적인 이슈까지 포함한 바람직한 미래에 대한 공론화 과정이 필요하다.

□ 빅데이터를 활용한 선진국 행정효율화 사례 확산과 정부3.0²⁾

미국, 영국 등 세계 주요 선진국들은 빅데이터를 기술의 혁신이자 새로운 문제해결의 방식으로 판단하면서 미래지향적인 사고로 발전해 나가고 있다. 이처럼, 빅데이터를 활용한 국가별 행정효율화 사례가 세계 곳곳에서 나타나고 있으며, 산적인 정부부처의 문제 해결에 빅데이터를 적극적으로 활용하는 국가의 정부와 그렇지 않는 국가의 정부 간에 격차 발생 예상된다.

우리 정부도 민간부문에 창의와 활력을 불어넣기 위하여 정보공개와 개방의 폭을 과감하게 확대하고 민간이 보다 쉽게 활용할 수 있도록 2013년 6월 『정부3.0』 추진 기본 계획을 발표하였다. 정부3.0 기조에 맞추어 건축·도시 분야 개방데이터와 소셜 데이터 등 빅데이터 기반 융합 방향 모색 필요하며, 특히 정보화시대에 빠르게 변화하는 건축·도시 정책의 수요와 변화를 진단하고 예측할 수 있는 방법론 개발이 시급한 실정이다.

연구목적

- 건축·도시 분야 빅데이터 활용 방안 모색
- 빅데이터를 활용한 건축도시 분야 정책개발 방법론 도출
- 증거기반 건축도시 분야 정책 개발체계 제시

2. 선행연구 현황 및 차별성

기존 연구는 크게 특정 분야 정책 개발을 위한 빅데이터 활용 방안 연구와 빅데이터를 활용한 지자체 정책 지원 연구로 구분할 수 있다. 주요 연구내용으로는 국토관리, 보건의료, 관광행정, 환경 등 여러 분야에서 정책 수요 및 여론을 파악하고, 정책을 수립하는데 있어 빅데이터의 활용 방안에 대한 연구가 수행되었으나, 건축 도시 분야에서 빅데이터의 정책적 활용에 대한 연구는 부재한 실정이다. 또한 빅데이터를 활용한 정책 수요 도출, 정책 방향 도출 등 정책개발 과정 중 일부분에 대한 활용방안 연구는 수행되었으나, 정책 개발을 위한 전 과정에서의 활용 체계를 제안한 사례는 미흡한 실정이다.

본 연구는 건축 도시 분야의 빅데이터 현황과 정책 활용 방안을 검토하는 연구로써, 정책 개발 과정별 빅데이터 활용방안을 모색하고, 이를 토대로 증거 기반 정책 개발의 체계를 마련하는 차별성이 있다.

[표 1-1] 선행연구 현황 및 차별성

구분	선행연구와의 차별성		
	연구목적	연구방법	주요연구내용
주요 선행 연구	-과제명 : 빅데이터(Big Data)를 활용한 부산시 도시관리 전략	-문헌 및 인터넷 자료 검색을 통한 국내외 빅데이터 관련 정책 및 활용사례 분석	-빅데이터 개념 및 분석 방법론 검토
	-연구자(년도) : 이원규, 오동하(2013)	-부산시 각 부서 내 정형, 비정형 데이터 자료 파악	-국내외 빅데이터 관련 정책 및 활용사례 분석
	-연구목적 : 빅데이터 활용 사례분석, 부산시 데이터 현황 파악, 전문가 의견 및 SNS 자료 등 각종 사례 분석을 통해 향후 부산시의 빅데이터 기반 스마트 도시관리 추진 전략 수립	-전문가 설문조사, FGI를 통해 부산시에 적용 가능한 빅데이터 유형 도출 -부산시 관련 SNS자료를 분석하여 시민의 시정에 관한 생각을 시계열적으로 도출	-부산시 공공데이터 현황 및 부문별 공공데이터 활용 사례 분석 -빅데이터 분석의 필요정도, 지자체 대응 정도 등을 도출하고, 부산시의 추세를 대상으로 Socialmatrix를 이용하여 사례분석 수행 -빅데이터 개념 정립 및 부산시에서 도시관리를 위한 빅데이터 활용전략 제시
	-과제명 : 빅데이터를 활용한 과학적·효율적 행정 구현방안	-국내외 논문 등 문헌 자료 활용을 통해 빅데이터 주제의 선진 사례 연구 분석	-빅데이터 정책현황 및 활용사례를 조사하여 산업통상자원부의 주요 정책방향 수립에 필요한 제언 도출
	-연구자(년도) : 박시룡 외(2013)	-주요 국가들의 공공부문 및 일반부문의 빅데이터 적용 사례 조사를 통한 활용 가능성 타진	-빅데이터의 실제 활용 가능성 점검을 위하여 축적된 공공데이터 현황 파악 및 실제 적용 가능한 업무 발굴
	-연구목적 : '정부 3.0'의 성공적인 정착을 위하여 산업통상자원부의 주요 업무 중 빅데이터를 활용하여 정책구현을 효율적으로 추진할 수 있는 분야를 발굴하고, 이를 실현하기 위한 구체적 방법론 제시	-산업통상자원부 주요 업무의 평가지표 및 성과지표 등을 분석하여 빅데이터 통합 가능성 타진 -전문가 자문을 통해 빅데이터 특징에 적합한 활용법 모색	-주요 업무 중 빅데이터 활용의 우선 과제를 선정하고 실제 이행을 위한 구체적 적용 방법론 제시 -성공적인 빅데이터 활용을 위한 선행 요건 도출
	-과제명 : 빅데이터를 활용한 환경분야 정책수요 분석	-빅데이터 개념과 활용 현황에 대한 국내외 문헌조사 및 사례조사 시행	-빅데이터의 개념 및 시장, 주요국의 빅데이터 정책 동향 파악
	-연구자(년도) : 이마숙, 이창훈, 김지연(2014)	-국내 환경분야 정책수요에 대한 소셜 빅데이터 분석	-환경분야에 빅데이터를 적용한 사례를 탐색하여 연구에 필요한 시사점 도출
	-연구목적 : 빅데이터를 활용하여 환경분야 정책수요를 분석하는 방법론을 수립하고 실질적 분석 결과를 제시함으로써 환경분야 정책수요 도출	-전문가 세미나 및 전문가 자문회의 개최 -수집된 소셜미디어 자료 처리 시 통계분석 및 계량 분석 병행	-수집 자료에 기반하여 연구의 목적에 부합하는 정책수요 방법론 수립 및 정책수요 분석 -소셜 빅데이터 분석(빈도분석, 키워드 분석, 감성 분석 등)을 통해 수집된 비정형 데이터와 정책수요 분석 결과를 바탕으로 향후 환경분야 소셜 빅데이터 활용방안 제시

선행연구와의 차별성			
구분	연구목적	연구방법	주요연구내용
본 연구	-고제명: '장부3.0도민맞춤형계획수립 용역 최종보고서' -연구지(년도): 전제남 외(2014) -연구목적: 도민맞춤형 정책수립을 위한 기초자료 제공 및 정책대응이 가능한 체계적·고충합적인 도민맞춤형 정책수립	-가존연구 검토 및 사례 분석 -경상북도청 홈페이지 내 민원문 및 포털사이트를 대상으로 경상북도청 연관 검색어를 통한 데이터 수집 -데이터 수집 및 정제는 데이터 일괄처리 솔루션인 The SCRM 이용 -텍스트 마이닝(Text Mining), 소셜네트워크분석(SNA) 중 의미 연결망(Semantic Network) 등 다양한 방법을 통한 분석 실시	-국내외 공공부문 빅데이터 사례 합의 도출 -빅데이터로서 경상북도 홈페이지 및 포털사이트를 분석하여 주요 단어 연관 정도 중심성, 위계 중심성 등 결과 제시 -경상북도 홈페이지 및 포털사이트의 사계절 분석을 통해 연도별 상위 주요 단어 빈도수 변화, 이슈 단어 빈도수 변화 등 다양한 결과 제시
	-고제명: 빅데이터를 활용한 관광행정 수요조사 및 정책 제언 연구용역 -연구지(년도): 조원섭 외(2014) -연구목적: 빅데이터를 활용하여 충북 관광자원에 대한 도내외 사람들의 관심사를 정량적으로 분석하여 충북 관광 정책 수립에 기초자료로 제공하고, 향후 정책 방향 설정에 기여	-자체 웹 크롤러를 이용하여 비정형 데이터 수집 -수집된 데이터를 활용하여 충북 관광의 이슈, 연관 키워드, 감성 분석 수행 -내국인 및 외국인 대상 설문조사를 시행하여 빅데이터 분석 결과 보완 -분석 결과는 빅데이터에 적절한 가시화 기법지도, 차트, 표 등을 활용하여 디스플레이 및 모바일 어플리케이션 개발	-관광산업 관련 민간 및 공공기관 보유 빅데이터 발굴 및 분석을 통한 관광정책 개발 및 활용 방안 제시 -과학적 관광행정 수요조사를 통한 충북 관광의 비전 및 정책 목표, 추진 전략 제시 -관광분야 각 영역별 온라인 트렌드 분석 및 중점 추진사업 제시 -충북의 관광자원을 온라인상의 이슈 분석을 통해 지역에 특화된 관광자원 개발 정책 방향 제시 -빅데이터 기반 서비스 방안 제시
	-고제명: 보건 의료 빅데이터 활용을 위한 기본 계획 수립 연구 -연구지(년도): 강희정 외(2015) -연구목적: 국내 보건 의료 데이터를 정책 사회적으로 활용할 수 있는 국가 기본 계획 수립	-윤호찰 및 웹(web) 탐색을 통한 연구 검토 및 사례 조사 -1년간 언론 기사 수집 및 텍스트 마이닝(Text Mining) 분석 실시 -전문가 자문회의 개최 -국외 방문 조사 실시 -전문가 Working Group 운영	-외국의 빅데이터 활용 사례 검토 및 전문가 워크숍 논의 결과를 바탕으로 보건 의료 빅데이터 활용의 비전 수립 -보건 의료 데이터 현황 검토를 통해 보건 의료 데이터 특성 및 활용 기능성 분석 -보건 의료 빅데이터 활용 클러스터와 중점 추진 사업을 선정하여 이에 대한 사업 계획 수립 및 시범 평가 등을 진행하여 최종 활용 사업 개발 -보건 의료 빅데이터 기술적 인프라 및 거버넌스 구축 전략 제시
	-고제명: 지방자치단체 정책자원을 위한 데이터 분석 운영 체계 연구 -연구지(년도): 이승환 외(2015) -연구목적: 지방자치단체 데이터 분석 체계를 검토하여 정부인을 마련함으로써 빅데이터를 활용한 지역 주민의 요구에 부응하는 정책을 수립하고, 이를 바탕으로 효율적인 데이터 기반의 지방자치단체 업무 추진 체계 확립	-현행 지방자치단체 데이터 분석 체계 조사 및 분석 -중앙부처와 지방자치단체, 지방자치단체 간 데이터 협업 추진 현황 및 문제점 도출 -국내 데이터 분석 및 협업 운영 체계 사례 조사 -해외 주요국의 데이터 분석 체계 관련 사례 조사	-지방자치단체의 데이터 분석 및 협업 체계 현황 조사 분석 -지방자치단체 데이터 운영 체계 및 데이터 활용 우수 사례 분석 -지방자치단체의 (빅)데이터 활용 방안 제고를 위한 개선 방안 제시 -자속가 능한 지방자치단체 데이터 분석 및 운영 체계 제도화 방안 제시
	-건축·도시 분야의 빅데이터 현황과 그 정책 활용 방안을 검토하는 연구로서, 소셜 데이터, 공간 데이터, 미디어, 학술 문헌 검색 등 현행 빅데이터를 종합하고, 건축·도시 정책의 수립부터 평가에 이르기까지 활용 방안을 모색하며, 이를 토대로 중·장기 정책 수행을 위한 빅데이터 기반의 방법론 수립	-빅데이터 선행연구 및 방법론 문헌 조사 -국내외 빅데이터 활용 사례 조사 -데이터 유형별 빅데이터 수집 -유형별 빅데이터 분석을 통한 활용 방안 연구 -방법론 수립 및 사례 적용, 현황 -전문가 자문 및 검토	-빅데이터 개념 및 이론적 배경 등 빅데이터 선행 이론 검토 -소셜 공간 미디어 문헌 검색 등 건축·도시 분야 빅데이터 현황 조사 -정책 문제 설정, 정책 결정, 정책 집행, 정책 평가 등 정책 개발을 위한 자료별 특성 연구 -빅데이터 분석을 위한 모델 설계 -정책 개발을 위한 빅데이터 활용 방안 제시

3. 연구내용 및 방법

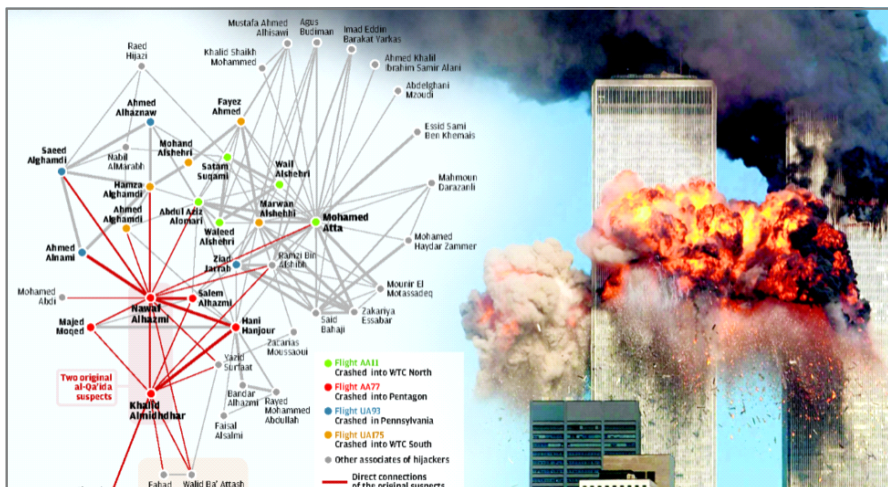
1) 주요 연구내용

□ 빅데이터 선행이론 검토

빅데이터의 개념 및 빅데이터가 우리 생활에 밀접하게 된 이론적 배경을 살펴본다. 이를 위해 빅데이터 관련 국내외 연구 문헌과 미래 정책 개발과 관련한 국내외 연구 문헌을 검토한다. 이를 바탕으로 국내외 빅데이터 활용 사례를 살펴봄으로써 빅데이터 관련 선행 이론에 대한 함의를 분석한다.

□ 건축·도시 분야 빅데이터 현황 조사

건축·도시 분야에서 활용할 수 있는 빅데이터는 다양한 형태로 나타난다. 기본적으로 건축물 대장 정보, 국가건물에너지 통합 정보 등 공간에 대한 데이터 조사 및 분석 방법이 존재하며, 트위터, 페이스북, 블로그 등과 같은 SNS 데이터를 활용하여 현황을 분석할 수 있다. 또한 보도자료, 언론 기사, 선거공약 등 미디어 데이터를 분석하는 방법, 논문·특허·보고서 등 학술 문헌 데이터를 조사 분석하는 방법 그리고 구글, 네이버 등 검색 데이터 조사 및 분석 방법이 있다.



[그림 1-2] 빅데이터 분석을 활용한 범죄수사분석(911테러)

출처 : <http://www.redflagnews.com/headlines/senators-warn-911-terror-network-still-active-in-us>

□ 정책개발을 위한 자료별 특성연구

정책개발을 위한 자료별 특성 연구로 크게 세 가지 유형의 빅데이터를 활용하고자 한다. 첫 번째는 여론 및 정서 탐색, 조기 정보 등 정책의제 설정을 위한 빅데이터이며, 두 번째는 실시간 반응, 참여와 협조, 공감대 형성 등 정책 결정을 위한 빅데이터이다. 마지막으로 집행 모니터링, 여론 추이 탐색 등 정책 집행과 평가를 위한 빅데이터를 활용하여 연구를 진행하고자 한다.

□ 빅데이터 분석을 위한 모델 설계 및 실증

빅데이터 분석을 위한 모델 설계 및 실증을 위해 다음과 같은 방법을 진행하고자 한다. 우선, 대상 정보 및 수집 범위를 설정하고, 대상 정보 수집에 대한 모듈 기획 및 설계를 진행한다. 그리고 수집 데이터 분석 방법의 설정 및 분석 알고리즘에 대한 조사 분석을 실시하여, 대상 정보 분석 모듈의 기획 및 설계를 진행한다. 이후 분석 수행 및 해석을 통해 문제점 파악 및 개선방안을 도출 한다.



[그림 1-3] 서울시 심야버스노선 결정과정

출처 : 이승환·김인철·권성수·윤익준·조희진(2015), 「지방자치단체 정책자원을 위한 데이터 분석 운영체계 연구」, 대구대학교 산학협력단

□ 정책개발을 위한 빅데이터 활용방안 제시

빅데이터 활용 모델 실증결과를 토대로 수정 및 보완한다. 이를 바탕으로 최종적으로 건축·도시 정책개발을 위한 빅데이터 활용방안을 제시한다.

2) 연구추진방법

□ 문헌조사

빅데이터 관련 문헌조사로 관련 이론 및 선행 연구 동향을 분석한다. 그리고 국내외 빅데이터 분석 방법론에 대한 조사 및 분석을 실시하고, 빅데이터와 관련된 건축·도시 정책 현황을 살펴본다.

□ 사례조사

사례 분석으로는 국내 및 해외에서 활용되는 정부 정책 개발에 대한 사례 조사를 실시한다. 또한 국내외 빅데이터 활용 우수 사례를 살펴봄으로써 추후 활용 방안에 대한 기초 자료를 수집한다.

□ 빅데이터 유형별 자료특성 조사

빅데이터는 여러 유형이 있으며, 각각의 유형별 특성이 다르다. 트위터, 페이스북, 블로그 등으로 대표되는 소셜 데이터, 건축물 대장 정보, 국가건물에너지 통합 정보 등 공간데이터, 보도자료, 언론기사, 선거공약 등 미디어데이터, 논문, 특허, 보고서 등 학술 문헌 데이터, 구글, 네이버 등 검색데이터 등 각각의 데이터 유형별 자료특성을 조사 분석한다.

□ 빅데이터 분석 방법론 조사

빅데이터는 전통적인 통계분석 방법으로 분석이 어려운 경우가 많으며, 이러한 빅데이터를 분석하기 위하여 여러 방법론이 대두되고 있다. 대표적인 빅데이터 분석방법론인 텍스트 마이닝 기법, 네트워크 분석, 공간 활용 분석 기법 등에 대한 조사를 수행한다.

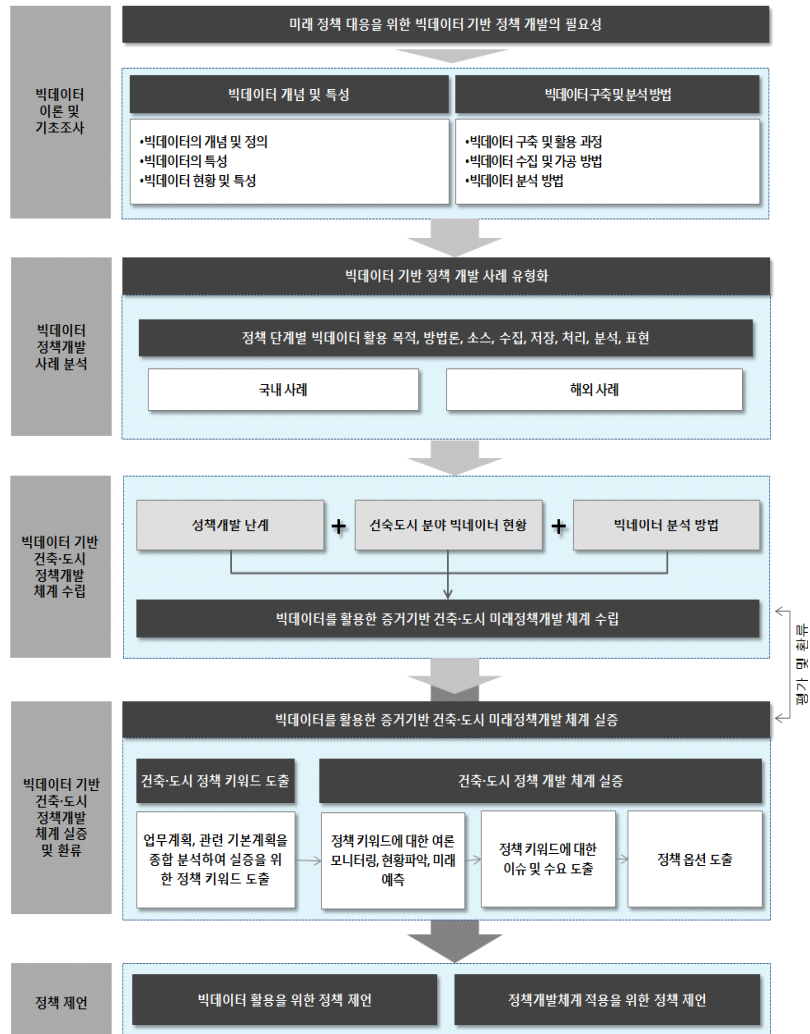
□ 사례 적용

빅데이터 사례, 유형, 분석 등의 선행 연구 결과를 토대로 빅데이터기반 건축·도시 미래정책 개발을 위한 체계를 수립하고, 수립된 체계의 실증을 위하여 사례분석을 수행한다. 사례분석을 위하여 대상 자료를 선정하고 이를 분석 및 해석하여 수립된 체계에 대한 평가와 환류 과정을 통하여 개발체계를 확정한다.

□ 전문가 자문

빅데이터의 유형 및 특성 연구, 개발체계의 수립과 활용방안 모색 등 연구의 진행 단계에서 공공·학계·실무 분야의 전문가 자문을 받아 단계별 추진방향을 점검한다.

□ 연구의 흐름도



[그림 1-4] 연구 흐름도

3) 연구의 범위

미래정책은 가까운 미래부터 30년 이상 먼 미래까지 범위가 매우 넓다. 이에 본 연구에서는 시간적 범위를 현재의 건축·도시 분야의 정책 키워드를 기준으로 이후 정책개발이 시급한 근 미래로 범위를 한정하였다. 연구의 내용적 범위는 정책과정 중의 정책개발에 해당하는 문제 형성 단계, 정책형성 단계에 빅데이터 활용 체계로 설정하였다.

제2장 빅데이터 관련 선행이론

1. 빅데이터 개념 및 특성
2. 빅데이터 분석 방법론 고찰

1. 빅데이터 개념 및 특성

1) 빅데이터 개념 및 정의

□ 빅데이터의 등장

정보통신기술 발달에 따라 관련 데이터가 급격히 증가하게 되었고 이러한 대규모 데이터의 활용방안에 대한 관심이 커지면서 ‘빅데이터’라는 용어가 등장하였다. 과거 1990년대에는 테라바이트 단위의 데이터를 데이터베이스에 저장하였지만, 2010년대에는 엑사바이트(1EB=1,000PB) 단위로 급속히 증가하고 있다¹⁾. 구글에 따르면 2003년까지 인류가 생산한 데이터가 총 5엑사바이트 수준인데, 이제는 같은 양을 단 하루 만에 생산²⁾하는 수준으로 데이터의 양이 증가하였다. 또한 Mckinsey에 따르면 매일 300억 개의 콘텐츠가 페이스북에서 공유되며, 전 세계 데이터는 매년 40%씩 증가³⁾하고 있다.

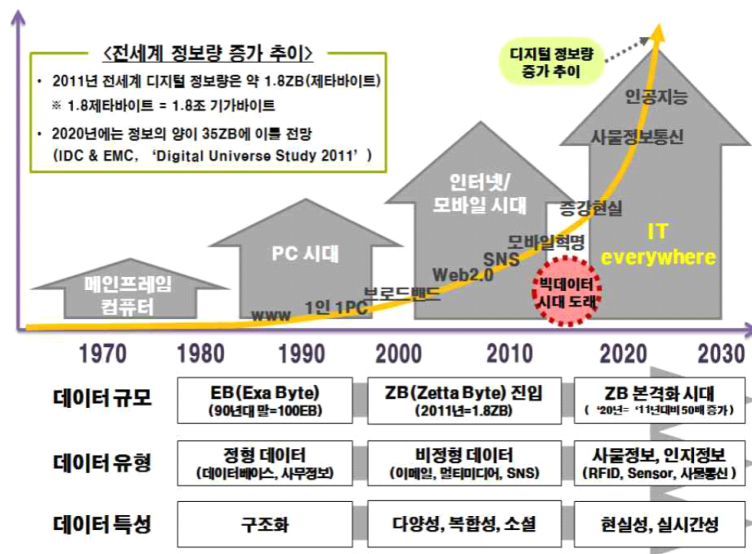
[표 2-1] 데이터 단위(바이트)

이름	킬로(kilo)	메가(mega)	기가(giga)	테라(tera)	페타(peta)	엑사(exa)	제타(zetta)	요타(yotta)
기호	k	M	G	T	P	E	Z	Y
크기	103	106	109	1012	1015	1018	1021	1024

1) 전도홍·김형래·김정래(2015), 「데이터 분석을 중심으로 한 빅데이터 입문」, 정익사. p.10

2) 한국소프트웨어기술인협회 빅데이터전략연구소(2016), 「빅데이터 개념」, 광문각. p.13

3) Manyika, James, Michael Chul, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh and Angela Hung Byers (2011), *Big Data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute. p.2



[그림 2-1] 빅데이터 개념의 등장

출처 : 정자선(2011), 「新가치창출 엔진, 빅데이터의 새로운 가능성과 대응전략」, 한국정보화진흥원, p.12

빅데이터라는 용어는 2012년 세계경제포럼에서 신기술 중 하나로 빅데이터가 선정되면서 통용되기 시작하였는데, ICT, IOT 기술 등이 보편화됨에 따라 소셜, 사물, 라이프로그 데이터 등이 결합되면서 빅데이터의 영향력이 증대되면서 등장했다. 기존 데이터 관리 시스템으로 다루기 어려운 비정형 데이터가 빠르게 증가하면서 기존의 데이터 저장·관리·분석 기법이 한계와 도전에 직면하게 되고 빅데이터라는 신기술이 부상하게 된 것이다. 또한 빅데이터의 가공과 분석에 따라 상황인식, 문제해결, 미래전망이 가능해지고 데이터가 경제적 자산과 경쟁력의 척도로 부각되면서 중요성이 날로 증가하고 있다.⁴⁾

□ 빅데이터의 정의

빅데이터에 대한 정의는 2001년 Gartner에서 정의한 데이터 규모(Volume), 빠른 속도(Velocity), 다양성(Variety)의 3가지 개념(3Vs)⁵⁾이 가장 대표적이며, 이는 데이터의 급성장에 따른 시대 변화를 나타낸다. Gartner는 2012년 빅데이터의 활용 및 목적의 개념까지 포함하여 다음과 같이 재정의 하였다.

“ 통찰력, 의사결정, 프로세스 최적화를 위해 비용효율적이고 혁신적인 정보처리 과정을 요하며, 3Vs(Volume, Velocity, Variety)를 갖는 정보 자산”⁶⁾

4) 이원규, 오동하(2013), 빅데이터(Big Data)를 활용한 부산시 도시관리 전략, 부산발전연구원, 6p.

5) Laney Doug (2001), 3D Data Management: Controlling Data Volume, Velocity, and Variety, META Group

이후 다양한 연구자들이 빅데이터의 특성에 가시성(Visibility), 유효성(Validity), 정확성(Veracity), 가치(Value) 등 다양한 V를 추가하면서 빅데이터에 대한 정의를 확대하였다. 우리나라 국가 정보화 전략위원회에서는 ‘대용량 데이터를 활용·분석하여 가치 있는 정보를 추출하고 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술⁷⁾’로 정의하고 있다. 이 외에 Mckinsey는 데이터 크기에 보다 초점을 맞추어 기존의 데이터베이스 관리 도구로 수집·저장·관리·분석하기 어려운 대용량 데이터 집합⁸⁾으로 정의하였으며, IDC에서는 다양한 유형의 대규모 데이터를 활용하여 저렴한 비용으로 새로운 가치를 창출하고, 데이터의 빠른 수집, 분석을 지원하도록 고안된 신기술과 아키텍처(컴퓨터 시스템 구성)⁹⁾로 정의하고 있다.

이러한 기존의 빅데이터의 정의를 종합해 보면, 빅데이터란 기존의 데이터에 비해 규모가 크고, 복잡 다양하며, 생성·처리 속도가 빨라 새로운 관리 및 분석 방법이 요구되는 데이터의 속성적 개념과 함께, 데이터를 활용·분석하여 가치 있는 정보를 추출하는 데이터 가치의 개념까지 포함하는 정보화 기술로 정의할 수 있다.

2) 빅데이터의 특성과 형태

빅데이터의 특성은 데이터 규모(Volume), 빠른 속도(Velocity), 다양성(Variety), 정확성(Veracity), 가치(Value)¹⁰⁾ 등 5가지로 구분하여 설명된다. 첫 번째, 규모(Volume)는 IT의 일상화에 따라 디지털 정보량이 기하급수적으로 폭증, 요타바이트(YB¹¹⁾)시대로 진입을 앞두고 있는 엄청난 정보의 양이며, 두 번째, 다양성(Variety)은 로그 기록, 소셜 데이터, 위치 정보, 카드사용내역 등 데이터 종류의 증가 및 비정형 데이터 유형의 다양하게 나타나는 빅데이터의 특성이다. 세 번째, 빠른 속도(Velocity)는 센서 정보, 스트리밍 정보 등 실시간 정보 증가, 실시간 정보 증가에 따른 데이터 생성 및 이동 속도 증가 그리고 실시간 데이터 활용을 위해 데이터 처리 및 분석 속도가 중요함을 의미하며 이 모든 것이 빠르게 나타나고 있

6) Beyer, Mark A. and Douglas Laney (2012), *The Importance of 'Big Data' : A Definition*, Gartner

7) 이각범(2011), 「빅데이터를 활용한 스마트 정부 구현(안)」, 국가정보화전략위원회

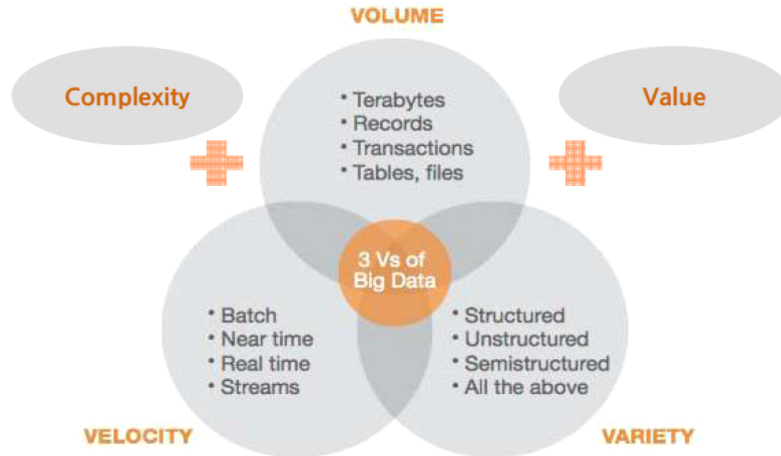
8) Manyika, James, Michael Chul, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh and Angela Hung Byers (2011), *Big Data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute

9) Vesset, Dan, Henry D. morris, Gard Little, Lucinda Borovick, Susan Feldman, Matthew Eastwood, Benjamin Woo, Richard L. Villars, Jean S. Boman, Carl W. Olofson, Steve Conway and Natalya Yezhkova (2012), *Worldwide Big Data Technology and Services 2012-2015 Forecast*, idc. p.1

10) 한국소프트웨어기술인협회 빅데이터전략연구소(2016), 「빅데이터 개론」, 광문각. p.24

11) 1000GB = 1TB, 1000TB = 1PB, 1000PB = 1EB, 1000EB = 1ZB, 1000ZB = 1YB

는 현상을 의미한다. 네 번째, 정확성(Veracity)은 신뢰도 높은 데이터 활용이 분석의 정확도에 영향을 미치는 것을 의미하며, 다섯 번째, 가치(Value)는 빅데이터가 추구하는 지향점을 가치 창출로 미래 지향적 특성을 담고 있다.



[그림 2-2] 빅데이터의 특성

출처 : 이마숙·이창훈·김지연(2014), 「빅데이터를 활용한 환경분야 정책수요 분석」, 한국환경정책·평가연구원

빅데이터는 자료의 형태에 따라 정형, 비정형, 반정형 데이터로 구분할 수 있다¹²⁾. 정형 데이터는 일정한 항목(필드)에 저장한 데이터를 의미하며 건축물 대장 정보, 토지거래 정보 등과 같이 데이터베이스로 관리가 가능한 반면, 비정형 데이터는 항목이 정해져 있지 않은 데이터로 블로그, 소셜 데이터, 뉴스미디어 등을 통해 생산된 데이터를 의미한다. 예를 들어 페이스북, 트위터, 유튜브 등에서 제공하는 데이터가 이에 속한다. 반정형 데이터는 일부 정형 데이터를 포함하는 경우인데, 예를 들어 메타데이터를 포함하는 문서, 사진 등이 여기에 속한다.

[표 2-2] 빅데이터 형태에 따른 유형 구분

구분	내용
정형 데이터 (structured data)	·정해진 구조로 고정된 필드에 저장되어 있는 데이터 ·예)관계형 데이터 베이스(RDBMS), 스프레드시트 등
반정형 데이터 (semi-structured data)	·고정된 필드에 저장되지는 않지만 메타데이터, 스미카 등을 포함하는 데이터 ·예) XML, HTML 문서 등
비정형 데이터 (unstructured data)	·미리 정해진 구조가 없고 고정된 필드에도 저장되지 않는 데이터 ·예) 텍스트 문서, 이미지, 동영상, 음성 등

출처 : 박시룡·안순권·장선화·이효정(2013), 「빅데이터를 활용한 과학적·효율적 행정 구현방안」, (주)서울경제신문 백상경제연구원. p.18

12) 김대중·윤서연(2013), 「빅데이터를 활용한 국토정책반응 모니터링 및 정책수요 예측 방안」, 국토연구원

이러한 빅데이터는 기존 데이터와 관리 방식에서 차이점이 존재한다. 기존의 대용량 데이터는 데이터관리시스템(DBMS)을 활용하여 처리하였다면, 빅데이터에서는 더 크고 다양한 데이터를 관리할 수 있는 새로운 기술 요구되고 있다. 대표적으로 Hadoop과 NoSQL기술이 기존 데이터 처리 방식과 대비되는 신기술이며, Hadoop은 파일시스템으로 정형, 비정형 구분없이 처리가 가능하다. NoSQL 역시 하나의 레코드에 정형, 비정형 데이터가 모두 들어갈 수 있도록 하는 데이터베이스 관리 방식 중 하나이다.

[표 2-3] 기존의 데이터 관리 방식과 빅데이터 관리 방식의 차이점

구분	기존	빅데이터
데이터 트래픽	·테라바이트 수준	·페타바이트수준 ·장기간의 정보 수집 및 분석 ·방대한 데이터 처리량
	·데이터가 수십억 건 이상되면 H/W를 더 많이 사 용해도 성능이 증가하지 않음	·H/W만 늘리면 성능 증가
데이터 유형	·정형데이터 중심	·비정형 데이터의 비중이 높음 ·처리의 복잡성 증대
프로세스 및 기술	·단순한 프로세스 및 기술 ·정형화된 처리/분석 결과 ·원인/결과 규모 중심	·다양한 데이터 소스 및 복잡한 로직 처리 ·데이터 처리 복잡도가 높아 분산 처리기술 필요 ·새롭고 다양한 처리방법 개발 필요 ·상관관계 규명 중심
	·RDB(Relational Database)에 근간	·Hadoop, NoSQL, CEP, R 등 개방형 소프트웨어

출처 : 노규성(2016), 「빅데이터와 공공 혁신 10대 사례」, 커뮤니케이션북스㈜, 남광우·신강원·김대중·신동빈·안종욱·김승범·최산화·장동인 (2014), 「특집 빅데이터와 스마트도시」, 「도시정보(대한국토도시계획학회 정보지)」, v.390, 대한국토도시계획학회를 종합하여 재작성

3) 빅데이터 종류와 특징

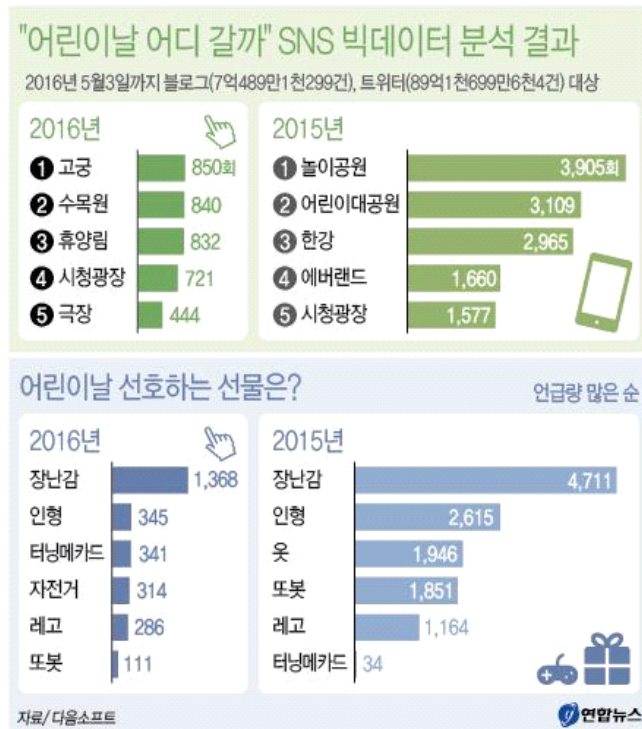
① 비정형, 반정형 데이터

□ 소셜 데이터

소셜 빅데이터는 사람들 간의 인적 네트워크를 기반으로 발생하는 텍스트, 이미지 등의 데이터를 말하며, 크게 SNS, 블로그, 커뮤니티 빅데이터로 구분할 수 있다.

SNS 빅데이터는 트위터, 페이스북 등 소셜네트워크 서비스를 통해 작성된 글과 사진, 기사 정보를 의미하며, 사람들 간의 위치, 관심사, 감정 등을 태그로 공유하기 때문에 심층적 분석이 가능하다. 페이스북은 SNS 중 가장 큰 규모의 서비스로 분석데이터로서 가치가 크나, 폐쇄적

공유가 많고 웹 크롤러의 접근을 막고 있어 직접적인 정보 수집에는 어려움이 있다. 트위터는 웹 크롤러를 통한 수집이 용의하고 공개를 기본으로 하고 있어 SNS 빅데이터 분석에 가장 많이 활용되고 있으나 페이스북에 비해 상대적으로 이용자가 작다. 블로그는 일상 정보와 개인의 특정 관심사가 주를 이루며, 전문가부터 일반인까지 다양한 주제가 공유되는 텍스트 데이터로 정보제공, 보도자료 공개, 논평부터 일상적인 공유와 광고에 이르기까지 다양한 목적의 활동이 포함되어 있다. 커뮤니티 빅데이터는 포털 등의 커뮤니티 사이트에서 특정 주제에 관심이 있는 사람들이 공유하는 텍스트 데이터로 관련 이슈에 대한 여론 파악에 활용도가 높다.

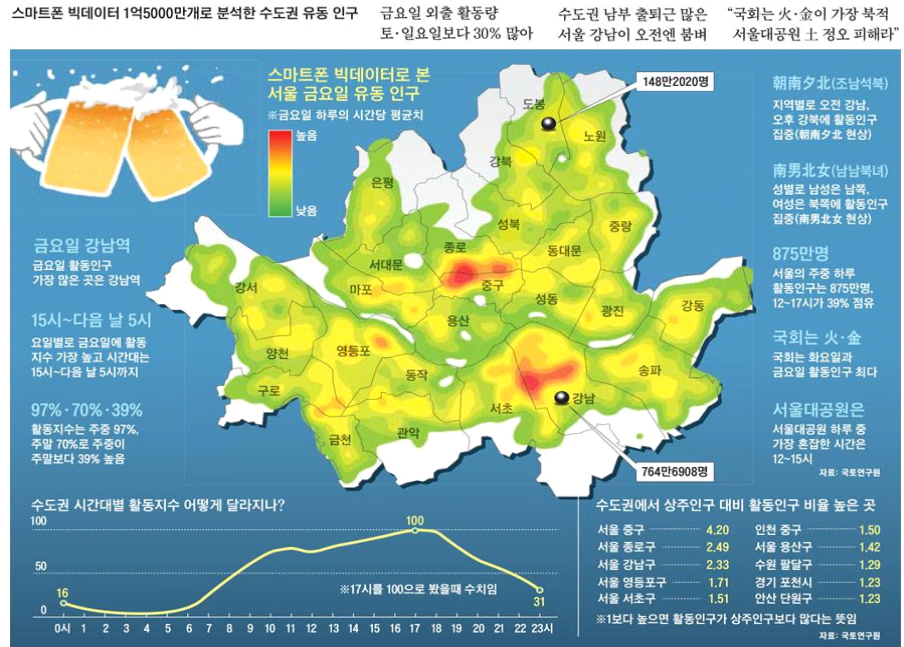


[그림 2-3] SNS 빅데이터 분석 사례
 출처: <http://www.yonhapnews.co.kr>

□ 로그 데이터

로그 데이터는 사람들이 일상생활에서 사용하는 모든 기기와 도구들이 남기는 디지털 흔적을 지칭하며, 기기와 도구의 종류에 따라 매우 다양하다. 빅데이터로 다루는 로그 정보는 신용카드 사용정보, 휴대전화로그정보, 교통카드 사용정보, 인터넷 사용정보, 차량 네비게이션 운행정보, CCTV 영상정보 등을 예로 들 수 있으며, 대표적으로 활용되는 정보는 신용카드 사용정보와 휴대전화로그정보이다.

기자국을 통해 취득되는 휴대전화기기의 위치 정보는 시간 및 공간에 따른 휴대전화가입자의 분포를 확인할 수 있으며, 이동통신사에서 일정 공간 단위로 개별 위치정보를 취합한 휴대전화 로그 데이터를 구축할 수 있다. 우리나라는 휴대전화 보급률이 매우 높아 사람들의 생활패턴을 실시간으로 파악할 수 있어 다양한 분야에 활용 가능하다.



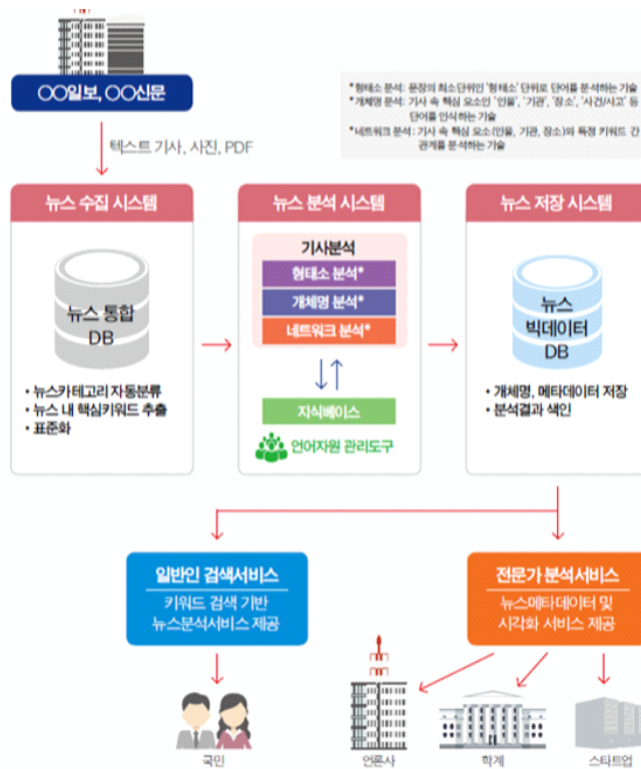
[그림 2-4] 휴대폰 빅데이터로 유동인구 분석 사례

출처: http://news.chosun.com/site/data/html_dir/2015/02/10/2015021000456.html

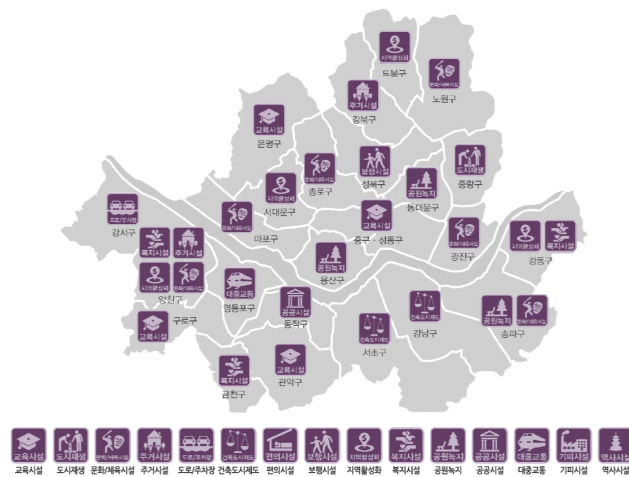
□ 미디어 데이터

미디어 데이터는 온라인 뉴스가 일반화 되면서 생산되는 신문사, 방송사, 포털 등의 뉴스 데이터로 보도한 기사 원문의 텍스트 데이터와 그와 관계된 데이터를 의미한다.

언론기사 빅데이터는 온오프라인 상의 언론사의 보도 데이터인 만큼 데이터의 신뢰도가 높아 양질의 분석이 가능하다. 그 외 해당 기사의 댓글의 수집과 분석을 통하여 특정 이슈에 관한 여론의 동향을 실시간으로 파악할 수 있다. 보도자료 빅데이터는 정부, 공공기관, 협회 등의 공공단체가 생산하는 정책자료로 언론기사 빅데이터와 유사한 성격을 갖고 있지만 상대적으로 수집과 가공이 용이하여 특정 주제나 정책의 효과 등을 보다 효율적으로 분석할 수 있다. 선거공약 빅데이터는 대통령, 국회의원, 지자체장, 지자체의원 등 선출직 공무원의 공약 데이터로 실질적인 지역의 민의를 종합한 데이터로 지역별 정책 수요 분석을 할 수 있다.



[그림 2-5] 언론기사 빅데이터 분석 시스템 사례
출처: <https://www.bloter.net/archives/254773>



[그림 2-6] 서울시 제20대 국회의원 건축도시정책 공약 현황
출처: 조영진·이은석·이경신(2016), 「건축도시정책동향」, v.39, 건축도시공간연구소, p.59

② 정형 데이터

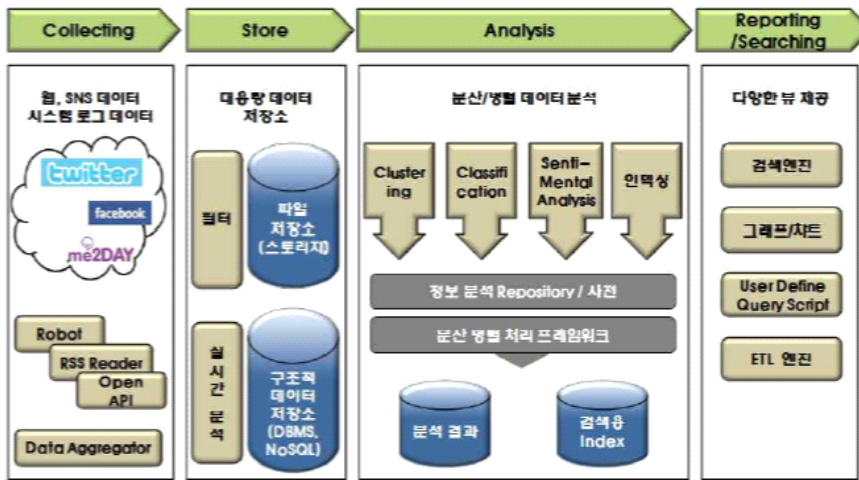
정형 데이터는 일정한 항목(필드)에 저장한 데이터를 의미하며, 빅데이터 등장 이전부터 전통적으로 다뤄져온 대부분의 데이터가 정형데이터라고 할 수 있다. 정형 빅데이터는 우리나라 통계청에서 구축 제공하고 있는 각 종 통계데이터, 건축행정정보시스템의 건축물 대장 정보 등과 학술 문헌·기술 문서·특허 정보 등의 서지 또는 색인자료 등을 예로 들 수 있다.

이러한 정형 데이터는 비정형·반정형 데이터에 비해 상대적으로 가공이 용이하고 다양한 통계분석이 가능하다. 빅데이터에 대한 관심이 높아지면서 기존에 단독으로 분석 사용되어 왔던 많은 통계들이 서로 다른 이 기종 통계와 결합하여 새로운 시사점을 만들어 내는 작업들이 활발하게 진행되고 있으며, 특히 기존의 정형 데이터와 비정형, 반정형 데이터와 연계를 통하여 새로운 전략을 만들어 내고 있다.

2. 빅데이터 분석 방법론 고찰

1) 빅데이터 구축 및 활용 과정

빅데이터 구축 및 활용은 데이터 수집, 데이터 가공, 데이터 분석, 시각화의 과정을 거친다. 데이터 활용 목적에 따라 데이터의 종류를 설정하고, 특성을 파악한 후 해당 데이터를 수집한다. 수집한 데이터를 특성에 따라 정형데이터의 경우 다른 데이터와의 매칭이 가능하도록 속성 값의 표준화를 처리하고, 반정형·비정형 데이터의 경우 전처리 과정을 통해 정형 데이터로 가공한다. 공간분석이 필요한 경우 위치정보를 추출하여 공간 정보화할 수 있다. 분석결과에 대하여 도표, 차트, 지도 등으로 시각화를 통해 이를 활용할 수 있다.



[그림 2-7] 빅데이터 구축 및 활용 과정

출처 : 차재필(2013), 「빅데이터 정책동향」, 한국교통연구원, 세미나 자료

2) 빅데이터 수집 및 가공 방법

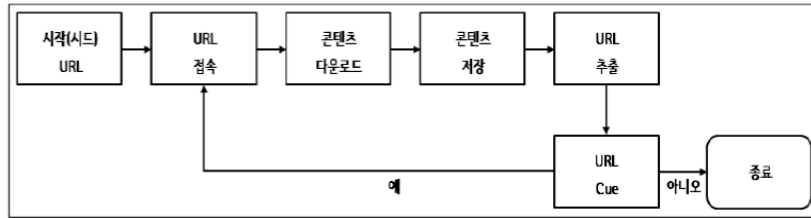
□ 데이터 유형별 수집 방법

정형데이터의 경우 API(Application Programming Interface)시스템으로 공개된 경우 규약에 따라 수집 및 활용할 수 있으며, 데이터 보유 기관과 협의가 된 경우 파일단위로 받아서 오프라인 형태로 연계하는 방법과 DB연계를 통해 온라인 실시간 연계로 구축이 가능하다. 그러나 트래픽, 시스템 과부하, 데이터 소유권 등의 문제로 인해 데이터 보유기관과 사전협의를 필요하다. 수집한 데이터는 각 기관의 데이터 구조 및 내용에 따라 다를 수 있으므로 활용성을 높이기 위해서 통합적 데이터 구축에 대한 고려가 필요하며, 데이터별 생산 및 갱신 주기를 고려하여 데이터 수집 시기와 방법을 결정해야 한다.¹³⁾

비정형 및 반정형 데이터에 대한 수집 방법은 웹 크롤러¹⁴⁾를 이용하여 인터넷에 공개된 다양한 웹정보를 자동으로 복사 및 수집하는 웹 크롤링 방법이 대표적이며, 연구문헌, 선거공약의 데이터 등 규모가 크지 않은 데이터의 경우 온라인 검색 및 다운로드 등을 통해 활용할 수 있다.

13) 김대중·윤서연(2013), 「빅데이터를 활용한 국토정책반응 모니터링 및 정책수요 예측 방안」, 국토연구원

14) 웹문서의 위치(URL)와 링크정보, 문서내용 등 각종 정보들을 수집해오는 컴퓨터 프로그램으로 스파이더, 로봇, 웹수집기, 로봇에이전트 등 다양한 이름으로 불림



[그림 2-8] 웹크롤러를 이용한 비정형 데이터 수집 과정

출처 : 김대중·윤서연(2013), 「빅데이터를 활용한 국토정책반응 모니터링 및 정책수요 예측 방안」, 국토연구원, p.53

[표 2-4] 데이터 수집 기술

구분	특징	비고
크롤링(Crawling)	·SNS, 뉴스, 웹 문서 등 인터넷 상에서 제공되는 웹정보 수집	웹문서 수집
FTP	·TCP/IP 프로토콜을 활용하는 인터넷 서버로부터 각종 파일을 수신	파일 수집
Open API	·서비스, 정보, 데이터 등을 어디서나 쉽게 이용할 수 있도록 개방된API 데이터 수집 방식 제공 ·다양한 어플리케이션을 개발할 수 있도록 개발자와 사용자에게 공개	실시간 데이터 수집
Rss(Really Simple Syndication)	·웹 기반 최신정보를 공유하기 위한 XML 기반 콘텐츠 배급 프로토콜	콘텐츠 수집
스트리밍(Streaming)	·인터넷에서 텍스트, 음성, 오디오, 비디오 데이터를 실시간으로 지속적으로 수집할 수 있는 기술	실시간 데이터 수집
Log Aggregator	·웹서버 로그, 웹로그, 트랜잭션 로그, 클릭로그, DB의 로그 등 각종 로그 데이터를 수집하는 오픈 소스 기술(Chukwa, Flume, Scribe 등)	로그수집
RDB Aggregator	·관계형 데이터베이스에서 정형데이터를 수집하여 HDFS(하둡 분산파일시스템)이나 HBase와 같은 NoSQL에 저장하는 오픈 소스 기술(Sqoop, Direct JDBC/ODBC 등)	RDB기반 데이터 수집

출처 : 양현철·김진철·신신애·김배현(2014), 「빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼」, 미래창조과학부, 한국정보화진흥원, p.14

□ 데이터 가공 방법

정형데이터의 경우 다양한 자료의 출처 및 특성을 쉽게 파악하고 효율적으로 데이터를 관리하기 위해 데이터 표준화 작업 필요하다. 즉, 데이터 출처 및 생산 시기 등 메타데이터를 활용한 표준화 작업이 수행되어야 한다. 비정형·반정형 데이터의 경우 방대한 양과 불규칙한 데이터 속성으로 전처리 과정이 필수적으로 요구된다. 따라서 대부분 텍스트 자료로 자연어처리 기술을 이용하여 형태소를 추출하여 가공해야 한다.

공간분석이나 지도로 시각화 작업이 필요한 경우, 공간정보화 과정이 필요하다. 예를 들어, 텍스트 데이터 내용에 지명이나 주소가 포함된 경우, 트위터 이용 시 저장된 위치의 정보, 사진에 저장된 위치 정보, 정형데이터에 포함된 주소 또는 경위도 좌표 등 모든 유

형의 데이터가 다양한 형태로 위치정보를 가지고 있을 수 있으므로, 이를 활용하여 표준화된 공간정보화(Geo-coding, Geo-tagging, Geo-parsing 등)가 가능하다.

데이터 가공 방식은 데이터 여과, 변환, 정제, 통합, 축소 등의 과정을 거친다. 필터링(Filtering)은 데이터의 오류 발견, 보정, 삭제 및 중복성 확인 등의 과정을 통해 품질을 향상시키는 방법으로 자연어 처리 기술, 기계학습 등의 추가 기술 적용 필요하며, 데이터 변환(Transformation)은 데이터 분석이 용이하도록 데이터 유형 등을 일관성 있게 변환하는 기술로 새로운 데이터 속성을 만들거나, 데이터 추세에서 벗어나는 값을 변환하고, 통계적 기법으로 정규화하는 등 다양한 변환 방법이 사용된다. 데이터 정제(Cleansing)는 수집된 데이터의 불일치성을 교정하기 위한 방식으로 결측치 처리 및 잡음 처리 기술이 활용되며, 데이터 통합은 출처가 다른 상호 연관성 있는 데이터들을 하나로 결합하는 기술, 데이터 축소는 분석에 불필요한 항목을 삭제하거나, 데이터 인코딩 등으로 데이터 축소하여 분석 효율성을 향상시키는 방법이다.¹⁵⁾

[표 2-5] 데이터 가공 방법

구분	특징
필터링(filtering)	·오류 발견, 보정, 삭제 및 중복성 확인 등으로 데이터 품질 향상
데이터 변환 (transformation)	·데이터 유형 변환 등 데이터 분석이 용이한 형태로 변환 ·정규화, 집합화, 요약, 계층 생성 등의 방법 활용
데이터 정제(Cleansing)	·결측치들을 채워 넣고, 이상치를 식별 또는 제거하고, 잡음 섞인 데이터를 평활화하여 데이터의 불일치성을 교정
데이터 통합(Integration)	·데이터 분석이 용이하도록 유사 데이터 및 연계가 필요한 데이터들을 통합하는 기술
데이터 축소(Reduction)	·분석 컴퓨팅 시간을 단축할 수 있도록 데이터 분석에 활용되지 않는 항목 등을 제거하는 기술

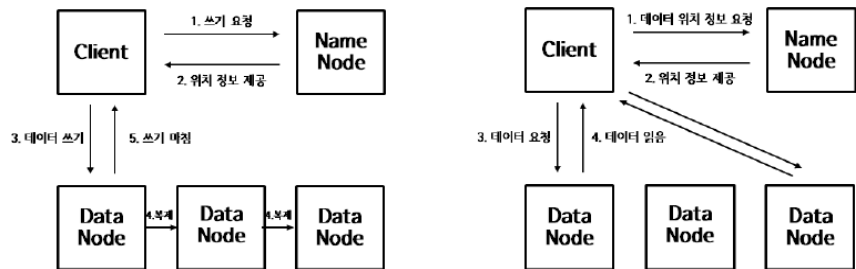
출처: 양현철·김진철·신신애·김배현(2014), 「빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼」, 미래창조과학부, 한국정보화진흥원 p.28

□ 저장 및 관리 방법

방대한 양의 정보를 신속하고 안정적으로 저장하고 처리하기 위해서 빅데이터 처리 기술이 필요하다. RDB 저장은 컬럼과 값을 코드화하거나, 데이터형을 변환 처리하여 테이블 형태로 저장하는 방법이며, NoSQL 저장방법은 RDB 저장 데이터를 컬럼과 값을 key와 value로 구분하여 저장하고, 비정형 데이터는 key-value로 저장하는 방법이다. 즉, 수집한 데이터가 RDB 테이블 형태의 데이터인 경우 key값을 추출하고 전체 데이터는 value에 저장하며, key값과 추출값 간의 관계를 파악할 수 있는 변환용 검증기가 부여된

15) 양현철·김진철·신신애·김배현(2014), 「빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼」, 미래창조과학부, 한국정보화진흥원.

다. xml, html등의 비정형 데이터인 경우, key값을 추출하고 데이터 내용은 value에 저장되며, 데이터 내용을 document기반으로 변환하여 저장할 수 있다.



[그림 2-9] HDFS의 데이터 쓰기 과정(좌)과 데이터 읽기 과정(우)
출처 : 김대중·윤서연(2013), 「빅데이터를 활용한 국토정책반응 모니터링 및 정책수요 예측 방안」, 국토연구원

분산파일 시스템은 대용량 파일을 여러 대의 컴퓨터에 나누어 저장하고 관리하는 기술로, 파일을 블록단위로 분할해 여러 데이터 노드에 나누어 저장하며, 각 파일의 기본 정보 및 블록의 위치 정보를 네임 노드에서 관리하도록 한다. 파일의 위치와 내용 정보에 대한 복사본을 여러 대의 컴퓨터에 분산하여 저장하기 때문에 검색 시간도 단축되고, 여러 곳에서 동시에 검색 요청이 들어와도 한 곳에 작업량이 집중되지 않는 장점이 있다.¹⁶⁾

[표 2-6] 데이터 저장 방식

구분	특징	비고
RDB	·관계형 데이터를 저장하거나, 수정하고 관리할 수 있게 해주는 데이터 베이스 ·SQL문장을 통하여 데이터베이스의 생성, 수정 및 검색 등 서비스 제공	oracle, mssql, mysql, sybase, MPPDB 등
NoSQL	·Not-Only-SQL의 약자이며, 비관계형 데이터 저장소로, 기존의 전통적인 방식의 관계형 데이터베이스와는 다르게 설계된 데이터베이스 ·테이블 스키마가 고정되지 않고 테이블간 조인 연산을 지원하지 않으며, 수평적 확장이 용이 ·key-value, document key-value, colum 기반의 NoSQL이 주로 활용 중	MongoDB, Cassandra, Hbase, Redis 등
분 산 파 일 시스템	·분산된 서버의 로컬 디스크에 파일을 저장하고 파일의 읽기, 쓰기 등과 같은 연 산을 운영체제가 아닌 API를 제공하여 처리하는 파일시스템 ·파일 읽기/쓰기 같은 단순 연산을 지원하는 대규모 데이터 저장소 ·범용 x86서버의 CPU, RAM 등을 사용하므로 장비 증가에 따른 성능 향상 용이 ·페타바이트 단위의 대용량데이터 저장 지원 용이	HDFS (Hadoop Distributed File System)

출처 : 양현철·김진철·신신애·김배현(2014), 「빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼」, 미래창조과학부, 한국정보화진흥원, p.38

16) 김대중·윤서연(2013), 「빅데이터를 활용한 국토정책반응 모니터링 및 정책수요 예측 방안」, 국토연구원

□ 빅데이터 분석 서비스 제공 플랫폼

다양한 민간 및 공공기관에서 앞서 제시한 데이터 수집 및 가공 기술을 확보하기 어려운 사용자들에게 쉽게 빅데이터를 처리하고 분석할 수 있도록 빅데이터 인프라 서비스를 제공하고 있다. 민간 업체의 경우, 포털사이트, 통신사, 신용카드 빅데이터 센터 등에서 주도적인 역할을 하고 있다.

네이버 데이터랩은 사업을 운영하고 있거나 창업을 계획 중인 예비 사업자들이 자신들의 데이터 경쟁력을 확보할 수 있도록 민간공공 데이터를 확인하고 분석할 수 있는 빅데이터 서비스를 제공하고 있으며, SK 텔레콤 지오비전 데이터는 SK그룹 및 분야별 전문 파트너사가 보유한 DB를 Mash-up하여 다양한 원천데이터의 Join이 가능한 공간데이터 서비스를 제공한다. 솔트룩스에서는 소셜 데이터를 실시간으로 수집 및 분석 하는 플랫폼을 개발하여 빅데이터 분석과 이에 따른 복잡한 패턴을 풀이하여 미래 예측까지 가능하도록 서비스를 제공하고 있으며, TIBUZZ는 온라인상의 방대한 데이터를 수집하고 분류, 분석 및 시각화를 통하여 인사이트를 제공하는 소셜 데이터 분석 서비스이다. BC카드 스토어 스토리는 창업에 앞서 과거 동일건물에서 업종별/매출추이를 확인할 수 있는 서비스이며, 신한카드 빅데이터 센터 역시 당사 기본정보를 기반으로 외부 정보와 결합하여 지역 경제, 추세분석, 전통시장 분석 등의 영역에 대하여 빅데이터 컨설팅을 제공하고 있다.

오프라인으로 제공되는 빅파이 센터는 통신데이터, 신용거래데이터, 경기도청 보유 공공 데이터 등을 제공하여 전문가의 협력 및 네트워킹을 통하여 아이디어 및 관련 기술 개발을 위한 오프라인 공간이다. 서울대학교 빅데이터 연구원은 빅데이터 관련 초학제적 연구 기관으로 플랫폼 기술, 예측과 시뮬레이션을 지원하는 복합적 분석기술, 다양한 응용 분야에 대한 사회문제를 해결하는 민간기관으로 잘 알려져 있다.

[표 2-7] 민간부문 빅데이터 분석 서비스 기관

기업	데이터	분석 내용
네이버 데이터랩	·데이터 융합 분석 : 개별 이용자가 가진 데이터를 검색 데이터에 융합해 분석하는 기능	·개인이나 사업자가 보유한 거래내역 등의 데이터를 네이버 검색에 데이터와 비교 가능, 새로운 사업 방향이나 아이디어를 모색하는 중소기업, 스타트업들에게 유용하도록 제공
	·지역 통계 : 네이버 지역 서비스에 등록된 업체와 해당 지역 내 모바일 검색 사용자 수를 비교해 업종별 분포도 제공	·공공기관이 제공하는 데이터도 네이버 지도에 융합되어 제공 ·국토교통부가 제공하는 실거래가 공개시스템과 연계하여, 읍/면/동 단위로 세분화된 아파트 실제 매매 가격과 전세, 월세 거래량 등을 네이버 지도상에서 제공
	·검색어로 알아보는 대한민국 - 2007년부터 현재까지 각 분야별, 이용자 라이프스타일에 따라 10	·관심있는 검색어에 대해 최소 3개월에서 최대 10년까지의 변화 추세 확인, 설정한 기간 내 해당 검색어의 최대 검색량을 100으로 산출한 상대적 수치가 제공

기업	데이터	분석 내용
	여 년간의 인기 검색어 데이터	·공연 분야의 인기 검색어에 오른 특정 뮤지컬이 최근 3개월 내에 얼마나 검색되었는지 그래프 형태로 제공하는 방식
SK텔레콤 지오비전 데이터	·위치정보(기지국 분석), 모바일 행태(네트워크 분석), 라이프스타일(고객 분석), 인구DB(주거/유통/상주), 공간DB(위치정보 연계), 상권DB, 매출분석(OK캐쉬백), 부동산정보(부동산114 시세분석)	·지오비전은 SK텔레콤의 데이터 비즈니스 플랫폼 사업으로 빅데이터를 기반으로 공공분야의 정책 수립 및 평가(관광/치안/교통/재해재난/인구조사 등)과 개인/민간 영역의 data 기반 경영(상권분석, gCRM, 타겟마케팅 등)을 지원 ·이와 연계하여, 데이터 시각화를 통한 기업의 빅데이터 활용과 인사이트 탐색을 지원하는 뉴스젤리와 함께 데이터 시각화 기술을 바탕으로 콘텐츠, 솔루션 개발
솔트룩스	·오투(O2) : 소셜 데이터 실시간 수집, 분석 플랫폼 ·디투(D2) : 빅데이터의 분석과 복잡한 패턴의 발견 실시간 빅데이터 분석 플랫폼	·빅데이터 수집과 변환, 검색, 분석, 시각화에 이르는 전 과정을 지원하는 'D2'를 활용, 고객의 소리(VOC)의 고객 상담과 질의응답(Q&A) 등록 내용, 이디스커버리(eDiscovery), 국방·안보 분야 등의 데이터를 실시간으로 색인하고 분석 ·블로그, 리뷰, 트위터, 뉴스 등 소셜 데이터와의 실시간 융합분석 실시간 인메모리 분석과 통계, 기계학습에 기반한 미래 예측까지 가능 ·뉴스, 블로그, 트위터와 같은 소셜 빅데이터의 실시간 수집, 검색과 고급 분석
TIBUZZ	·인스타그램, 페이스북(페이지), 트위터 등 SNS와 블로그, 커뮤니티사이트, 유튜브 등의 온라인 데이터를 분석	·브랜드 분석(신규캠페인 제안을 위한 브랜드 및 경쟁사 분석), 위기 관리(실시간 부정 이슈파악, 즉각적인 대응가이드 마련), 상품개발(업종 트렌드 및 소비자 관심사 분석) 등 각각의 비즈니스 목적에 따라 데이터 분석 제공 ·ex. 유해/스팸/광고성 데이터 필터링, 이벤트일 전후 데이터 비교 분석, 제귀 키워드 추가를 통한 데이터 정제 등
BC카드 스토어스토리	·점포이력 서비스 - 창업에 앞서 과거 동일건물에서 업종별/매출추이를 확인할 수 있는 서비스	·카드사용 내역 등을 확인하여 창업에 앞서 특정 건물의 업소 개폐업 이력과 추정 매출을 확인할 수 있는 점포이력 서비스를 제공, 또한 주소 입력과 업종 선택만으로 상권/입지특성, 추정매출, 임대시세 등을 확인 최적의 창업지역과 업종을 선택할 수 있는 점포평가 서비스 제공
신한카드 빅데이터 센터	·카드사 기본정보 : 지역경제소비 트렌드 분석, 내·외지인 소비성향 비교분석, 외국인 소비분석 등을 바탕으로 공공기관 및 지자체에 정책 지원	·지역경제(지역 소비현황 분석 및 내·외지인 분석을 통한 지역경제 현황분석 및 관광활성화 정책 지원), 축제분석(축제권역별 소비현황 분석 및 연계지역 분석을 통한 축제활성화 정책 및 연계관광정책 지원), 전통시장(전통시장의 소비규모 정량적 측정 및 재방문, 활성화 정책 지원), 장애인/대중교통 정책 개선 지원 등 다양한 분야의 분석 지원 서비스를 제공
빅파이 센터 (오프라인)	·통신데이터, 신용거래데이터, 경기도청 보유 공공 데이터 등을 제공하며, 이를 분석할 다양한 빅데이터 분석환경을 제공	·창업을 위한 빅데이터 분석이 필요한 창업 예정자, 빅데이터 관련 교육이나 세미나 공간 수요, 자신의 데이터와 다양한 데이터 분석을 통한 개인 사업을 위한 시민들에게 빅데이터 분석 인프라 제공, 사용자 유형별 컨설팅 제공
서울대학교 빅데이터 연구원	·빅데이터 관련 초학제적 연구기관으로 플랫폼 기술, 예측과 시뮬레이션을 지원하는 복합적 분석기술, 다양한 응용 분야에 대한 사회문제 해결	·빅데이터 인프라기술 분과, 법 정책 분과, 미래산업경제분과, 사회복지 분과, 방송 문화 스포츠 분과, 보건의료 분과 등 서울대 단과대학과 연계하여 연구진을 구성, 최근 발간된 분석 보고서에 따르면 정부 행정 통계를 이용한 자영업 창업 및 생존률 분석, 빅데이터를 이용한 보건의료복지 고위험, 고비용 집단의 선별과 예측 모형 개발 등 광범위한 분야에서 데이터 분석을 통한 연구 지원 사업 수행

공공기관은 다양한 기관에서 빅데이터와 관련된 센터를 운영 중이다. 정부통합전산센터 헤안은 각 부처 및 기관의 공공데이터를 수집·보유함으로써 언제나 필요한 데이터를 획득하고 분석할 수 있는 공통기반 인프라 서비스를 제공하고 있으며, K-ICT 빅데이터 센터는 창업과 사업화 촉진, 중소기업의 활용 확산 등 빅데이터를 기반으로 창조경제 지원 및 국민 행복 실현을 목표로 주요 사업을 추진 중에 있다.

또한 공공데이터 포털은 지방자치단체별 파일데이터, 오픈API 데이터를 제공함으로써 이를 교육, 국토관리, 산업고용, 사회복지 등에 활용하고 있으며, 한국교통연구원은 교통카드 데이터, 내비게이션 데이터, SK T-map 실시간 교통정보 데이터 등을 활용하여 교통 DB를 구축하고 국가 교통조사 및 수요 분석을 수행하여 미래 교통의 방향을 제시하고 있다.

한국도로공사의 경우, 교통량, 통행 속도, 통행시간 등 하루 2억 건의 교통 데이터를 생성, 보유하고 있으며, 이를 통해 신규 콘텐츠를 개발하고 미래 교통 발전을 위한 연구를 위해 데이터를 제공하기도 한다. 서울특별시 빅데이터 캠퍼스는 대기오염, 일자리 감소, 고령화, 저출산 등의 사회문제 등을 해결하기 위해 서울시와 시민, 시민사회가 함께 참여한 기관으로 사회 문제에 대한 솔루션 서비스를 제공하고 분석 결과를 공유함으로써 보다 나은 사회를 위한 다양한 서비스를 제공하고 있다.

[표 2-8] 공공부문 빅데이터 분석 서비스 기관

기업	데이터	분석 내용
정부통합 전산센터 해안	· 범정부 빅데이터를 수집·저장·융합하는 공통 기반 플랫폼 · 정부 업무데이터, 공공데이터 포털 · 트위터, 블로그, 커뮤니티 등 민간 정보	· 주요 사회이슈 분석: 실시간 웹소셜 데이터로부터 사회적이슈가 있는 키워드를 선정 및 분석하여 빅데이터 기반 공공행정 정책 수립 기반 제공 · 지자체 이슈 분석 지원: 사회적 파장이 예상되는 이슈 및 정책 등을 SNS, 언론 등 빅데이터 분석을 통해 사전에 인지·공유하여 지자체별 정책 조율 및 선제적 대응이 가능하도록 지원 · 전문분석 서비스: 공무원이나 기관이 직접 빅데이터를 수집·저장·분석·시각화 등을 할 수 있도록 빅데이터 분석 도구와 저장 공간 등 분석 환경 제공
K-ICT 빅데이터 센터	· 국토, 교통 IT서비스, 산업 경제, 의료 보건, 방송 통신 등 국내외 빅데이터 관련 사업, 사이트와의 협조 및 자체 수집을 통한 활용도 높은 데이터를 수집하여 공개	· 창업, 사업화 촉진 - 벤처, 스타트업의 빅데이터 솔루션 및 응용서비스 개발 등 창업 사업화 지원 · 중소기업 활용 확산 - 데이터 소재정보 확인, 사용가능 여부 확인, 저작권 범위 및 권리 관계 확인, 데이터 융합에 따른 권리 귀속, 저작권 부당 활용 및 손해배상 문제 등 빅데이터 기반 창조경제 지원 및 국민 행복 실현을 목적으로 주요사업 추진
공공데이터 포털	· 파일데이터(19,623) - 서울 교통 통계정보, 대전 서구 무더위 쉼터 등 · 오픈API(2,207) - 월별 집합건물 사업장 공장을 현황, 주택 시장 분기별 분석자료 등 · 표준데이터(34) - 전국지역 특화자료, 박물관미술관 정보 등	· 지방자치단체별 파일데이터, 오픈API 데이터를 제공함으로써 이를 교육, 국토관리, 산업고용, 사회복지 등 활용사례를 공유하고 각각의 공공데이터를 시각화하여 어플리케이션, 문서 등으로 제공 · ex. 고양시 음식물류 쓰레기 배출 현황, 대규모 환경기준 초과횟수, 서울시 CCTV 현황 등의 자료를 지도 상으로 제공
한국교통 연구원	· 교통카드 데이터, 내비게이션 데이터, SK T-map 실시간 교통정보 데이터 등을 활용	· 국가교통DB 구축 및 교통투자평가 연구, 국가교통조사·수요분석 연구, 국가교통 DB·통계분석 연구 등 미래 교통의 방향 제시 · 교통카드 및 내비게이션 데이터를 교통정책에 어떻게 활용할 수 있는지에 대해 분석
한국도로공사	· 교통량, 통행 속도, 통행시간 등 하루 2억 건의 교통 데이터 생성, 보유 - 신규 콘텐츠 개발, 미래 교통 발전을 위한 연구를 위해 데이터 개방	· 고속도로 교통 데이터를 제공함으로써 연후 또는 명절기간 통행 예상 시간 & 빠른길 안내 시스템 개발, 휴게소의 남녀 이용 비율 분석을 통한 남녀 화장실 비율 결정, SK텔레콤과 협업을 통해 SK T-Map 데이터를 제공하고 보다 정확한 예측 시스템 개발
서울특별시 빅 데이터 캠퍼스	· 서울시유동인구 정보, 택시 승하차 정보, 기상 정보 등의 빅데이터와 서울시의 공간 정보를 활용한 수치지도, 행정 데이터, 분석데이터 조회	· 대기오염, 일자리 감소, 고령화, 저출산 등의 사회문제 등을 해결하기 위해 서울시와 시민, 시민사회가 함께 참여, 업종별 소매업 매출액의 결정요인, 기후 변화 취약성 지도 작성, 미세먼지 및 대기오염에 따른 공원 이용 감소 패턴 등 분석 솔루션 서비스를 제공하고 분석결과를 공유함으로써 사회 문제 해결

3) 빅데이터 분석 방법

빅데이터 분석은 기존 통계적 분석에 데이터 마이닝, 텍스트 마이닝, 소셜네트워크 분석 등의 추가적 분석 방법론을 활용하여 분석한다.

[표 2-9] 빅데이터 분석 방법

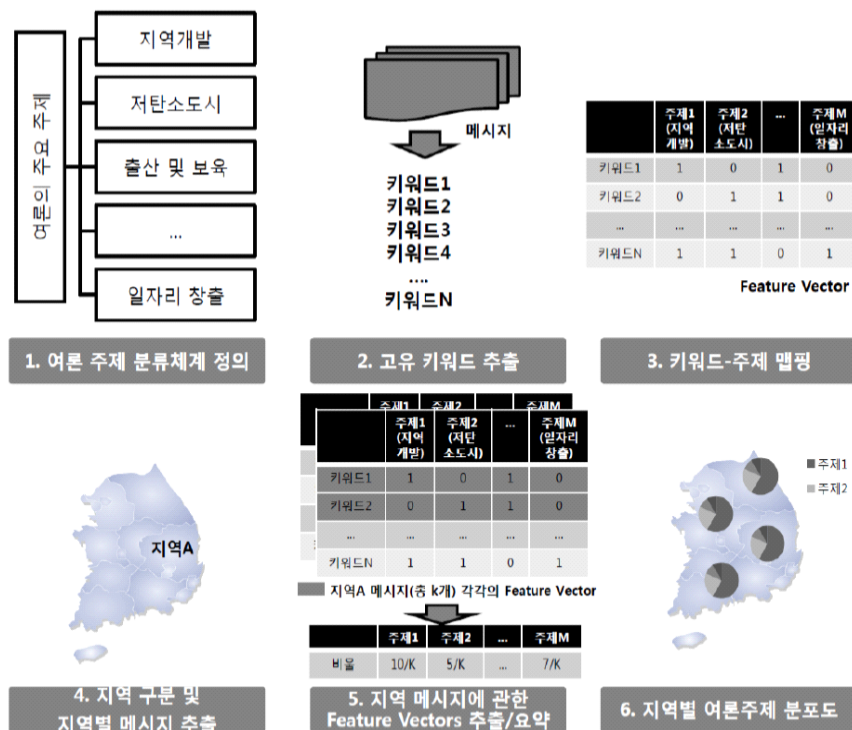
구분	내용
통계적 분석	·전통적인 분석 방법으로 주로 수치형 데이터에 대하여 확률을 기반으로 어떤 현상의 추정, 예측을 검증하는 기법
기술통계	·대표적으로 평균(산술평균, 중앙값, 최빈값), 분산, 표준편차 등
상관분석	·두 변수 간에 어떤 선형적 관계를 갖고 있는 지를 분석하는 방법
회귀분석	·독립변수와 종속변수 사이의 상관관계에 따른 선형적 관계식(수학적 모델)을 구하여 어떤 독립변수가 주어졌을 때 이에 따른 종속변수를 예측 ·또한 이 수학적 모델이 얼마나 잘 설명하고 있는지 판별하기 위한 적합도를 측정하는 분석 방법
분산분석	·두개 이상 다수의 집단을 비교하고자 할 때 집단 내의 분산, 총 평균과 각 집단의 평균의 차이에 의해 생긴 집단 간 분산의 비교를 통해 만들어진 F분포를 이용하여 가설검정을 하는 방법
주성분분석	·다양한 변수들에 대해 분석하는 다변량 분석으로 많은 변수들로부터 몇 개의 주성분들을 추출하는 방법
데이터 마이닝	·대용량의 데이터로부터 패턴인식, 인공지능 기법 등을 이용하여 숨겨져 있는 데이터간의 상호 관련성 및 유용한 정보를 추출하는 기술
예측	·대용량 데이터 집합 내의 패턴을 기반으로 미래를 예측(예: 수요 예측)
분류	·일정한 집단에 대한 특정 정의를 통해 분류 및 구분을 추론(예: 이탈한 고객)
군집화	·구체적인 특성을 공유하는 자료들을 분류 ·미리 정의된 특성에 대한 정보를 가지지 않는다는 점에서 분류와 다름(예: 유사 행동 집단의 구분)
패턴분석	·동시에 발생한 사건간의 상호연관성을 탐색(예: 장바구니의 상품들의 관계)
순차패턴분석	·연관 규칙에 시간의 개념을 반영하여 시계열에 따른 패턴들의 상호연관성을 탐색 (예: 금융상품 사용에 대한 반복 방문)
텍스트 마이닝	·텍스트 마이닝은 텍스트 기반의 데이터로부터 새로운 정보를 발견할 수 있도록 정보 검색, 추출, 체계화, 분석을 모두 포함하는 Text-processing기술 및 처리 과정
소셜네트워크 분석	·대용량 소셜 미디어를 언어분석 기반 정보추출을 통해 이슈를 탐지하고, 시간의 경과에 따라 유통되는 이슈의 전체과정을 모니터링 하고 향후 추이를 분석하는 기술
공간분석	·특정 데이터를 좌표화하여 지도위에 시각화하는 것을 바탕으로 공간 데이터와 속성 데이터를 이용하여 새로운 가치를 발견하는 분석 방법

출처 : 양현철 외(2014), 빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼(Version 1.0), 미래창조과학부, 한국정보화진흥원, 빅데이터전략센터

□ 군집 분석(클러스터 분석)¹⁷⁾

클러스터 분석(군집분석)은 비슷한 특성을 가지는 개체들의 집단을 발굴하는데 사용하는 기법이며, 분류체계 기반 여론 트렌드 분석은 여론 주제들의 분류체계를 먼저 정의하고, 수집한 메시지의 고유 키워드들과 여론 주제들을 매칭하여 분석하는 기법이다.

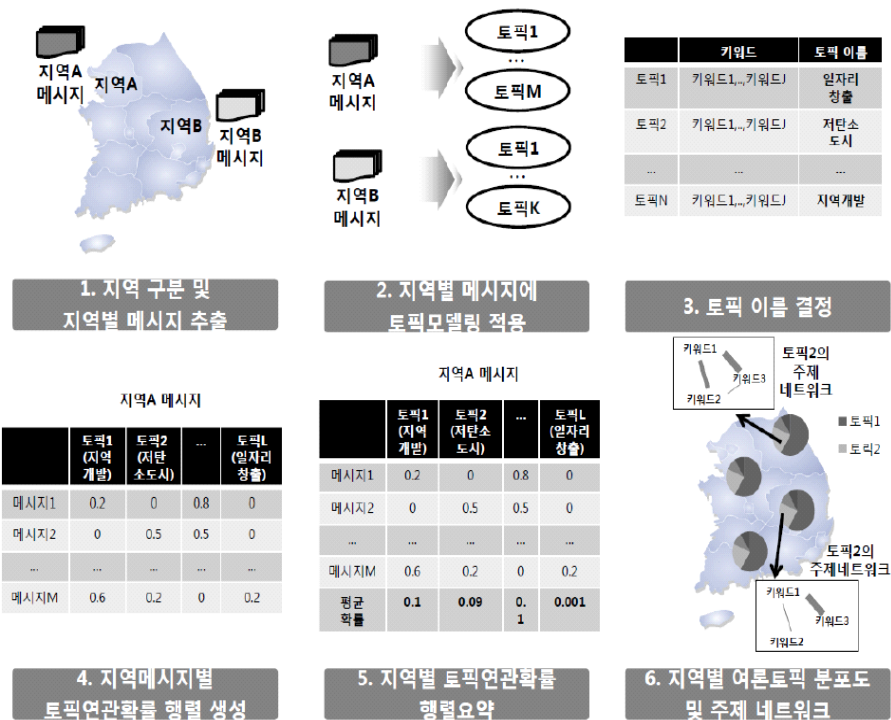
토픽모델링 기반 여론 트렌드 분석은 문서 내 단어들의 출현 빈도를 통계적으로 분석하여 전체 입력 문서를 관통하는 잠재적인 주제들을 추출하고, 추출된 주제들의 상관관계나 시간적 변화를 파악하는 텍스트 마이닝 기법이다. 분류체계 기반은 기존의 혹은 연구자에 의한 주제 분류체계를 활용하는 반면, 토픽모델링 기반은 전체 키워드간의 통계적 분석을 통해 단어 집합들을 찾아내 주제를 선정하는 특징이 있다.



[그림 2-10] 분류체계 기반 여론 토픽 분석의 원리

출처 : 황명화(2014), 「여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구」, 국토연구원. p.55

17) 황명화(2014), 「여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구」, 국토연구원. pp.55-59

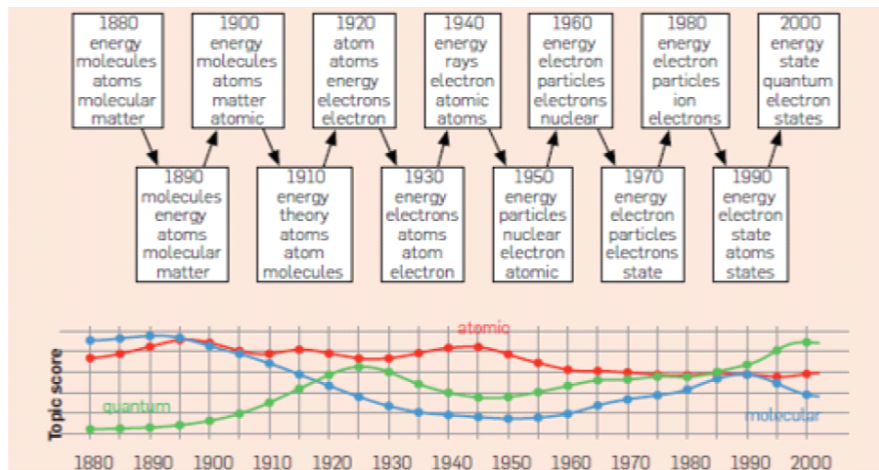


[그림 2-11] 토픽모델링 기반 여론 내용 분석의 원리

출처 : 황명화(2014), 「여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구」, 국토연구원. p.56

이처럼, 군집 분석(클러스터분석)에는 다양한 분석기법이 있으나, Latent Dirichlet Allocation(LDA) 방법이 주로 통용되고 있다. LDA기법은 텍스트 자료들은 알려지지 않은 여러 주제로 구성되어 있고, 이 주제들은 다양한 단어들로 표현되며, 구성 비율이 다르고, 이 비율을 결정하는 확률 분포가 존재한다는 것을 전제로 한다. 이 전제를 바탕으로 문서 내 주제 구성 비율을 결정짓는 확률 분포를 추정하고, 그 과정에서 전체 텍스트 자료를 주도하는 주제들과 그 주제들을 구성하는 주요 단어 집합들을 찾아내는 분석 방법이다.¹⁸⁾

18) 황명화(2014), 「여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구」, 국토연구원

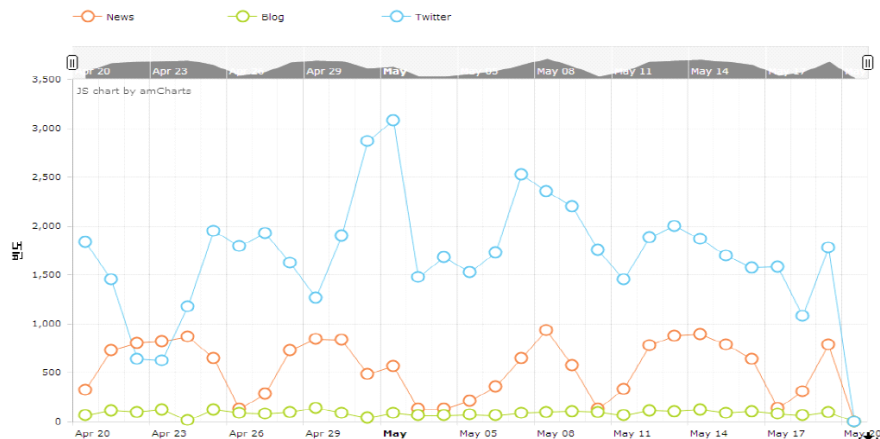


[그림 2-12] 토픽모델링을 활용한 Science 논문의 내용의 경향 분석

출처 : Blei, David M. (2012), "Probabilistic Topic Models", COMMUNICATIONS OF THE ACM, v.55(4), ACM. pp.77-84

□ 텍스트 마이닝¹⁹⁾

텍스트 마이닝 기법은 비/반정형의 텍스트 데이터에서 자연어 처리로 형태소 분석을 통해 의미있는 정보를 추출하고 텍스트 간의 연계성 파악 등을 분석하는 방법이다. 빈도 분석은 키워드 발견 빈도를 통해 주제에 대한 관심도를 분석하는 방법으로 키워드 분석으로 지칭하기도 하며, 특징기 간동안 발견빈도 변화를 통해 해당 주제에 대한 관심도 변화를 살펴보는 추이분석에 대한 수행이 가능하다. 또한 서로 다른 키워드에 대한 빈도를 비교하여 관심도에 대한 차이도 확인 가능하다.

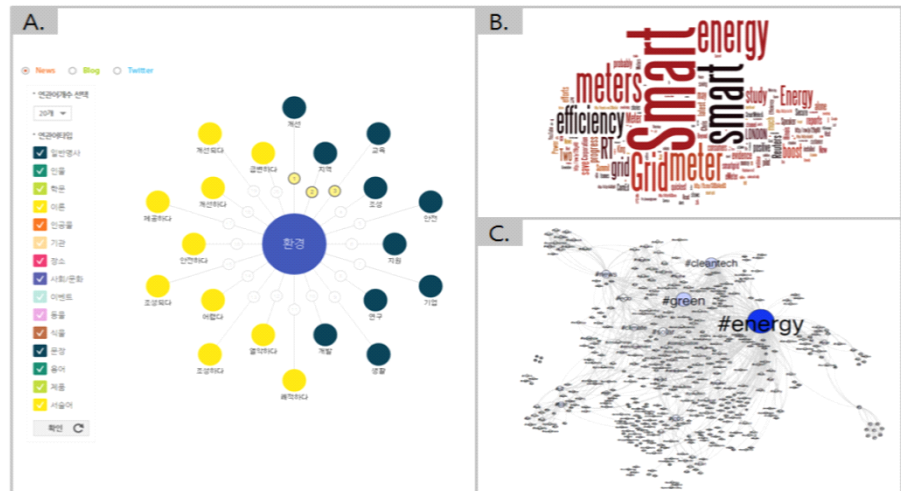


[그림 2-13] 환경 키워드에 대한 빈도 분석 사례

출처 : 이미숙·이창훈·김지연(2014), 「빅데이터를 활용한 환경분야 정책수요 분석」, 한국환경정책·평가연구원

19) 이미숙·이창훈·김지연(2014), 「빅데이터를 활용한 환경분야 정책수요 분석」, 한국환경정책·평가연구원 pp.38-42

연관키워드 분석은 데이터 내에서 관심 키워드를 포함하여 함께 나타나는 키워드를 추출해 해당 키워드와 어떤 주제를 연결시켜 생각하는지 분석하는 방법이다. 분석 목적과 활용 엔진에 따라 연관키워드를 일반명사, 인물, 장소, 서술어 등으로 그룹화할 수 있으며, 연관 분석은 연관 키워드에 대한 빈도로 나타내거나, 네트워크 분석을 통해 빅데이터 분석 자료를 제공하기도 한다.

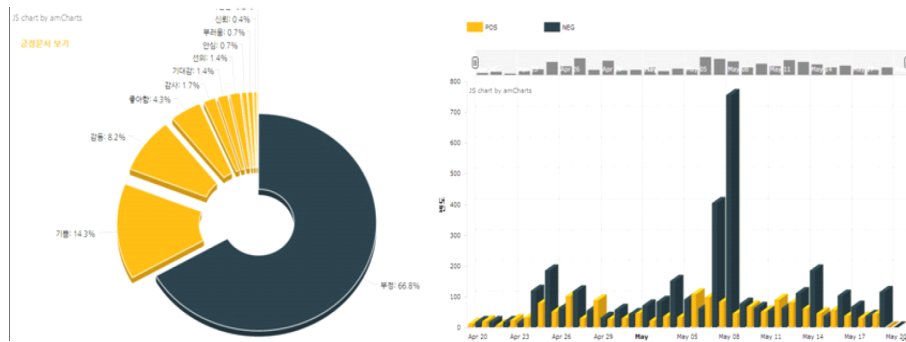


주: (A) '환경' 키워드에 대한 연관어 분석. 빈도순 넘버링 방식.
 (B) 'Smart Grid' 키워드에 대한 연관어 분석. Word Cloud 이미지 방식.
 (C) 'Energy' 키워드에 대한 해시태그 연관어 분석. 연관어 네트워크 방식.

[그림 2-14] 다양한 연관어 분석 및 시각화 사례

출처 : 이미숙·이창훈·김지연(2014), 「빅데이터를 활용한 환경분야 정책수요 분석」, 한국환경정책·평가연구원

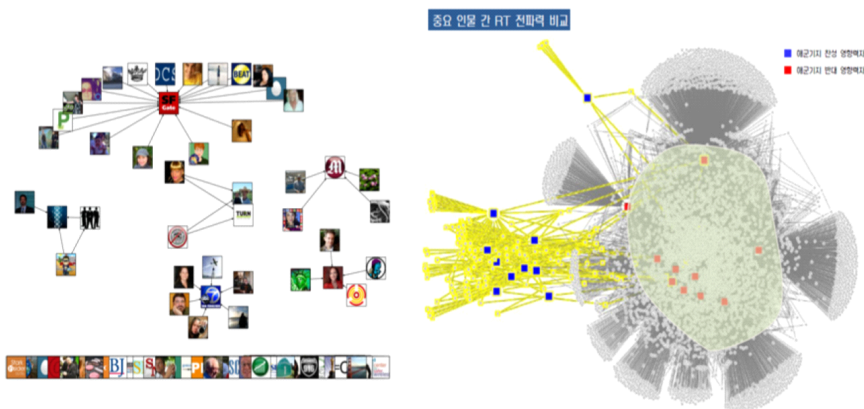
감성분석은 좀 더 솔직한 의견을 자유롭게 개진하는 트위터, 페이스북, 블로그 등의 정보 특성을 이용해 긍정 또는 부정 감성이 발현되는 빈도로 여론 파악(오피니언 마이닝, 평판분석이라는 용어로도 사용)하는 분석 방법이다. 학습단계를 통해 구축된 감성 사전을 바탕으로 문장 안에 있는 형태소별 감성을 추출하여 해당 문장의 감성 정도를 산정한다. 분석 시스템에 따라 긍정 또는 부정으로 나누어지기도 하고 여기에 중립 감성이 추가되는 경우, 긍정 또는 부정을 기쁨, 감동, 신뢰, 감사, 슬픔, 두려움, 싫어함, 실망 등 감정을 나타내는 의미로 세분화하는 경우도 존재한다.



[그림 2-15] 감성 분석 예시

출처 : 이미숙·이창훈·김지연(2014), 「빅데이터를 활용한 환경분야 정책수요 분석」, 한국환경정책·평가연구원. p.40

□ 소셜네트워크 분석



[그림 2-16] 네트워크 분석 사례(좌: 트위터상에서 ‘스마트그리드’를 언급한 주요 영향력자 간의 관계 / 우: 국내 트위터 사용자들 주요 인물의 리트윗 전파력)

출처 : (좌) Russell et al.(2011), 9p / (우)사이람(2014), 13p 이미숙·이창훈·김지연(2014)에서 재인용

소셜네트워크 분석은 소셜미디어 사용자 간의 네트워크, 또는 문서, 키워드 간의 상관관계를 분석해 특정 이슈가 어떻게 발현되어 전파 되는지를 보여주는 기법²⁰⁾으로, 트위터의 경우 리트윗 관계, 키워드의 동 출현 문서 간의 네트워크 분석을 통하여 인물, 키워드 간의 영향력과 상관관계를 파악하는 기법이다.

20) 상동, p.42

제3장 빅데이터 기반 정책개발 사례

- 1. 정책의 개념 및 과정
 - 2. 국내 사례
 - 3. 국외 사례
-

1. 정책의 개념 및 과정

1) 정책의 개념

정책이란 학자마다 여러 가지로 정의하고 있으나 이를 정부가 수립하는 정책이라는 측면으로 본다면 “바람직한 사회 상태를 이룩하려는 정책 목표와 이를 달성하기 위해 필요한 정책수단에 대하여 권위 있는 정부기관이 공식적으로 결정할 기본 방침¹⁾”으로 정의할 수 있다. 건축·도시 정책 역시 이러한 일반적인 개념 하에 분야별 정책으로 이해할 수 있다.

정책이 가지는 성격은 총 4가지로 정리할 수 있으며, 당위성, 정치성·권력성, 미래지향성, 영향의 양면성이다. 첫째 당위성은 ‘마땅히 있어야 할 것’, ‘당연히 바람직한 것’을 찾아서 구현시키려는 의도이며, 둘째 정치성·권력성은 정책의 행동의 주체가 정부 또는 공공기관으로 정치성과 권력성을 내포하고 있기 때문이다. 셋째 정책의 미래지향성은 당면한 현재의 문제뿐만 아니라, 미래의 바람직한 사회를 목표로 하고 있기 때문이며, 마지막으로 영향의 양면성은 장래의 바람직한 상태를 이루기 위한 의도적인 행동으로 국민들에게 상반되는 영향을 끼치는 경우 발생할 수 있기 때문이다.²⁾

로위(Lowi)는 정책을 배분 정책(Distributive Policy), 규제 정책(Regulatory Policy), 재분배 정책(Redistributive Policy), 구성 정책(Constitutional Policy)으로 분류하였다.

1) 정정길·최종원·이시원·정준금·정광호(2014), 「정책학원론」, 서울: 대명출판사. p.35

2) 정상현(2001), 「정책학의 이해」, 신아출판사. pp.13-33

배분 정책은 국민들에게 권리나 이익, 또는 재화나 서비스를 배분하는 내용을 지닌 정책으로 19세기 미국의 국유지 불하정책, 하천·항만사업, 연구개발사업, 군수품 구매 등의 내용을 지닌 정책이다. 규제 정책은 개인이나 일부집단에 대해 재산권 행사나 행동의 자유를 구속·억제하여 반사적으로 다른 사람들을 보호하려는 정책으로서 기업 간의 불공정 경쟁 및 과대광고의 규제 등이 해당한다. 재분배 정책은 고소득층으로부터 저소득층으로의 소득이전을 목적으로 하는 정책으로, 누진세에 의하여 고소득층으로부터 보다 많은 조세를 징수하고 저소득층에게 사회보장지출을 하여 소득의 재분배를 도모하는 정책을 말한다. 마지막으로 구성 정책은 헌정수행에 관한 운영규칙에 관련된 정책을 지칭한다.³⁾

알몬드(G.A. Almond)와 파웰(B.G. Powell)은 정책을 추출 정책과 상징 정책 두가지로 분류하였다. 추출 정책은 정부가 국내의 환경으로부터 병역·조세 등과 같은 인적·물적 자원의 일부를 뽑아내는 정책을 말하며, 상징 정책은 국민의 국가정체성이나 국가의 국제적 위상을 높이는 것을 내용으로 하는 정책을 지칭한다.⁴⁾

2) 정책의 과정

일반적으로 정책과정이란 정책을 개발하고, 개발된 정책을 집행하고, 집행된 정책을 평가하는 일련의 단계를 지칭한다. 정책과정론(policy process theory)은 정책문제의 인지로부터 목표설정·대안의 분석·결정·합법화·집행·평가의 과정을 거쳐 정책이 종결되기까지의 과정을 다룬 이론이라고 할 수 있다. 이러한 정책과정은 정치적 성격을 띠고 정치적 역동성을 지닌다. 하지만 현실의 정책과정은 이러한 합리적 순서에 의해 이루어진다고 보는 견해는 미미하다고 할 수 있다. 현실에서는 이전의 결정패턴에 수정을 가하면서 진행하거나, 또는 어떤 정책을 세우는 여러 조건의 조합이 중요하다는 등의 견해가 있다.⁵⁾

건축·도시 분야의 정책 역시 정책과정에 따라 순서대로 이루어지는 경우는 많지 않다. 정책개발 단계에서 결정된 정책이라도 정책집행 단계에서 문제점이 발견되어 정책개발 단계의 정책대안을 변경하거나 수정하는 경우도 빈번하다. 그러나 정책과정은 정책단계에 따라 필요한 행위를 구분하고 해당 행위에 필요한 자원을 구분할 수 있어 대단히 유용하다.

이러한 정책과정을 단계별로 구분하는 모델은 학자별로 다양하다. 정책학의 아버지라 불리는 Lasswell(1956)은 정책과정을 7단계로 구분하였다. 7단계는 각각 정보를 수집·분석하는

3) 남기범(2006), 「현대 정책학 강의」, 한국학술정보, pp.40-43

4) 정정길·최종원·이시원·정준금·정광호(2014), 「정책학원론」, 서울: 대명출판사, pp.60-64

5) 행정학용어 표준화연구회(2010), 「이해하기 쉽게 쓴 행정학용어사전」, 새정보미디어

정보과정, 정책대안을 작성하는 건의과정, 대안 중 최종안을 선정하는 처방과정, 최종안을 시범 시행하는 발동과정, 전면 시행하는 적용과정, 성공여부를 판정하는 평가과정, 정책의 수정과 폐기 등을 결정하는 종결과정이다. 이러한 Lasswell의 정책과정은 현재의 여러 정책 과정모델의 근거로 활용되고 있다.⁶⁾

현재 정책학에서 가장 대표적으로 활용하고 있는 정책과정 모형은 Anderson(1975)의 모형으로 총 5단계로 구성되어 있으며, 문제형성, 정책형성, 대안채택, 정책집행, 정책평가 단계이다. 이를 기초로 하여 현행 건축·도시 분야의 특징을 반영하면 다음과 같이 정리할 수 있다.

① 문제형성(problem identification and agenda formation)

정책의 문제를 인식하는 단계로서 정책당국에서 해결해야하는 문제를 선정하는 단계이다. 건축·도시 분야의 정책에 대한 문제는 국민의 실생활과 직간접적으로 많은 관련이 있어 정책을 결정하는 정책결정자들 역시 관심이 크다. 그러나 건축·도시 분야에 제기되는 문제는 수없이 많아 이를 모두 정책의제로 정할 수는 없다. 이에 건축·도시 분야 정책당국에서 보다 문제의 심각함을 인지하고 해결해야하는 문제를 선정하는 단계이다.

정책의제 선정을 위해서는 건축·도시 분야에 정책에 대한 국민들의 인식·의견에 대한 여론모니터링과 국민들의 현재 추진 중인 관련 정책에 대한 불만 및 개선사항에 대한 민원조사 등의 의견수렴이 필요하다. 또한 인허가 현황, 방재 현황 등의 건축·도시 분야의 각종 지표에 대한 현황 분석과 정책방향, 정책공약, 국내외 환경 등에 관한 조사도 필요하다. 이러한 여러 조사와 분석을 통해서 정책이슈와 수요를 도출해서 정책의제를 선정하는 것이 건축·도시 분야 정책의 문제형성 단계이다.

② 정책형성(formulation)

정책형성 단계는 선정된 정책의제를 해결할 수 있는 실현가능한 대안을 만드는 단계이다. 건축·도시 분야의 정책 대안은 법률의 제·개정, 기준(가이드라인)·표준 등 제도마련, 재정지원, 시범사업 시행, 정책 홍보 등 정책의제에 따라 다양하다. 건축·도시 분야의 정책형성 단계는 이러한 다양한 정책 대안을 개발하는 단계이다.

건축·도시는 하나의 문제를 해결하는 다양한 정책대안이 있을 수 있어 이 단계에서는 선정된 문제를 해결할 수 있는 모든 형태의 대안을 도출하는 것이 중요하다. 이를 위해서는 문제 해결을 위한 다양한 정보의 수집이 필요하다. 또한 수집된 정보와 데이터에 기

6) 한국행정학회(2013), “행정학전자사전: 정책과정론”, 한국행정학회,
http://www.kapa21.or.kr/data/kapa_dictionary_view.php?num=957.(검색일: 2017.09.01)

초하여 정책대안을 구성하는 것이 중요하다. 과거에는 정책대안을 전문가의 의견을 토대로 만드는 경우가 많았다. 그러나 건축·도시 분야가 의견기반 정책에서 증거기반 정책으로 진화하기 위해서는 정책형성 단계에서 정보와 데이터를 근거로 정책대안을 마련하는 것이 중요하다.

③ 대안채택(adoption)

대안채택 단계는 정책형성 단계에서 문제를 해결하기 위하여 개발한 여러 정책대안 중 최종안을 채택하는 단계이다. 또한 이 단계에서 최종안이 집행될 수 있도록 관계 기관에서 의결하는 등의 합법성을 부여한다.

다 수의 대안 중에서 어떠한 대안을 최종안으로 정하기 위해서는 여러 대안의 평가과정을 거쳐야 한다. 전문가집단의 자문, 이해관계자를 위한 공청회, 대안에 대한 여론, 정부의 정책방향 등을 통하여 평가를 거치며, 특히 건축·도시 분야는 비용효과분석 등 다양한 영향평가 방법론이 개발되어 있어 이를 통하여 타당성을 검토할 수 있다. 결정된 최종안에 대해서 최종안에 따라 예산을 편성하는 과정이 필요하다.

④ 정책집행(implementation)

채택된 최종안을 정부에서 실행에 옮기는 것이 정책집행 단계이다. 특히 건축·도시 분야는 중앙부처에서 직접 사업을 시행하는 정책이 다 수 있어 대안채택 단계에서 결정된 예산에 따라 인적·물적 자원을 배분하는 것이 중요하다. 또한 효과적인 정책집행을 위해서는 문제형성 단계에서 발굴된 문제를 해결하기 위해서 개발되고 최종 채택된 대안이 차질 없이 진행될 수 있도록 모니터링 등을 통해서 일정과 내용을 관리하는 것이 필요하다.

⑤ 정책평가(evaluation)

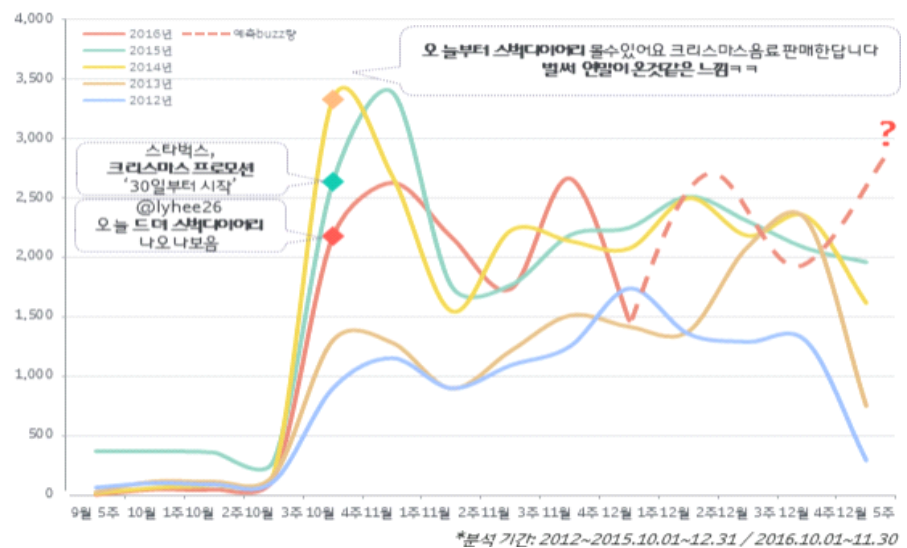
이는 정책과정 중 문제형성 단계에서 선정된 정책의제가 정책집행단계를 거쳐 해당 정책이 효과적이었는지를 판단하고 성공과 실패의 원인을 찾는 단계이다. 이 단계에서는 해당 정책집행으로 문제가 해결되었는지 아니면 다른 원인이 작동한 것은 아닌지 등 정확한 인과관계에 대한 평가가 중요하다.

건축·도시 분야의 정책은 단기간에 정책집행의 효과가 도출되는 것도 있지만, 대다수의 정책 효과는 시간이 소요되는 것이 많다. 따라서 정책평가를 기간에 따라 여러 번 수행할 필요성이 있으며, 특히 대상지 기반의 사업의 경우 장기적인 모니터링을 필요로 한다. 이러한 정책평가 단계에서 다시 문제가 도출되어 문제형성 단계로 정책과정이 반복되는 평가 후 환류 과정을 거치는 경우도 빈번하다.

2. 국내 사례

1) SK텔레콤, 인터넷·SNS 여론분석 시스템 ‘스마트 인사이트’

SK텔레콤의 ‘스마트 인사이트’는 인터넷 및 SNS에 올라오는 온라인 여론을 실시간으로 모니터링하여 분석하는 시스템으로 지난 2011년에 개발 되었다. 주로 기업의 마케팅을 지원하는 실시간 빅데이터 분석 플랫폼으로 사용자가 원하는 키워드를 시스템에 입력 하면 사전에 수집된 SNS, 블로그, 카페, 게시판, 뉴스 등의 각 종 온라인 채널의 데이터 와 검색 시점의 추가 수집 데이터를 종합 분석하여, 해당 키워드의 빈도, 연관어, 감성 분석 등을 수행하여 결과를 사용자에게 제공하는 유료시스템이다.



[그림 3-1] 스마트 인사이트 분석 예시

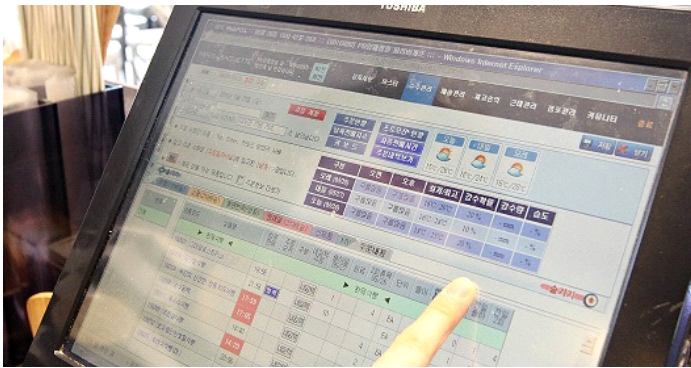
출처 : <http://www.smartinsight.co.kr/pr/industryReport/industryReport.do>

‘스마트 인사이트’는 온라인 여론을 분석하여 기업, 브랜드, 상품 등에 대한 대중의 관심도를 분석하여 제공하며, 사용자는 기존 설문조사나 인터뷰조사를 통해서 알 수 있었던 새로운 상품 및 시장에 대한 조사 결과를 온라인을 통하여 실시간으로 구득할 수 있다. 또한 기업에 대한 평판을 실시간으로 모니터링하여 기업의 위기관리를 보다 빠르게 효율적으로 수행할 수 있다. ‘스마트 인사이트’는 마케팅 분야의 빅데이터 전문가가 아니더라도 빅데이터를 활용할 수 있도록 지원하는 플랫폼이라는 데에 의의가 있다.

건축도시 분야 정책 활용의 관점에서 살펴보면 정책과정 중 문제형성 단계에서 SNS, 블로그, 카페, 게시판, 뉴스 등의 빅데이터 가능성을 확인할 수 있다. 특히 SNS 빅데이터를 활용한 여론모니터링은 기존의 설문조사 등을 보완 또는 대체할 수 있을 것으로 기대되며, 또한 정책평가 단계에서 해당 정책에 대한 여론의 평가에도 활용 가능할 것으로 판단된다.

2) 파리바게뜨, 기상과 제품별 판매량을 분석한 날씨 판매지수⁷⁾

최근 국내에서는 케이웨더, 웨더뉴스 등 기상 정보 제공업체에서 고정적으로 날씨 정보를 대중에게 제공하고 있다. 또한 기업에서는 기상 정보를 분석하여 이를 적극적으로 마케팅에 활용하는 사례가 증가하고 있다. 날씨가 서로 다른 지역은 지역의 기후에 적합한 제품을 군을 선별하여 전략 상품으로 활용하는 등 기상 데이터 분석을 통한 생산관리를 체계화하고 효율성 제고를 위하여 빅데이터를 활용한다.



[그림 3-2] 파리바게뜨 계산대(POS)에 제공되는 날씨지수
출처 : <http://www.bizwatch.co.kr/pages/view.php?uid=23379>

이러한 기상 데이터를 활용한 대표적인 사례인 파리바게뜨는 최근 5년간 전국 169개 지점의 일별 매출과 기상자료를 통계적 기법으로 지수화한 날씨판매 지수를 개발하였다. 날씨 판매 지수는 날씨에 따른 판매율을 나타내는 지수로 가맹점에서는 이를 바탕으로 판매량을 예측한다. 이는 주문량을 조절할 수 있는 찬스로스(판매할 제품이 없어 발생하는 손실)을 방지하고 재고부담을 줄일 수 있는 장점이 있다. 기상관측 자료와 10억 건 이상의 점포별 상품 판매 데이터를 실시간으로 분석하여, 전국 3,100여 개 파리바게뜨 점포 단말기에 제공된다. 날씨를 예측하고, 날씨별 판매가 높았던 빵을 파악하여 점포별

7) 윤미영(2013), 「더 나은 미래를 위한 데이터 분석」, 한국정보화진흥원

주문량을 조절하도록 하였다. 또한 파리바게트 매장의 단말기 화면에 ‘일별 날씨 판매지수 최대 변동’이라는 항목을 표시하였다. 예를 들어, 생크림 케이크 항목 옆에 ‘토요일 50.15%, 일요일 27.15%’ 같은 숫자가 표기된다. 이는 매장 주변 날씨 예보, 요일 등의 요소를 종합했을 때 최근 2주 평균보다 그만큼 매출이 늘어날 전망이다라는 뜻으로, 날씨 지수를 도입한 지 한 달 만에 조리빵 매출이 30% 증가했다.

해당 사례에서 활용한 데이터는 기상 데이터와 기업의 판매 및 재고 데이터이며, 이를 융합하여 새로운 대안을 만들었다는 점을 주목할 수 있다. 비록 민간기업의 활용사례이지만 빅데이터를 활용하여 다수의 정책대안을 도출하는 정책형성 단계의 빅데이터 활용 사례로 볼 수 있다.

3) 서울시 심야버스 노선수립 지원시스템⁸⁾

한정된 재원과 토지로 대도시의 교통 혼잡은 점차 심각해지고 있다. 시민들은 보다 안전하고 편리하게 대중교통을 이용하길 원하며, 날로 발전하는 첨단 기술과의 접목을 통해 교통시스템의 개선이 필요한 상황이다. 이에 서울시에서는 시민들의 이용 패턴을 분석, 이용자의 안전과 편의를 도모하고 운영 및 이용 효율을 극대화하기 위한 목적으로 서울시 심야버스 노선 수립 지원시스템을 구축하였다.

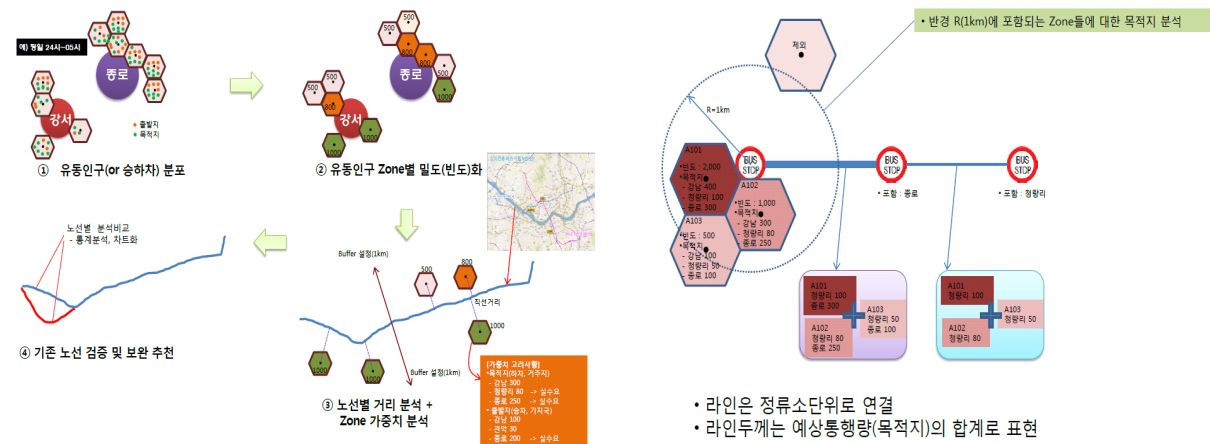
공공기관인 서울시가 보유한 심야택시 승하차 데이터 5백만 건과 민간회사인 KT의 통화량 데이터 30억 건을 융합, 택시 승하차 위치를 통해 심야 대중교통 수요를 추정하고, 청구지 주소 및 심야 시간대 휴대전화 위치 정보를 활용하여 심야 유동인구를 추정하였다. 서울 전역을 1km 반경의 1250개 셀 단위로 나눠서 유동인구·교통수요량을 시각화하고, 기존 버스노선과 시간·요일별 유동인구·교통수요 패턴 분석으로 노선부근 유동인구 가중치를 계산하여 최적의 노선과 배차간격을 도출하였다.

이처럼 서울시 버스정책과가 수립한 노선 안을 빅데이터 분석 결과를 반영하여 수정하였고, 기존의 노선 안에서 남부터미널, 건대 입구 등 유동 인구가 많은 지역을 거쳐 가도록 노선을 변경하였다. 이러한 분석을 통하여 수립된 서울시 심야버스 지원시스템은 심야시간 택시 이용 대신 심야버스의 이용으로 1인당 약 7,050원짜의 택시비 절감, 1년으로 환산하면 162억 원의 절약 효과가 나타났으며, 연간 24억 원의 유류비 절감과 연간 498톤의 이산화탄소 배출이 억제되는 효과가 나타났다. 행정 데이터만 활용하던 과정 정책에서 민·관이 구축한 빅데이터를 융합하여 정책을 수립했다는 데 의미가 크며, 향

8) 한중수(2013), “서울시, 빅데이터로 심야버스 노선 구축”, 뉴스1. <http://news1.kr/articles/1219217>. (검색일: 2017.06.01)

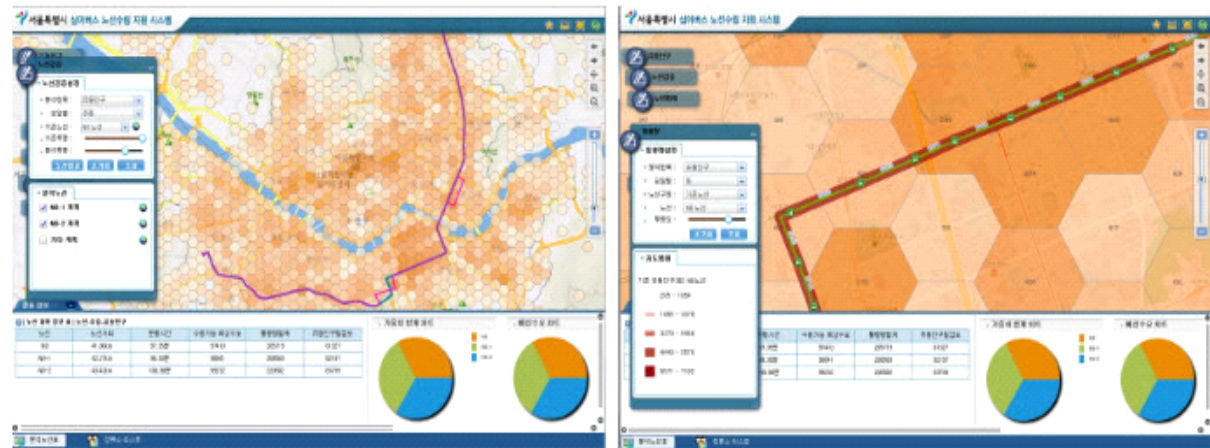
후 시정 전반에 걸쳐 빅데이터를 다양하게 접목시킬 수 있는 가능성을 제공하였다는데 큰 의의가 있다.

서울시 심야버스 노선수립사례는 공공정책 분야의 대표적인 빅데이터 활용사례로 알려져 있으며, 비정형 로그데이터인 휴대전화 데이터와 택시승하차 데이터를 결합하여 유동인구를 추정하고 이를 기반으로 정책을 수립하였다. 정책과정으로 보면 정책형성과 대안채택 단계에 빅데이터를 활용한 사례이다.



[그림 3-3] 서울시 심야버스 노선 검증 및 통행량 산출 알고리즘

출처 : <https://zeronova.kr/2013/08/07/seoul-bus-route-optimization/>



[그림 3-4] 빅데이터 기반 심야버스 노선수립 지원시스템 인터페이스

출처 : <http://www.esri.com/support/professional-service/case-studies/>

4) 경기도 소셜 미디어 분석을 통한 관광 트렌드 파악⁹⁾

2014년 경기도는 내국인이 가장 많이 찾는 국내 여행지 1위를 차지했지만, 여행 만족도, 재방문 의향 및 추천 의사가 전국 평균보다 낮게 나타났다. 따라서 경기도에서는 경기도 관광의 질적 개선을 위하여 소셜 미디어에서 거론되는 관광 관련 빅데이터 분석을 통하여 여론을 파악하고 경기도 관광 개선 정책 및 새로운 사업 마련에 활용될 수 있는 기초 데이터를 구축하고자 하였다. 또한 분석 결과를 반영한 경기도의 관광 발전 방안을 제시 하며, 관광 경쟁력 강화를 위해 대중의 관심사와 수요에 대한 이해를 기반으로 경기도 관광 트렌드를 분석하고자 하였다.

데이터의 수집은 트위터에서 1년간 작성된 약 1천만 건 가량의 관광 관련 트윗을 수집하여 분석하고, 홍보 및 스팸 트윗을 최대한 제거하여 일반 이용자들의 관심을 파악하였다. 이를 관심지수와 공감지수 측정을 통하여 키워드를 분석하고, 경기도 세부 지역별, 이슈별 분석을 수행하여 관광 활동에 대한 홍보 강화의 필요성을 도출하였다.

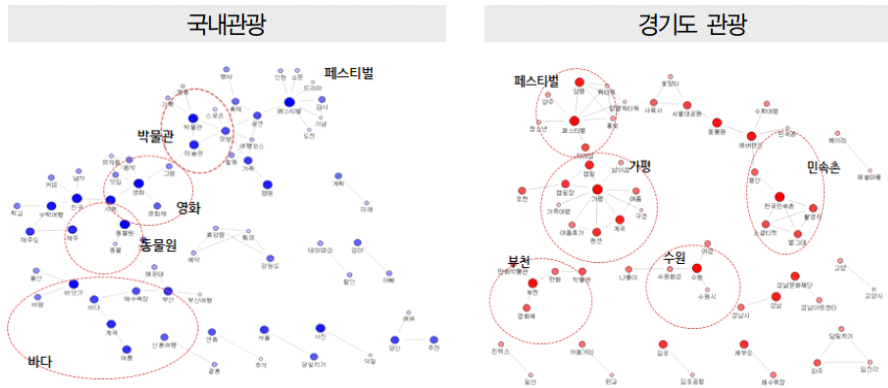


[그림 3-5] 경기도 소셜 미디어 분석 과정

출처 : 이수진·전유나(2015), 「소셜미디어 분석을 통한 경기도 관광 트렌드 연구」, 경기연구원

이러한 분석을 통하여 경기도에 대한 전반적인 관광 동향 뿐만 아니라 관광 정책, 자원 등 세부 이슈에 대한 관심을 파악하였고, 관광과 관련된 트윗에 나타난 연관 키워드를 관계망으로 시각화하였다. 이를 바탕으로 경기도 관광에 대한 SNS 마케팅의 활성화와 페스티벌 및 캠프 관련 자원의 질적 개선 및 홍보 강화, '모험' 트렌드를 반영한 야외 체험 관광 콘텐츠의 개발, 예술문화자원 및 상업시설 강화 등의 개선점을 도출하였다.

9) 이수진·전유나(2015), 「소셜미디어 분석을 통한 경기도 관광 트렌드 연구」, 경기연구원, pp.1-154



[그림 3-6] 국내외 관광 연관 키워드 관계도
출처 : 이수진·전유나(2015), 「소셜미디어 분석을 통한 경기도 관광 트렌드 연구」, 경기연구원

본 사례는 지자체의 대표적인 빅데이터인 소셜 데이터인 SNS 데이터를 활용하여 여론 모니터링을 통하여 문제점을 발굴하고, 감성분석, 네트워크분석 등을 활용하여 정책대안을 개발하였다. 정책과정으로 구분해 보면 경기도 관광에 대한 문제점을 발굴하는 문제형성 단계와 대안을 도출하는 정책형성 단계에 빅데이터를 활용한 사례이다.

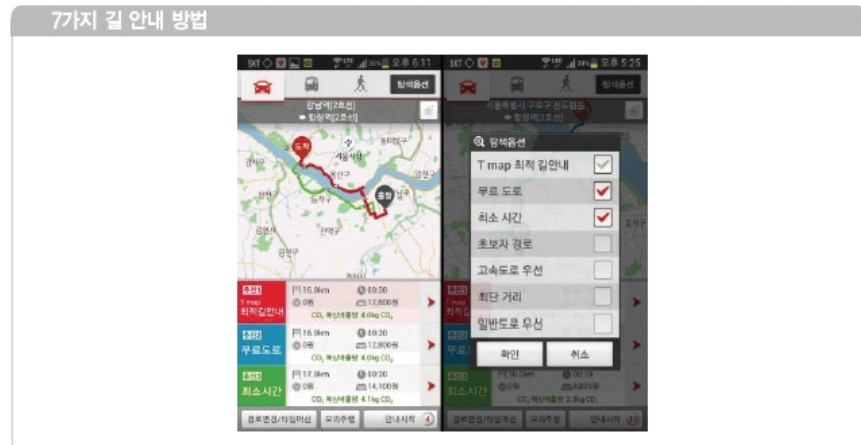
5) SK텔레콤 T맵을 통한 실시간 교통 정보 서비스¹⁰⁾

2001년 서비스를 시작한 T맵은 우리나라 대표적인 길안내 서비스이다. 스마트폰의 보급 활성화에 따라 지능형 서비스 프레임워크(IFS)를 통한 개인화 데이터 축적과 활용을 통해서 서비스를 하고 있다. 이러한 데이터의 수집 방법은 콜택시, 유료 운반 차량, 고속 버스 등에 위성 위치 확인 시스템을 장착하고 T맵 내비게이션을 통하여 실시간으로 전국 도로의 교통 정보를 수집한다. 5분 단위 실시간 정보를 활용하여 도착 예상 시간을 정확히 예측한다. 지도와 길 안내 프로그램을 별도의 하드웨어에 설치하는 기존 내비게이션과 달리 SK텔레콤 서버의 슈퍼컴퓨터가 실시간 정보를 기반으로 목적지까지의 최단 거리를 계산하여 정확한 도착 시간을 제공, 이용자의 편의를 제공하고 있다.

이러한 SK T맵 서비스는 단순히 네비게이션 서비스 이외에도 수집되는 데이터베이스를 통하여 기업 내부 마케팅으로 활용이 가능하고, 내/외부 데이터 결합을 통한 새로운 수익 모델의 확보가 가능하다. 이를 위하여 정보 수집 단계에서부터 경영 관리 정보, 고객 의견 및 청구 정보, 통화 로그, 무선 데이터 로그, 애플리케이션 로그, 외부 정보 등을 구

10) 윤미영(2013), 「더 나은 미래를 위한 데이터 분석」, 한국정보화진흥원

측 분석 할 수 있도록 구성하고 있으며, 기업의 정형·비정형 데이터를 이원화하여 정형 데이터를 위한 통합 모델링을 구축하고, 비정형 데이터 중복을 최소화하여 이원 저장소 간의 연계 분석이 가능하다.



[그림 3-7] 7가지 길 안내 방법

출처 : 블로터닷넷, 2013.1.14, <http://www.bloter.net/archives/140459>

국내 대표적인 스마트폰 기반 내비게이션 서비스인 T맵은 대표적인 빅데이터인 로그데이터를 활용하여 서비스를 수행하고 있으며, 로그데이터 중 실시간 교통데이터 로그를 공간 정보화하여 이를 반영하여 활용하고 있다. 플랫폼 기반의 서비스로 정책과정으로 분류하자면 전과정이 포함된다고 할 수 있다. 다만 사용자 서비스의 관점에서 다양한 로그데이터(실시간교통정보)를 수집하여 대안(실시간경로)을 도출하고 있어 정책과정 중 정책형성 단계에 시사점이 있다.

6) 국민권익위원회 민원정보 분석시스템¹¹⁾

민원 발생 시 문의 해야하는 기관과 연락처를 파악하기 어렵고 전화 상담, 민원 제기 등의 절차가 까다로워 만족도가 떨어지는 등의 문제에 대응하여, 대국민 서비스의 활성화와 신뢰도 및 만족도 향상을 위해 범정부 차원에서 통합된 민원 서비스 체계를 구축할 필요성이 제기되었다. 또한 공공 서비스에 대한 국민의 요구 수준 상승 및 수요자 중심의 서비스, 능동적 서비스 같은 정부 서비스 패러다임 변화에 대한 대응체계 마련이 필요한 시점에서 정부의 민원 서비스 내용을 보다 다양화, 고도화하는 것이 필수 과제로 대두되었고, 빅데이터를 활용하여 민원 정책에 관한 분석 서비스를 개발하였다.

11) 윤미영·권정은(2012), 「빅데이터로 진화하는 세상」, 한국정보화진흥원

이에 국민권익위원회는 온라인 국민소통시스템을 위해 2010년 온라인 국민소통시스템을 오픈하고, 2011년 11월에 개편하여 국민과 정부 간의 가교 및 소통 촉진을 위한 민원 정보 분석을 추진하였다. 홈페이지 민원, 제안, 콜센터 상담 등 1일 11,000여 개의 데이터를 수집하여, 키워드별 추이 분석 및 의미기반의 다차원 분석을 통한 개선사항 조기발견 및 선제적 대응을 목적으로 반복적 민원의 사전 예방 및 대책을 마련하고자 하였다. 과거 민원 발생 현황 등을 분석한 월별, 지역별 분야별 주요 민원 3~6개를 선정하여 캘린더 형태로 제작, 주요 민원 지도를 토대로 각 기관이 자체 민원 유형을 추가하여 상시 활용하고자 하였다.

그 결과 서비스 접근의 편리성 제고와 정부와 국민 간의 능동적 상호 작용을 확대, 언제, 어디서나 정부와의 상담이 가능한 서비스 제공 체계를 구축하였고, 정보의 획득 및 이슈 민원 분석 시간의 단축 등을 통한 업무 효율화 등 긍정적 효과를 가져왔다.

국민권익위의 민원정보분석시스템은 기존 민원데이터를 빅데이터 분석기법을 활용하여 시사점을 도출하고 이를 통하여 민원서비스의 개선과 민원 업무 효율을 높인 사례이다. 활용 데이터인 민원데이터는 정형데이터와 비정형데이터가 복합적으로 구성된 데이터로 볼 수 있다. 이는 민원의 발생 지역, 민원 유형 등은 정형데이터 이지만 민원의 내용은 자유로운 텍스트 형태로서 비정형데이터로 볼 수 있다. 정책과정으로 구분하면 민원데이터에 기초하여 문제를 찾고 이를 종합하였다는 측면에서 문제형성 단계의 빅데이터 활용 사례이다.



[그림 3-8] 민원정보분석시스템 지역별 민원 지도
출처: 윤미영·권정은(2012), 「빅데이터로 진화하는 세상」, 한국정보화진흥원

7) 서울시 관광사업¹²⁾

서울시 산하 공공기관인 서울관광마케팅에서는 기존 외국인 관광객을 대상으로 설문조사에만 의지하던 서울시 외국인 관광 정책 결정의 한계를 극복하기 위해 이동통신과 카드사 빅데이터를 활용해 객관적으로 측정된 정보를 관광정책에 활용하고자 하였다. 관광 서비스 개발을 위하여 국적별 외국인 관광객의 입국 후 이동경로를 파악하고, 외국인 관광객이 국내에서 사용하는 로밍 통화 데이터와 신용카드 이용 데이터를 활용하여 각 이동경로에서의 행태 및 소비금액 등을 파악하였다. 관광객의 동선 정보와 결합된 소비 패턴을 분석하고 해당 정보를 Open API Web 기반 시각화 솔루션을 구축하여 데이터를 분석하였다. 이를 통하여 K-pop 등 한류스타의 팬들을 대상으로 관광객의 패턴 정보를 활용하여 서비스를 제안¹³⁾하는 등 분석결과를 바탕으로 실제 관광 사업에 적용하고 실행할 수 있는 방안을 제시하였다.



[그림 3-9] K-pop 스타(EXO) 팬을 위한 관광 서비스 지도

출처 : http://korean.visitseoul.net/tours/K-Pop-EXO-KR_/10457

12) 서울관광마케팅(2014), “[소식] 서울관광마케팅, SK텔레콤, 신한카드와 함께 관광 빅데이터 사업 추진”, <http://www.seoulwelcome.com/kr/?p=5091>.(검색일: 2017.07.01)

13) http://korean.visitseoul.net/tours/K-Pop-EXO-KR_/10457

본 사례는 빅데이터 유형 중 로그데이터를 활용한 것으로 휴대전화 로그데이터와 신용카드 사용데이터를 활용하여 이동동선과 소비행태를 분석하여 새로운 관광루트를 발굴하였다. 정책과정으로 보면 다수의 관광거점을 발굴하고 이를 빅데이터를 활용하여 대안을 개발하고, 최적의 대안을 채택한 것으로 정책형성과 대안채택 단계의 빅데이터 활용 사례로 볼 수 있다.

8) 국내 사례 종합 및 시사점

국내 사례를 종합하면 다양한 유형의 빅데이터가 실제 사업과 정책개발에 활용되고 있음을 확인할 수 있다. 전통적인 빅데이터인 소셜 데이터는 물론이거니와 기관 또는 기업 내부의 기존 정형데이터도 프로젝트의 성격에 따라 다양하게 활용되고 있다. 분야별 주제와 이슈에 대한 여론 수렴과 시장의 동향 분석을 위해서는 소셜 데이터가 주로 활용되고 있으며, 특정 주제에 대한 지역적/공간적 문제도출과 대안개발을 위해서는 로그데이터 중 휴대전화, 신용카드 데이터가 많이 사용되고 있다. 그 외 기존 보유 또는 구축하고 있는 텍스트데이터 등에 빅데이터 분석 방법을 적용하여 새로운 시사점을 도출하는 사례도 확인할 수 있다.

정책과정 단계별로 빅데이터 활용 사례를 구분하면 대부분의 사례가 문제형성과 정책형성, 대안채택 과정에 빅데이터를 중점적으로 활용하고 있다. 여론의 동향을 빅데이터로 파악하고 모니터링을 통하여 현재 상황에 대한 진단을 하는 문제형성 단계와 빅데이터 분석을 통하여 다양한 대안을 개발하는 정책형성 단계에 시사점이 있다.

SKT의 스마트 인사이트는 소셜 데이터를 실시간으로 분석하여 여론 동향을 파악하고, 국민권익위는 기존의 민원데이터를 텍스트 마이닝을 통하여 문제점을 도출하는 등 정책과정 중 문제형성 단계에서 빅데이터 활용 가능성을 보여준다. 또한 서울시 심야버스, SKT T맵 등 나머지 사례에서는 로그데이터를 중심으로 정책대안을 만들고 또한 다양한 대안 중 최적의 대안을 탐색하는 정책형성과 대안채택 단계에서 정량적 의사결정을 위한 빅데이터 적용방안을 확인 할 수 있다.

경기도는 소셜 데이터 분석을 통하여 경기관광의 문제점과 대안을 포괄적으로 탐색하였고, 서울시는 휴대전화·신용카드 등 로그데이터를 활용하여 대안(관광코스)을 개발하고 최적의 대안을 찾았다. 이 사례들에서 확인할 수 있듯이 관광분야라는 동일한 주제라도 정책과정과 달성하려는 목적에 따라 활용 빅데이터가 달리 적용됨을 알 수 있다.

[표 3-1] 국내 빅데이터 기반 정책개발 사례 종합

활용사례	빅데이터 유형	빅데이터 활용 정책과정
- 트위터 연관 키워드 : 관광 여론 모니터링 (경기연구원)	소셜 데이터	문제형성, 정책형성
- 인터넷 및 SNS 실시간 동향 : 시장조사 및 위기관리 (SKT)	소셜 데이터	문제형성
- 민원 상담 키워드 : 월별, 지역별 민원 파악 (권익위)	정형데이터 + 텍스트데이터	문제형성
- 교통카드 및 휴대전화 위치정보 : 버스 노선 수립 (서울시 심야버스)	로그데이터	정책형성, 대안채택
- 교통 빅데이터 구축 : 실시간 혼잡구간 정보 제공 (SK T맵)	로그데이터	정책형성
- 지점별 매출 및 기상 자료 : 날씨에 따른 수요 예측 (파리바게뜨)	정형데이터 + 로그데이터	정책형성
- 휴대전화로밍, 신용카드 : 관광 코스 개발 (서울시 관광)	로그데이터	정책형성, 대안채택

3. 국외 사례

1) 뉴욕 범죄현황 통계분석 및 예측 시스템 컴프스탯(CompStat)¹⁴⁾

컴프스탯은 COMPUter STATistics의 약자로 경찰에서 범죄 다발지점을 인지하고 우선적으로 해결하는 것을 목적으로 과거에 발생한 범죄들을 지도상에 표시하여 향후 범죄 발생 지역을 예측하는 시스템이다. 시스템에 포함된 정보는 범죄패턴을 파악할 수 있도록 범죄유형, 희생자, 사건 발생 일시 등으로 향후 전산화된 지도를 통해 범죄를 예방하고자 하였고, '실시간 범죄센터'의 운영으로 다양한 범죄 정보 데이터베이스로부터 빠르게 정보를 추출하여 현장에 있는 일선 경찰관들에게 필요한 정보를 제공하는 데이터 허브로서의 기능을 수행한다.

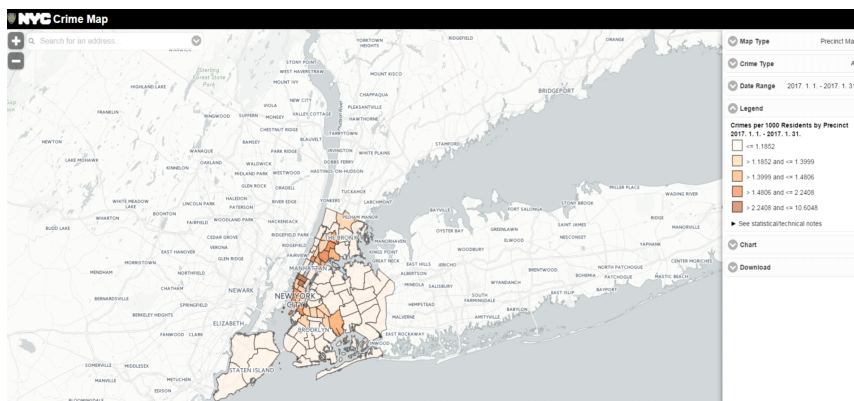
뉴욕의 실시간 범죄센터는 뉴욕 경찰국 본청에서 직접 관리하며 40여 명 이상의 수사관과 민간 분석가들에 의해 24시간 동안 운영된다. 5백만 건 이상 뉴욕 주의 범죄 기록, 가석방, 보호관찰 파일, 2천만 건 이상의 뉴욕시 범죄자 고발, 긴급신고, 소환 기록 등을 통해 실시간으로 위험을 분석하여, 범죄를 예방하고자 하였다. 이를 활용하여 분산되어 있던 뉴욕시 경찰국의 범죄 관련 데이터를 통합하는 과정에서 일선 경찰관들이 참여하는 서비스를 구성하였고, 이는 2017년 현재 NYC Crime Map이라는 이름으로¹⁵⁾ 제공되

14) 김창수·송민정(2014), 「빅데이터 경영론」, 학현사, pp.144-149

15) NYC Crime map, <https://maps.nyc.gov/crime/>. (검색일: 2017.0607)

고 있으며, 뉴욕 시민에게 범죄 발생 현황 및 통계 자료를 공개하고 지역구별 평균, 개별 범죄, 히트맵 등 범죄 지도를 제공한다. 1994년 활용 이후 미국 전역에서 2009년까지 범죄가 19.4% 감소하였지만, 컴프스택의 효과로 뉴욕시는 세 배 이상의 64.5%가 감소하였다. 컴프스택은 1994년 뉴욕시가 도입한 이후 보스턴, 필라델피아, 마이애미, 뉴올리언스 등 다른 미국 내 타 대도시에서도 도입을 시작하였다.

본 사례의 활용 빅데이터는 일정한 형식이 있는 정형데이터이다. 또한 해당 정형데이터는 공간정보(경위도 좌표)를 포함하고 있어 공간빅데이터로 구분 할 수 있다. 이러한 공간빅데이터를 자체적으로 개발한 분석 알고리즘을 통하여 사전에 범죄발생 가능성이 높은 지역을 선별하고 있어 정책형성과 대안채택 단계의 빅데이터 활용 사례로 볼 수 있다.



[그림 3-10] NYC Crime Map
출처 : <https://maps.nyc.gov/crime/>

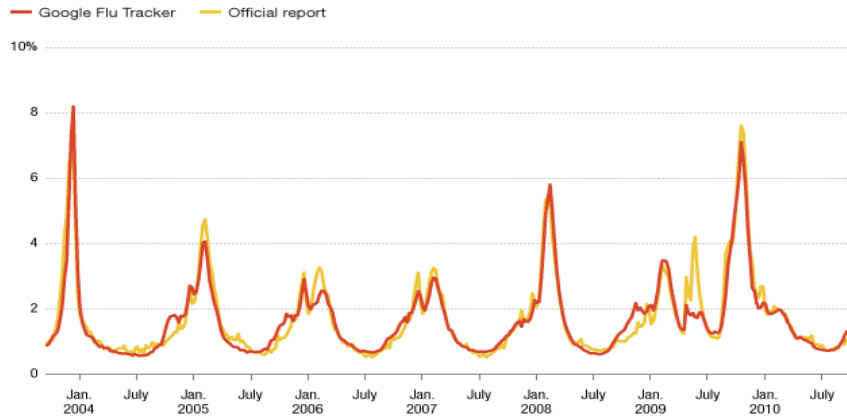
2) 구글 검색어 분석을 통한 독감 유행 예측 시스템 Google Flu Trends¹⁶⁾

2008년 구글은 독감, 인플루엔자 등 독감과 관련된 검색어 쿼리의 빈도를 조사하여 '구글 독감 트렌드(Google Flu Trends)'라는 독감 확산 조기 경보체계를 마련하였다. 해당 지역에서 '실제 병원을 방문한 환자의 수'와 '구글에서 독감 관련 키워드를 검색한 사용자의 수'의 관계를 설정하고, 알고리즘을 개발하여, 지역별로 매우 높음, 높음, 보통, 낮음, 매우 낮음 등 5단계로 독감 정보를 제공하였다.

'구글 독감 트렌드'의 결과를 미국 질병통제본부의 데이터와 비교해 본 결과, 검색 빈도 및 실제 독감증세를 보인 환자 숫자 사이에 매우 밀접한 상관관계가 확인되었고, 미국

16) Google, "Google Flu Trends", <https://www.google.org/flutrends/about/>.(검색일:2017.06.07)

보건 당국보다 한발 앞서 시간 및 지역별 독감 유형 정보를 제공하는 것으로 알려져 검색 엔진의 검색데이터를 활용한 대표사례로 알려졌다.



[그림 3-11] Google Flu Trends 시계열 예측

출처 : <http://www.npr.org/sections/health-shots/2014/03/13/289802934>

그러나 실제 독감 발병과 단순 우려로 인한 검색을 구분하지 못하는 한계¹⁷⁾ 등 예측력에 대한 문제가 학계와 언론을 통하여 확산되었다. 구글은 2015년에 해당 서비스의 대중 서비스를 중단하고, 관련 단체가 신청하면 해당 자료를 제공하는 제한적 서비스로 전환하였다. 현재 ‘구글 독감 트렌드’의 데이터는 많은 질병관련 단체에 제공되어 보다 예측력이 높은 독감예보 모델을 만드는 것에 기여하고 있다.

이러한 검색엔진의 검색데이터는 비정형인 로그데이터의 하나로 검색 정보에 사용자의 현재 상태나 상황에 관한 내용이 담겨 있고, 실시간으로 무수히 많은 누적되어 사회적인 변화나 흐름을 파악하는 것이 가능하다. 따라서 정책과정의 문제형성 단계인 여론탐색과 동향파악에 활용 가능성이 크다.

3) 미국 영화사 SNS 분석을 통한 영화 흥행 예측¹⁸⁾

과거 영화의 흥행 여부를 판단하기 위해 비용을 들여 여론조사를 실시했던 할리우드 영화사들이 최근 트위터, 페이스북 등 소셜 미디어 분석으로 대체하고 있다. 할리우드 영

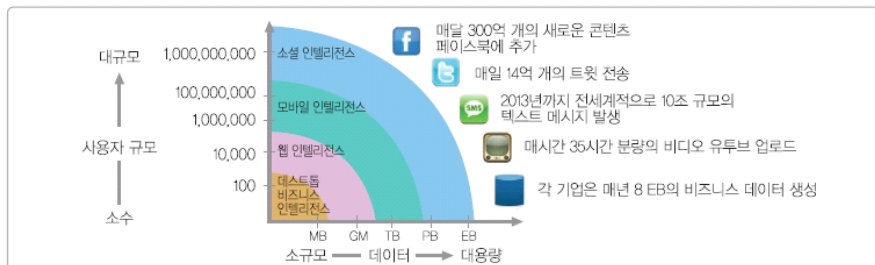
17) Butler, Declan (2013), “When Google got flu wrong”, Nature, <http://www.nature.com/news/when-google-got-flu-wrong-1.12413>.(검색일:2017.06.08)

18) Garrahan M.(2012), “Hollywood turns to tweet analysis”, Financial Times, <https://www.ft.com/content/9c810f04-907e-11e1-9e2e-00144feab49a>.(검색일: 2017.06.08)

회사에서 트위터 분석 업무를 맡고 있는 벤 칼슨은 “트위터를 통해 분석한 영화 ‘트와일라잇’의 예상 흥행 실적은 약 1억 4,000만 달러였는데 실제로 1억 3,850만 달러로 매우 근접했다.”고 언급하였다. 최근 여론 분석 기술로 주목받고 있는 오피니언 마이닝은 소셜 미디어와 웹사이트 등에 나타난 여론과 의견을 분석, 유용한 정보로 재가공하는 기술이다. 이는 사용자가 게재한 의견과 감정을 나타내는 패턴을 이용해 특정 오브젝트에 대한 의견이 긍정인지 또는 중립, 부정인지 찾아내는 기술로써 분석을 바탕으로 영화 흥보의 전략 수립, 흥행 실적에 대한 예측으로 활용되었다.

각각의 영화사들은 개봉 전 SNS를 면밀히 검토하여 영화에 대한 고객들의 관심 정도를 파악하고 흥보 예산을 책정한다. 영화를 본 사람들이 SNS에 관람평을 올리면 잠재 고객들이 이를 보고 영화 관람 여부를 결정하기에, 영화 흥행에 대한 SNS의 영향력은 점차 확산되고 있다. 2010년 리메이크된 ‘나이트메어’는 트위터에 올라온 부정적인 반응 때문에 개봉 후 유료 관객이 50%가 감소하는 등 SNS의 영향력은 점차 증가하는 추세이다.

이러한 미국 영화사의 영화 흥행 예측 사례는 소셜 데이터를 활용하여 실시간으로 사용자의 의견을 분석하여 동향파악과 모니터링에 활용한 것이다. 정책과정 관점에서는 문제형성 단계의 여론 동향파악과 정책평가 단계의 정책 모니터링에 활용 가능성을 보여주는 사례이다.



[그림 3-12] 소셜네트워크 데이터량

출처 : IDG Korea(2012), 「빅 데이터의 이해」, IDG Tech Report

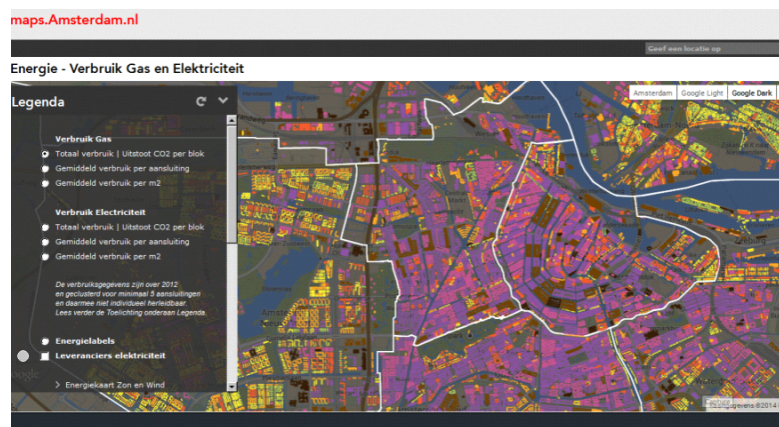
4) 암스테르담 스마트 시티 프로젝트 에너지 아틀라스(Energy Atlas)¹⁹⁾

에너지 아틀라스는 EU의 TRANSFORM 프로젝트(저탄소 프로젝트)의 일환으로 도시 전역의 에너지 사용량 절감 및 신재생 에너지 생산 필요성 등이 제기되었고, 도시 전역의 에너지 사용량을 파악하여 에너지를 절약하고, CO2발생을 줄이며, 신재생 에너지 사

19) amsmarterdam city, “A tool that stimulates the use of renewable energy”, <https://amsterdamsmartcity.com/projects/energy-atlas>. (검색일: 2017.06.09)

용을 증대하고자 도입되었다. 데이터 수집은 암스테르담 시에서 제공하는 Open data와 Liander, Waternet의 에너지 회사, NUON, TNO의 주택협회, AFWC 암스테르담연맹과 같은 협력 파트너들이 제공한 데이터를 활용하였다.

에너지 아틀라스는 인구, 주택 크기, 가스 및 전기 사용량, 태양열 생산에너지, 풍력 생산 에너지 등 지역의 특성을 나타낼 수 있는 통계데이터와 에너지 사용량 등의 측정 데이터 복합 맵핑을 통해 사용자의 목적에 따라 개인화된 화면을 제공한다. 수집한 데이터를 암스테르담 지역의 지도 위에 맵핑하여 도시의 에너지 사용에 대하여 직관적으로 파악할 수 있도록 하였다. 2~3개 항목을 동시에 맵핑하는 등의 인터랙티브 맵을 제공하여 사람들이 원하는 정보를 보다 쉽게 이해할 수 있도록 도움을 주고 있다.



[그림 3-13] Use of Gas and Electricity

출처 : https://maps.amsterdam.nl/energie_gaselektra/?LANG=ko

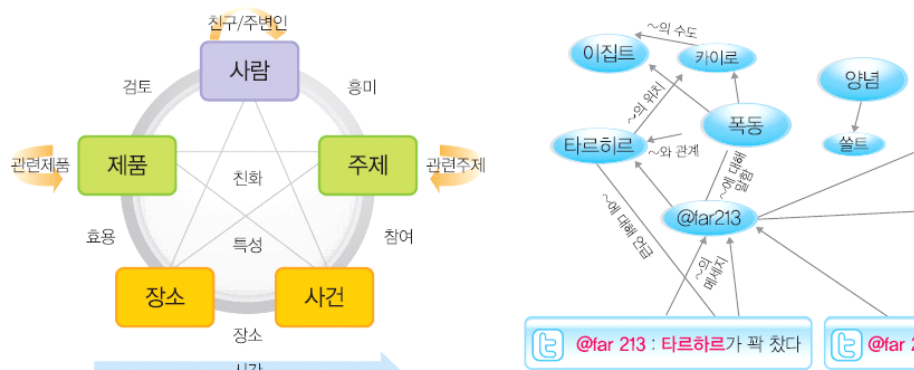
현재 에너지 아틀라스를 통해 지역의 폐열을 사용할 수 있는 가능성을 확인하고 시와 기업들이 이를 활용할 수 있는 방안을 논의 중이다. 신재생에너지 시스템으로의 전환 수단으로 빅데이터를 분석 및 활용한 프로젝트는 시민들에게 자신의 에너지 소비 및 사용에 대한 이해를 증진시키고 기업들에게 신재생에너지 자원 및 기반 정보를 제공하였다는 데 의의가 있다.

에너지 아틀라스는 에너지 공급정보와 건축물 현황정보를 결합하고, 공공과 민간이 보유하고 있는 데이터를 융합하여 하나의 시스템으로 구성하고, 서로 다른 정형데이터를 공간정보기반으로 결합하여 지도기반 정보서비스를 제공한 사례이다. 해당 시스템은 정책개발의 여러 분야에 활용할 수 있지만 기본적으로 에너지정책의 문제점을 직관적으로 탐색할 수 있도록 지원한다는 측면에서 문제형성 단계에 활용가능성을 보여준다.

5) 월마트, 소셜 데이터 분석을 통한 소비자 패턴 조사²⁰⁾

월마트는 2011년 소셜 미디어 회사 코스믹스(Kosmix) 인수를 통해 본격적으로 소셜네트워크와 관련한 빅데이터 도구와 콘텐츠를 개발하였다. 이를 통하여 유통과 전자상거래(e-commerce) 간의 시너지를 발휘하여 단순한 소매 영업점을 뛰어넘는 높은 투자 수익률을 거두었고, 각 분야의 특징을 파악하여 사업에 반영하였다.

2012년 인터넷 쇼핑물 ‘월마트닷컴’에 적용한 소셜 게놈(Social Genome)은 ‘트위터’나 ‘페이스북’ 같은 소셜네트워크 서비스 이용자가 자주 쓰는 단어 · 사람 · 제품 · 장소 · 사건을 검색하고 분석해 상품 판촉에 활용한다. 단순히 단어를 찾아내는 데 머무르지 않고 검색한 말과 이용자 간 맥락을 살핀 결과를 제공하는, 이른바 ‘시맨틱(Semantic) 검색’ 체계다. 월마트가 소셜 게놈을 시험했더니 ‘검색한 뒤 상품을 산 비율’이 기존 인터넷 검색 체계보다 10~15% 상승했다. 검색 결과를 클릭한 양도 20% 늘어 소셜 게놈 체계가 상대적으로 소비자의 검색 기대치에 가까운 결과를 제공했음을 보여 줬다. 월마트는 소셜 게놈 서비스에 적합한 검색 엔진인 ‘폴라리스(Polaris)’까지 자체 개발했다.²¹⁾



[그림 3-14] 소셜 게놈(Social Genome) 개념과 분석 사례

출처: 윤미영·권정은(2012), 「빅데이터로 진화하는 세상」, 한국정보화진흥원

또한 온라인에서의 지역별 고객 반응을 오프라인 매장으로 연결 가능하도록 하였으며, 시시각각 변화하는 소비자의 패턴을 분석하여 적재적소에 필요한 물품을 제공함으로써 불필요한 재고 낭비를 방지하고, 고객 만족도 향상으로 이어져 기업 발전에 선순환적인 역할을 하고 있다.

20) 윤미영·권정은(2012), 「빅데이터로 진화하는 세상」, 한국정보화진흥원

21) 한국정보통신기술협회(2013), “네이버 지식백과: 소셜 게놈”, ICT 시사용어 300.

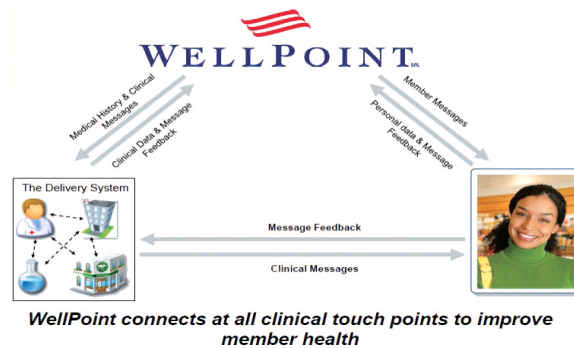
<http://terms.naver.com/entry.nhn?docId=3587142&cid=59277&categoryId=59280>.(검색일: 2017.07.01)

월마트의 소셜 데이터 기반 사례는 빅데이터를 활용하여 온·오프라인 연계방안을 제시한 것으로, 정책개발 관점에서 보면 여론 모니터링을 위하여 소셜 데이터를 수집하여 빈도와 감성 분석으로 결과를 해석하는 것보다 맥락을 파악(시맨틱)하면 보다 실용적인 결과를 얻을 수 있음을 보여준다. 정책과정으로 보면 여론 모니터링을 통한 문제형성 단계와 정책평가 단계의 빅데이터 활용 사례이다.

6) 미국 웰포인트 빅데이터 기반 의료정보 제공 시스템²²⁾

미국 의료보험회사인 웰포인트(WellPoint)는 IBM의 왓슨 솔루션을 도입하여 회사에 등록된 건강보험 자료와 3,420만 명에 대한 환자 정보를 통합 분석하였다. 환자의 증상, 면담 결과, 진단 결과 등 진료 내역에 대한 모든 정보를 저장하고 2억 페이지에 해당하는 자료를 검색 분석하여 3초 안에 모든 결과가 도출되도록 하였다. 다양한 환자의 의료기록 수집을 통해 데이터를 분석하여 최적의 진단 및 환자 치료 가이드라인을 제시하고, 불필요한 진단 및 치료 억제로 시간과 비용을 절약할 수 있도록 하였다.

High Level Flow of Health Information



[그림 3-15] 웰포인트 의료정보 제공 체계

출처 : 윤미영·권정은(2012), 「빅데이터로 진화하는 세상」, 한국정보화진흥원

이는 환자의 상황에 맞는 최선의 치료 방법을 도출하여 불필요한 치료를 줄여 환자 및 의료보험 회사의 진료비 낭비를 방지하였다. 또한 고령화 사회로 접어들면서 노인들의 만성 질환을 체계적으로 관리, 고령층에 대한 효과적인 진료 서비스를 제시하는 등 빠르게 변화하는 의료 환경에서의 진단 및 치료 방법을 환자, 의료보험 회사 및 의사가 공유함으로써 환자에게 최적의 치료 서비스를 제공하고, IT 기술을 응용한 의료 기기의 보급, 원격 치료 가능으로 환자 중심의 의료 사장에 대한 개편 가능성을 제기하였다.

22) 윤미영·권정은(2012), 「빅데이터로 진화하는 세상」, 한국정보화진흥원

웹포인트 사례는 우리나라와 같이 건강보험이 전국민에게 적용되고 해당 기록이 공공에서 일원화되어 관리되고 있는 나라에서 더욱 시사점이 크다. 의료 빅데이터는 개인정보 누출의 위험으로 상대적으로 수집이 어렵지만 우리나라의 경우 이미 정형데이터로 수집과 가공이 되어 있어 익명화에 대한 기술개발과 관련제도 정비가 선행된다면 건축·도시 분야에서도 건강한 공간환경 조성을 위한 활용가치가 높은 빅데이터이다. 정책과 정 관점에서 보면 정형데이터(의료데이터)를 분석하여 다 수의 대안(치료방법)을 도출하고 최적의 대안을 찾는 다는 점에서 정책형성과 대안채택 단계에 시사점이 있다.

7) 브라질 리우데자네이루 재해예방 통합관리 지능형 운영센터²³⁾

브라질 리우데자네이루는 환경, 교통, 치안 등 도시에서 발생하는 모든 정보를 통합하여 스마트 도시 구현과 도시에서 발생하는 각종 사건, 사고에 대비하여 도시 보안 향상을 목적으로 재해예방 통합관리 지능형운영센터를 설립하였다.

2014년 월드컵과 2016년 올림픽을 준비하기 위해 도시 내 30여 개 기관의 정보와 프로세스를 단일체제로 통합하고 자연재해, 교통, 전력공급 등을 24시간 감시하는 시스템을 도입하였다. 이러한 데이터는 30여 개에 이르는 시 정부 산하부서와 기관이 공유한 다양한 정보를 통합·연동하여 갑작스러운 홍수나 산사태와 같은 자연재해 뿐 아니라 대규모 행사, 교통관계까지 통합 제어가 가능하도록 계획하였다. 예를 들어, 강 유역의 지형 측량 자료, 강수량 통계, 레이더 사진 등의 데이터와 교통체증, 정전 사태 등 도시에 영향을 미치는 데이터를 분석하여 48시간 전에 교통 상황을 예측할 수 있다.

초기에는 홍수 등 비상사태를 예측하기 위해 구축되었지만, 도시의 다양한 이벤트와 해변에서 열리는 파티, 스포츠 경기 중 팬들의 돌발 행동, 교통사고 등에 이르기까지 범위가 확대되었고, 시민들에게 페이스북과 트위터를 통해 현재 날씨와 교통상황 등의 정보를 실시간으로 제공하고 있다. 또한 콘서트, 축구경기, 축제 등 특별한 행사 시, 리우데자네이루의 홍수 및 산사태 예측에 변동이 있을 경우, 시청 공무원과 간접 대응팀에 통보된다.

이러한 시스템을 통하여 응급상황 대응시간이 약 30% 가량 개선되었고, 사망자 수도 10% 감소하였으며, 2011년 대비 16% 많은 900여 명의 노상 방뇨자를 적발했다. 지능형 운영센터의 미션은 실시간 모니터링 및 분석을 위해 다양한 데이터를 통합하는 것으로, 비상사태나 대규모 행사 개최 시 실시간으로 정보를 취합, 분석한 자료를 기반으로

23) 윤미영(2013), 「더 나은 미래를 위한 데이터 분석」, 한국정보화진흥원

도시 관리자들이 각종 상황을 파악하고 결정하는 도시의 컨트롤 타워로서 역할을 수행한다.



[그림 3-16] 리우데자네이루 지능형운영센터

출처 : 윤미영(2013), 「더 나은 미래를 위한 데이터 분석」, 한국정보화진흥원

대표적인 스마트시티 사례인 브라질 리우데자네이루는 빅데이터가 스마트시티에 어떻게 활용되는 지를 잘 보여준다. 기존 시가 보유한 관련 데이터를 통합하고, 추가로 도시 곳곳에 설치된 센서(CCTV 등)가 수집한 정보를 통합하여 분석을 수행한다. 분석 수행 결과는 의사결정을 위한 다양한 대안으로 만들어지며 이를 토대로 도시 관리자가 상황을 파악하고 의사를 결정한다. 또한 실행된 내용에 대해서 다시 모니터링이 시행되고 이를 통해서 정책과정이 피드백 된다. 이렇게 스마트시티에서 빅데이터는 도시 운영의 모든 분야에 활용된다. 따라서 정책과정의 모든 단계에서 빅데이터를 활용하는 모델이며, 활용하는 빅데이터 유형도 매우 다양하다. 다만 리우데자네이루의 경우 재해예방에 보다 특화되어 있어 도시내의 재해취약점을 발굴하고, 이에 대한 대안을 개발하는데 빅데이터가 집중적으로 활용되고 있다. 따라서 주로 활용되는 빅데이터는 공간빅데이터이며, 정책과정 중 문제형성과 정책형성 단계에 시사점이 많은 사례 이다.

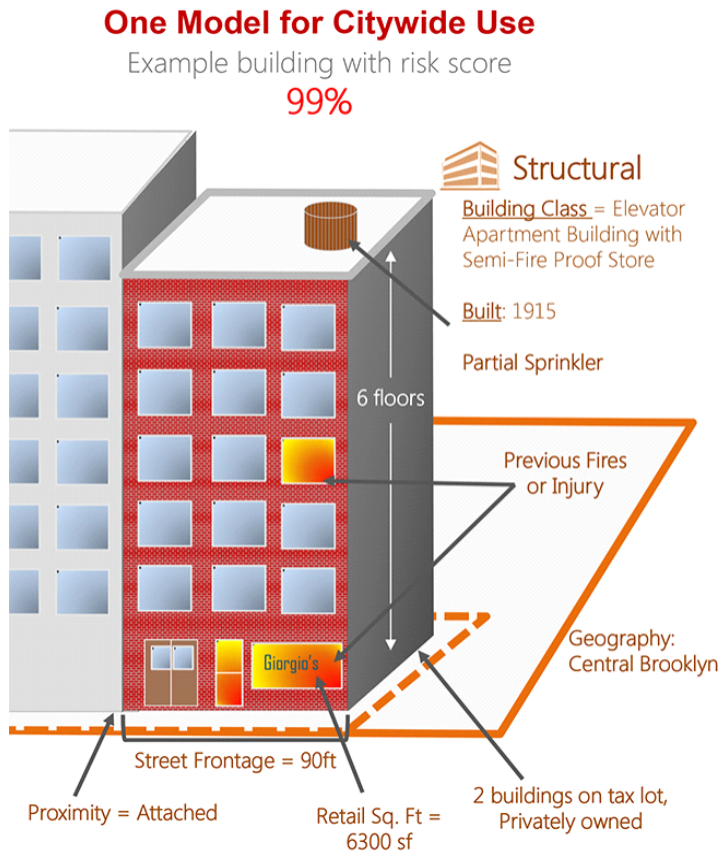
8) 뉴욕시 소방서(FDNY)의 RBIS & FireCast²⁴⁾

기존에 NYPD는 건축물의 소방점검 결과를 문서로 작성하여 기존의 건물 검사 결과를 종이에 작성하여 보관 및 관리해왔다. 그러나 이러한 문서는 뉴욕시 관내 소방서간 정보 공유가 되지 않는 등 여러 가지 문제점이 있었다. 이러한 점을 보완하기 위하여

24) Heaton, Brian(2015), "New York City Fights Fire with Data", <http://www.govtech.com/em/safety/New-York-City-Fights-Fire-Data.html>. (검색일: 2017.06.20)
Urban Omnibus(2014), "The Analytics of Fire", <http://urbanomnibus.net/2014/06/the-analytics-of-fire/>. (검색일: 2017.06.20)

RBIS(Risk-Based Inspection System:위험기반 검사 시스템)과 FireCast의 개발이 시작되었다.

FireCast는 RBIS의 핵심모듈로 5개의 뉴욕시 관내 기관들로부터 받은 데이터에 기초하여 26가지 검사를 수행하여 60여 개의 위험요소로 구성하는 알고리즘이며, 이를 바탕으로 화재에 보다 취약한 건축물의 목록을 생성한다. 수집하는 데이터는 건축물유형, 거주자수, 용도, 건축시기, 면적, 층수, 전기설비, 스프링클러 개수, 엘리베이터 유무 등 건축물 특성정보와 불법개조, 소음 등의 행동적 단서 또는 위반행위까지 포함하고 있으며, 기존 화재 이력 역시 중요한 변수로 포함되어 있다. 현재 RBIS는 FireCast 2.0이 탑재되어 운영되고 있으며, FireCast는 RBIS가 위험을 점수화(risk-scoring)하는 엔진 역할을 수행하며, 조사관이 매일 점검 결과를 RBIS에 입력하면, RBIS가 위험 점수를 책정하여 건물별 위험성 프로파일을 생성한다.



[그림 3-17] Example of a high risk building according to FireCast
출처 :<http://urbanomnibus.net/2014/06/the-analytics-of-fire/>

RBIS는 뉴욕시의 약 35만 개의 건물에 대해 화재에 취약한 건물을 추적하여 우선적으로 소방서에서 점검해야 할 건축물의 우선순위 선정에 도움을 준다. FDNY가 접근 할 수 있는 소프트웨어를 제공하며, FDNY 대원이라면 누구나 건축물과 구조물의 최신 점검 내역조회가 가능하다. 또한 FDNY의 소방점검의 전체 검사 워크플로우를 자동화하여 마지막 검사 날짜와 같은 최신 정보 조회를 가능하도록 개발되었다.

FDNY의 RBIS는 뉴욕시의 데이터 공유 플랫폼인 DataBridge에 연결되어 도시계획 부서의 건축물 정보와 데이터를 공유하고 있다. 이러한 공유 플랫폼을 활용하여 소방점검 시 발견되는 건축물 정보에 대한 오류를 수정 제안하는 등의 피드백을 통하여 데이터 오류를 상호 보완하여 보다 완성도 높은 데이터베이스를 구축하고 있다. DataBridge는 뉴욕시장 직속의 정보센터인 MODA(The Mayor's Office of Data Analytics)가 구축 운영하고 있으며, 뉴욕시의 20개 기관이 각각 운영하고 있는 데이터베이스를 자동적으로 연결해 주는 플랫폼이다.

FDNY의 빅데이터 활용 사례는 건축물 관련 데이터를 활용하여 화재위험도를 정량적으로 측정하였다는 측면에서 의미가 있다. 건축물의 속성정보와 화재발생정보 등의 서로 다른 분야에서 관리해온 정형데이터를 건축물 단위로 통합하고 이를 통해서 새로운 정보를 만들어 낸 것이다. 정책과정에서 보면 잠재적인 화재발생 가능성이 높은 건축물을 빅데이터를 활용하여 발굴하였다는 측면에서 문제형성 단계의 빅데이터 활용 사례라 볼 수 있다.

9) 국외 사례 종합 및 시사점

국외 사례도 국내 사례와 동일하게 다양한 유형의 빅데이터가 활용되고 있지만, 세부적으로 보면 활용 데이터 종류에 차이가 있다. 국외의 경우 범죄 데이터와 의료 데이터를 활용한 사례를 다 수 찾을 수 있으나, 국내는 해당 자료를 보유하고 있는 경찰청과 건강보험공단의 일부 사례를 제외하고는 거의 없다. 국내의 경우 범죄 피해자와 병원의 환자에 대한 개인정보보호와 해당 데이터 분석 결과가 미치는 영향에 대한 대상 지역주민의 우려가 크기 때문이다. 반대로 외국의 경우 휴대전화와 신용카드 정보를 활용한 빅데이터 사례가 많지 않다. 이는 개인정보에 대한 국가별 인식의 차이로 볼 수 있다.

국내 사례에서도 확인된 것처럼 여론의 동향과 모니터링을 위해서는 소셜 데이터를 주로 활용하고 있다. 또한 세부 주제와 이슈에 대응하여 구체적인 문제점 발굴과 이에 대

한 다양한 대안을 개발하기 위해서 정형데이터와 로그데이터의 활용이 많은 것으로 확인 되었다. 그리고 정책과정의 단계별로 빅데이터 활용 실태를 보면 정책집행 단계를 제외하고 거의 모든 정책과정에 빅데이터를 활용하고 있음을 확인할 수 있다.

국외 사례에서는 자체적인 알고리즘을 개발하여 빅데이터를 분석 활용하는 사례가 많이 나타났다. 뉴욕시의 CompStat은 기존 범죄데이터를 종합하여 범죄발생지역을 사전에 예측하는 알고리즘을 개발하였고, 구글은 검색데이터 기반의 독감 예측 알고리즘, 월마트의 소셜 계층, FDNY의 FireCast 등 대다수의 사례에서 기존의 빅데이터 분석방법론에 구축 데이터의 특성에 기초한 자체적인 알고리즘을 구축하여 운영하고 있다.

FDNY의 사례에서 뉴욕시 MODA의 DataBridge는 정부가 보유하고 있는 공공빅데이터의 구축 및 활용에 큰 시사점을 제시하고 있다. 우리나라 정부에서도 데이터공개와 활용정책에 많은 비용과 시간을 투자하고 있다. 그 결과 많은 양의 데이터가 민간에 공개되고 있으나, 각 부처별·기관별 구축하고 있는 데이터가 매칭 되지 않아서 실제 활용에 한계가 있다. 이러한 공공 빅데이터의 활용을 높이기 위해서는 뉴욕의 DataBridge와 같은 서로 다른 데이터 간에 이를 연계할 수 있는 시스템이 반드시 필요하다.

[표 3-2] 국외 빅데이터 기반 정책개발 사례 종합

활용사례	빅데이터 유형	정책과정
범죄발생 위치정보 : 범죄발생 지역 예측 및 대응(뉴욕시, CompStat)	정형 데이터	정책형성, 대안채택
검색엔진 조회수 : 검색데이터 기반 독감 유행 예측(구글)	로그 데이터	문제형성
트위터 실시간 동향 : 영화 흥행 예측 및 마케팅 예산 책정(미국 영화사)	소셜 데이터	문제형성, 정책평가
에너지 정보 + 건축물 정보 : 에너지 관련 정책 수립(암스테르담)	정형 데이터	문제형성
트위터 연관 키워드 : 소비자 수요 패턴 파악 (월마트)	소셜 데이터	문제형성, 정책평가
의료 데이터 : 최적의 진단 및 치료 방법 파악(웰포인트)	정형 데이터	정책형성, 대안채택
도시 및 재해관리 데이터 : 재해예방 통합관리(리우데자네이루)	정형 데이터	문제형성, 정책형성
건축물 정보 + 화재 정보 : 소방점검 우선순위 도출(뉴욕소방서)	정형 데이터	문제형성

제4장 빅데이터 기반 건축·도시 정책 개발 체계 수립

1. 건축·도시 분야 정책 현황

2. 빅데이터 기반 건축·도시 정책개발 체계 수립

1. 건축·도시 분야 정책 현황¹⁾

2017년 3월 현재 우리나라 건축·도시 분야 정책은 국토교통부에서 총괄하고 있으며, 그 중 국토도시실 내의 도시정책관실, 건축정책관실에서 담당하고 있다. 도시정책관실은 도시정책과, 도시재생과, 도시경제과, 녹색도시과, 건축정책관실은 건축정책과, 녹색건축과, 건축문화경관과로 구성되어 각 분야별 정책을 개발, 집행, 모니터링 하고 있다.

국토교통부에서 홈페이지를 통하여 제시하고 있는 건축·도시 정책의 주요 주제는 5가지이며 ‘수요자 중심의 토지이용체계 확립 및 살고 싶은 도시공간 구현’, ‘재해로부터 안전한 도시 조성’, ‘건축물 안전관리 및 건축투자 활성화’, ‘녹색건축 활성화 및 아름다운 국토경관 조성’ 5가지이다.²⁾

본 연구에서는 건축정책관실 및 도시정책관실의 2017년 업무계획을 바탕으로 건축·도시 정책을 종합한다.

1) 본 과제 수행 중 2017년 5월10일 새로운 정부의 출범으로 도시재생기획단이 신설되는 등 국토교통부 조직의 변화가 있으며, 정책 또한 신정부의 공약사업이 반영되어 일부 변화가 있다. 따라서 본문의 내용과 현재 정책내용이 다소 차이가 있을 수 있으나, 연구 목적이 세부적인 정책을 개발하는 것이 아니라 정책 개발 체계를 수립하는 것이기 때문에 자료 수집 시점인 2017년 3월을 기준으로 연구가 진행되었다.

2) 국토교통부, “도시/건축 정책”, http://www.molit.go.kr/USR/WPGE0201/m_35312/DTL.jsp.(검색일: 2017.03.12)

1) 건축정책관실 2017년 업무계획 분석³⁾

국토교통부의 2017년 업무계획의 주요과제는 경제활성화, 주거복지, 생활 안전 등 총 6가지로 구성되어 있다. 그 중 건축정책관실의 업무는 5개의 주요과제에 포함되어 있으며, 10개의 세부추진과제로 구성되어 있다.

□ 전문인력의 육성 통한 일자리 창출

주요과제 중 '경제살리기'에 해당하며 설계공모 사업 및 수상전을 통해 신진건축사를 발굴하고, 도시재생 앵커시설 건축 등에 신진건축사 참여기회를 확대하고, 건축설계 및 감리 품질향상을 위한 건축사 대가기준 현실화 추진하고자하는 계획이다.

- 주요 정책 키워드 : 신진건축사, 건축사 대가기준

□ 지진 안전 관련 정보 공개 및 모니터링 체계 구축

주요과제 '국민생활 안전 강화'의 지진·풍수해 등 자연재해 대비 부분에 해당하며, 신규건축물의 내진설계 적용여부를 건축물대장에 공개하고, 기존건축물의 내진설계 적용여부도 공개토록 추진하며, 건축물표지판(준공식)에도 내진등급 및 내진능력을 표시할 계획이다.

- 주요 정책 키워드 : 건축물내진설계, 건축물표지판 내진등급

□ 내진 보강 및 설계기준 개선

주요과제 '국민생활 안전 강화'의 지진·풍수해 등 자연재해 대비 부분에 해당하며, 건축물의 내진성능 개선을 위하여 신축건축물의 내진규정을 현재 연면적 500㎡이상이거나 3층 이상에서 연면적 200㎡이상이거나 2층이상 건축물로 강화하고, 기존건축물에 대한 인센티브(건폐율·용적률 10%내 완화) 제공을 통해 지진 대응역량 강화할 계획이다. 또한 내진설계기준 개선을 위하여 안전처의 '내진설계기준 공통적용사항'(17.3 공표)에 따라 도로·철도·공항 등 시설물별 내진설계기준 개선을 추진할 계획이다.

- 주요 정책 키워드 : 내진규정강화, 내진보강 인센티브, 시설물 내진설계기준

□ 도시방재체계 구축 및 건축물 안전관리 강화

주요과제 '국민생활 안전 강화'의 생활공간·건설현장·시설물 안전 부분에 해당하며, 건축물 구조안전(시트법), 승강기(승강기관리법), 전기설비(전기사업법), 가스설비(도시

3) 2017년 국토교통부 업무계획에서 도시정책관실, 건축정책관실에 요청하여 해당 정책관실의 주요 업무를 취합하여 정리

가스사업법), 소방설비(소방시설관리법), 정화조(하수도법) 등의 개별법에 따라 유지관리 되는 건축물의 통합관리 체계 마련을 위한 건축물 유지관리법 제정을 추진할 계획이다. 지역별 안전관리 강화를 위하여 건축별을 개정하여 건축 인허가시 내진·화재기준 등 안전 관련 전문적 검토를 지원하는 지역건축안전센터의 도입을 추진할 계획이다. 건축물자재를 불시 점검하는 건축안전모니터링 지속적으로 수행하고, 점검 점검 대상을 건축구조, 건축자재 점검뿐만 아니라 공사단계별 감리업무 수행실태(허가권자 감리자 지정 제도 이행여부 등)로 확대할 예정이다. 마지막으로 증축수리 과정에서 건축기준을 인지하지 못하여 발생하는 불법행위 근절을 위하여 '건축안전 App' 개발할 계획이다.

- 주요 정책 키워드 : 건축물 유지관리법, 지역건축안전센터, 건축안전모니터링, 건축안전 App

□ 건축자산 향유 기반 마련

주요과제 '교통 편리와 생활품격'중 살기 좋은 생활환경 조성 부분에 해당하며, 건축자산의 활성화를 위하여 건축자산 진흥구역 확대를 위한 특례(건폐율 상향 등) 부여를 추진하고 「건축자산 진흥구역 지정·운영 가이드라인」 마련할 계획이다. 또한 한옥의 체계적 관리 지원을 위하여 「한옥 공사비 가이드라인」을 수립(17.7) 하고 「한옥마을 특성을 반영한 지구단위계획 수립 안내서」 제작할 계획이다.

- 주요 정책 키워드 : 건축자산 진흥구역, 한옥마을 지구단위계획

□ 비시가화지역 및 중단 건축물의 체계적 관리

주요과제 '교통 편리와 생활품격'중 살기 좋은 생활환경 조성 부분에 해당하며, 공사중단 건축물의 보상 기준을 마련하고, 취득기준 다양화(경매 등)를 통해 '16년 선도사업(4개소) 성과를 가시화하고, 2차 사업(2개소) 착수할 계획 이다. 또한 금융권·건설사 등 민간참여 확대를 위해 정기설명회 개최, 정비사업 수수료 기준 개선 등 사업 확산 체계 구축할 계획이다.

- 주요 정책 키워드 : 공사 중단 건축물 보상기준, 사업 확산 체계

□ 도시내 공공서비스에 대한 기능 조정

주요과제 '함께 사는 사회'중 사회적 약자 보호 부분에 해당하며, 공공노후건축물 리뉴얼사업의 '16년 착수한 선도사업 성과를 가시화하고, 2차 선도사업에 결합건축 제도 등 다양한 개발 방식 활용을 추진할 계획이다.

- 주요 정책 키워드 : 공공노후건축물 선도사업, 선도사업 개발 방식

□ 스마트기술을 활용한 지속가능한 도시공간 조성

주요과제 '국토교통의 미래' 중 국토교통 신성장동력의 발굴·육성 부분에 해당하며, 제로에너지 건축의 활성화를 위하여 인증제를 시행하고, 온실가스 배출권 및 전력(신재생 에너지) 거래 활성화를 통해 경제적 인센티브 효과 확대할 계획이다. 또한 2020년 공공기관 제로에너지 건축 의무화에 대비하여 시장형 공기업 시설을 조기 의무화하고, 2017년 하반기에 입주예정인 서울 노원 임대주택 실증단지를 대표브랜드화 할 계획이다.

- 주요 정책 키워드 : 제로에너지 건축 인증제, 온실가스 배출권 및 전력 거래, 공공기관 제로에너지 건축, 홍보확산 대표브랜드

□ 건축물의 에너지성능 개선

주요과제 '국토교통의 미래' 중 기후변화 등 환경 이슈에 대한 대응력 제고 부분에 해당하며, 그린리모델링 사업은 노후 건축물을 다수 보유한 중앙 행정기관 등에 기획지원 사업을 우선지원하고, 2018년 이후 시공 기술지원 등을 연계할 계획이다. 민간 그린리모델링 이자 지원사업의 경우 사업자를 50개 이상 추가 지정(현재 373개)하고, 교육 및 기술지원 통한 활성화 유도할 계획이다. 공공건축물 에너지소비량 공개제도는 지자체 및 공공기관 경영평가 등에 반영할 수 있는 평가항목 개발하고, 2018년 이후 반영 추진할 계획이다. 에너지소비총량설계를 확대하여 건축물 설계시 총량평가 기준을 만족하면 사양별 기준을 면제하여 '에너지소비총량제'로 전환을 유도할 계획이며, 총량평가 적용 대상도 단계적 확대 추진할 계획이다. 패시브건축물 수준으로 지역별 단열기준을 강화하여 에너지절감형 건축물 보급 확산할 계획이다.

- 주요 정책 키워드 : 그린리모델링사업, 그린리모델링이자지원사업, 공공건축물 에너지소비량 공개제도, 에너지소비총량제설계, 지역별 단열기준

□ 각종 신고 및 조회 서비스 개선

주요과제 '국토교통의 미래' 중 일하는 방식의 혁신 부분에 해당하며, 주소이동 시 전입 신고로 건축물대장, 부동산등기부 주소변경 일괄처리하고 표시변경등기 전자촉탁을 시행할 계획이다. 또한 건축법 개정을 통하여 본인 또는 사망자 소유 건축물현황을 제공하여 대국민 서비스 개선할 계획이다.

- 주요 정책 키워드 : 건축물대장·부동산등기부, 건축물현황 대국민 서비스

2) 도시정책관실 2017년 업무계획 분석⁴⁾

도시정책관실의 업무는 5개의 국토교통부 주요과제에 포함되어 있으며, 17개의 세부추진과제로 구성되어 있다.

□ 토지이용체계 등 규제 합리화

주요과제 '경제 살리기' 중 국토의 경쟁력 강화 부분에 해당하며, 토지이용규제기법법을 개정하여 지역·지구 신설 및 관리 강화, 지정권자의 주기적 재검토 등 토지이용체계의 간소화 추진할 계획이다. 또한 용도지구와 유사한 지역·지구 상호간의 중첩 지정을 방지하고, 기존 화장시설, 공동묘지, 자연장지, 장례식장을 장사시설로 통폐합 하는 등 도시계획시설 종류 통폐합 등도 추진 예정이다. 또한 관광지에 입지 가능한 건축물 범위를 확대하고, 해안권에 숙박시설 등의 입지를 허용하는 해양관광진흥지구를 도입할 계획이다. 그리고 개발제한구역 내의 대지화 되어 있는 도로용지에 물건적치와 노외주차장 설치, 소규모 도정시설 허용 등 주민불편을 개선할 예정이며, 주민지원사업으로 생활비용 보조를 받는 대상자 범위를 확대할 계획이다.

- 주요 정책 키워드 : 토지이용체계, 도시계획시설 통폐합, 해양관광진흥지구 도입, 개발제한구역 주민불편 개선, 주민지원사업 생활비용 보조

□ 잠재력 있는 국토공간의 체계적 활용

주요과제 '경제 살리기' 중 국토의 경쟁력 강화 부분에 해당하며, 미준공 토지구획정리사업의 현황을 조사하고, 지구별 현안사항·사업여건 분석 등을 거쳐 공공주도형 정상화 방안 마련할 계획이며, 지자체 전수조사 후 수익성·권리관계 등을 분석하여 LH 주도 시범사업을 추진할 예정이다. 장기 미집행 도시계획시설의 집행계획 수립시기를 조기화하고, 10년 내 사업시행 등 의무화 등을 통해 과다한 도시계획시설 지정을 방지할 계획이며, 시설해제 이후 해당 부지의 계획적 개발·관리가 가능하도록 성장관리방안 수립 등 다양한 관리방안 마련할 예정이다.

- 주요 정책 키워드 : 미준공 토지구획정리사업, 장기 미집행 도시계획시설, 도시계획시설 지정

□ 경제기반 도시재생 민간참여 및 금융지원 확대

4) 2017년 국토교통부 업무계획에서 도시정책관실에 요청하여 해당 정책관실의 주요 업무를 취합하여 정리

주요과제 '경제 살리기' 중 국토의 경쟁력 강화 부분에 해당하며, 도시재생 사업 추진에 민·관 네트워크 구축을 위하여 민간업계 유형별(유통사·은행·증권사·건설사 등) 민·관 협의체를 구성·운영하여 도시재생 민간참여 확대할 계획이다. 또한 2016년에 리츠설립 및 기금지원 등이 착수된 천안 외에 청주, 서울, 대구 등에도 도시재생 민간참여사업 확대를 추진할 예정이다.

- 주요 정책 키워드 : 도시재생 민·관 네트워크 구축, 도시재생 민간참여사업

□ 도시와 조화로운 노후산단 재생사업 추진

주요과제 '경제 살리기' 중 국토의 경쟁력 강화 부분에 해당하며, 산단의 쇠퇴원인과 입지 특성 등을 분석하여 유형화하고 도시계획과 연계한 산단재생 고도화 기법 개발하고 재생계획 및 시행계획의 타당성 검증을 위한 평가기준을 마련하여 단계별 평가 실시 등 사업 관리강화를 추진할 계획이다.

- 주요 정책 키워드 : 산단재생 고도화, 산단 재생 평가기준, 사업 관리강화

□ 도시재생 등과 연계한 지역 밀착형 일자리 창출

주요과제 '경제 살리기' 중 청년창업 등 일자리 창출 지원 부분에 해당하며, 도시재생 사업을 통한 사회적경제 육성을 위하여 집수리, 유휴공간 활용 등 도시재생 관련 소규모 회사를 예비사회적기업으로 지정하고 지원을 확대할 예정이다. 빈집·빈점포 활용을 위하여 지자체와 민간의 창의성을 활용하여 지역을 활성화하는 시범사업을 추진하고, 성공모델을 확산시킬 계획이다. 또한 빈집·빈점포 활용에 경험이 풍부한 청년기획자 중심으로 창업아이템 기획, 투자자 모집, 컨설팅 등을 실시하여 지역청년들의 소규모 창업으로 연결시킬 계획이다. 지하상가를 활용하여 입점 가능용도(現 상점, 음식점 등)를 스타트업 사무실, 전시공간, 임대형 창고 등으로 확대하여 일자리 창출을 지원할 계획이다.

- 주요 정책 키워드 : 소규모회사 예비사회적기업, 빈집·빈점포 활용 시범사업, 지역청년 소규모 창업

□ 스마트시티 해외진출 지원체계 및 네트워크 강화

주요과제 '경제 살리기' 중 국토교통산업의 전략적 해외진출 부분에 해당하며, 스마트 시티 관련 해외 석학, 국제기구, 글로벌 선도기업, 지자체 등이 참여하는 '글로벌 스마트 시티 주간'(Korea Smart Cities Week) 개최할 계획이다. 스마트시티 해외 네트워크 강화를 위하여 중동·중남미 등 주요 진출 대상국과 MOU를 체결하고, 바르셀로나 등 선진도

시, WB 등 국제기구와의 협력도 강화할 예정이다. 스마트시티 통합정보망을 통하여 해외사업 정보를 일괄 제공하고, 참여 의사가 있는 공공·민간기업간 정보 교류를 위해 '통합 정보망' 운영을 내실화할 계획이다. 스마트시티 추진단 소속 3개 분과위원회(도시인프라/에너지·환경/ICT 솔루션) 활동의 내실화로 민간기업의 쌍방향 참여 유도할 예정이며, 해외 외빈 등에 대한 국내 기술 홍보를 위해 투어프로그램 대상지를 확대하고, 전용 홍보관 운영(서울 강남구) 등을 추진할 계획이다.

- 주요 정책 키워드 : 글로벌 스마트 시티 주간, 주요 진출 대상국 MOU, 선진도시·국제기구, 스마트시티 통합정보망, 스마트시티 추진단 분과위원회, 국내 기술 홍보, 전용 홍보관

□ 용산공원 조성 및 민영공원 도입 등 도시공원 활성화

주요과제 '교통 편리와 생활품격' 중 살기 좋은 생활환경 조성 부분에 해당하며, 용산공원을 현재·미래세대의 휴식공간 및 자연생태공간으로 거듭날 수 있도록 내실 있는 조성계획을 추진할 예정이다. 조경기본계획('17.6)에 따른 조경지원센터 등 지정기준 마련을 통하여 민간소유 공원에 공원시설과 별도의 수익시설의 추가설치를 허용하는 민영공원제도 도입을 추진할 계획이다.

- 주요 정책 키워드 : 용산공원 조성계획, 민영공원제도

□ 비시가화지역의 계획적 개발 및 관리

주요과제 '교통 편리와 생활품격' 중 살기 좋은 생활환경 조성 부분에 해당하며, 비시가화지역의 그 동안의 개발 실태, 문제점 등을 전반적으로 점검하고, 국토계획법 시행령 개정을 통하여 비시가화지역의 체계적 관리방안 마련할 계획이다. 또한 성장관리방안 등 기존 제도 활용도 확대(모니터링 강화, 절차간소화 등) 및 개발행위 기준 강화 등을 통해 계획적인 개발 유도할 계획이다.

- 주요 정책 키워드 : 비시가화지역 실태 점검, 비시가화지역 관리방안

□ 각종 신고 및 조회 서비스 개선

주요과제 '국토교통의 미래' 중 일하는 방식의 혁신 부분에 해당하며, 국민 편의성 제고를 위해 인터넷 기반 개발행위 인허가시스템의 구축하고, 전국으로 확대 보급할 계획이다.

- 주요 정책 키워드 : 인터넷 기반 개발행위 인허가시스템

□ 달동네 등 취약지역 지원 확대

주요과제 ‘함께 사는 사회’중 소외지역 지원 확대 부분에 해당하며, 달동네 등 주거취약 지역 환경개선 및 일자리·복지 등 자활을 지원하는 새뜰마을사업 확대할 계획이다. 또한 취약계층에 대한 집수리 봉사, 재능기부 등 공공기관·민간기업·봉사단체 등의 사회공헌 사업과의 연계를 확산할 계획이다.

- 주요 정책 키워드 : 새뜰마을사업 확대, 사회공헌사업 연계

□ 지역밀착형 도시재생 사업 활성화

주요과제 ‘함께 사는 사회’중 소외지역 지원 확대 부분에 해당하며, 도시재생사업 성과 확산을 위하여 선도사업의 성과를 가시화하고, 단계별 추진과정과 노하우를 백서화하여 타 지자체로 확산을 유도할 계획이다. 또한 2차지구에 대한 활성화모형을 확정하여 사업을 본격 추진하고, 쇠퇴도 개선효과 분석을 위한 성과측정 지표도 개발할 예정이다. 문체부와 함께 쇠퇴지역(5개소)에 찾아가는 문화컨설팅을 시행하고, 법무부 등과는 범죄예방사업 연계지원 등 부처 간 협업사업을 지속적으로 발굴할 계획이다. 도시재생활성화계획 수립 시 타 법에 따른 의제사항을 확대하고, 대도시는 자체 수립권한을 부여(50만 미만은 도지사 승인)하는 등 절차 간소화를 추진할 계획이다. 지역 쇠퇴분석 고도화를 위해 도시재생정보체계를 개편하고, 총괄사업관리자 제도 도입을 통해 효과적 사업추진을 지원할 예정이다. 공동체 활성화를 위하여 도시재생사업 지역 내 젠트리피케이션 관련 가이드라인을 마련하고, 공동이용시설에 대한 사용료 경감 근거를 신설할 예정이다.

- 주요 정책 키워드 : 도시재생사업 선도사업 성과, 2차지구 사업, 성과측정 지표, 부처 간 협업사업, 도시재생활성화계획 수립 간소화, 도시재생정보체계 개편, 총괄사업관리자 제도, 젠트리피케이션 가이드라인

□ 도시 내 공공서비스에 대한 기능 조정

주요과제 ‘함께 사는 사회’중 사회적 약자 보호 부분에 해당하며, 포용적 도시의 조성을 위하여 국공립어린이집 등 사회취약계층을 위한 공공서비스 제공시설에 도시계획 특례(용적률 등)를 확대하고, 포용성 관련 사회경제적 지표를 개발하여 도시지속성 평가에 반영할 계획이다. 또한 지역별 특성(노인, 외국인 비율 등)에 맞는 포용 정책이 확대될 수 있도록 포용도시 조성의 원칙과 방향 등을 도시계획에 포함시킬 예정이며, 차별 없는 공공서비스를 지원할 수 있는 도시 공간(건강도시, 노인도시 등)을 조성할 수 있는 인센티

브제도 등을 마련할 예정이다.

- 주요 정책 키워드 : 사회취약계층 공공서비스 제공시설, 포용성 지표 개발, 포용도시 도시계획, 공공서비스 인센티브제도

□ 계획 수립 과정에 주민 참여 확대

주요과제 '함께 사는 사회' 중 갈등관리 및 지역·주민의 참여 확대 부분에 해당하며, 지역 사회의 공동체 복원 및 활성화를 지원할 수 있도록 도시계획 수립 과정에 주민참여를 확대하고, 공청회 방식이외에 주민 주도의 도시계획(경관·역사·방재 등)을 유도할 수 있도록 참여 방안(온-오프라인, 주민협의체 구성 등)을 다양화할 계획이다.

- 주요 정책 키워드 : 도시계획 주민참여, 주민 참여 방안 다양화

□ 도시방재체계 강화 및 건축물 안전 모니터링 사업 확대

주요과제 '국민생활 안전' 중 생활공간·건설현장·시설물 안전 부분에 해당하며, 재해취약성분석 지침 개정을 통하여 재해취약성 분석, 도시복원력, 재해정보공유체계 도입 등 종합적 방재체계 구축 추진할 계획이다.

- 주요 정책 키워드 : 종합적 방재체계

□ 스마트기술을 활용한 지속가능한 도시공간 조성

주요과제 '국토교통의 미래' 중 국토교통 신성장동력의 발굴·육성 부분에 해당하며, 스마트시티 조성을 위하여 도시 플랫폼 하에서 에너지, 물관리, IoT 등 요소기술을 연계하여, 국민 정책 체감도를 높일 수 있도록 체계를 개편하고, 조성(건설+운영)과 산업진흥을 포함하도록 법체계를 개편하고, 국내 도시에 대한 스마트시티 수준 진단 및 인증제도 도입할 계획이다. 특화단지 구축(4개), 서비스 경진대회, 전문인력 양성 등으로 新산업을 지원하고, 스마트시티 안전망 5대 서비스도 확산보급(6개 지자체) 할 계획이다.

- 주요 정책 키워드 : 스마트시티 조성 플랫폼 체계, 법체계, 수준 진단 및 인증제도, 스마트시티 신산업, 스마트시티 안전망 서비스

□ 분야별 핵심 R&D 선정·추진

주요과제 '국토교통의 미래' 중 미래를 대비하는 국토교통 계획 수립 부분에 해당하며, 신흥국 스마트시티 시장 진출을 위해 기존 U-City 고도화 R&D에 반영하여 단기적으로 실용화 가능한 패키지를 개발하는 해외진출 플랫폼 구축 R&D 시행할 계획이다.

- 주요 정책 키워드 : 스마트시티 해외진출 플랫폼

3) 건축·도시 분야 2017년 정책 종합

건축정책관실의 2017년 업무계획을 내용분석한 결과 10개의 세부추진과제는 28개의 세부내용으로 분석되었다. 28개의 세부내용을 국토교통부 2017년 주요과제별로 살펴보면 경제살리기 2건, 안전강화 9건, 교통편리/생활품격 4건, 함께 사는 사회 2건, 국토교통의 미래 6건, 환경이슈 대응력제고 5건으로 나타났다. 28개의 세부내용 중 안전강화가 9건으로 가장 높게 나타나 업무계획을 토대로 해석하면 2017년에 건축분야에서 가장 많은 정책은 건축물 안전강화 분야라는 것을 확인할 수 있다.

건축정책관실의 2017년 업무계획 세부내용을 정책단계로 구분해 보면 문제형성 단계에 해당하는 정책세부내용은 없으며, 정책형성 및 대안채택 단계는 13건, 정책집행 단계는 12건, 정책평가 단계는 3건으로 분석되었다.⁵⁾ 따라서 2017년 3월 현재 건축분야 정책은 정책형성 및 대안채택과 정책집행 단계에 집중되어 있음을 확인할 수 있다.

[표 4-1] 건축분야 정책 내용 분석(2017년 건축정책관실 업무계획)

세부정책	세부실행내용	정책단계*	주요주제
전문인력의 육성 통한 일자리 창출	-신진건축사 발굴(공모전 등)	c	경제
	-건축사대가기준 현실화	b	살리기
지진 안전 관련 정보 공개 및 모니터링 체계 구축	-건축물 내진설계적용여부 공개	c	국민생활
	-건축물표지판 내진등급 및 내진능력 표시	c	안전 강화
내진 보강 및 설계기준 개 선	-신축건축물 내진규정 강화	c	국민생활
	-기존건축물 내진보강 시 인센티브 제공	c	안전 강화
	-시설물 내진설계기준 개선	c	
도시방재체계 구축 및 건축물 안전관리 강화	-건축물유지관리법 제정	b	국민생활
	-지역건축안전센터 도입	b	안전 강화
	-건축안전모니터링수행 및 점검대상 확대	d	
	-건축안전App개발	b	
건축자산 향유 기반 마련	-건축자산진흥구역 지정운영가이드라인 마련	b	교통 편리
	-한옥마을특성반영 지구단위계획수립안내서제작	b	생활품격
비시가화지역 및 중단 건축물의 체계적 관리	-공사중단건축물 보상기준마련	b	교통 편리
	-사업확산체계 구축	b	생활품격

5) 세부추진 내용에 따라 주제와 내용이 다양한 과제가 구성되어 있는 경우가 많아 일부는 정책형성 및 대안채택 단계에 있고, 일부는 정책집행 단계에 있는 경우도 나타났다. 세부추진 내용을 분석하여 보다 가까운 쪽의 정책의 단계로 구분하였다.

세부정책	세부실행내용	정책단계*	주요주제
도시내 공공서비스에 대한 기능 조정	-공공노후건축물 선도사업 성과가시화	d	함께 사는 사회
	-선도사업에 다양한 개발방식 활용추진	b	
스마트기술을 활용한 지속가능한 도시공간 조성	-제로에너지건축 인증제 시행	c	국토교통 미래
	-온실가스배출권 및 전력거래 인센티브효과확대	d	
	-공공기관 제로에너지건축 시장형공기업 조기의무화	c	
	-홍보확산을위한 대표브랜드 발굴	b	
건축물의 에너지성능 개선	-그린리모델링사업 기획지원사업 우선지원	c	환경 이슈 대응력제고
	-그린리모델링이자지원사업 사업자추가지정 및 교육· 기술지원	c	
	-공공건축물 에너지소비량공개제도 평가항목개발	b	
	-에너지소비총량제설계 확대	b	
	-지역별 단열기준 강화	b	
각종 신고 및 조회 서비스 개선	-전입신고시 건축물대장·부동산등기부주소변경 일괄 처리	c	국토교통 미래
	-소유건축물현황 대국민서비스 개선	c	

※ a.문제형성, b.정책형성 및 대안채택, c.정책집행, d. 정책평가

도시정책관실 2017년 업무계획의 17개 세부추진과제를 분석해 보면 총 53개의 세부내용으로 분석되었다. 53개의 세부내용을 국토교통부 2017년 주요과제별로 살펴보면 경제살리기 24건, 안전강화 1건, 교통편리/생활품격은 5건, 함께 사는 사회 17건, 국토교통의 미래 6건으로 나타났다. 54개의 세부내용 중 절반에 가까운 24건이 경제살리기에 해당하는 정책으로, 2017년 도시분야에서 중점적으로 추진하는 정책은 경제살리기 분야라는 것을 확인할 수 있다.

도시정책관실의 2017년 업무계획 세부내용을 정책단계로 구분해 보면 문제형성 단계에 해당하는 정책세부내용은 건축분야와 동일하게 없으며, 정책형성 및 대안채택 단계는 27건, 정책집행 단계는 24건, 정책평가 단계는 2건으로 분석되었다.⁶⁾ 따라서 2017년 3월 현재 도시분야 정책 역시 정책형성 및 대안채택과 정책집행 단계에 집중되어 있음을 확인할 수 있다.

6) 세부추진 내용에 따라 주제와 내용이 다양한 과제가 구성되어 있는 경우가 많아 일부는 정책형성 및 대안채택 단계에 있고, 일부는 정책집행 단계에 있는 경우도 나타났다. 세부추진 내용을 분석하여 보다 가까운 쪽의 정책의 단계로 구분하였다.

[표 4-2] 도시분야 정책 내용 분석(2017년 도시정책관실 업무계획)

세부정책	세부실행내용	정책단계*	주요주제
토지이용체계 등 규제 합리화	-토지이용체계 간소화	b	경제 살리기
	-도시계획시설 종류 통폐합	b	
	-해양관광진흥지구 도입	c	
	-개발제한구역 주민불편 개선	b	
	-주민지원사업 생활비용 보조 대상자 범위 확대	c	
잠재력 있는 국토공간의 체계적 활용	-미준공토지구획정리사업 현황조사	b	경제 살리기
	-공공주도형 정상화 방안 마련	b	
	-장기미집행도시계획시설 집행계획 수립시기 조기화	c	
	-과다한 도시계획시설지정 방지	b	
경제기반 도시재생 민간참 여 및 금융지원 확대	-도시재생 민·관네트워크 구축	c	경제 살리기
	-도시재생 민간참여사업 확대	c	
도시와 조화로운 노후산단 재생사업 추진	-산단재생 고도화기법 개발	b	경제 살리기
	-산단재생 평가기준마련 및 사업관리 강화	b	
도시재생 등과 연계한 지역 밀착형 일자리 창출	-소규모회사 예비사회적기업 지정 및 지원	c	경제 살리기
	-빈집·빈점포활용 시범사업 추진	b	
	-지역 청년 소규모창업 확산	b	
	-지하상가 입점가능 용도 확대	c	
스마트시티 해외진출 지원체계 및 네트워크 강화	-글로벌스마트시티주간 개최	c	경제 살리기
	-주요진출대상국 MOU 체결	c	
	-선진도시·국제기구 협력 강화	b	
	-스마트시티 통합정보망 개선	c	
	-스마트시티추진단분과위원회 활동 내실화	c	
	-스마트시티 국내기술 홍보대상자 확대	c	
	-스마트시티 전용홍보관 운영	c	
용산공원 조성 및 민영공원 도입 등 도시공원 활성화	-용산공원 조성계획 추진	c	교통 편리 생활품격
	-민영공원제도 도입	c	
비시가화지역의 계획적 개발 및 관리	-비시가화지역 실태점검	b	교통 편리 생활품격
	-비시가화지역 관리방안 마련	b	
각종 신고 및 조회 서비스 개선	-인터넷 기반 개발행위 인허가시스템 구축	c	국토교통 미래
달동네 등 취약지역 지원 확대	-새들마을사업 확대	b	함께 사는 사회
	-사회공헌사업과의 연계	b	
지역밀착형 도시재생사업 활성화	-도시재생사업 선도사업성과 가시화	d	함께 사는 사회
	-도시재생사업 백서 제작	d	
	-도시재생 2차지구사업 본격 추진	c	

세부정책	세부실행내용	정책단계*	주요주제
	-성과측정 지표 개발	b	
	-부처간 협업사업 진행 및 발굴	c	
	-도시재생 활성화계획 수립 간소화	b	
	-도시재생 정보체계 개편	b	
	-홍괄사업관리자제도 도입	c	
	-젠트리피케이션 가이드라인 마련	b	
도시 내 공공서비스에 대한 기능 조정	-사회취약계층 공공서비스 제공시설 도시계획특례 확대	c	함께 사는 사회
	-포용성관련 사회경제적지표 개발	b	
	-포용도시조성 원칙·방향 도시계획에 포함	c	
	-차별없는 공공서비스 인센티브제도 마련	b	
계획 수립 과정에 주민 참여 확대	-도시계획 수립과정 주민참여 확대	b	함께 사는 사회
	-주민참여 방안 다양화	b	
도시방재체계 강화 및 건축 물 안전 모니터링 사업 확대	-종합적 방재체계 구축	b	국민생활 안전 강화
스마트기술을 활용한 지속 가능한 도시공간 조성	-스마트시티 조성플랫폼 체계개편	c	국토교통 미래
	-스마트시티 관련 법체계개편	b	
	-스마트시티 수준진단 및 인증제도 도입	b	
	-스마트시티 신산업 지원	b	
	-스마트시티 안전망서비스 확산·보급	c	
분야별 핵심 R&D 선정추진	-스마트시티 해외진출 플랫폼구축 R&D시행	c	국토교통 미래

※ a.문제형성, b.정책형성 및 대안채택, c.정책집행, d. 정책평가

2017년 국토교통부의 업무계획에 기초하여 건축·도시 분야 정책을 종합하여 보면 먼저 양적으로는 세부실행내용을 기준으로 건축분야는 10개 세부추진과제에 28개 세부추진 내용으로 분석되었고, 도시분야는 17개의 세부추진과제에 53개 세부추진내용으로 분석되었다. 각각의 세부추진내용은 실제 실행하는 개별정책으로 볼 수 있어, 건축·도시 분야를 종합하면 2017년도에 추진 예정인 정책은 총 81개로 나타났다.

정책 내용적으로는 추진 예정 정책을 국토교통부의 2017년 추진 주요주제로 정책내용을 기준으로 종합하면, 건축분야의 경우 경제살리기 2건, 안전강화 9건, 교통편리/생활 품격 4건, 함께 사는 사회 2건, 국토교통의 미래 6건, 환경이슈 대응력제고 5건, 도시분야는 경제살리기 24건, 안전강화 1건, 교통편리/생활품격은 5건, 함께 사는 사회 17건, 국토교통의 미래 6건으로 나타났다. 건축분야의 경우 건축물의 안전강화 정책, 도시분야의 경우 경제살리기 정책에 초점이 맞춰져 있는 것을 확인할 수 있다.

추진 예정 정책을 정책단계로 살펴보면 건축분야는 정책형성 및 대안채택 단계 13건, 정책집행 단계 12건, 정책평가 단계 3건으로 나타났다. 도시분야의 경우 정책형성 및 대안채택 단계는 27건, 정책집행 단계는 24건, 정책평가 단계는 2건으로 분석되었다. 건축·도시분야를 종합하면 정책형성 및 대안채택 단계 40건, 정책집행 단계 36건, 정책평가 단계 5건으로 나타났다.

2. 빅데이터 기반 건축·도시 정책개발 체계 수립

정책개발(policy development)이란 정책문제를 해결하기 위해 정책 수단을 개발하는 것을 말한다. 즉 정책개발이란 정책문제와 목표들을 명확히 규정하고 모형을 개발한 다음 각 대안들의 비용·편익 및 효과 등의 비교분석을 통해 정책의 효율성·실행가능성·대표성·논리성 등을 향상시키려는 일련의 활동을 말한다.⁷⁾

정책과정 중 정책개발의 범위를 정리하면 정책과정의 5단계 중 정책집행 이전 단계인 문제형성, 정책형성, 대안채택 3단계에 해당된다. 따라서 본 절에서는 정책의 과정에서 앞에서 분석한 빅데이터의 활용과 건축·도시 분야의 정책과 연계하여 정책을 개발하는 일련의 체계를 수립한다.

1) 문제형성 단계 - 정책의제 선정

정책과정 중 문제형성 단계에서는 특정 주제·이슈에 대한 대중의 의견과 여론을 파악하기 위한 여론 모니터링이 필수적이다. 전통적으로는 이를 위하여 설문조사 등을 통하여 여론을 파악하여 왔다. 소셜 빅데이터, 로그 빅데이터 등의 여론 모니터링을 통하여 문제를 탐색하는 사례가 많이 나타나고 있다.

빅데이터를 활용한 증거기반 정책개발을 위해서는 정책의제를 선정하는 초기 단계부터 빅데이터를 활용하는 것이 중요하다. 이는 향후 후속 단계에서 동일한 데이터를 활용하여 분석 알고리즘과 분석의 해상도를 달리하여 대안개발과 정책평가에 이르기까지 동일 프로세스를 활용할 수 있기 때문이다.

7) 이종수(2009), “네이버 지식백과: 정책개발”, 행정학사전, 대영문화사.
<http://terms.naver.com>(검색일:2017.06.21)

건축·도시 분야 2017년 업무계획 분석결과 문제형성 단계에 있는 정책은 없는 것으로 파악되었다. 이는 업무계획의 특성상 관련 기본계획과 여론 동향, 정책결정자의 의지 등에 의하여 이미 도출되어 있는 경우가 많기 때문인 것으로 판단된다. 그러나 기본계획의 수립, 여론의 동향 파악 등에 빅데이터의 활용가치가 높은 것으로 사례 분석을 통하여 확인한 만큼 건축·도시 분야에서 정책의제 선정과정에 빅데이터의 활용가치가 낮은 것으로 확대해석은 경계하여야 한다.

□ 소셜 빅데이터

소셜 빅데이터는 여론 모니터링을 위한 설문조사를 대체 또는 보완할 수 있는 도구로 많이 활용되고 있으며, 다양한 이용자의 의견들을 직접 수렴하여 여론을 파악하기에 좋은 데이터이다. 국내외 사례에서도 확인한 것처럼 영화의 흥행, 상품 구매자의 소비 수요 패턴, 특정 지역의 관광지의 여론, 특정 키워드의 SNS 동향 등 다양한 이슈에 대한 동향 파악과 문제점 발굴에 활용되고 있다.

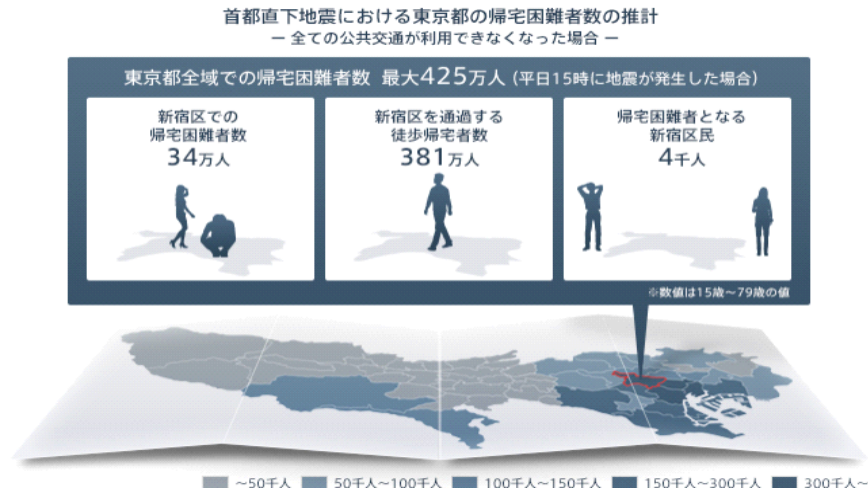
소셜 빅데이터 분석은 비정형 텍스트 분석 방법론을 적용한 연관 키워드의 시계열 분석을 주로 사용한다. 소셜 빅데이터를 활용한 여론 모니터링의 가장 큰 장점은 여론의 변화를 실시간으로 모니터링 할 수 있어 이를 통한 정책 대응을 신속히 할 수 있다는 것이다. 그러나 수집하는 소셜 빅데이터의 이용자의 성별, 연령별 계층 분포에 따라 편향된 결과가 도출될 가능성이 있어 데이터 수집 채널을 다양화할 필요가 있다. 트위터 등의 전통적인 SNS 이외에 블로그 및 뉴스 댓글 등 다양한 매체로부터 소셜 데이터 수집 및 활용이 가능하며, 우리나라의 경우 SNS 보다 상대적으로 포털 이용자가 더 폭넓은 계층을 아우르는 경향이 있다.

□ 로그 빅데이터

로그 빅데이터의 대표적인 유형은 사람들의 이동에 대한 로그인 휴대전화와 신용카드·교통카드 데이터와 사람들의 관심사에 대한 로그인 인터넷 검색엔진 등의 검색데이터로 볼 수 있다.

사람들의 통행량을 실시간으로 정확한 지점까지 확인할 수 있는 휴대전화 등의 빅데이터는 건축·도시 정책에 활용가치가 높다. 휴대폰 빅데이터는 통신사에서 제공하는 활동 인구 데이터를 정책 수요 도출 및 모니터링에 활용하는 데이터로, GPS를 이용한 상세 이동 데이터를 건축·도시 현황 파악에 활용한 사례도 존재한다. 특히 신용카드 빅데이터

와 연계하면 가공 정도에 따라 단순 문제 도출부터 정책 대안 개발까지 활용가치가 높다. 그러나 휴대전화나 신용카드의 데이터는 구득이 어렵고, 개인정보 보호를 위한 익명화의 수준에 따라 분석의 해상도가 달라지며, 높은 비용을 지불해야 하는 등 실제 활용하기 위해서는 사전 준비가 필요하다.



[그림 4-1] 방재계획 수요 도출 휴대폰 빅데이터 활용 사례

출처 : https://www.nttdocomo.co.jp/corporate/disclosure/mobilespatial_statistics/experiment/index.html

검색 빅데이터는 검색엔진에 사람들이 입력하는 검색어의 추의를 통하여 대중의 주요 관심사를 손쉽게 알아볼 수 있는 장점이 있다. 특히 구글, 네이버 등 국내 검색 점유율이 높은 검색엔진에서 특정 검색어 통계를 알 수 있도록 시스템을 제공하고 있어 특정 정책에 대한 여론의 추의를 확인하는 데 용이하다. 구글의 경우 이러한 검색 트렌드를 가지고 독감 유행을 예측하는 등 실시간 모니터링을 통하여 동향 변화를 감지할 수 있어 문제 형성 단계에 유용하게 활용할 수 있다.

[표 4-3] 문제형성 단계 빅데이터 활용 예시

단계별 데이터		목적	방법	결과
소셜 데이터	트위터	영화 흥행 여부 판단	감성분석(오피니언 마이닝)	영화 홍보전략 수립, 흥행 실적 예측
	트위터	관광 관련 여론 동향 분석	연관키워드, 키워드 빈도	경기도 관광 정책 방향 설정
	트위터	상품 관련 온라인 동향 분석	빈도분석, 연관어, 감정분석	기업 홍보 방향 설정에 활용
로그 데이터	검색엔진로그	독감 확산 조기 경보체계 마련	관련 검색어, 시간대, 위치 정보를 활용하여 트렌드 분석	시간 및 지역별 독감 유행 정보 제공

단계별 데이터		목적	방법	결과
정형 데이터	에너지 공급 소비정보, 건축물정보	에너지 관련 정책 개발	공간분석	에너지 관련 정책 대안 발굴
	재난재해관련 공공데이터	도시 재난재해 통합 관리	재해 통합 정보 시스템 구축	통합관리로 재해 감소
	건축물정보, 화재정보 등	소방점검 우선순위 도출	데이터 연계 및 통합 및 분석 알고리즘 개발 및 적용	소방업무 효율화 및 효과적인 화재 대응
	민원데이터	민원의 사전예방 및 대책 마련	키워드별 추이 분석, 의미기반 클러스터링 분석	민원정보 분석 시스템으로 민원 대응 및 민원지도 제공

2) 정책형성 단계 - 정책대안 개발

정책형성 단계의 빅데이터 활용은 다양한 데이터를 결합 및 연계하여 정책 대안을 만드는 것에 초점이 맞춰져 있다. 기존의 정책대안은 전문가 의견이나 정책입안자의 견해에 따라 개발되는 경우가 많이 있다. 이러한 전문가의 의견은 주관적인 경우가 많아 정책대안이 실효적이지 못한 경우가 있다. 이에 빅데이터를 이용한 정책 대안 개발은 보다 정량적이고 객관적인 데이터에 기반을 둔다.

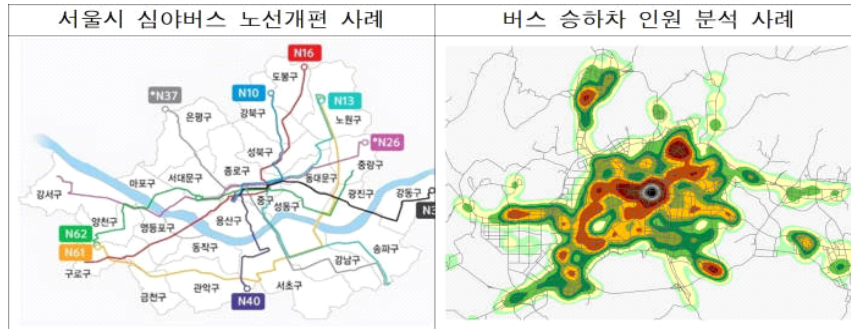
정책 대안 개발은 정형데이터의 활용 사례가 많으며, 이는 기존 정보를 보유하고 있는 기업·기관에서 보유 및 구축하고 있는 데이터를 기반으로 빅데이터 분석 기술을 도입하여 새로운 대안을 만들어 내는 경우가 많기 때문이다

건축·도시 분야 2017년 업무계획 분석결과 정책형성 및 대안채택 단계에 있는 정책은 세부 실행 내용 81건 중 40건이다. 해당하는 세부 실행 내용의 키워드를 보면 ‘기준 현실화’, ‘법제정’, ‘센터 도입’, ‘가이드라인 마련’, ‘안내서제작’, ‘App개발’, ‘기준 마련’, ‘활용 추진’, ‘브랜드 발굴’, ‘체계 구축’, ‘항목 개발’, ‘설계 확대’, ‘기준 강화’, ‘체계 간소화’, ‘종류 통합’, ‘주민불편 개선’, ‘현황 조사’, ‘기법 개발’, ‘관리 강화’, ‘사업 추진’, ‘협력 강화’, ‘사업 확대’ 등을 확인할 수 있다. 이는 어떤 대상을 ‘개발’, ‘제작’, ‘확대’, ‘강화’, ‘개발’, ‘조정’ 하는 것이며, 이에 대한 구체적인 방법론(대안)은 개발 예정에 있는 상황이다. 따라서 대상이 확정이 되어 있는 경우이기 때문에 해당 대상과 관련한 데이터를 구축·활용하는 것이 중요하다.

□ 소셜 빅데이터

문제형성 단계에서 소셜 빅데이터가 여론의 동향을 측정하는데 활용되었다면, 정책형성 단계에서 소셜 빅데이터는 연관 키워드 분석 등 분석을 보다 고도화 하여 문제에 대한 대안을 만드는데 활용할 수 있다. 경기도 관광 활용사례에서 연관 키워드를 활용하여 경기도 관광의 새로운 코스(대안)을 발굴하는 것을 예로 들 수 있다.

□ 로그 빅데이터



[그림 4-2] 서울시 및 대전시의 교통카드 빅데이터 활용 사례

출처 : <http://www.etnews.com/20160525000173>

휴대폰 빅데이터는 공간적 계획이 중요한 건축·도시 정책개발에 활용도가 높고, 활동인구 분포를 고려한 건축·도시 정책 대안 개발에 중요한 역할을 제공한다. 교통카드 및 신용카드 데이터 또한 시간대별 이동패턴 및 업종별 경제활동을 위치정보와 연계하여 분석할 수 있는 장점이 있다. 특히 특정 지역의 유동 인구와 소비 정보를 실시간으로 파악할 수 있어 다양한 정책 대안을 만들 수 있다. 서울시 심야버스와 SK T맵, 서울시 관광 등 로그데이터를 활용하여 다양한 대안을 만들어 활용한 사례이다.

□ 미디어 빅데이터

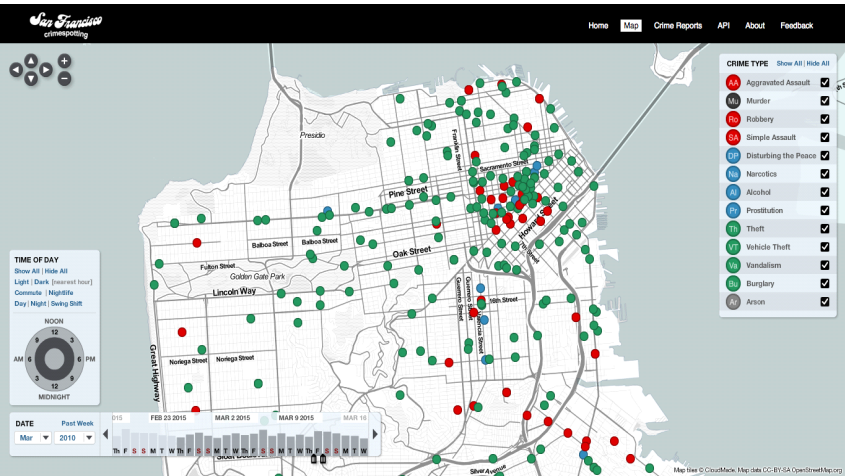
미디어 데이터 중 선거공약 빅데이터는 대상 지역의 요구를 수렴한 자료로써 선거 유형에 따라 국민 또는 지역주민의 정책 이슈와 수요 파악이 용이하다. 특히 지역의 현안에 대한 다양한 대안을 제시하고 있는 경우가 많아 정책대안 개발에 바로 활용할 수 있다는 측면에서 유용한 빅데이터이다.

□ 정형 빅데이터

기존 빅데이터 활용 사례에서 확인할 수 있듯이 정책형성 단계에서 가장 많이 활용하고 있는 빅데이터는 기존 보유하고 있는 정형데이터이다. 정형 빅데이터는 기존 기업·기관 등에서 보유하고 있는 데이터이지만 이를 분석 및 활용할 기술이 부재로 단순 활용에 그쳐왔다. 그러나 빅데이터 분석기술이 발전하고 증거기반의 정책대안개발에 대한 사회적 요구가 높아지면서 이를 활용한 사례가 크게 증가하고 있다.

정형 빅데이터는 기본적으로 해당 기업·기관에 의하여 관리된 데이터이기 때문에 비정형·반정형 빅데이터에 비해 가공이 용이하다는 장점이 있으며, 또한 자료 수집이 편리

하다. 그러나 통상 하나의 정형데이터를 활용하는 경우보다는 서로 다른 정형데이터를 결합하여 사용해야 하는 경우가 많아 이를 연계하는 방안을 찾는 것이 중요하다. 위치별 범죄발생자료(미국), 의료 데이터(웹포인트), 도시 재해 관련 데이터(리우데자네이루), 지점별 매출 자료(파리바게트) 등 활용되는 정형데이터는 대상이 되는 사업과 정책에 다양하다. 이러한 정형 빅데이터를 위치기반으로 가공하여 지도에 표출하여 정책대안을 개발하는 사례가 많으며, 특히 지도상에 표시할 수 있는 모든 데이터를 특정 지점 기반으로 분석하여 대안을 도출하는 공간 분석은 건축·도시의 정책개발에 활용도가 크다.



[그림 4-3] 미국 샌프란시스코 범죄지도
출처: <http://stamen.com/work/crimespotting>

[표 4-4] 정책형성 단계 빅데이터 활용 예시

단계별 데이터		목적	방법	결과
소셜데이터	트위터	여론동향 분석	네트워크 분석, 군집 분석	관광 콘텐츠 개발
로그 데이터	택시 승하차, 휴대전화 데이터	심야 대중교통 수요 측정	공간분석, 승하차 공간정보화, 심야시간 위치 공간분석	심야버스 노선 대안 개발
	네비게이션 데이터	실시간 혼잡구간 정보 제공	공간분석 자체알고리즘 개발	실시간 빠른 길 서비스 개발
	휴대전화, 신용카드 데이터	외국인 서울 관광 루트 개발	휴대전화, 신용카드 데이터	외국인 관광 루트 대안 개발
정형 데이터	범죄현황 데이터 (범죄 기록, 긴급신고 등)	범죄 우발지역을 파악하여 예방	범죄 데이터의 공간정보화 및 데이터 시각화	범죄 다발 지역 공개
	건강보험자료, 환자 정보	환자 상황에 맞는 맞춤형 진료 제공	인공지능 시스템으로 최적의 진단 및 치료 가이드라인 제시	2억 페이지에 해당하는 자료를 검색, 3초 안에 결과 제시
	재난재해관련 공공데이터	도시 재난재해 통합 관리	재해 통합 정보 시스템 구축	재난 대응 시뮬레이션 개발
정형 +로그	기상정보 + 판매정보	날씨에 따른 판매량 패턴을 파악 공급량 조절	통계적 분석으로 날씨 판매지수 개발	판매수요 예측 및 생산 관리

3) 대안채택 단계 - 최적 대안 선정

정책과정 중 대안채택 단계는 정책형성 과정에서 도출된 다양한 대안 중 최적의 대안을 찾는 것이다. 따라서 대안채택 단계의 빅데이터 활용은 이전의 단계보다 알고리즘 개발을 통하여 대안을 채택하는 등 빅데이터를 보다 분석적으로 접근하는 경우가 많다. 또한 빅데이터를 구축·분석의 경제성과 효과성을 높이기 위하여 정책형성 단계의 다수 대안 개발과 동시에 수행하는 것이 바람직하다.

빅데이터를 활용하지 않아도 통상 건축·도시 정책개발에 있어서도 대안채택 단계를 별도로 구성하는 경우는 많지 않다. 앞서 건축·도시 분야 2017년 업무계획 분석결과에서도 확인한 것과 같이 대안을 개발하는 과정에 최종대안을 선정하는 과정도 포함한다.

□ 로그 빅데이터

휴대전화 빅데이터는 대안채택 단계에서 활용은 유동인구 정보를 분석을 기반으로 알고리즘을 개발하여 다수의 대안 중 최적의 대안을 찾는 것에 다수 활용되고 있다. 기존 사례를 보면 다수의 개발된 심야 버스 노선 중 최종적으로 어느 노선이 더 합리적인 것인가를 시뮬레이션 하는 데이터로 활용되었고, 다수의 관광 코스 중 어떠한 코스를 최종적으로 개발 및 안내할 것인가를 결정하는 데이터로도 유용하게 활용되었다.

□ 정형 빅데이터

해외 사례에서 살펴본 범죄 빅데이터 활용 사례와 같이 단순히 범죄 데이터를 시각화하여 지도에 표출하고 이를 통하여 범죄다발지점을 도출하는 것은 정책형성 단계에 대안 개발에 도움을 주는 것이라 할 수 있으며, 이러한 빅데이터 기반에 범죄예측 알고리즘을 개발하여 향후 범죄가 발생할 지점을 예측하는 것은 대안채택 단계의 빅데이터 활용이라 할 수 있다. 의료 빅데이터를 기반으로 특정 증상의 환자에 대한 맞춤형 진단을 하는 것 역시 비슷한 맥락이다.

[표 4-5] 대안채택 단계 빅데이터 활용 예시

단계별 데이터	목적	방법	결과
로그 데이터	택시 승하차, 휴대전화 데이터	심야 대중교통 수요 측정	자체 분석 알고리즘 개발
	휴대전화, 신용카드 데이터	외국인 서울 관광 루트 개발	휴대전화, 신용카드 데이터
정형 데이터	범죄현황 데이터 (범죄 기록, 긴급신고 등)	범죄 우발지역을 파악하여 예방	범죄 발생 예측 알고리즘 개발
	건강보험자료, 환자 정보	환자 상황에 맞는 맞춤형 진료 제공	최적 진단 알고리즘 개발
			범죄발생 예상 지역 집중 순찰
			맞춤형 진단 사례 제공

4) 빅데이터 기반 건축도시 정책개발 체계

빅데이터 기반 건축도시 정책개발 체계는 앞에서 정리한 문제형성 단계, 정책형성 단계, 대안채택 3단계로 구성하였다. 각 단계별 빅데이터 활용 체계는 ‘질의개발’, ‘분석내용 결정’, ‘데이터 종류 및 특성 파악’, ‘분석방법 설계’, ‘데이터 수집 및 가공’, ‘분석 및 결과해석’의 6단계로 구성하였다. 빅데이터 분석체계는 목적(질의)에 따라 수집하는 데이터의 종류 등에 차이가 있는 것이며 기본적인 빅데이터 분석의 틀은 동일하다.

□ 질의개발

문제형성 단계의 질의는 여론 모니터링을 위한 질의가 대표적이다. 여론 모니터링을 위한 질의개발은 해당 주제에 대한 일반 국민, 이해관계자에 대한 의견 청취 내용으로 구성되며, 기존 설문지 문항 개발과 유사한 과정으로 볼 수 있다. 예를 들면 ‘현재 국민들이 건축물에서 어떠한 종류의 불안감을 느끼는가?’가 문제형성 단계의 질의라 할 수 있다.

정책형성 단계의 질의는 해당 문제에 대한 현황(실태)에 대한 질의이다. 예를 들면 문제형성 단계에서 ‘건축물 범죄예방’이 시급한 문제로 선정되었다면, 정책형성 단계에서는 ‘현재 강력범죄가 다발하고 있는 지역은 어디인가?’, ‘범죄가 발생한 건축물의 유형은 어떠한가?’와 같은 질의들을 도출할 수 있다.

대안채택 단계의 질의는 대안에 따른 경제적·사회적 효과를 검토할 수 있는 질의라 할 수 있다. 대안별로 정책실행에 따른 비용의 규모와 그에 따른 효과를 정량화를 위한 것으로 예를 들면 ‘다가구·다세대 주택의 침입방어 성능 강화’가 상기한 정책형성 단계의 대안으로 도출되었다면 질의는 ‘전국 다가구·다세대 주택의 침입방어 성능 강화를 위한 비용소요와 그에 따른 효과는 어떠한가?’가 될 수 있다.

□ 분석내용 결정

질의개발을 완료한 후 어떠한 내용을 분석할 것인지 분석내용을 결정하는 과정을 거친다. 질의에 따른 적절한 범위의 내용을 선정하는 것이 시행착오를 줄일 수 있다. 분석 내용이 적을 경우 해당 질의에 대한 답을 도출하기 어려운 경우가 발생하며, 분석 내용이 많을 경우 비용과 시간이 많이 소요된다.

문제형성 단계의 경우 전반적인 국민 여론을 분석할 것인지, 전문가 집단의 의견을 파악할 것인지 등의 대상을 결정하여야 하며, 정책형성 단계의 경우 질의에 따라 대안 도출

이 가능한 내용 범위를 설정한다. 대안채택 단계는 경제성, 효과성, 실현가능성 등에 대한 분석내용을 결정한다.

□ 데이터 종류 및 특성 파악

분석 내용을 결정한 후 해당 내용에 관한 어떠한 데이터가 있는지 파악 한다. 분석 내용에 따라 소셜 데이터, 로그 데이터, 정형데이터 등 다양한 빅데이터를 활용할 수 있으며, 실제 구득이 가능한 데이터를 중심으로 데이터의 특성을 파악한다.

[표 4-6] 빅데이터 기반 건축·도시 정책 개발 단계별 활용 데이터 및 활용 예시

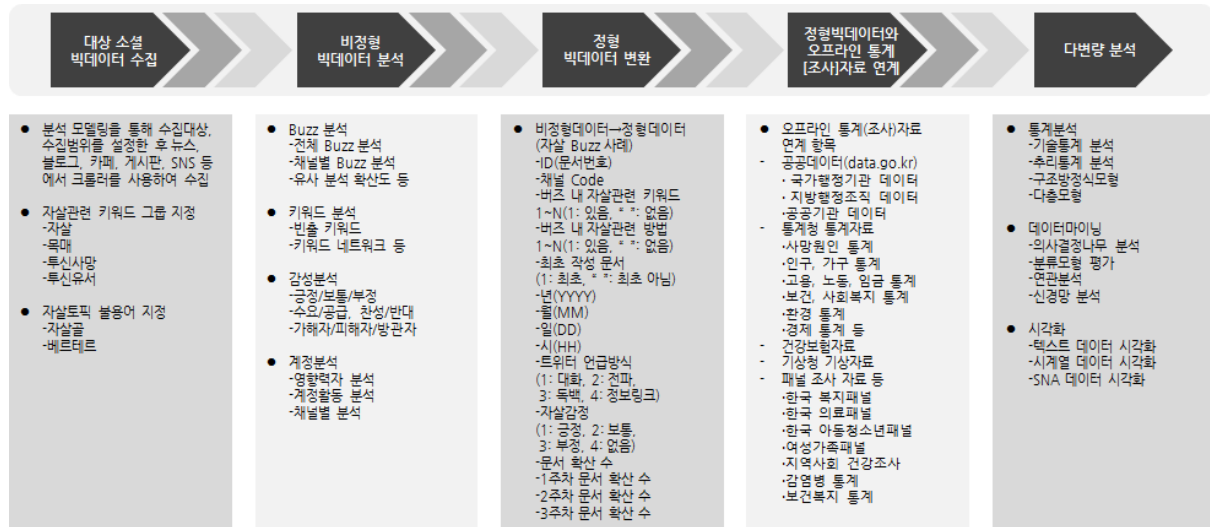
데이터	데이터 수집 방법	정책 개발 예시
소셜 미디어 데이터	트위터, 블로그, 페이스북, 특정 커뮤니티 홈페이지 게시판,	크롤링, Rss(Really Simple Syndication), 스트리밍
	신문사, 방송사, 포털 등의 뉴스, 뉴스 댓글, 보도자료	·(문제형성/정책형성) 관광 여론 분석을 통한 관광 정책 방향 설정 ·(문제형성/정책형성) 지자체 민원데이터 분석을 통한 정책 이슈 도출 및 정책 대안 마련
	선거공약	중앙선거관리위원회 홈페이지 policy.nec.go.kr
	민원 정보	국민신문고 www.epeople.go.kr, 지자체 홈페이지 민원 게시판
로그 데이터	검색엔진로그	Log Aggregator
	신용카드 사용정보	카드사 구매
	휴대전화 로그정보	통신사 구매
	교통카드 사용 정보	국토교통부 교통카드빅데이터 통합정보시스템
	네비게이션 운행정보	·(정책형성) 실시간 빠른 길 찾기 서비스
	CCTV 영상정보	관계기관 협의
정형 데이터	건축물에너지사용량정보	국토교통부 건축데이터 민간개방 시스템
	건축행정정보	open.eais.go.kr
	국가관리 통계의 건설, 주택, 토지 관련 통계	kosis.kr
	지가, 주택가격 등 부동산 정보	부동산 통계정보시스템 www.r-one.co.kr
	공간정보	·(문제형성/정책형성) 조류 이동 경로 공간분석을 통한 조류독감 발생지점 도출 및 방역시스템 위치 선정
	-통계청 데이터 GIS DB	통계지리정보서비스 sgis.kostat.go.kr
	-건물, 행정경계, 지형도, 수치지 등 국가 공간정보	국가공간정보포털 market.nsdi.go.kr
	-공간 및 위치 정보의 유통 및 판매	

데이터	데이터 수집 방법	정책 개발 예시
교통정보 -도로교통공단이 제공하고 있는 교통사고 정보 지역검색 -국가교통전반데이터 -고속도로 관련 데이터 -경기도 버스정보시스템	교통사고 분석 시스템 taas.koroad.or.kr/api 국가교통DB 포털 www.ktdb.go.kr 고속도로 공공데이터 포털 data.ex.co.kr 경기 버스정보 시스템www.gbiss.go.kr	·(문제형성/정책형성) 보행자 안전 정책 대안 발굴
정형 데이터	인문사회통계	
-국가 관리 통계	kosis.kr	
-서울시 관리 통계 및 서베이 결과	stat.seoul.go.kr	
-주민등록인구통계	www.mois.go.kr/	
재난재해관련 공공데이터	재난안전데이터포털 data.mpss.go.kr	·(정책형성) 재난 대응 시뮬레이션
기상정보	기상자료개방포털 data.kma.go.kr	
보건의료 공공데이터	국민건강보험공단, 건강보험심사평가원, 국립암센터, 한국보건사회연구원, 한국보건산업진흥원, 질병관리본부 협의	·(정책형성/대안채택) 인공지능 진단 시스템
복지 관련 공공데이터	사회보장정보시스템, 지역보건의료정보시스템, 사회서비스전자바우처, 보육통합정보시스템, 사회복지시설정보시스템, 취약계층 지원시스템	
기타 행정정보		
-서울시 공공데이터	data.seoul.go.kr	
-서울시 정보공개	opengov.seoul.go.kr	
-경기도 공공데이터	data.gg.go.kr	
-중앙부처 공공데이터	www.data.go.kr	
-정부기관선정 핵심지표	www.index.go.kr	
-공공기관 업무 수행 정보	www.open.go.kr	

□ 분석 방법 설계

분석 대상 데이터를 결정한 후에는 어떠한 분석 방법을 활용할지를 설계한다. 빅데이터 분석방법에 대해서는 본 연구의 2장에서 제시하였으며, 빅데이터 분석 기술이 빠르게 발전하고 있어 새로운 방법론에 대한 조사도 필요하다. 또한 빅데이터 분석을 위하여 새로운 시스템을 개발하는 것보다 기존에 개발되어 있는 유무상의 플랫폼 서비스를 활용하는 것이 보다 빠르고 경제적이기 때문에 빅데이터 분석 서비스의 활용에 대한 고려도 필요하다.

무엇보다 빅데이터 분석에서 가장 중요한 부분은 이기종의 데이터를 통합하여 분석하는 방법을 설계하는 것이라 할 수 있다. 아래의 도표는 비정형 빅데이터인 소셜 데이터와 기존의 정형데이터를 통합하여 분석하는 방법 예시이다.



[그림 4-4] 소셜빅데이터와 정형데이터를 통합한 분석 절차 및 방법 예시
출처: 송태민·송주영(2015), 『빅데이터 연구 한 권으로 끝내기』, 한나래아카데미, p.31

□ 데이터 수집 및 가공

데이터의 수집방법은 정형데이터와 비정형·반정형 데이터로 구분할 수 있다. 정형데이터는 엑셀과 같은 파일로 수집하거나, 오라클·SQL 등 데이터베이스를 바로 연결하여 수집한다. 비정형데이터는 웹 크롤러 등을 이용하여 공개된 자료를 자동으로 수집한다. 자세한 기술적인 수집 및 가공 방법은 본 연구의 2장에서 제시하였다.

건축·도시 분야 정책개발을 위한 빅데이터 분석의 데이터 수집 및 가공은 공개된 공공데이터에 대한 조사를 선행하고, 각 부처별 관련 데이터 보유현황을 파악하는 것이 중요하다. 또한 건축물과 토지에 관련하여 공개가 되지 않은 데이터라도 개인정보보호법 등 법률에 저촉되지 않는다면 원칙적으로 제공받을 수 있다. 다만 공공에서 주기적으로 공개하지 않는 데이터의 경우 수집시간이 상당기간 소요될 수 있어 이를 감안하여 데이터 수집 및 가공 기간을 정하는 것이 좋다.

□ 분석 및 결과해석

빅데이터 분석 방법은 통계적 분석, 데이터 마이닝, 텍스트 마이닝, 소셜네트워크 분석, 공간분석 등으로 다양한 분석 방법이 있으며, 본 연구의 2장에서 여러 분석 방법을 정리하였다. 빅데이터 분석이라고 해서 기존의 통계분석 방식을 배제하는 것이 아니라 기존 것처럼 분석 대상 및 목표에 따라서 새로운 분석방법 및 알고리즘을 개발하여 적용하기도 하며 이러한 경우에는 연구개발에 소요되는 시간과 비용을 충분히 고려하여야 한다.



[그림 4-5] 빅데이터 분석을 통한 정책 개발 과정

제5장 빅데이터 기반 건축도시 정책 개발 체계 실증

- 1. 대상 정책 선정
 - 2. 문제형성 단계 실증
 - 3. 정책 형성 단계
 - 4. 평가 및 환류
-

1. 대상 정책 선정

1) 대상 정책(건축물 유지관리 정책) 선정 배경

최근 국민 관심도가 증가하고 있고, 국토교통부 주요 업무계획에 포함되어 있는 건축물 유지관리 정책을 빅데이터 기반 정책개발 체계 실증 대상으로 선정하였다.

최근 건축물 붕괴, 화재 등의 안전사고와 태풍, 홍수, 지진 등의 자연재해로 인한 피해, 생활공간에서 발생하는 범죄로 건축·도시 공간의 안전 관리 정책에 대한 중요성이 대두되고 있으며, 폭염·폭한 일수가 증가하면서 냉난방비 절감을 위한 건축물 유지관리방안에 대한 일반 국민의 관심도 증가하고 있다. 또한 전 세계적 경기 침체로 건축물을 신축하기 보다 기존 건축물을 리모델링하여 사용하려는 수요가 증가하면서 건축물 유지관리 방안은 더욱 중요한 정책 이슈가 되고 있다.

이에 발 맞춰 국토교통부에서도 그린리모델링, 공공노후 건축물 리뉴얼, 건축물 유지관리법 제정, 건축물 안전관리 등 건축물 유지관리 정책을 2017년 주요 정책과제로 선정하여 추진 중에 있다.

2) 건축물 유지 관리 정책의 범위

먼저 국토교통부의 2017년 업무계획에 포함되어 있는 건축물 유지관리 관련 정책은 지진 등 건축물 안전 관리, 에너지 성능 관리, 건축물 현황 정보 관리 등으로 구분된다.

[표 5-1] 건축분야 정책 내용 분석(2017년 건축정책관실 업무계획)

구분	세부정책	세부실행내용	주요주제
안전관리	지진 안전 관련 정보 공개 및 모니터링 체계 구축	-건축물 내진설계적용여부 공개	국민생활
		-건축물표지판 내진등급 및 내진능력 표시	안전 강화
	내진 보강 및 설계기준 개선	-신축건축물 내진규정 강화	국민생활
		-기존건축물 내진보강 시 인센티브 제공	안전 강화
		-시설물 내진설계기준 개선	
	도시방재체계 구축 및 건축물 안전관리 강화	-건축물유지관리법 제정	국민생활
		-지역건축안전센터 도입	안전 강화
		-건축안전모니터링수행 및 점검대상 확대	
-건축안전App개발			
에너지 성능 관리	건축물의 에너지성능 개선	-그린리모델링사업 기획지원사업 우선지원	환경 이슈
		-그린리모델링이자지원사업 사업자추가지정 및 교육·기술지원	대응력제고
		-공공건축물 에너지소비량공개제도 평가항목개발	
		-에너지소비총량제설계 확대	
		-지역별 단열기준 강화	
		건축물 현황 정보 관리	각종 신고 및 조회 서비스 개선
-소유건축물현황 대국민서비스 개선			

건축물 유지관리 정책의 수단으로 활용되고 있는 ‘건축물 유지관리 점검 매뉴얼’, ‘도시 기후변화 재해 취약성 분석 매뉴얼’, ‘공동주택 관리업무 진단 매뉴얼’에서 다루고 있는 주요 검토 항목을 분석하였다. 각 항목들을 분류해보면, 건물과 대지의 관계에 따른 관리 항목, 건물의 높이 및 형태, 구조안전, 화재안전, 설비, 에너지 및 친환경 성능 관리 등으로 구분된다.

[표 5-2] 건축물 유지관리 정책 범위

대 항목	소항목	세부항목
대지	대지의 안전 등	- 옹벽과 건축물과의 안전 확인 - 지반침하 여부
	대지의 조경	- 조경면적 확보 여부 - 조경시설 성능 유지여부
	공개공지 등의 확보	- 공개공지 성능 유지여부 - 공개공지 면적확보 여부
	대지와 도로의 관계	- 대지가 도로와 접하고 있는지 여부

대 항목	소항목	세부항목
	건축선에 따른 건축제한	- 건축물, 캐노피, 발코니, 노대 및 담장의 건축선 내 유지여부
	공동주택 토목시설	- 배수시설, 도로, 옹벽, 법면, 담장, 비상급수시설
	공동주택 어린이놀이터	- cctv, 안전표지판, 놀이시설물 파손여부, 바닥완충제 훼손여부, 바닥면 용기·침하균열, 파고라·벤치 파손여부 등
	공동주택 조경시설	- 휴게시설, 자전거 보관소, 재활용품장, 운동기구, 조형물, 수목, 지피식물 등
높이 및 형태	건폐율	- 건폐율 유지여부
	용적률	- 용적률 유지여부
	대지 안의 공지	- 공지의 폭 유지여부 - 공지성능 유지여부
	높이 제한	- 높이제한 유지여부
	일조 등의 확보를 위한 높이 제한	- 지붕변경, 증축으로 인한 일조권 유지여부
	외부형태의 원형 유지	- 광고물 등에 의한 디자인 훼손 여부 - 미관지구 안에서의 외관변경 여부
	공동주택 건축시설	- 동별 관리상태, 옥상, 동지하, 지하주차장 등
구조안전	구조내력 등	- 주요구조부 변형 및 균열 여부(육안점검 가능부분) - 옥상의 구조적 안전여부 - 내진설계 적용 여부
화재안전	복도 · 계단 · 출입구	- 복도, 계단, 출입구의 성능 유지여부
	옥상광장	- 옥상광장의 피난성능 유지여부
	방화구획	- 방화문, 방화셔터 등의 성능 유지여부 - 방화구획 적합 여부
	경계벽 · 칸막이벽	- 경계벽 및 칸막이벽의 변경 등 방화성능 유지여부
	그 밖의 피난설비	- 배연설비의 성능 유지여부
	내화구조	- 내화구조의 성능 유지여부
	방화벽	- 방화벽의 성능 유지여부
	방화지구 안의 건축물	- 외벽의 성능 유지여부 - 창호의 성능 유지여부
	내부 마감재료	- 내부마감의 방화성능 유지여부
	외벽 마감재료	- 외부마감의 노후화 및 마감재 탈락 여부
	지하층	- 지하층의 소방설비 성능 유지여부 - 지하층 피난구, 피난계단의 성능 유지여부
	공동주택 소방시설	
건축설비	급수설비	- 급수설비 성능 유지여부
	배수설비	- 배수설비 성능 유지여부
	냉방설비	- 냉방설비 성능 유지여부

대 항목	소항목	세부항목
	난방설비	- 난방설비 성능 유지여부
	환기설비	- 환기설비 외관 유지여부 - 환기설비 성능 유지여부 - 공기조화설비 외관 유지여부
	피뢰설비	- 피뢰설비 성능 유지여부
	방송수신설비	- 방송설비 성능 유지여부
	전기설비 설치공간	- 전기설비 설치공간 확보 여부
	승강기	- 승강기 성능 유지여부
	공동주택 승강기, 난방, 전기시설	
	에너지 및 친환경 관리 등	- 단열성능 유지여부(출입문, 창호, 외벽 등) - 결로 발생 여부 - 창호 기밀성 성능 유지여부
	친환경건축물 인증	- 친환경건축물 인증의 유지관리 여부
	지능형건축물 인증	- 지능형건축물 인증의 유지관리 여부
	에너지 효율등급 인증	- 에너지효율등급 인증의 유지관리 여부
	공동주택 행정, 회계 업무 등	- 관리규약, 장기수선계획 및 총당금, 하자보수 및 하자 보수충당금, 관리 관리 등

위의 내용을 종합하면 건축물 유지관리 정책의 범위는 건축물에서의 안전에 관한 사항, 설비 성능에 관한 관리, 허가기준 준수 여부, 에너지 성능 관리, 건물의 수명, 유지관리 비용, 공동주택 생활환경 관리로 구분할 수 있다.

[표 5-3] 건축물 유지관리 정책 분야와 제도 현황

유지관리 정책 분야		유지관리 제도 현황	
대상	세부 내용	대상	세부 내용
안전	건축물 내 추락 등 안전사고 / 붕괴, 내진 등 구조안전 / 화재 안전 / 홍수, 태풍, 폭염 등 기후 변화 취약성 / 범죄 안전	공동주택 관리 제도	공동주택 안전점검 / 장기수선계획, 장기수선충당금 / 하자보수충당금 / 주택관리사, 주택관리업자 / 입주자대표회의, 공동주택 관리기구 / 공동주택 관리 분쟁조정
설비 성능	건축물 설비 성능 관리	시설물안전점검제도	안전점검 및 정밀안전진단 유지관리업자, 안전진단전문기관
허가기준 준수여부	건축물의 높이 등 허가기준 준수 유지 여부 관리 대지와와의 관계 준수 유지 여부	소방시설 설치 유지 및 안전 관리제도	소방시설 자체점검제도 / 소방특별조사 / 소방시설관리사, 소방시설관리업
에너지성능 및 친환경	단열 및 기밀 성능, 결로 / 실내 공기질 / 친환경 소재 / 이산화탄소 배출 저감	건축물 정보 관리	건축물 생애이력 관리시스템
장수명	가변성, 수리용이성, 내구성	기본계획	시설물의 안전 및 유지관리 기본계획
유지관리 비용	공동주택 관리비 / 전기세, 수도세, 가스비 등 시설자산	유지관리시스템	BIM을 통한 유지관리 기타 유지관리시스템
공동주택 생활환경	공동주택 관리 조직 분쟁조정(하자보수, 층간소음 등)		

2. 문제형성 단계 실증

1) 질문 도출

건축물 유지관리 정책의 이슈와 수요를 도출하기 위해 도출한 주요 질문은 아래와 같다.

- 건축물 유지관리 정책에 대한 관심도가 높은 분야는 무엇인가?
- 건축물 유지관리 정책에 대한 긍부정 여론은 어떠한가?
- 건축물 유지관리에 대한 국민 관심도는 어떻게 변화되고 있는가?

2) 분석내용 설정

위의 주요 질문을 바탕으로 유지관리 정책에 대한 일반 국민 여론 분석을 목적으로 관련 키워드의 사용 빈도와 연관 키워드를 분석하여 국민이 가장 관심을 갖는 이슈가 무엇인지 도출하고자 한다. 또한 시계열 추이 분석을 통해 각 이슈에 대한 여론의 변화에 대해 살펴보고 감성 분석을 통해 각 이슈에 대한 긍부정 평가를 살펴보고자 한다. 즉, 키워드 빈도분석, 연관키워드 분석, 시계열 추이 분석, 감성 분석을 주요 내용으로 선정하였다.

3) 데이터 종류 및 특성 파악

여론 분석에 주로 활용되는 데이터는 소셜 미디어 데이터이다. 본 연구에서는 대국민 여론 분석을 위해 국내에서 가장 많이 사용되고 있는 네이버 뉴스, 블로그(네이버, 다음 티스토리), 트위터 게시글 정보를 활용하였다.

소셜 미디어 데이터를 활용하기 위해서는 분석하고자 하는 내용을 검색하기 위한 키워드 설정이 중요하며, 데이터 수집 및 분석에 소요되는 시간과 데이터 신뢰도를 고려하여 수집 기간을 선정해야 한다. 또한 많은 양의 데이터를 분석하고자 하는 경우, 데이터 수집을 위한 크롤링 기술 등이 필요한데, 이러한 경우 개인이 데이터를 수집하여 분석하기에는 어려움이 있다. 따라서 간단한 분석 서비스를 제공하는 온라인 플랫폼을 이용하거나 전문 분석 기업에 의뢰하는 것이 효율적이다.

4) 분석 방법 설계

분석의 주요내용, 데이터 종류 및 특성을 고려하여 검색 키워드를 선정하고, 데이터 수집 기간, 수집 채널, 분석 기술을 선정하는 등 분석 방법을 설계 하였다.

먼저 검색 키워드는 앞서 살펴본 건축물 유지관리 정책 범위에 따라 키워드를 선정하고, 샘플 데이터 수집을 통해 정책 세부 검색 키워드를 조정하는 과정을 거쳤다.

[표 5-4] 건축물 유지관리 분야 검색 키워드 선정

정책 분야 키워드	검색 키워드	정책 분야 키워드	검색 키워드
안전	·안전사고 ·붕괴 ·화재	에너지성능 및 친환경	·결로현상 ·실내공기질 ·친환경인테리어
허가기준 준수여부	·불법개조 ·방조개조 ·불법건축물	장수명	·가변성 ·수리용이성
설비 성능	·설비 교체	유 지 관 리 비용	·관리비 ·전기세, 수도세, 가스비
		공 동 주 택 생활환경	·공동주택 하자보수 ·공동주택 층간소음
			·단열 ·에너지절약
			·내구성
			·난방비, 냉방비 ·시설자산
			·공동주택 분쟁

안전, 허가기준 준수여부, 설비성능, 에너지 성능, 장수명, 유지관리 비용, 공동주택 생활환경 7가지 분야 중 의견제시 문건 빈도가 많고, 정책적 문제 해결이 중요하다고 판단 되는 분야를 선정하여 세부적인 키워드 분석을 수행하였다.

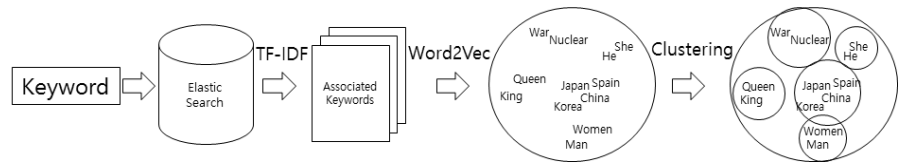
데이터 수집 기간은 시계열 분석이 가능하도록 5년 이상의 기간을 선정하였으며, 데이터 수집에 소요되는 시간, 데이터의 질과 양을 고려하여 최종적으로 2014년 4월1일부터 2017년 4월 30일로 선정하였다. 네이버 뉴스 및 블로그의 경우 검색 건수는 많아도 실제 확인 가능한 콘텐츠는 4,000건으로 제한되어 있어 검색 기간 단위를 가능한 줄여야 하는 한계가 있었다.

데이터 수집 채널은 여론의 참여가 가장 활발한 네이버, 다음, 티스토리 블로그와 트위터, 그리고 다양한 언론사의 뉴스를 모아 제공하는 네이버 뉴스로 선정하였다.

데이터 수집 기법으로 크롤링 시스템을 통해 자동 수집하는 방법을 선택하였으며, 수집된 데이터를 활용하여 키워드별 빈도수와 시계열 추이 분석, 키워드에 대한 긍부정 빈도수와 시계열 추이 분석, 연관 키워드 분석, 데이터 수집 채널별 키워드 분포 분석을 수행하는 것으로 계획 하였다.

소셜미디어 데이터의 대표적 분석 기법인 텍스트 마이닝 기법을 적용하였으며, 연관 키워드 분석과 키워드에 대한 긍부정 분석 과정을 구분하여 진행하였다.

먼저 연관 키워드 분석은 통계 기반(TF-IDF)으로 키워드를 추출한 후 벡터화하여 클러스터링 하는 과정을 거친다.



[그림 5-1] 텍스트 마이닝 과정
출처: ㈜아크릴(2017), 내부자료

연관 키워드 분석의 첫 번째 단계는 문서 내에서 키워드를 추출하는 과정이다. ‘필요성’, ‘방향’ 등과 같이 중요도가 낮은 키워드를 제외하고 의미 있는 키워드를 추출하기 위해 통계적 기술이 필요한데, 특정 키워드가 특정 문서에서 등장하는 횟수(Term Frequency)와 문서군 내에서 등장하는 횟수(Inverse Document Frequency)를 고려하는 TF-IDF(Term Frequency-Inverse Document Frequency)기법을 적용하였다. 또한 키워드 필터링을 위해 특정 문서에서 키워드의 최소 등장 횟수와 문서군에서 키워드를 포함하는 문서의 최대 개수 기준을 설정하였다.

```

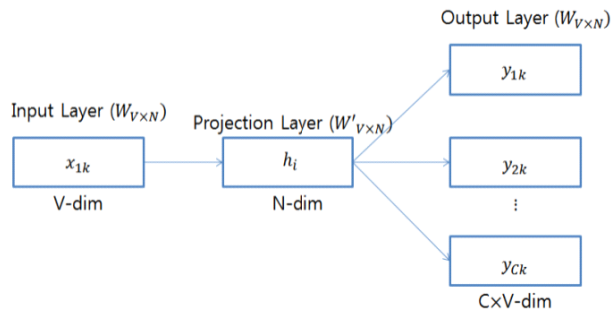
"terms" : {
  "가치" : {
    "doc_freq" : 746,
    "ttf" : 1093,
    "term_freq" : 3,
    "score" : 18.439026
  },
  "검증" : {
    "doc_freq" : 294,
    "ttf" : 400,
    "term_freq" : 3,
    "score" : 21.226295
  },
  "결과" : {
    "doc_freq" : 3607,
    "ttf" : 5030,
    "term_freq" : 3,
    "score" : 13.714495
  },
  "누구" : {
    "doc_freq" : 2569,
    "ttf" : 3418,
    "term_freq" : 3,
    "score" : 14.732239
  },
  "매도" : {
    "doc_freq" : 326,
    "ttf" : 594,
    "term_freq" : 2,
    "score" : 13.944894
  },
  "서적" : {
    "doc_freq" : 144,
    "ttf" : 215,
    "term_freq" : 2,
    "score" : 15.571346
  },
  "수익" : {
    "doc_freq" : 751,
    "ttf" : 1639,
    "term_freq" : 5,
    "score" : 30.698355
  },
  "시작" : {
    "doc_freq" : 4818,
    "ttf" : 7383,
    "term_freq" : 4,
    "score" : 17.128342
  }
}

```

[그림 5-2] TF-IDF를 이용하여 추출한 키워드와 키워드 중요도 수치 예시
출처: ㈜아크릴(2017), 내부자료

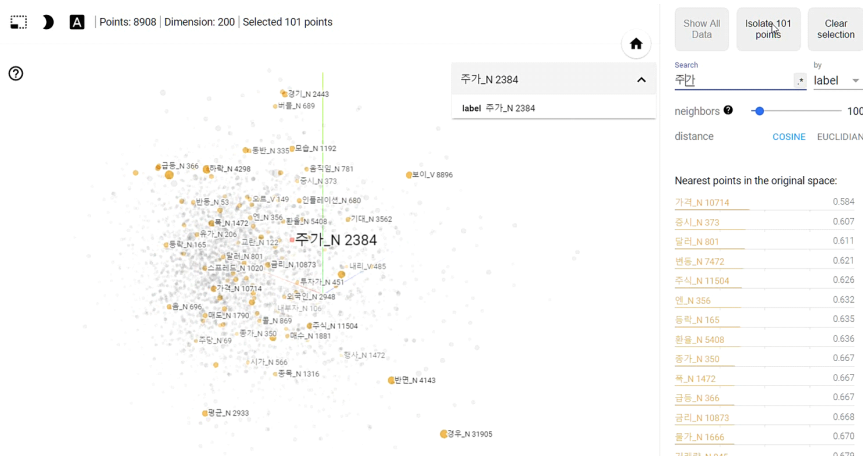
두 번째 단계로, 추출된 키워드를 벡터 형태로 수치화해야 하는데, 이때 한 단어로부터 여러 단어를 추측하는 Skip-gram모형을 적용한 Word2vec 기법을 이용하였다.¹⁾ 이러한 과정이 필요한 이유는 한글 특성상 한 개의 단어가 여러 의미를 포함하기도 하고, 하나의 의미에 사용되는 단어가 다양하기 때문이다. 즉, 문서에서 추출된 중요도가 높은 키워드에 대해 같은 의미를 포함하는 키워드를 구분하기 위함이다. 특정 키워드의 연관 키워드를 벡터화한 데이터는 “키워드_형태소단어ID” 형태로 나타나며 TF-IDF 값과 특정 키워드와의 연관성 정도 값을 갖는다.

1) 이진욱 외(2017), 감성분석과 Word2vec을 이용한 비정형 품질 데이터 분석, 한국품질경영학회, 45(1), 117-127p



[그림 5-3] Skim-gram 모형 개념도

출처: 이진욱·유국현·문병민·배석주(2017), “감성분석과 Word2vec을 이용한 비정형 품질 데이터 분석”, 『품질경영학회지』, v.45(1), 한국품질경영학회

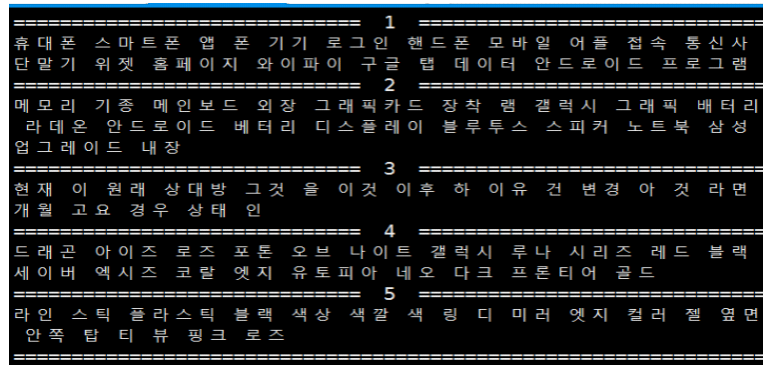


[그림 5-4] 키워드 벡터화 예시(‘주가’라는 특정 키워드에 대한 연관키워드 들을 벡터화 한 결과)

출처: ㈜아크릴(2017), 내부자료

마지막으로 벡터화된 키워드 정보를 바탕으로 의미가 비슷한 키워드끼리 분류하기 위해 클러스터링 분석 방법을 적용하였다. 클러스터링 기법에는 centroid model, connectivity model, density model 과 같은 많은 모델들과 기법들이 있으나 Word2vec를 이용해 벡터화한 데이터의 특성상 의미나 쓰임이 비슷한 단어끼리 비슷한 위치에 위치하도록 단어가 비교적 고르게 되어 있어 중심이 되는 키워드와 비슷하게 위치한 단어들을 묶어주는 centroid model이 적합하다.

키워드에 대한 감성 분석(극부정분석)을 위해서는 문장에 긍정/중립/부정으로 감성태깅된 감성 사전을 구축하고, 분석 대상이 되는 텍스트 정보의 형태소를 분석한 후 이를 벡터화하고, 감성사전과 텍스트 정보를 연결하는 과정을 거쳐야 한다.



[그림 5-5] 키워드 클러스터링 결과 예시(‘갤럭시’라는 특정 키워드에 대한 연관키워드 분류)
출처: ㈜아크릴(2017), 내부자료

먼저 감성사전은 소셜 데이터 분석 전문 기업인 ㈜아크릴에서 구축한 데이터베이스를 이용한다. 감성사전에는 세종코퍼스와 웹에서 수집한 약 350만 문장(세종 코퍼스, 구글 플레이 스토어 댓글, 네이버 구매후기 등)으로 구성되어 있고 각 문장은 긍정/중립/부정으로 감성태깅되어 있다.

감성사전과 분석 대상이 되는 텍스트를 연결하기 위해서는 형태소를 분석하고 벡터화해야 한다. 웹에 공개되어 있는 형태소 분석기²⁾ 중 웹에서 크롤링한 데이터 분석에 성능이 좋은 Twitter형태소 분석기를 적용하였다. 형태소 분석된 데이터를 감성사전에 연결하기 위해서는 벡터화 과정이 필요하며 연관어 분석에서 사용한 Word2vec기법을 적용하였다.

기포N 자두N 부활V 오르V 입술N 자국N 먼저M 입안N 흐르V 들어가V 한M 오금N 맥주N 해N 적시N 전N 예리X 들V 한N 내려놓V 넣킨N 들V 입N 누르V 입N 가장자리N 흘러나오V 맥주N 회V 만N 냅킨N 얼룩N 남기V 얼룩N 부분N 부분N 빨강V 핏자국N 번지V 있V 맞은편N 통유리N 너머N 회고무래X 하늘N 봉숭이N 물빛N 물들V 있V 예리X 하늘N 냅킨N 번갈V 보V 페인트N 자트N 흰색N 명암N 색_N 겨우M 찾아내V 핏빛N 봉숭이N 물빛N 오직M 명암N 강도N 구분N 지V 하V 캔자M 이M 두M 색N 어울리V 낫칠V 음V 느끼V 지V 기억N 한M

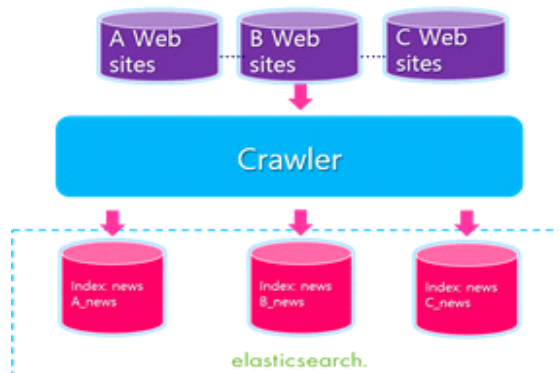
[그림 5-6] 형태소 분석 결과 예시
출처: ㈜아크릴(2017), 내부자료

벡터화된 텍스트 정보와 기 구축된 감성사전을 연결하기 위해서는 신경망 모델이 많이 사용되는데, 다양한 신경망 모델 중 LSTM(Long Short-Term Memory) 모델과 FNN(Feed Forward Neural Network)모델을 융합하여 사용하였다.

2) komoran : www.shineware.co.kr / twitter : <https://github.com/kkma> : kkma.snu.ac.kr

5) 데이터 수집 및 가공

데이터 수집은 빅데이터 분석 전문 기업인 ㈜아크릴에서 개발한 웹 정보 수집 시스템을 활용하였다. 시스템은 자바 기반으로 구현되어 있으며, 크롤러를 통해 복수의 검색 채널에 있는 텍스트 정보를 수집해 분산 검색 엔진인 엘라스틱서치(Elasticsearch)에 저장된다. 또한 환경설정을 통해 검색 url, 크롤링 필터링 규칙, 사이트 방문 규칙, Seed수, Thread 수 등 사이트에 맞는 다양한 옵션 설정이 가능하다.



[그림 5-7] 웹 정보 수집 시스템 구조

출처: (주)아크릴(2017), 내부자료

웹 정보 수집 시스템의 결과에는 문서 번호, 제목, url, 게시일, 수집날짜, 문서 요약, 카테고리, 상세 카테고리, 본문 정보가 포함되어 있으며 기사의 본문 정보는 형태소 분석기, 감성인식 엔진을 통해 분석되고 분석 결과도 함께 저장된다. 대부분의 웹 사이트에서 로봇 방문자를 차단하는 기능을 가지고 있는데, 그로 인해 크롤러의 접근이 어려운 경우를 대비하여 자동 비주기적으로 크롤링 하는 기술을 적용하였다. 또한 데이터 수집에 소요되는 시간을 단축하기 위해 Apach Zookeeper기반의 분산처리 시스템을 적용, 다수의 서버를 구축하도록 하여 빠른 속도로 수집이 가능하도록 하였다.

Field	Description
articleId	사이트 내 article ID
title	Article 명
url (id)	article 주소
publishTime	article 게재 날짜
crawlTime	크롤링 날짜
abstract	Article 요약(optional)
category	Article 의 카테고리(optional)
subCategory	Article 의 상세 카테고리(optional)
article	기사 본문

[illegible]

[그림 5-8] 웹 정보 수집 시스템 결과

6) 분석 및 결과해석

□ 분석 대상 선정

앞서 분석 방법 설계에서 밝힌 바와 같이, 유지관리 정책 7개 분야에 대한 사전 검색을 통해 국민적 관심도가 높고 정책적 문제해결이 시급하다고 판단되는 ‘안전’ 분야를 주요 분석 대상으로 선정하였다. 검색 문건 수가 가장 많은 분야는 유지관리비용과 에너지 성능 및 친환경 분야 였으나, 유지관리비용과 관련한 문건에서는 개인의 의견이 포함되어 있다기 보다 유지관리 비용을 언급하는 내용이 많아 분석 대상에 적합하지 않은 것으로 판단 되었으며, 에너지 성능 및 친환경 분야의 경우 광고글이 대부분이고 실제 개인 의견과 관련한 게시글은 적은 것으로 판단되었다.

[표 5-5] 분석 대상 선정 결과

분야	검색 키워드	버즈량			관련 원문 내 주요 키워드
		합계	네이버뉴스	네이버블로그	
안전	건축물 안전사고	21,114	10,247	10,867	철거, 공사, 해체, 방치건축물, 해빙기
	건축물 지진	19,059	9,085	9,974	내진설계, 지진안전성
	건축물 붕괴	15,200	7,767	7,433	해빙기, 노후, 지붕, 주차장, 화재, 태풍
	건축물 침수	2,471	1,064	1,407	차수판, 태풍, 장마철
	건축물 화재	52,245	15,347	36,898	고층건축물, 대형건축물, (준)불연자재, 스프링클러
	건축물 범죄	8,989	4,424	4,565	공/폐건축물, 지하주차장, 셉테드
	소계			119,078	
설비성능	건축물 설비교체	2,736	928	1,808	노후
허가기준 준수여부	건축물 불법개조	1,653	551	1,102	증/개축, 창고, 용도변경, 방조개기, 복층개조
	방조개기	1,390	430	960	원룸, 고시원, 화재, 주차공간, 전입신고불가, 월세
	건축물 공개공지	1,660	660	1,000	용적률/높이제한 완화, 불법사유화
	불법건축물	31,359	15,708	15,651	불법증축, 주차장 불법 건축물, 농지, 임야, 불법용도변경
	건축물 지반침하	1,645	847	798	연약부지, 해빙기, 균열
	소계			37,707	
에너지 성능 및 친환경	결로현상	41,131	1,908	39,223	결로방지 단열재
	실내공기질	19,107	10,790	8,317	친환경페인트, 천연벽지
	친환경인테리어	169,138	14,348	154,790	바닥, 벽지, 창호, 등 건축자재, 에코인테리어 지원
	단열	275,363	36,027	239,336	보온, 난방비, 준불연단열재
	건축물에너지절약	8,587	4,002	4,585	에너지소비총량제, 에너지절약계획, 녹색건축인증, 건축물에너지효율등급
	소계			513,326	※ 광고성의 버즈가 대부분
장수명	건축물가변성	417	69	348	장수명(長壽命) 주택 건설·인증제도, 석고보드, 노후건축물

분야	검색 키워드	버즈량			관련 원문 내 주요 키워드
		합계	네이버뉴스	네이버블로그	
	건축물수리용이성	94	25	69	
	건축물내구성	7,358	1,405	5,953	
	소계			7,869	
유지관리 비용	관리비	835,084	71,185	763,899	체납, 자동납부
	건물 전기세	17,298	385	16,913	
	건물 수도세	55,304	334	54,970	
	건물 가스비	6,097	102	5,995	
	건축물 난방비	2,362	550	1,812	
	건축물 냉방비	318	75	243	
	건축시설자산	19,899	7,488	12,411	건축자산
	소계			936,362	※ 건축물 유지관리의 측면보다는 사용자 개인의 비용을 계시하는 경우가 많아 유의미한 결과를 얻기 어려움
공동주택 생활환경	공동주택 하자보수	5,040	1,400	3,640	공동주택 하자보수 관리 교육, 하자보수보증, 공동주택하자보증 사전알리미
	공동주택 층간소음	10,150	4,748	5,402	옆집 벽간 소음, 조례, 공동주택층간소음맞춤형 서비스
	공동주택 분쟁	11,804	5,716	6,088	공동주택분쟁상담실, 공동주택품질검수단, 입주민, 하자
	소계			26,994	

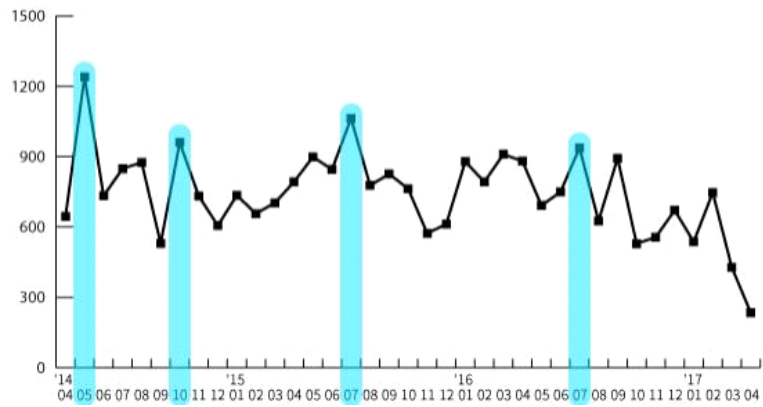
□ 안전사고에 대한 키워드 분석

안전 분야 중 안전사고, 지진, 붕괴, 침수, 화재, 범죄 6개 분야를 주요 검색 키워드로 하였으며 각 검색 키워드에 대한 분석 결과는 다음과 같다.

먼저 건축물 안전사고에 대한 전체 버즈량은 27,476건이며, 월 평균 버즈량은 743건이다. 또한 2014년 5월, 10월, 2015년 7월, 2016년 7월에 버즈량이 급증하는 것으로 나타났다.

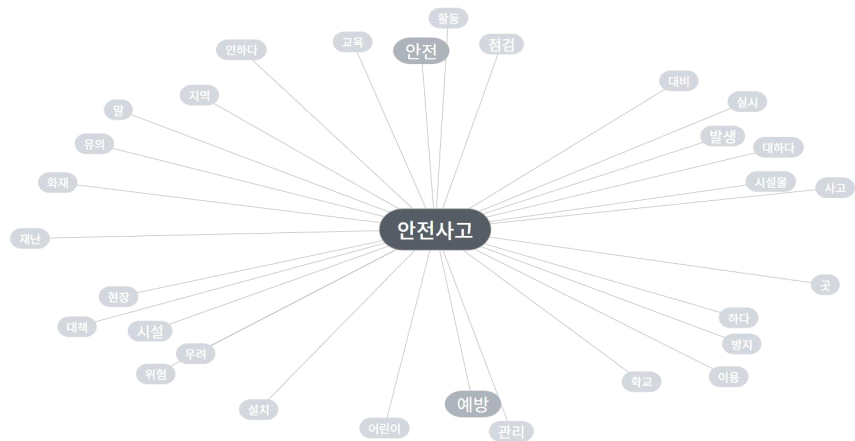


[그림 5-9] '안전사고'에 대한 키워드 분석 결과



[그림 5-10] '안전사고'에 대한 시계열 빈도

안전사고와 관련한 연관 키워드는 안전(41.0%), 예방(30.8%), 관리(19.0%), 시설(18.8%), 발생(18.5%), 점검(17.1%), 대하다(13.2%), 하다(12.3%), 시설물(11.9%), 사고(10.0%)로 나타났으며, 데이터 수집 기간 직전에 발생한 세월호 참사의 영향으로 유추되는 어린이, 학교 등의 키워드와 방지, 유의, 재난, 대책 등 안전사고 발생의 사전 대비와 관련한 키워드가 함께 언급 되었다.



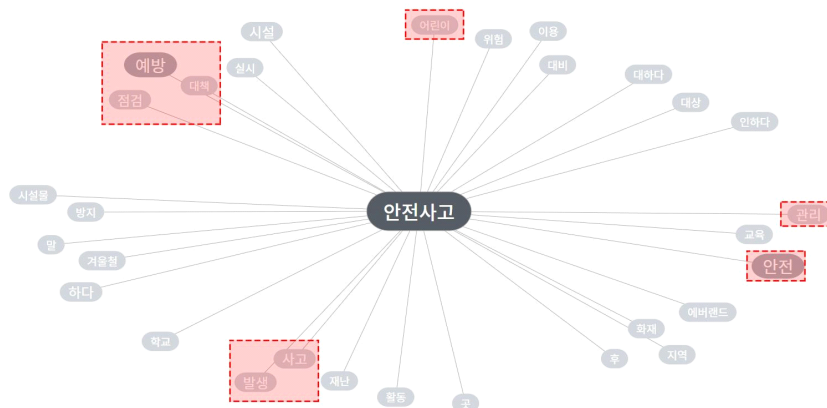
[그림 5-11] '안전사고'에 대한 연관키워드

안전사고와 관련하여 가장 많은 게시글이 나타난 채널은 네이버블로그로 총 17,006건(61.89%)이었고, 뉴스 9,386건(34.16%), 트위터 1,084건(3.9%)으로 나타났다. 채널별 출현 빈도수가 높았던 연관키워드는 차이가 없는 것으로 나타났으며, 안전, 예방, 시설, 점검 등의 연관키워드 빈도수가 높았다.

Blog 전체보기 61.9% (전체 27,476 중 17,006)				News 전체보기 34.2% (전체 27,476 중 9,386)				Twitter 전체보기 3.9% (전체 27,476 중 1,084)			
NO	연관키워드	비즈량	여론감성	NO	연관키워드	비즈량	여론감성	NO	연관키워드	비즈량	여론감성
1	안전	(-)37572	긍정 2033 (26.85%) 중립 4928 (65.08%) 부정 611 (8.07%)	1	안전	(-)3310	긍정 763 (23.05%) 중립 2210 (66.77%) 부정 337 (10.18%)	1	시설	(-)465	긍정 55 (11.83%) 중립 369 (79.35%) 부정 41 (8.82%)
2	예방	(-)5449	긍정 1363 (25.01%) 중립 3721 (68.29%) 부정 365 (6.70%)	2	관리	(-)2720	긍정 467 (17.17%) 중립 2060 (75.74%) 부정 193 (7.10%)	2	시설물	(-)408	긍정 33 (8.09%) 중립 362 (88.73%) 부정 13 (3.19%)
3	발생	(-)3277	긍정 667 (20.35%) 중립 2118 (64.63%) 부정 492 (15.01%)	3	예방	(-)2632	긍정 582 (22.11%) 중립 1867 (70.93%) 부정 183 (6.95%)	3	예방	(-)381	긍정 37 (9.71%) 중립 316 (82.94%) 부정 28 (7.35%)
4	시설	(-)3210	긍정 731 (22.77%) 중립 2112 (65.79%) 부정 367 (11.43%)	4	시설물	(-)2187	긍정 332 (15.18%) 중립 1715 (78.42%) 부정 140 (6.40%)	4	안전	(-)378	긍정 44 (11.64%) 중립 292 (77.25%) 부정 42 (11.11%)
5	점검	(-)2901	긍정 602 (20.75%) 중립 2090 (72.04%) 부정 209 (7.20%)	5	곳	(-)1770	긍정 237 (13.39%) 중립 1420 (80.23%) 부정 113 (6.38%)	5	관리	(-)344	긍정 24 (6.98%) 중립 314 (91.28%) 부정 6 (1.74%)
6	대하다	(-)2475	긍정 573 (23.15%) 중립 1633 (65.98%) 부정 269 (10.87%)	6	발생	(-)1743	긍정 319 (18.30%) 중립 1147 (65.81%) 부정 277 (15.89%)	6	점검	(-)305	긍정 31 (10.16%) 중립 259 (84.92%) 부정 15 (4.92%)
7	하다	(-)2183	긍정 569 (26.07%) 중립 1273 (58.31%) 부정 341 (15.62%)	7	유역	(-)1572	긍정 208 (13.23%) 중립 1292 (82.19%) 부정 72 (4.58%)	7	곳	(-)279	긍정 18 (6.45%) 중립 259 (92.83%) 부정 2 (0.72%)
8	관리	(-)2152	긍정 530 (24.63%) 중립 1403 (65.20%) 부정 219 (10.18%)	8	점검	(-)1488	긍정 269 (18.08%) 중립 1112 (74.73%) 부정 107 (7.19%)	8	유역	(-)248	긍정 18 (7.26%) 중립 230 (92.74%) 부정 0 (0.00%)
9	실시	(-)1861	긍정 318 (17.09%) 중립 1434 (77.06%) 부정 109 (5.86%)	9	시설	(-)1487	긍정 272 (18.29%) 중립 1057 (71.08%) 부정 158 (10.63%)	9	배려다	(-)175	긍정 3 (1.71%) 중립 172 (98.29%) 부정 0 (0.00%)
10	사고	(-)1801	긍정 365 (20.27%) 중립 1169 (64.91%) 부정 267 (14.83%)	10	하다	(-)1144	긍정 214 (18.71%) 중립 757 (66.17%) 부정 173 (15.12%)	10	실시	(-)140	긍정 11 (7.86%) 중립 119 (85.00%) 부정 10 (7.14%)

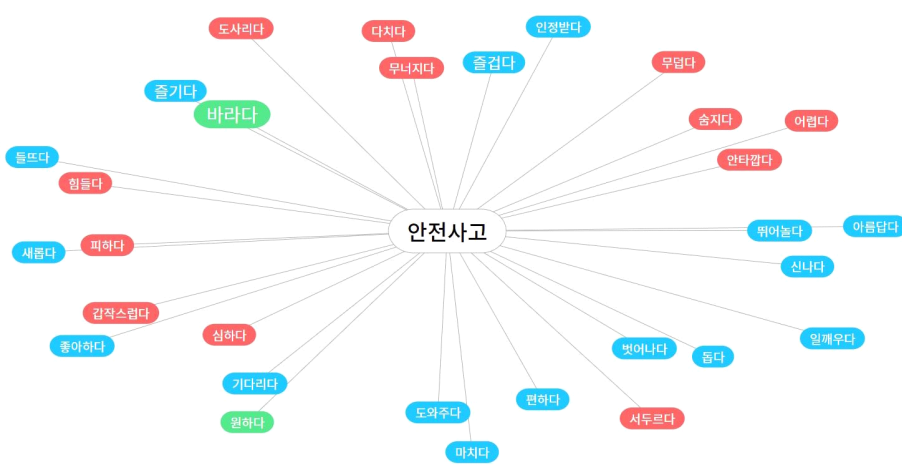
[그림 5-12] '안전사고'에 대한 데이터 수집 채널별 연관 키워드 순위

버즈량이 급증하는 2014년 5월(1,240건)과 10월(960건)은 각각 세월호 참사와 성남 판교에서 발생한 환풍구 추락사고의 영향인 것으로 해석되는데, 5월에는 단순히 건축물에 대한 안전 뿐 아니라 생활 환경 전반에서의 안전에 대한 관심이 높아지면서 안전사고 버즈량도 급증하는 것으로 보인다. 10월에는 사전에 예방할 수 있었던 인재라는 여론이 많아 예방, 점검, 대책, 관리 등의 키워드가 많이 언급 되었다. 또한 2015년 7월, 2016년 7월에 나타난 버즈량 급증은 장마철에 발생하는 안전사고에 대한 여론이 대부분이었다.



[그림 5-13] 버즈량이 급증한 2014년 10월의 게시글에서 나타난 연관 키워드

안전사고에 대한 감성 분석 결과, 긍정16.75%, 중립 72.83%, 부정 10.42%의 비율로 나타났다. 중립을 제외하고는 긍정 감성이 부정 감성보다 조금 높았다. 긍정 감성어 상위에는 일깨우다, 도와주다, 벗어나다 등 안전사고 이후 안전의 중요성 또는 해결에 대한 키워드가 나타났다. 부정 감성어 상위에는 도사리다, 숨지다, 무너지다, 다치다, 심하다 등 안전사고 발생과 관련된 키워드가 나타났다.



[그림 5-14] ‘안전사고’에 대한 감성 분석 결과(빨강: 부정, 파랑: 긍정, 녹색: 중립)

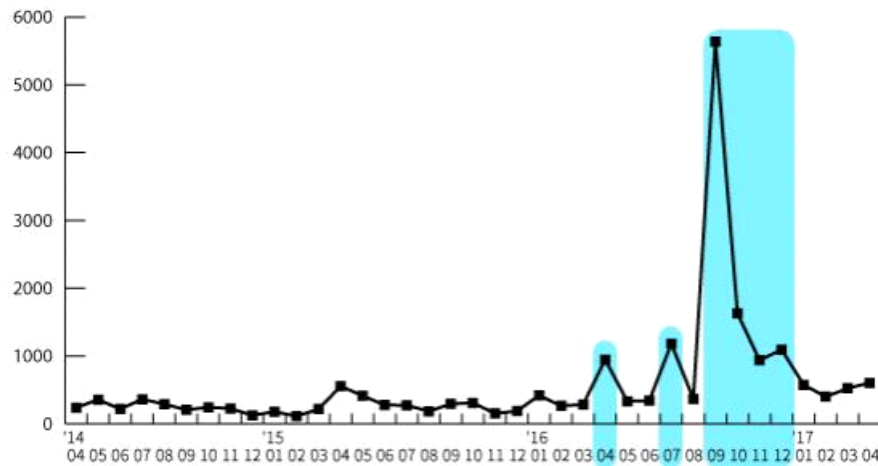
안전사고는 건축물에서 뿐 아니라 생활환경 전반에서 발생하는 경우도 많아 전체적으로 높은 버즈량을 유지하며 특정 이슈 발생 시 버즈량이 급증하는 것으로 나타났다. 뒤 이어 분석하게 될 침수, 지진, 화재, 붕괴 등과 연관성이 높은 키워드들이 나타났으며 세월호 참사 이후 안전교육, 예방을 위한 대책 등에 관심이 높은 것으로 보인다.

□ 지진에 대한 키워드 분석

지진에 대한 전체 버즈량은 20,677건이며 월 평균 버즈량인 559건 기준으로 일정한 버즈량을 유지하다 2016년 4월과 7월, 9~12월에 평균 버즈량을 넘어서는 현상이 나타났다.

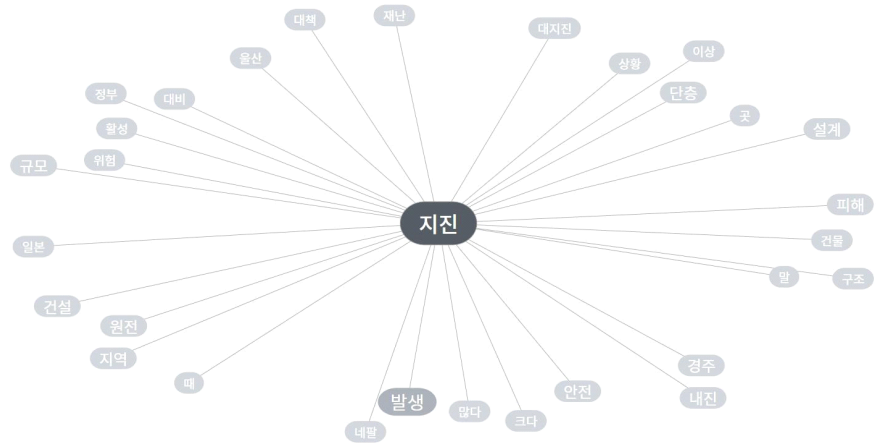


[그림 5-15] '지진'에 대한 키워드 분석 결과



[그림 5-16] '지진'에 대한 시계열 빈도

지진과 관련한 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%)으로 나타났으며, 그 외 지진이 일어난 지역명(울산, 네팔, 일본)이나 건축과 관련된 단어(건물, 구조, 설계, 건설), 지진과 관련한 학술적 용어(단층, 규모, 대지진), 해결 방안 관련 단어(재난, 대책, 정부, 대비)등이 언급되었다.



[그림 5-17] '지진'에 대한 연관키워드

지진과 관련하여 가장 많은 버즈가 발생한 채널은 블로그로 총 12,499(60.45%)건이었고, 뉴스7,354건(35.57%), 트위터 824건(3.99%) 순으로 나타났다. 블로그의 상위 키워드는 발생, 규모, 피해, 설계, 내진 등 지진 발생에 대한 키워드와 예방 대책 관련 키워드가 함께 나타났으며, 뉴스의 상위 키워드로는 발생, 규모, 원전, 안전, 피해 등 지진 발생과 관련된 정보에 대한 내용이 주로 나타났다. 트위터 상위 키워드의 경우 건설, 원전, 지역, 단층, 대비 등으로 타 채널에 비해 다양한 관점에서의 키워드를 보여주었다.

Blog 전체보기 60.4% (전체 20,677 중 12,499)				News 전체보기 35.6% (전체 20,677 중 7,354)				Twitter 전체보기 4.0% (전체 20,677 중 824)			
NO	연관키워드	버즈량	여론감성	NO	연관키워드	버즈량	여론감성	NO	연관키워드	버즈량	여론감성
1	발생	(-)1751	긍정 179 (10.22%) 중립 1247 (71.22%) 부정 325 (18.56%)	1	발생	(-)1613	긍정 174 (10.79%) 중립 1097 (68.01%) 부정 342 (21.20%)	1	건설	(-)358	긍정 29 (8.10%) 중립 246 (68.72%) 부정 83 (23.18%)
2	규모	(-)982	긍정 105 (10.69%) 중립 698 (71.08%) 부정 179 (18.23%)	2	규모	(-)895	긍정 86 (9.61%) 중립 623 (69.61%) 부정 186 (20.78%)	2	원전	(-)346	긍정 19 (5.49%) 중립 281 (81.21%) 부정 46 (13.29%)
3	피해	(-)834	긍정 140 (16.79%) 중립 502 (60.19%) 부정 192 (23.02%)	3	원전	(-)802	긍정 125 (15.59%) 중립 478 (59.60%) 부정 199 (24.81%)	3	한중일	(-)116	긍정 0 (0.00%) 중립 116 (100.00%) 부정 0 (0.00%)
4	설계	(-)802	긍정 154 (19.20%) 중립 545 (67.96%) 부정 103 (12.84%)	4	안전	(-)790	긍정 167 (21.14%) 중립 500 (63.29%) 부정 123 (15.57%)	4	지역	(-)89	긍정 7 (7.87%) 중립 73 (82.02%) 부정 9 (10.11%)
5	내진	(-)795	긍정 150 (18.87%) 중립 546 (68.68%) 부정 99 (12.45%)	5	피해	(-)786	긍정 117 (14.89%) 중립 507 (64.50%) 부정 162 (20.61%)	5	고리	(-)83	긍정 8 (9.64%) 중립 62 (74.71%) 부정 13 (15.66%)
6	안전	(-)790	긍정 142 (17.97%) 중립 540 (68.35%) 부정 108 (13.67%)	6	내진	(-)735	긍정 132 (17.96%) 중립 450 (61.22%) 부정 153 (20.82%)	6	단층	(-)80	긍정 6 (7.50%) 중립 54 (67.50%) 부정 20 (25.00%)
7	경주	(-)693	긍정 90 (12.99%) 중립 450 (64.94%) 부정 153 (22.08%)	7	경주	(-)714	긍정 87 (12.18%) 중립 485 (67.93%) 부정 142 (19.89%)	7	대비	(-)76	긍정 0 (0.00%) 중립 73 (96.05%) 부정 3 (3.95%)
8	지역	(-)676	긍정 80 (11.83%) 중립 451 (66.72%) 부정 145 (21.45%)	8	지역	(-)663	긍정 61 (9.20%) 중립 465 (70.14%) 부정 137 (20.66%)	8	발생	(-)76	긍정 4 (5.26%) 중립 61 (80.26%) 부정 11 (14.47%)
9	건설	(-)658	긍정 163 (24.77%) 중립 356 (54.10%) 부정 139 (21.12%)	9	설계	(-)617	긍정 122 (19.77%) 중립 383 (62.07%) 부정 112 (18.15%)	9	대책	(-)69	긍정 1 (1.45%) 중립 68 (98.55%) 부정 0 (0.00%)
10	크다	(-)636	긍정 94 (14.78%) 중립 389 (61.16%) 부정 153 (24.06%)	10	단층	(-)571	긍정 62 (10.86%) 중립 394 (69.00%) 부정 115 (20.14%)	10	규모	(-)69	긍정 9 (13.04%) 중립 39 (56.52%) 부정 21 (30.43%)

[그림 5-18] '지진'에 대한 데이터 수집 채널별 연관 키워드 순위

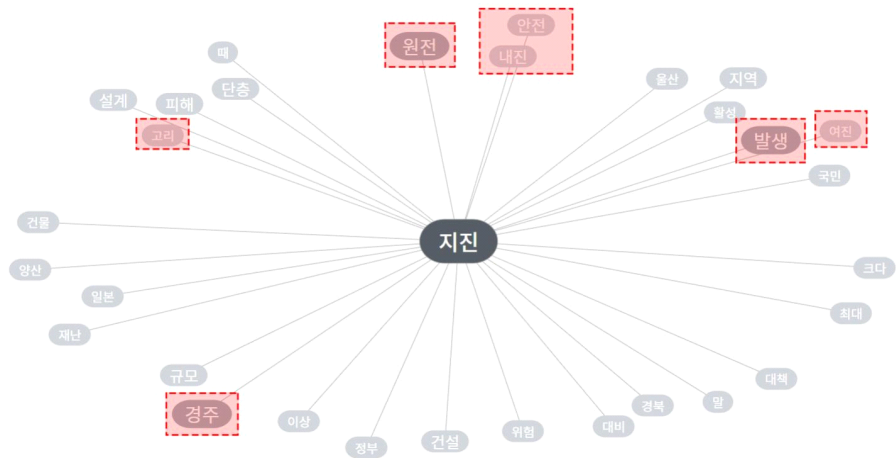
A hub-and-spoke diagram illustrating the concept of earthquakes ('지진'). The central node is '지진'. It is connected to numerous peripheral nodes, which are grouped into five color-coded categories:

- Red Dashed Boxes (Specific Events):**
 - 일본 지진 (Great East Japan Earthquake)
 - 일본 쓰나미 (Great East Japan Tsunami)
 - 구마모토 지진 (Great East Japan Earthquake)
- Pink Solid Boxes (General Concepts):**
 - 피해 (Casualties/Damage)
 - 발생 (Occurrence)
- Grey Rounded Rectangles (Other Categories):**
 - 원인 (Causes):** 원판 이동, 판 경계, 대륙, 내진, 모도.
 - 영향 (Effects):** 설계, 단층, 피해.
 - 관련 분야 (Related Fields):** 정부, 사람, 쿠마, 크다, 건물, 건설, 배, 도로, 구조, 지원.
 - 지역 (Locations):** 발, 안전, 대책, 규모, 국, 지역, 우, 구마모토, 대지진.

지진

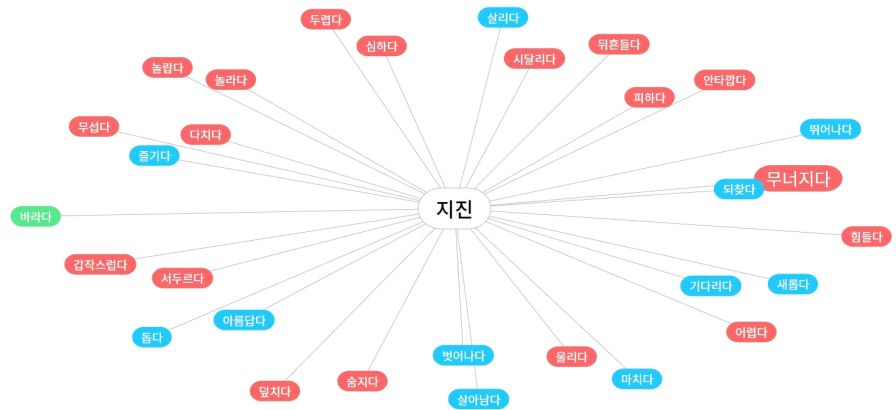
- 발생
- 양바다
- 발생
- 밀집
- 조사
- 이상
- 건물
- 정도
- 피해
- 환만도
- 평가
- 원전
- 안전
- 규모
- 부산
- 고리
- 위험
- 일본
- 건설
- 설계
- 크다
- 울산
- 지역
- 단층
- 중단
- 북한

102 빅데이터를 활용한 건축·도시 미래정책 개발체계 연구



[그림 5-21] 2016년 9~12월 연관 키워드

지진에 대한 감성 분석 결과, 긍정10.34%, 중립 71.32%, 부정 18.34%의 비율로 나타났으며, 중립 감성을 제외하면 부정 감성이 긍정 감성의 2배 가까이 높았다. 긍정 감성이 상위에는 지진 발생 후 복구, 지진 대응 대책관련 키워드인 새롭다, 돕다, 아름답다, 뛰어나다, 벗어나다, 살아남다, 되찾다 등이 나타났으며, 부정 감성이 상위에는 무너지다, 뒤 흔들다, 힘들다, 심하다, 다치다, 숨지다 등 지진이 야기한 현상들에 대한 키워드가 나타났다.



[그림 5-22] '지진'에 대한 감성 분석 결과(빨강: 부정, 파랑: 긍정, 녹색: 중립)

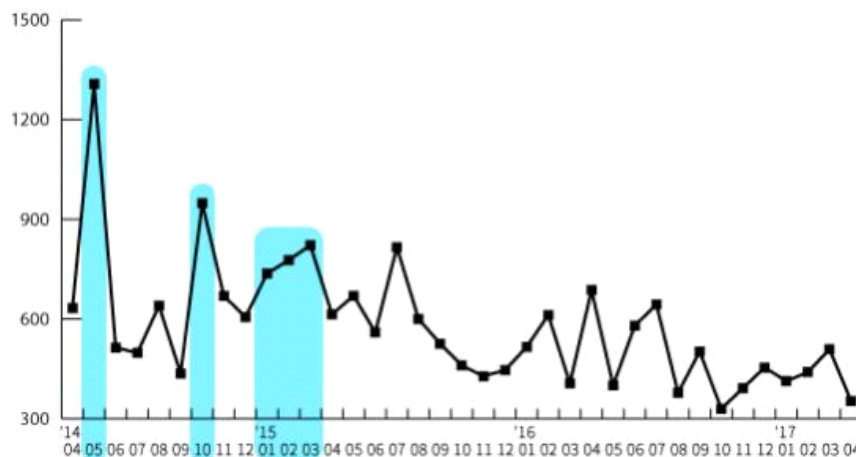
지진에 대한 키워드 분석 결과, 국내에서 전에 없던 큰 규모의 지진이 여러번 발생하면서 방재 대책에 대한 국민적 요구가 높아졌다고 할 수 있다. 특히 지진에 취약한 건물의 내진 설계 등 건축도시 분야에서의 정책 마련이 주요한 이슈로 도출 되었다.

□ 붕괴에 대한 키워드 분석

건축물 붕괴에 대한 전체 버즈량은 21,323건이며, 월 평균 버즈량은 576건이나 버즈가 발생할 때와 그렇지 않을 때의 편차가 큰 것으로 나타났다. 또한 2014년 5월, 10월 그리고 2015년 1~3월의 버즈량이 급증하였다.

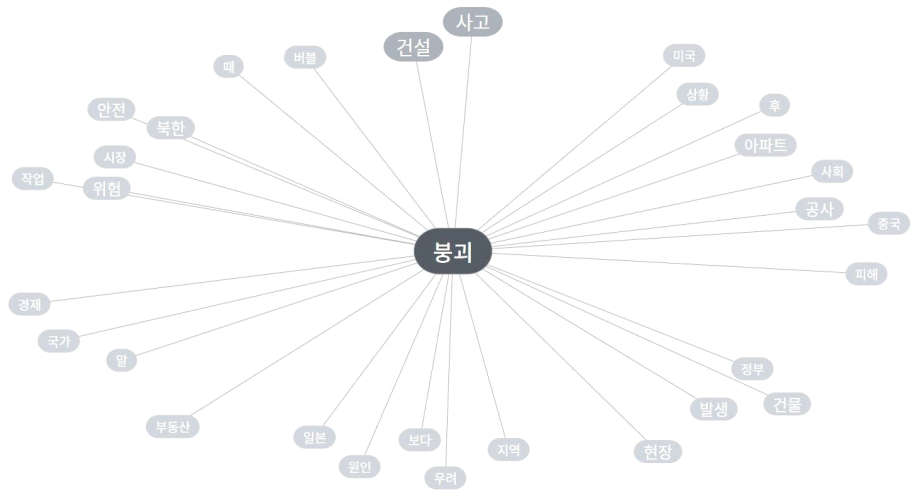


[그림 5-23] '붕괴'에 대한 키워드 분석 결과



[그림 5-24] '붕괴'에 대한 시계열 빈도

붕괴와 관련한 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%)로 나타났으며, 그 외에 작업, 피해, 우려, 원인 등의 키워드가 함께 언급되었다.



[그림 5-25] '붕괴'에 대한 연관키워드

붕괴와 관련하여 가장 많은 버즈가 발생한 채널은 블로그로 총 16,184건(75.9%)이었으며, 뉴스 4,456건(20.9%), 트위터 683건(3.2%)으로 나타났다. 채널별 출현 빈도수가 높았던 연관키워드는 사고, 현장, 건설, 북한 등으로 거의 유사하게 나타났다.

Blog 전체보기 75.9% (전체 21,323 중 16,184)				News 전체보기 20.9% (전체 21,323 중 4,456)				Twitter 전체보기 3.2% (전체 21,323 중 683)			
NO	연관키워드	버즈량	여론감성	NO	연관키워드	버즈량	여론감성	NO	연관키워드	버즈량	여론감성
1	사고	(-)1580	긍정 166 (10.51%) 중립 926 (58.61%) 부정 488 (30.89%)	1	사고	(-)1282	긍정 115 (8.97%) 중립 766 (59.75%) 부정 401 (31.28%)	1	건설	(-)322	긍정 32 (9.94%) 중립 195 (60.56%) 부정 95 (29.50%)
2	건설	(-)1234	긍정 206 (16.69%) 중립 704 (57.05%) 부정 324 (26.26%)	2	현장	(-)710	긍정 58 (8.17%) 중립 415 (58.45%) 부정 237 (33.38%)	2	사고	(-)171	긍정 8 (4.68%) 중립 115 (67.25%) 부정 48 (28.07%)
3	발생	(-)1068	긍정 138 (12.92%) 중립 704 (60.86%) 부정 280 (26.22%)	3	건설	(-)677	긍정 88 (13.00%) 중립 405 (59.82%) 부정 184 (27.18%)	3	아파트	(-)129	긍정 6 (4.65%) 중립 73 (56.59%) 부정 50 (38.76%)
4	북한	(-)1024	긍정 194 (18.95%) 중립 546 (53.32%) 부정 284 (27.73%)	4	발생	(-)663	긍정 57 (8.60%) 중립 440 (66.37%) 부정 166 (25.04%)	4	현장	(-)106	긍정 6 (5.66%) 중립 66 (62.26%) 부정 34 (32.08%)
5	공사	(-)771	긍정 78 (10.12%) 중립 464 (60.18%) 부정 229 (29.70%)	5	아파트	(-)566	긍정 67 (11.84%) 중립 357 (63.07%) 부정 142 (25.09%)	5	던	(-)70	긍정 4 (5.71%) 중립 65 (92.86%) 부정 1 (1.43%)
6	위험	(-)761	긍정 109 (14.32%) 중립 472 (62.02%) 부정 180 (23.65%)	6	공사	(-)525	긍정 72 (13.71%) 중립 286 (54.48%) 부정 167 (31.81%)	6	북한	(-)68	긍정 9 (13.24%) 중립 39 (57.35%) 부정 20 (29.41%)
7	현장	(-)739	긍정 77 (10.42%) 중립 451 (61.84%) 부정 205 (27.74%)	7	안전	(-)518	긍정 85 (16.41%) 중립 322 (62.16%) 부정 111 (21.43%)	7	평양	(-)67	긍정 3 (4.48%) 중립 40 (59.70%) 부정 24 (35.82%)
8	안전	(-)720	긍정 131 (18.19%) 중립 441 (61.25%) 부정 148 (20.56%)	8	위험	(-)433	긍정 62 (14.32%) 중립 280 (64.67%) 부정 91 (21.02%)	8	연천	(-)64	긍정 0 (0.00%) 중립 63 (98.44%) 부정 1 (1.56%)
9	미국	(-)698	긍정 142 (20.34%) 중립 335 (47.99%) 부정 221 (31.66%)	9	북한	(-)384	긍정 60 (15.63%) 중립 219 (57.03%) 부정 105 (27.34%)	9	주소	(-)60	긍정 0 (0.00%) 중립 60 (100.00%)
10	건물	(-)640	긍정 91 (14.22%) 중립 320 (50.00%) 부정 229 (35.78%)	10	건물	(-)344	긍정 36 (10.47%) 중립 177 (51.45%) 부정 131 (38.08%)	10	시장	(-)53	긍정 0 (0.00%) 중립 37 (69.81%) 부정 16 (30.19%)

[그림 5-26] '지진'에 대한 데이터 수집 채널별 연관 키워드 순위

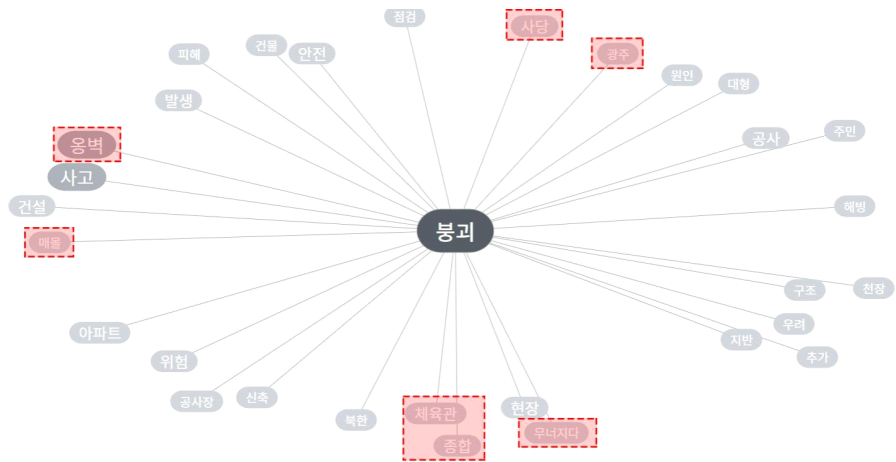
버즈량이 급증하는 2014년 5월에는 북한의 고층 아파트가 붕괴되면서 버즈량이 1,307건으로 급증하였으며 북한, 평양, 평양시, 평천구역, 안산, 아파트, 고층 등 붕괴 건물에

[illegible]

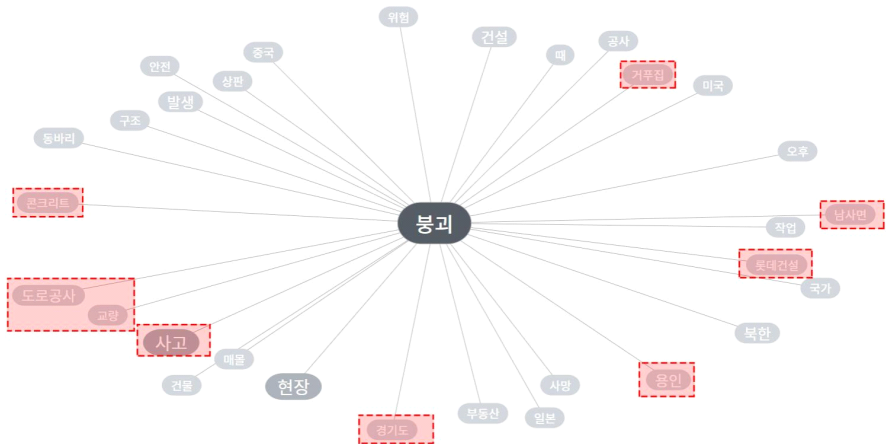
```

graph LR
    붕괴((붕괴)) --- 성수대교_붕괴[성수대교 붕괴]
    붕괴 --- 공사[공사]
    붕괴 --- 터널[터널]
    붕괴 --- 월드[월드]
    붕괴 --- 구조[구조]
    붕괴 --- 관교[관교]
    붕괴 --- 관철[관철]
    붕괴 --- 시공[시공]
    붕괴 --- 문명[문명]
    붕괴 --- 환[환]
    붕괴 --- 배[배]
    붕괴 --- 건설[건설]
    붕괴 --- 망[망]
    붕괴 --- 무너지다[무너지다]
    붕괴 --- 막다[막다]
    붕괴 --- 댐[댐]
    붕괴 --- 건물[건물]
    붕괴 --- 미국[미국]
    붕괴 --- 관교[관교]
    붕괴 --- 코스피[코스피]
    붕괴 --- 안전[안전]
    붕괴 --- 부실[부실]
    붕괴 --- 성수대교[성수대교]
    붕괴 --- 관리[관리]
    붕괴 --- 사고[사고]
    붕괴 --- 보타[보타]
    붕괴 --- 위험[위험]
    붕괴 --- 후[후]
    붕괴 --- 발생[발생]
    붕괴 --- 전[전]
    붕괴 --- 현장[현장]
    붕괴 --- 복합[복합]
    붕괴 --- 시공[시공]
  
```

106 빅데이터를 활용한 건축·도시 미래정책 개발체계 연구

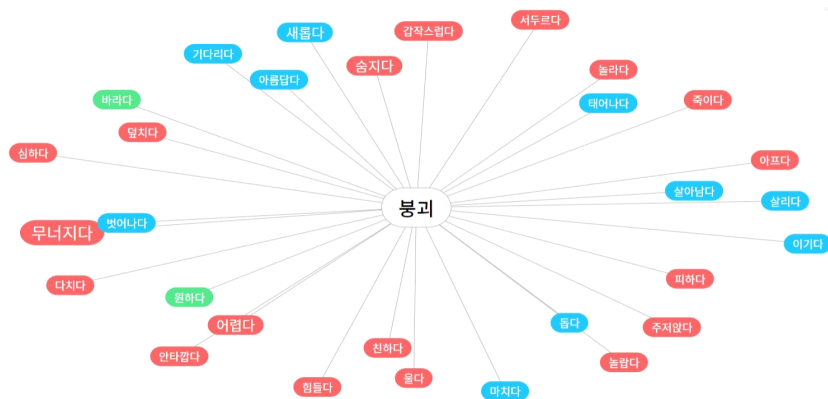


[그림 5-29] 2015년 1~2월 연관 키워드



[그림 5-30] 2015년 3월 연관 키워드

붕괴에 대한 긍부정 여론은 긍정10.04%, 중립 67.18%, 부정 22.79%로 나타났으며, 중립 감성을 제외하고 부정 감성이 긍정 감성의 2배 이상으로 높은 비율을 차지하는 것으로 나타났다. 긍정 감성어 상위에는 붕괴 현장 수습과 관련한 돕다, 살리다가 나타났으며 부정 감성어 상위에는 무너지다, 숨지다, 어렵다, 갑작스럽다, 다치다 등 붕괴 현장을 표현하거나 그 피해와 관련된 키워드가 나타났다.



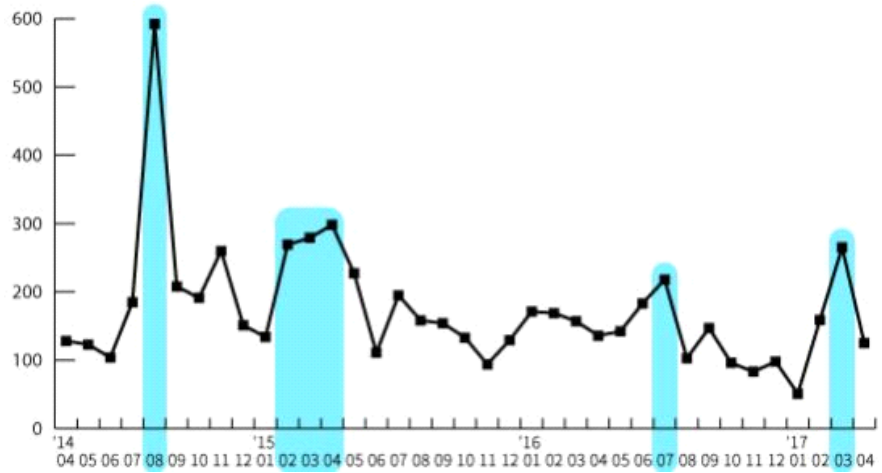
건축물 붕괴 사고가 발생할 때 마다 버즈량이 증가는 양상을 보이며 안전, 위험, 사고, 피해, 정부, 우려, 원인 등의 키워드로 미루어 보아 붕괴 원인 규명과 예방 대책에 대한 관심이 높다고 볼 수 있다. 유사 사고가 반복되고, 사고 발생 시 마다 대책 마련에 대한 키워드가 높은 빈도로 등장하는 것으로 보아, 붕괴 사고 예방을 위한 대책이 아직 미흡하다고 판단된다.

□ 지반침하에 대한 키워드 분석

지반침하에 대한 전체 버즈량은 6,420건이며 월 평균 버즈량은 174건이나 시기에 따라 버즈량이 높은 폭의 차이를 보이는데, 2014년 8월과 2015년 2~4월, 2016년 7월과 2017년 3월 네 차례 정도의 버즈량 급증 구간이 나타났다.

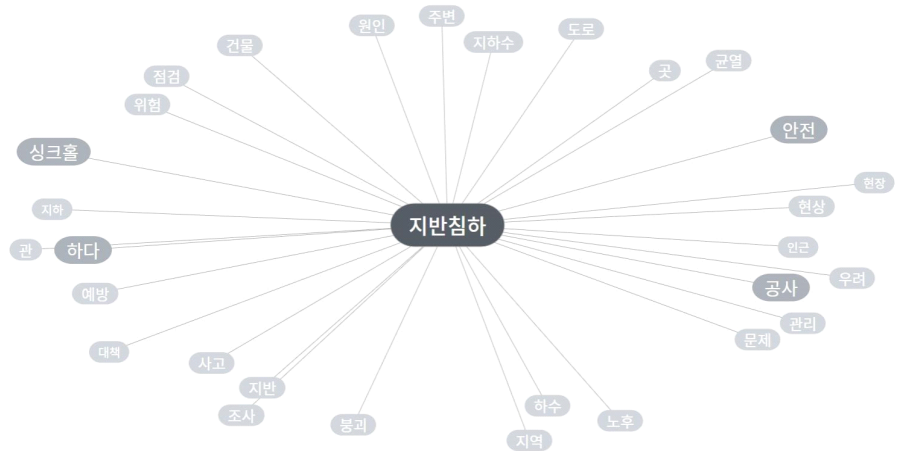


[그림 5-32] ‘지반침하’에 대한 키워드 분석 결과



[그림 5-33] '지반침하'에 대한 시계열 빈도

지반침하와 함께 등장하는 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%), 점검(7.8%), 사고(7.4%) 순으로 높은 빈도를 보였으며, 그 외에 지하수, 원인, 지하, 붕괴, 도로, 균열 등 지반침하의 원인이나 지반 침하 시 나타나는 현상 등에 대한 키워드와 예방, 관리, 우려, 대책 등의 지반 침하 대책 마련 요구의 여론을 보여주는 키워드가 함께 언급되었다.



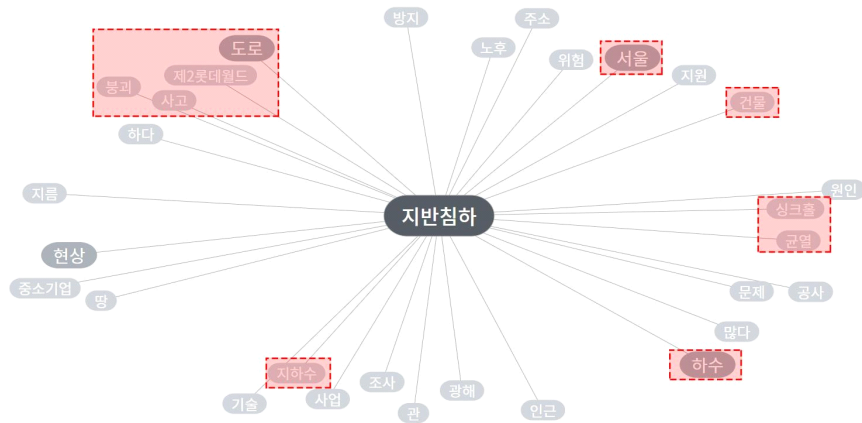
[그림 5-34] '지반침하'에 대한 연관키워드

지반침하와 관련하여 가장 많은 버즈가 발생한 채널은 블로그로 총 2,910건(45.33%) 이었고, 뉴스 2,153건(33.54%), 트위터 1,357건(21.14%) 으로 3개 채널에서 고르게 나타났다.

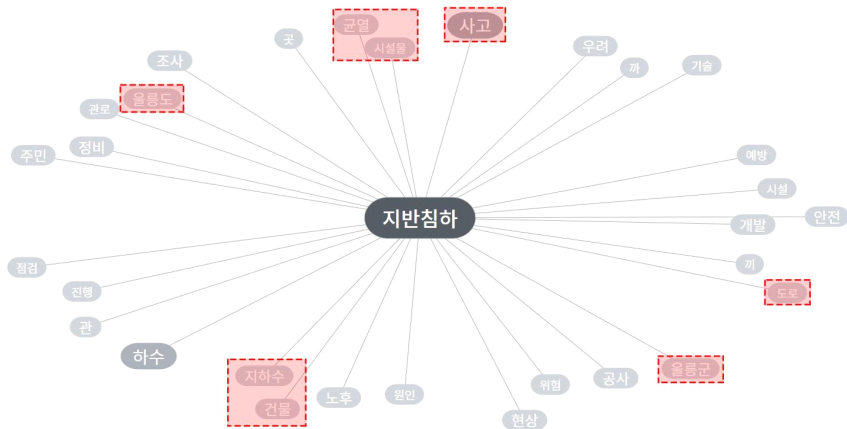
버즈량이 급증하는 2014년 8월에는 송파구 일대에서 싱크홀이 발생하면서 592건의 버즈량이 나타났는데, 싱크홀, 동공 등의 지반침하 현상에 대한 키워드와 석촌호수, 제2롯데월드, 잠실 등 싱크홀 발견 장소에 대한 키워드, 그리고 지하수, 공사, 지하철도 등 싱크홀의 원인으로 추측되는 키워드들이 언급되었다. 2015년 2~4월에는 용산과 신촌 등 서울 시내 곳곳에서 싱크홀이 발견되면서 846건의 버즈량이 나타났는데, 용산, 신촌, 강남 등 지반침하 현상이 발견된 지역과 도로, 공사, 토사, 공사장, 사고, 지반, 침하, 싱크홀 등 발생 현장 관련 키워드가 언급되었다. 2016년 7월에는 제2롯데월드 인근에서 또 싱크홀이 발견되면서 버즈량이 218건으로 급증하였는데, 붕괴, 싱크홀, 균열, 사고 등 싱크홀 관련 키워드와 안전이 우려되는 도로, 건물 등도 함께 나타났다. 2017년 3월에는 울릉도에 지반침하 현상이 발생하여 265건의 버즈량이 나타났는데, 시설물, 건물, 균열, 도로 등 피해 현황에 대한 키워드와 원인으로 유추되는 지하수 등이 키워드로 나타났다.

[그림 5-35] 2014년 8월 연관 키워드

[그림 5-36] 2015년 2~4월 연관 키워드



[그림 5-37] 2016년 7월 연관 키워드

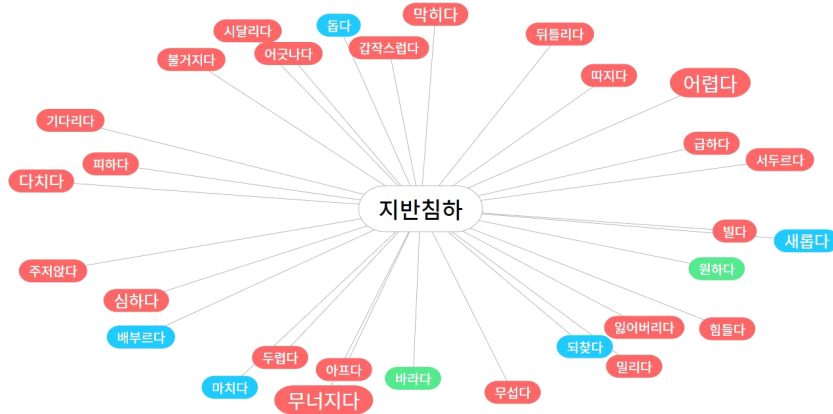


[그림 5-38] 2017년 3월 연관 키워드

지반침하에 대한 감성 분석 결과, 긍정 8.72%, 중립 72.35%, 부정 18.93%로 중립 감성의 비율이 높았으며, 중립 감성을 제외하고 부정 감성이 긍정 감성보다 2배 높게 나타났다. 긍정 감성어 상위에는 마치다가 있었으며, 부정 감성어 상위에는 무너지다, 어렵다, 심하다, 다치다, 주저앉다, 뒤틀리다 등 지반침하 현상에 대한 부정적인 내용에서 추출된 키워드가 나타났다.

지반 침하에 대한 국민 관심도는 2014년 8월 송파구 일대 싱크홀이 발견되면서 증가하기 시작하였다. 이후로 사건 발생에 따라 증감 폭이 크게 나타나긴 했으나, 2014년 8월 이전보다는 버즈량이 꾸준히 증가했다고 볼 수 있다. 또한 지하수에 의한 지반침하 현상

인 경우가 많아 해빙기나 비가 많이 오는 시기에 관심도가 높아지는 양상을 보였다. 시설, 건물, 건축물 등이 꾸준히 연관 키워드로 등장하는 것으로 보아 지반침하 시 시설물, 건축물에 미치는 영향에 대한 관심이 높은 것으로 보이며, 지반 침하에 대응한 건축물 안전관리 방안 대책이 필요한 것으로 판단된다.



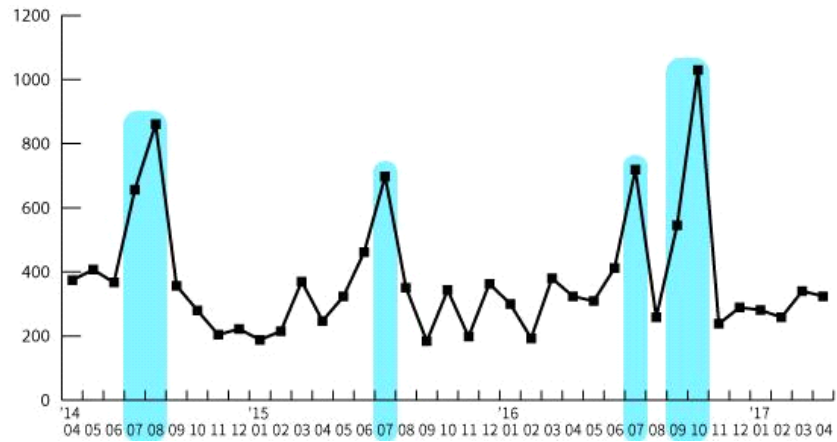
[그림 5-39] '지반침하'에 대한 감성 분석 결과(빨강: 부정, 파랑: 긍정, 녹색: 중립)

□ 침수에 대한 키워드 분석

침수에 대한 전체 버즈량은 13,851건이며, 월 평균 374건 기준으로 증감 폭이 큰 것으로 나타났으며, 매년 강우량이 많아지는 7월 전후로 버즈량이 급증하는 패턴을 보인다.

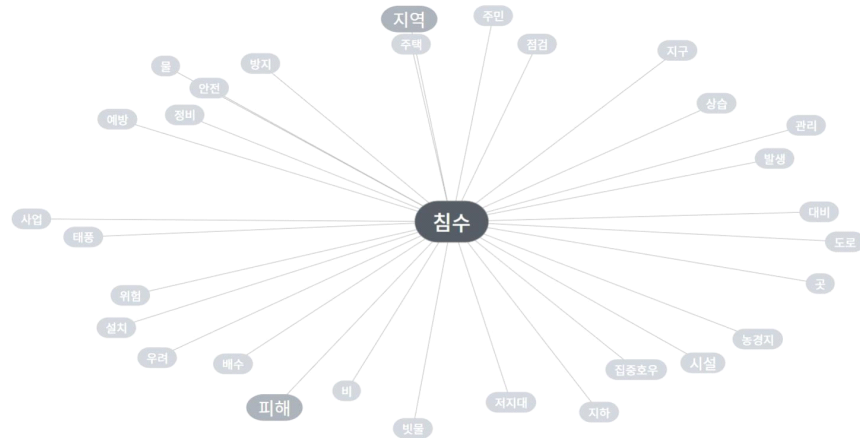


[그림 5-40] '침수'에 대한 키워드 분석 결과



[그림 5-41] '침수'에 대한 시계열 빈도

침수와 관련한 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%)으로 나타났으며, 그 외에 침수 발생 원인에 대한 키워드(태풍, 집중호우), 침수 발생 장소(주택, 지하, 저지대), 침수 대비 시설(배수, 정비, 점검, 관리, 설치)등과 관련된 키워드들이 함께 언급되었다.



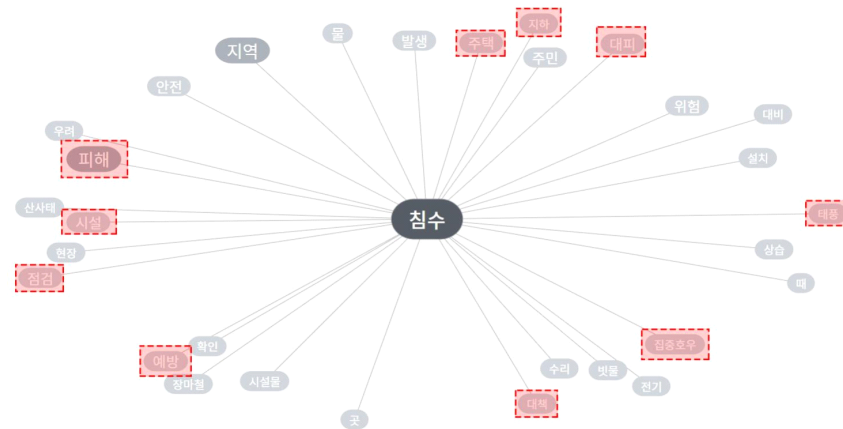
[그림 5-42] '침수'에 대한 연관키워드

침수와 관련한 버즈는 블로그에서 총 9,719건(70.17%)으로 가장 높게 나타났으며 뒤이어 뉴스3,720건(26.86%), 트위터 412건(2.97%) 순으로 나타났다. 블로그와 뉴스의 상위 키워드는 피해, 지역, 시설, 발생, 예방 등 침수 발생과 예방에 대한 내용이었으며, 트위터 상위 키워드는 블로그, 뉴스와 동일한 키워드도 있지만 침수 피해 방지시설사업과 관련한 신청, 접수, 과천시 등도 높게 나타났다.

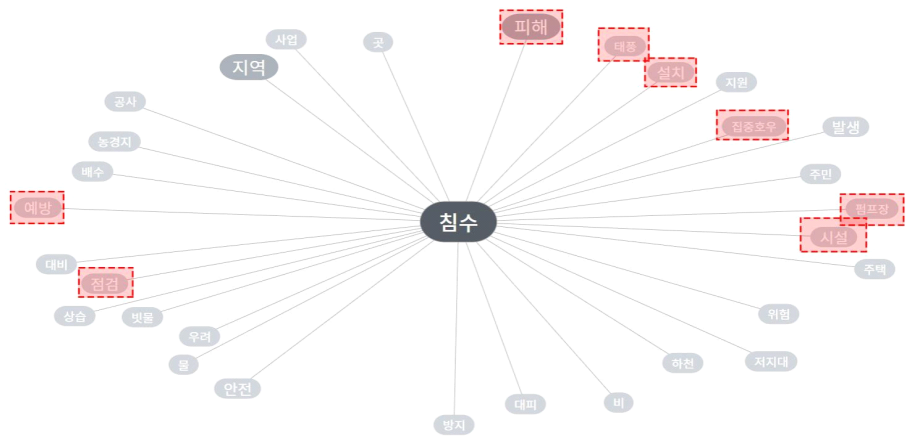
Blog 전체보기 70.2% (전체 13,851 중 9,719)				News 전체보기 26.9% (전체 13,851 중 3,720)				Twitter 전체보기 3.0% (전체 13,851 중 412)			
NO	연관키워드	비즈량	여론감성	NO	연관키워드	비즈량	여론감성	NO	연관키워드	비즈량	여론감성
1	피해	(-)2035	긍정 347 (17.05%) 중립 1360 (66.83%) 부정 328 (16.12%)	1	피해	(-)1617	긍정 310 (19.17%) 중립 1105 (68.34%) 부정 202 (12.49%)	1	시설	(-)222	긍정 16 (7.21%) 중립 188 (84.68%) 부정 18 (8.11%)
2	지역	(-)1718	긍정 333 (19.38%) 중립 1175 (68.39%) 부정 210 (12.22%)	2	지역	(-)979	긍정 215 (21.96%) 중립 648 (66.19%) 부정 116 (11.85%)	2	피해	(-)157	긍정 9 (5.73%) 중립 137 (87.26%) 부정 11 (7.01%)
3	시설	(-)1270	긍정 245 (19.29%) 중립 879 (69.21%) 부정 146 (11.50%)	3	시설	(-)781	긍정 166 (21.25%) 중립 526 (67.35%) 부정 89 (11.40%)	3	방치	(-)126	긍정 2 (1.59%) 중립 124 (98.41%) 부정 0 (0.00%)
4	발생	(-)756	긍정 94 (12.43%) 중립 513 (67.86%) 부정 149 (19.71%)	4	사업	(-)517	긍정 213 (41.20%) 중립 276 (53.38%) 부정 28 (5.42%)	4	사업	(-)95	긍정 2 (2.11%) 중립 93 (97.89%) 부정 0 (0.00%)
5	설치	(-)689	긍정 116 (16.84%) 중립 522 (75.76%) 부정 51 (7.40%)	5	발생	(-)456	긍정 78 (17.11%) 중립 311 (68.20%) 부정 67 (14.69%)	5	신천	(-)83	긍정 1 (1.20%) 중립 81 (97.59%) 부정 1 (1.20%)
6	물	(-)678	긍정 105 (15.49%) 중립 422 (62.24%) 부정 151 (22.27%)	6	주택	(-)399	긍정 51 (12.78%) 중립 270 (67.67%) 부정 78 (19.55%)	6	접수	(-)81	긍정 0 (0.00%) 중립 81 (100.00%) 부정 0 (0.00%)
7	사업	(-)615	긍정 228 (37.07%) 중립 335 (54.47%) 부정 52 (8.46%)	7	예방법	(-)390	긍정 91 (23.33%) 중립 272 (69.74%) 부정 27 (6.92%)	7	과천시	(-)79	긍정 0 (0.00%) 중립 77 (97.47%) 부정 2 (2.53%)
8	예방법	(-)581	긍정 148 (25.47%) 중립 397 (68.33%) 부정 36 (6.20%)	8	상습	(-)379	긍정 131 (34.56%) 중립 211 (55.67%) 부정 37 (9.76%)	8	상습	(-)52	긍정 2 (3.85%) 중립 49 (94.23%) 부정 1 (1.92%)
9	상습	(-)547	긍정 142 (25.96%) 중립 339 (61.97%) 부정 66 (12.07%)	9	설치	(-)342	긍정 75 (21.93%) 중립 244 (71.35%) 부정 23 (6.73%)	9	설치	(-)48	긍정 3 (6.25%) 중립 44 (91.67%) 부정 1 (2.08%)
10	배수	(-)482	긍정 105 (21.78%) 중립 310 (64.32%) 부정 67 (13.90%)	10	집중호우	(-)331	긍정 51 (15.41%) 중립 241 (72.81%) 부정 39 (11.78%)	10	주소	(-)43	긍정 0 (0.00%) 중립 43 (100.00%) 부정 0 (0.00%)

[그림 5-43] '침수'에 대한 데이터 수집 채널별 연관 키워드 순위

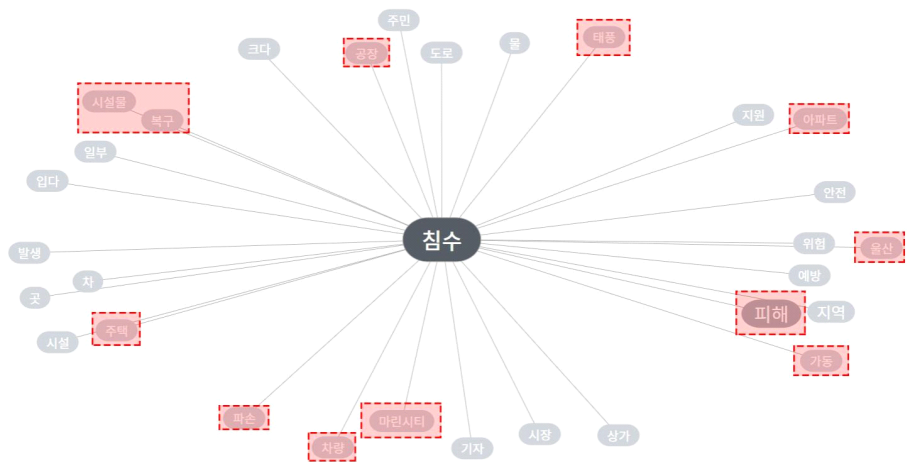
태풍과 집중호우가 발생하는 여름철에 비즈니스가 급증하는 양상을 나타내는데, 2014년 7~8월(각 656건, 861건)에는 침수의 원인인 집중호우, 침수 현장과 관련된 주택, 지하, 피해, 대피 등이 나타났으며, 그 외에 대책, 시설, 점검, 예방 등 침수 대비와 관련된 키워드가 언급되었다. 2015년 7월(698건)에는 침수 예방 시설인 빗물 펌프장의 사전 점검과 관련한 키워드가 새롭게 등장하였으며, 2016년 7~10월(각 719건, 544건, 1,030건)에는 기록적인 폭우가 내리면서 피해가 컸던 마린시티, 울산, 차량, 아파트, 공장, 주택, 파손 등 피해와 관련된 키워드가 많이 언급 되었다.



[그림 5-44] 2014년 7~8월 연관 키워드



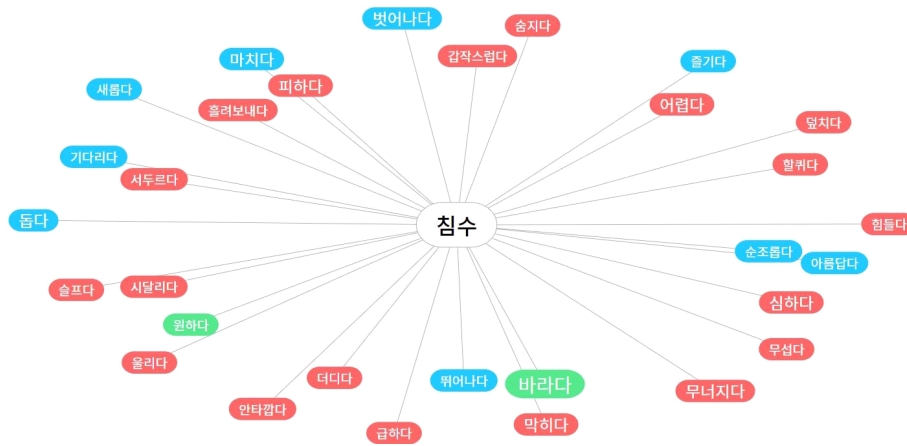
[그림 5-45] 2015년 7월 연관 키워드



[그림 5-46] 2016년 7~10월 연관 키워드

침수에 대한 감성분석 결과, 긍정12.55%, 중립 75.01%, 부정 12.45%로 중립 감성을 제외한 긍/부정 감성이 거의 비슷하게 나타났다. 긍정 감성어 상위에는 벗어나다, 마치다, 돕다, 피하다, 순조롭다 등 침수 피해 복구와 관련된 콘텐츠에서 추출된 키워드가 나타났다. 부정 감성어 상위에는 침수에 의한 피해 현황을 나타내주는 무너지다, 어렵다, 막히다, 심하다, 숨지다 등의 키워드가 추출되었다.

침수는 매년 여름 발생하는 이슈이다 보니 침수 예방이나 피해와 관련된 안전, 시설물, 수리, 설치, 대비, 배수, 상습, 저지대, 하천, 방지, 복구, 주택, 상가, 주민 등 다양한 키워드가 나타나고 있다. 특히 최근 들어 집중호우가 증가하면서 빗물펌프장 등 우수 관리를 위한 키워드가 등장하는 것으로 보아 우수 관리 정책의 변화가 필요한 것으로 판단된다.



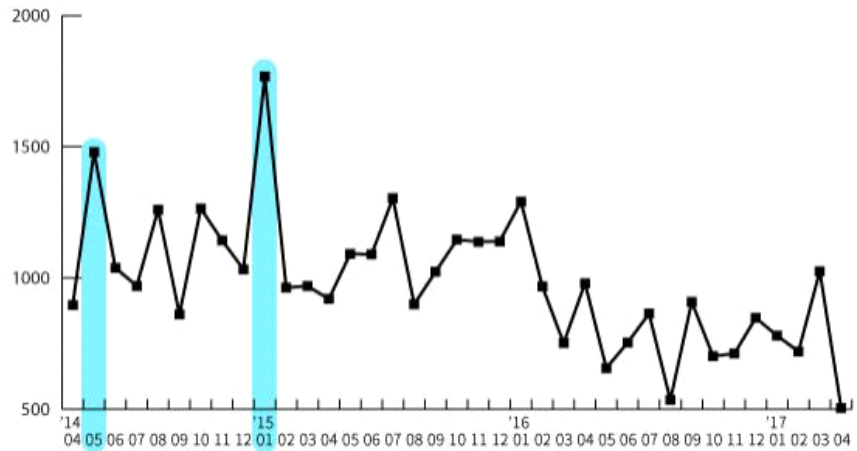
[그림 5-47] '침수'에 대한 감성 분석 결과(빨강: 부정, 파랑: 긍정, 녹색: 중립)

□ 화재에 대한 키워드 분석

화재에 대한 전체 버즈량은 36,403건, 월 평균 버즈량은 984건 이었으며, 버즈량 급증 현상은 2014년 5월, 2015년 1월 두차례 나타났다.

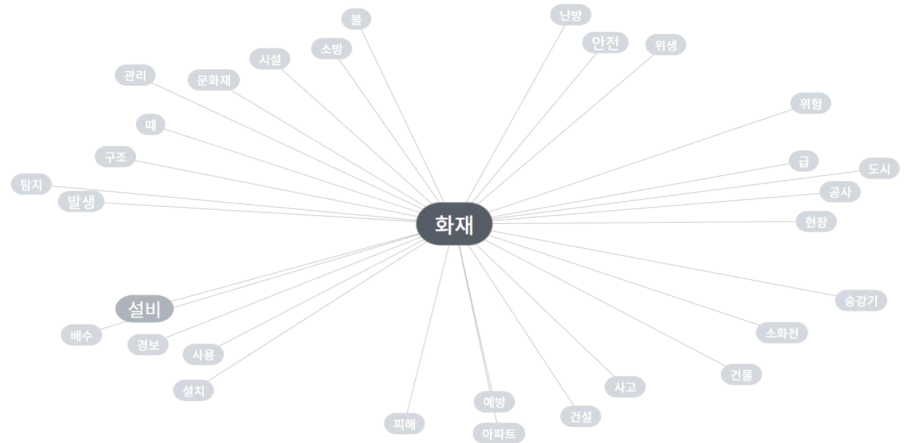


[그림 5-48] '화재'에 대한 키워드 분석 결과



[그림 5-49] '화재'에 대한 시계열 빈도

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%), 소방(2.9%), 시설(2.9%) 등으로 나타났으며 설비가 가장 상위권에 나타나는 것으로 보아 화재 예방을 위한 설비시설에 대한 관심이 높은 것으로 보인다.



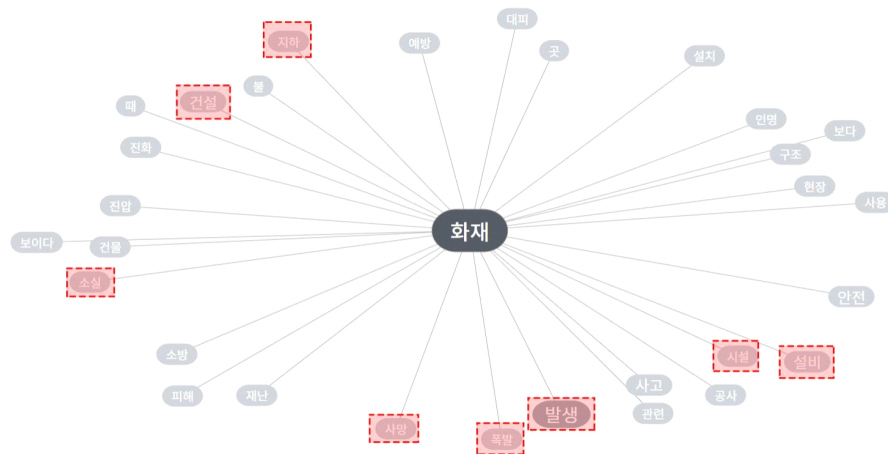
[그림 5-50] '화재'에 대한 연관키워드

화재와 관련하여 가장 많은 버즈가 발생한 채널은 블로그로 총 29,947건(82.27%)이었고, 뉴스 6,113건(16.79%), 트위터 343건(0.94%)이었으며, 채널별 연관키워드는 차이가 없는 것으로 나타났다.

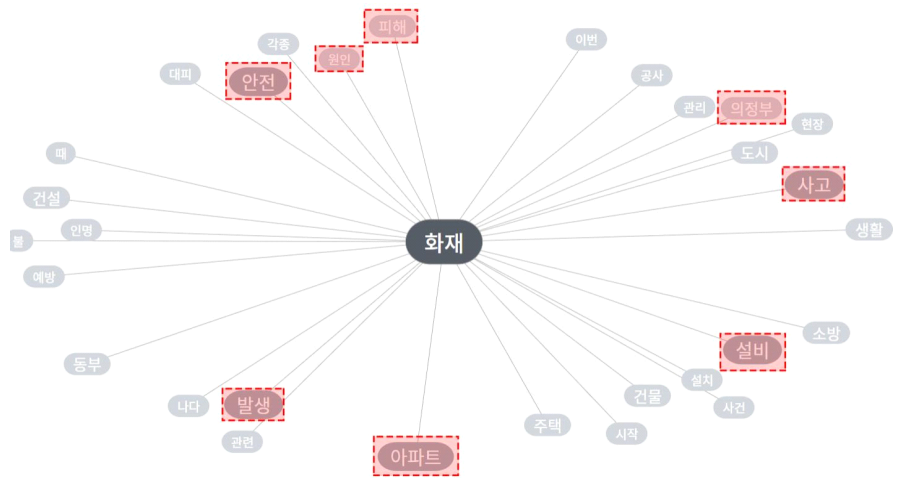
Blog 전체보기 82.3% (전체 36,403 중 29,947)				News 전체보기 16.8% (전체 36,403 중 6,113)				Twitter 전체보기 0.9% (전체 36,403 중 343)			
NO	연관키워드	비즈량	여론감성	NO	연관키워드	비즈량	여론감성	NO	연관키워드	비즈량	여론감성
1	설비	(-)5770	긍정 97 (1.68%) 중립 5569 (96.52%) 부정 104 (1.80%)	1	발생	(-)916	긍정 91 (9.93%) 중립 558 (60.92%) 부정 267 (29.15%)	1	건설	(-)161	긍정 21 (13.04%) 중립 127 (78.88%) 부정 13 (8.07%)
2	발생	(-)2182	긍정 277 (12.69%) 중립 1378 (63.15%) 부정 527 (24.15%)	2	안전	(-)732	긍정 116 (15.85%) 중립 506 (69.13%) 부정 110 (15.03%)	2	안전	(-)61	긍정 11 (18.03%) 중립 48 (78.69%) 부정 2 (3.28%)
3	안전	(-)1658	긍정 377 (22.74%) 중립 1067 (64.35%) 부정 214 (12.91%)	3	사고	(-)558	긍정 73 (13.08%) 중립 359 (64.34%) 부정 126 (22.58%)	3	현장	(-)59	긍정 7 (11.86%) 중립 42 (71.19%) 부정 10 (16.95%)
4	사고	(-)1203	긍정 179 (14.88%) 중립 748 (62.18%) 부정 276 (22.94%)	4	건설	(-)494	긍정 78 (15.79%) 중립 339 (68.62%) 부정 77 (15.59%)	4	주소	(-)37	긍정 0 (0.00%) 중립 1 (2.70%) 부정 36 (97.30%)
5	건물	(-)1189	긍정 173 (14.55%) 중립 705 (59.29%) 부정 311 (26.16%)	5	현장	(-)395	긍정 45 (11.39%) 중립 260 (65.82%) 부정 90 (22.78%)	5	의뢰	(-)35	긍정 4 (11.43%) 중립 31 (88.57%) 부정 0 (0.00%)
6	건설	(-)1129	긍정 195 (17.27%) 중립 708 (62.71%) 부정 226 (20.02%)	6	소방	(-)272	긍정 69 (25.37%) 중립 158 (58.09%) 부정 45 (16.54%)	6	소방	(-)33	긍정 10 (30.30%) 중립 23 (69.70%) 부정 0 (0.00%)
7	위생	(-)1080	긍정 18 (1.67%) 중립 1056 (97.78%) 부정 6 (0.56%)	7	건물	(-)270	긍정 32 (11.85%) 중립 167 (61.85%) 부정 71 (26.30%)	7	충남도	(-)32	긍정 4 (12.50%) 중립 28 (87.50%) 부정 0 (0.00%)
8	급	(-)919	긍정 29 (3.16%) 중립 882 (95.97%) 부정 8 (0.87%)	8	삼성전자	(-)264	긍정 10 (3.79%) 중립 229 (86.74%) 부정 25 (9.47%)	8	현대건설	(-)30	긍정 2 (6.67%) 중립 26 (86.67%) 부정 2 (6.67%)
9	승강기	(-)873	긍정 10 (1.15%) 중립 853 (97.71%) 부정 10 (1.15%)	9	피해	(-)260	긍정 54 (20.77%) 중립 149 (57.31%) 부정 57 (21.92%)	9	패	(-)27	긍정 0 (0.00%) 중립 27 (100.00%) 부정 0 (0.00%)
10	배수	(-)861	긍정 13 (1.51%) 중립 843 (97.91%) 부정 5 (0.58%)	10	LG	(-)259	긍정 21 (8.11%) 중립 198 (76.45%) 부정 40 (15.44%)	10	위원회	(-)26	긍정 5 (19.23%) 중립 21 (80.77%) 부정 0 (0.00%)

[그림 5-51] '화재'에 대한 데이터 수집 채널별 연관 키워드 순위

버즈량이 급증하는 2014년 5월(1,480건), 2015년 1월(1,767건)에는 각 고양터미널 화재와 의정부 도시형 생활주택 화재가 발생하였다. 두 시기에 발생, 설비, 사고 등 화재 발생과 관련한 키워드가 공통적으로 언급 되었으며, 2015년 1월 의정부 화재의 경우 130명의 사상자를 낸 큰 규모의 화재이다 보니 원인, 피해, 인명, 진화, 진압 등 수습과 관련한 키워드가 많이 언급 되었다.

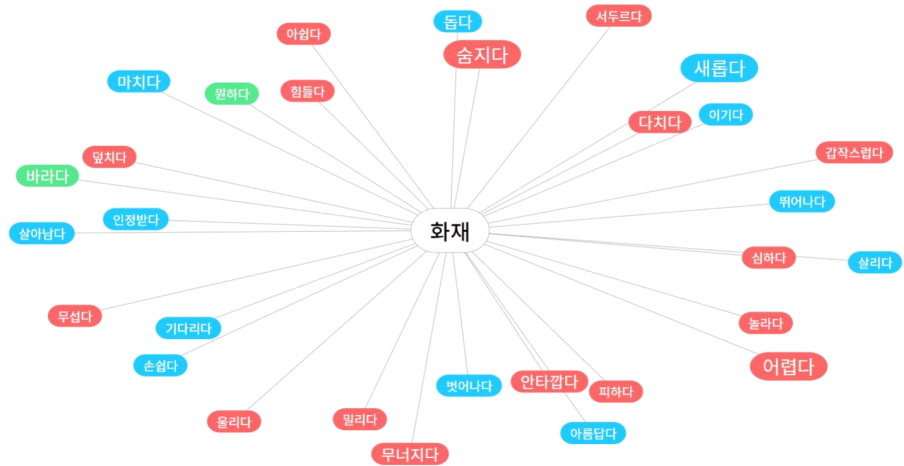


[그림 5-52] 2014년 5월 연관 키워드



[그림 5-53] 2015년 1월 연관 키워드

화재에 대한 감성 분석 결과, 긍정 7.94%, 중립 81.34%, 부정 10.71%로, 중립 감성을 제외하면 부정 감성이 조금 더 높은 것으로 나타났다. 긍정 감성어 상위에는 돕다, 뛰어나다, 벗어나다 등 화재 발생 이후 복구와 관련된 키워드가, 부정 감성어 상위에는 숨지다, 무너지다, 안타깝다, 어렵다, 다치다 등 화재로 인한 피해를 나타내는 키워드가 나타났다.



[그림 5-54] '화재'에 대한 감성 분석 결과(빨강: 부정, 파랑: 긍정, 녹색: 중립)

앞서 분석한 안전사고, 지진, 붕괴, 침수에 비해 건축물 관련한 키워드가 많이 등장하며, 설비, 구조, 탐지, 경보, 설치 등 화재 발생 예방을 위한 유지관리 방안에 대한 관심이 높다고 할 수 있다.

□ 결과해석

전반적으로 가장 많은 버즈가 발생한 키워드는 화재로 여론의 관심이 가장 높다고 볼 수 있으며, 뒤이어 높은 버즈를 발생하고 있는 안전사고와 붕괴에도 여론의 관심이 높다고 볼 수 있다. 지반침하하는 붕괴의 원인이 되는 경우가 많아 자체 버즈량은 낮은 수준이나 여론의 관심이 낮다고 보기는 어렵다. 지진의 경우 평균 버즈량이 매우 낮은 수준이나 경주에서 지진이 발생했던 2016년 하반기 버즈량이 폭증했으며, 이후 월 평균 버즈량이 증가한 것으로 보아 역대 최고 규모의 경주 지진 이후 여론의 관심이 증가했다고 볼 수 있다. 침수의 경우 강수와 연관이 높다 보니 비가 많이 오는 3분기의 버즈량이 상승하며 키워드 중 유일하게 시계열적 패턴을 보이고 있다.

키워드 별 감성 분석 결과, 긍정 감성 수치가 높은 키워드는 안전사고와 침수 두 개 뿐이었으며 침수의 경우 긍/부정 감성 수치가 거의 같은 수준으로 실질적으로 긍정 감성이 더 높게 나타난 키워드는 안전사고만 해당한다고 볼 수 있다. 전체 키워드의 감성 분석 수치는 긍정 감성 11.15%(14,070건), 중립 감성 74.30%(93,728건), 부정 감성 14.55%(18,352건)으로 중립 감성을 제외한 결과를 보았을 때 부정 감성이 더 높은 것으로 나타났다. 연관 키워드와 원문 콘텐츠의 내용으로 미루어 보아 부정 감성의 경우 현상 자체에 대한 의견과 사전에 대비하지 못한 것에 대한 의견이 많은 것으로 유추되며 긍정 감성의 경우 현상 자체에 대한 긍정적인 의견보다 현상 해결에 대해 긍정 반응을 나타내는 것으로 보인다.

버즈가 발생한 원문 콘텐츠들을 확인해 보았을 때, 원인이나 사실에 대한 내용은 보도성의 콘텐츠에서 많이 발생되었으며 예방 대책과 관련한 내용은 보도성 콘텐츠에서도 발생하긴 했지만 여론을 보여줄 수 있는 내용의 콘텐츠에서도 많이 발생된 것으로 나타난다. 버즈의 발생이 계절적 영향을 받는 경우 특정 기간을 전후로 많은 콘텐츠가 생성되었으며 나머지의 경우 대부분 특정 이슈가 발생했을 때에 콘텐츠가 생성되는 것으로 나타났다. 모든 키워드에서 가장 많은 버즈를 발생한 콘텐츠는 블로그 콘텐츠였으며 블로그 콘텐츠가 많다는 것은 여론의 관심이 높다고도 볼 수 있다. 트위터의 경우 키워드 마다 편차가 큰 편이며 중복된 내용이 많거나 무의미한 데이터가 많은 경우도 있지만 뉴스나 블로그에 비해 다양한 시각의 여론이 존재하기 때문에 적은 버즈라도 참고할 필요가 있다.

본 연구 결과 데이터를 보면 건축물 유지관리에 대한 관심은 붕괴, 침수, 화재 등 특정 사건이 발생했을 때 높은 관심을 보이는 것으로 나타났지만 사건 발생 시 마다 예방, 대책 등에 관한 키워드가 등장하는 것으로 보아 건축물 안전 측면에서의 유지관리 정책이 미흡한 것으로 판단된다. 건축물에서 발생하는 사건·사고가 꾸준히 발생하고 있고, 이에

대한 대책 마련이 필요한 만큼, 과거 사건·사고 데이터 분석을 통해 해결의 실마리를 찾아가야 할 것으로 판단된다.

같은 현상이라도 분야별 시각에 따라 다양한 버즈가 발생하는 만큼 건축 관련 분야의 객관적인 인사이트 도출이 가능하도록 여론 데이터 분석 결과와 관련 정형데이터 분석 결과를 함께 활용한다면 보다 정교한 정책 수립이 가능할 것으로 기대된다. 또한 지방자치단체에서도 지역 기반 여론 데이터 분석 결과를 활용하여 정교한 정책적 대응이 가능할 것으로 기대된다.

3. 정책 형성 단계

1) 문제 형성 단계에 따른 정책 형성 방향

앞서 수행한 문제 형성단계에서의 소셜 데이터 분석을 통해 건축물 유지관리 정책 중 화재 안전에 대한 국민 관심도가 큰 것으로 파악되었으며, 이를 바탕으로 건축물 화재 안전 정책 개발을 위한 데이터 분석을 수행하였다.

2) 질의개발

건축물 화재 안전 정책 개발을 위해 도출한 주요 질문은 아래와 같다.

- 최근 5년간 화재 발생 건수, 인명피해, 재산피해 등 화재 발생 현황은 어떠한가?
- 공간적 특성, 지역, 계절, 화재 특성별(발화효인, 열원 등) 화재 발생 경향은 어떠한가?
- 화재가 많이 발생하는 건축물 유형은 무엇이며, 건축물 속성별 화재 발생 현황은 어떠한가?

3) 분석내용 결정

건축물에서 발생한 화재의 공간 및 건축·도시적 속성을 살펴보고, 다양한 화재 발생 유형을 빅데이터를 통해 분석하여 건축물 화재 안전 정책 아이디어를 도출하는 것을 목표로 한다. 또한 화재 발생 건수, 인명피해, 재산피해 등의 증가 및 감소 추세를 화재발생의 공간적 특성 및 계절별, 건축물 특성별, 화재 특성별 경향을 분석하여, 건축물 화재에 효율적으로 대응할 수 있는 소방·방재 전략을 도출할 수 있음을 실증한다.

4) 데이터 종류 및 특성 파악

국가화재정보시스템 화재통계는 전국에서 발생한 화재의 발생일시, 장소, 원인, 특징, 피해 등을 전수 조사한 빅데이터로, 2007년 이후 발생 화재 통계를 국가화재정보센터 사이트³⁾를 통해 공개하고 있다. 화재 특성으로 발화열원, 발화요인, 장소 및 발화지점, 피해(인명, 재산, 소실면적) 등의 정보가 포함된다.

본 분석에서는 전체 화재통계 중 최근 5년(2012년~2016년)간 발생한 건축물 화재발생에 대한 데이터를 대상으로 하였다. 임야나 차량 화재 등은 건축물과 관련 없는 화재로 제외하고 건축물 및 해당 필지에 발생한 화재를 분석하였다.

화재 정보와 연계하여 사용할 수 있는 정보 중에는 건축물 속성정보를 알 수 있는 건축물 대장정보와 날씨에 따른 화재 발생 특성을 볼 수 있는 기상정보, 인구 특성에 따른 지역별 화재발생 특성을 보기 위한 인구총조사 정보가 있다. 건축물 대장 정보는 건축물에 대한 연면적, 구조, 용적률, 층수, 용도 등의 정보를 포함하며, 기상청 지상기상관측 통계⁴⁾ 자료는 전국 45개 지점의 월평균 기상 데이터를 제공한다. 인구총조사 자료 중 인구, 가구 및 주택 통계를 통해 지역별 인구수 정보를 알 수 있다.

5) 분석 방법 설계

□ 건축물 화재 발생 통계 분석

개인정보 보호를 위해 화재 발생장소는 법정동 단위로 제공받아 분석을 진행하였다. 다른 통계자료는 대부분 행정구역(시군구, 행정동) 단위로 집계되어 있어, 함께 연계하여 분석시 지난 5년간 행정구역의 변화를 반영하여 과거 화재 발생 위치를 법정동 단위로 현재 행정구역에 맞게 집계하였다. 다만 한 법정동이 여러 행정구역에 걸치는 경우 실제 화재 발생 위치를 알 수 없어, 대응되는 행정구역 각각에 화재건수를 균등하게 배분하여 분석하였다.

먼저 화재 정보를 활용하여 전국 연도별 화재발생 건수, 피해 현황을 분석하고, 건축물 용도별, 발화지점별, 발화요인 및 열원별 화재 발생 현황을 분석하였다.

3) 소방청, 소방청국가화재정보센터, <http://www.nfds.go.kr>(검색일: 2017.08.01)

4) 기상청, "지상기상관측",
<http://sts.kma.go.kr/jsp/home/contents/statistics/newStatisticsSearch.do?menu=SFC&MNU=MNU>(검색일: 2017.08.01)

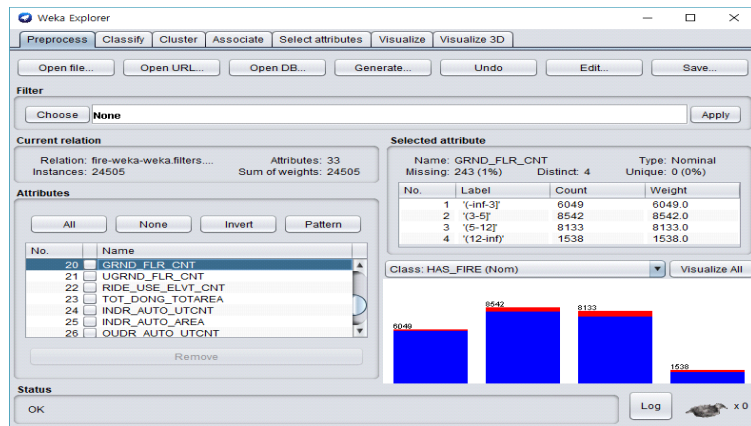
그리고 날씨에 따른 화재 발생 현황 분석을 위해 기상청 지상기상관측 통계 자료의 월평균 기상 데이터와 인접 법정동의 월별 화재 데이터를 연계하여 분석 하였다.

또한 지역별 화재 발생 특성을 파악하기 위해 시도별, 시군구별, 읍면동(법정동)별 화재 발생 현황을 분석하고, 시도별, 시군구별로는 인구수와의 상관관계를 분석하였다.

□ 건축물 대장 및 화재 정보 연계 빅데이터 분석

개별 건축물 단위에서 건축물의 특성에 따른 화재 발생 경향을 분석하기 위하여, 건축물 대장과 화재통계를 연계하여 빅데이터 분석을 수행하였다. 화재 발생건수 기준 상위 5개 시군구 중 수도권 1곳(서울시 강남구)과, 비수도권 1곳(경상남도 김해시)를 분석 대상으로 설정하여, 서로 다른 조건에서 건축물 속성과 화재 발생 사이의 관계를 보고자 하였다. 최근 5년간 강남구에서는 총 24,505개 건축물 중 1,560곳에 화재가 발생하였고, 김해시는 66,690개 건축물 중 1,138곳에 발생한 것으로 집계되었다.⁵⁾

본 분석에서는 건축물대장에 수록된 다양한 속성이 건축물 화재 발생에 미치는 영향을 분석하기 위하여, 빅데이터에 대한 기계학습 알고리즘 분석을 수행할 수 있는 통합환경인 Weka⁶⁾를 사용하고, 여러 분석 알고리즘 중 연관규칙(Association Rule) 분석을 선택하여 수행하였다.



[그림 5-55] 기계학습 및 빅데이터 분석 프로그램 Weka

출처: <http://www.cs.waikato.ac.nz/ml/weka/>

5) 한 건축물에 여러 번 화재가 반복하여 발생하는 경우가 무시되기 때문에 시군구 통계의 화재 발생건수보다 적게 나타난다.

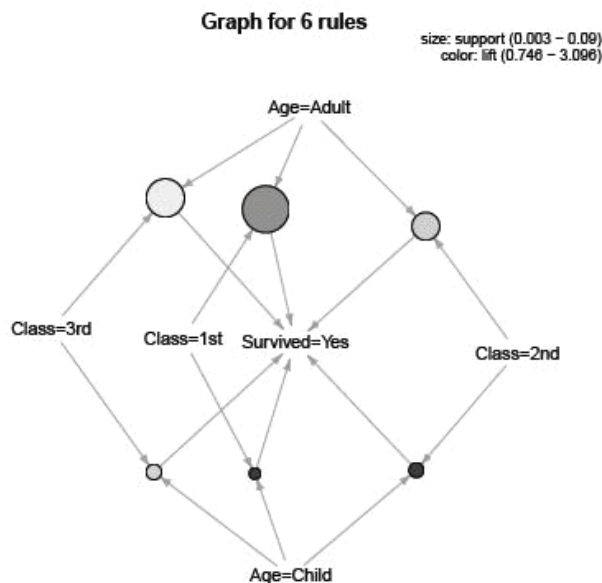
6) Witten, Ian H., Eibe Frank, Mark A. Hall and Christopher Pal (2016), Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann.

WEKA, "Weka 3: Data Mining Software in Java", <http://www.cs.waikato.ac.nz/ml/weka/>. (검색일: 2017.08.02)

연관규칙 분석은 공통적으로 나타나는 원소들을 찾고 한 원소가 다른 원소의 포함 여부에 주는 영향을 분석하는 기계학습 알고리즘이다. 이러한 과정은 데이터 외에 추가적인 입력 없이 자동으로 이루어지기 때문에, 수많은 규칙 중에서 결과에 가장 큰 영향을 미치는 규칙을 자동으로 찾아내는 것도 가능하다. 건축물대장과 같이 한 속성이 여러 값을 갖는 경우에도, 각각의 값을 독립적인 원소로 놓아 분석할 수 있다.

예를 들어, 타이타닉 생존 데이터에 대한 연관규칙 분석을 통해 연령과 선실 등급이 생존 여부에 미치는 영향을 분석할 수 있다. 1등칸의 아동 승객의 경우 살아남은 승객의 수로 따지면 3등칸 아동보다 적지만, 전체 1등칸 아동 대비 생존률은 3등칸 아동의 생존률보다 훨씬 높다는 것을 파악할 수 있다.

본 연구에서 사용한 연관규칙 분석 알고리즘은 Apriori⁷⁾로, 분석 속도가 느린 편이나 데이터 규모가 크지 않기 때문에 큰 문제가 되지 않으며, 결과 변수를 설정하고 해당 변수에 미치는 영향만을 골라 분석(Class Association Rule)할 수 있다는 장점 때문에 선택하였다.



[그림 5-56] 타이타닉 생존 데이터에 대한 연관규칙(Association Rule) 분석 결과
출처: <http://www.rdatamining.com/examples/association-rules>

7) Agrawal, Rakesh and Ramakrishnan Srikant (1994), "Fast algorithms for mining association rules", *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB*. pp.487-499. Santiago, Chile.(1994.9.12.-15.)

건축물 화재에 대한 클래스 연관규칙 분석 결과에서 후행조건은 화재 발생 (HAS_FIRE=1)으로 고정된다. 선행조건인 규칙에 해당되는 건축물 수와 그 중에서 화재가 발생한(선·후행조건에 모두 해당) 건축물 수는 연관규칙의 평가 척도를 측정하는데 활용된다.

컴퓨터 알고리즘은 데이터의 의미를 알지 못하고 산술적인 분석을 수행하기 때문에, 선행조건과 후행조건 사이에 직접적인 인과관계가 있어 ‘당연한’ 발견과 사람이 인과관계를 추론하기 어려운 ‘놀라운’ 발견을 구분할 수 없다. 예를 들어 본 연구의 연관규칙 분석 결과, 2012년 이전 건축허가를 받거나 2014년 이전 사용승인을 받은 것을 화재 발생에 영향을 미치는 변수로 도출하였는데, 이는 최근 5년간 화재 발생 여부에 건축물이 사용된 기간이 영향을 미친다는 당연한 결과로 화재안전 정책 수립과는 관련없는 내용이다. 따라서 본 연구에서는 분석 결과를 그대로 사용하지 않고, 연구자가 먼저 내용을 검토하고 의미 있는 결과만 추리는 과정을 거쳤다.

6) 분석 및 결과해석

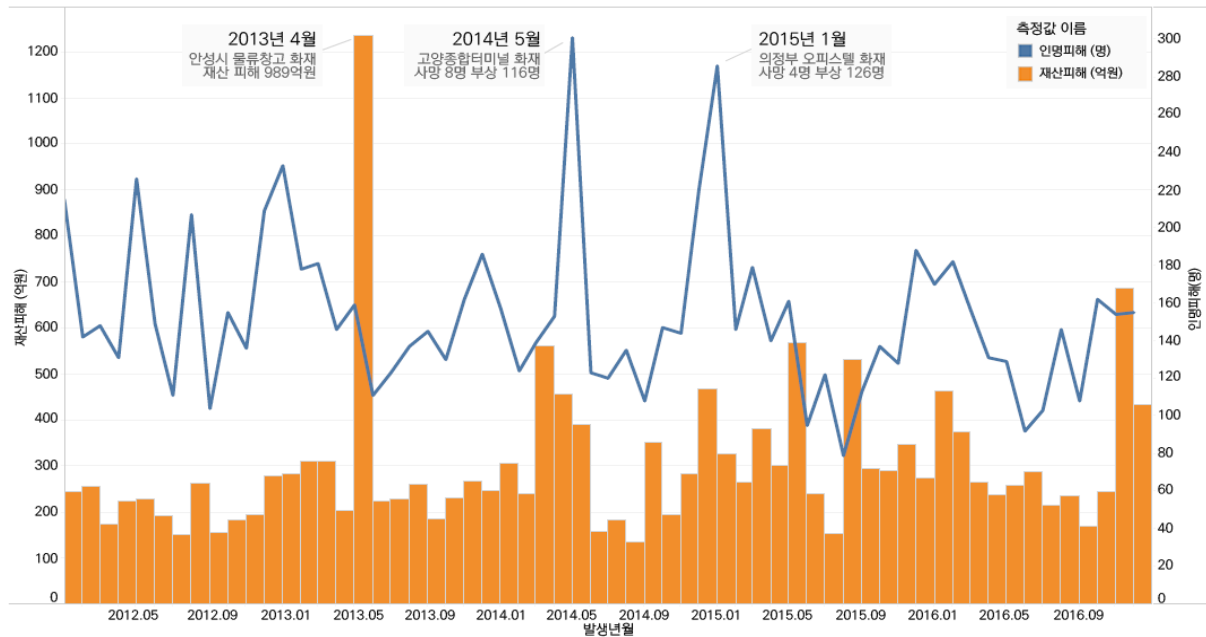
⑥ 화재 발생 통계

□ 화재 발생 및 피해 추세

전국에서 건축물 화재는 연평균 약 2만 6천 건의 화재가 발생하여, 사망 250여명, 부상 1,600여명, 재산피해 3천억여원 등의 피해를 발생시키고 있다. 특히 대형 화재사고가 1~2년에 한번씩 발생하여 전체 피해의 상당 부분을 차지하는데, 한 예로 2013년 발생한 안성시 물류창고 화재는 재산 피해 989억원을 내며 해당 연도 피해액의 1/4에 달하였다. 이러한 변동으로 인해 연간 화재 발생건수 및 피해 규모에서 통계적으로 유의미한 추세를 발견할 수 없었다.

[표 5-6] 전국 연도별 화재 발생, 인명 및 재산피해 통계

연도	2012	2013	2014	2015	2016
발생건수	26,800	25,775	25,950	26,446	27,353
인명피해(명)	1,933	1,891	1,871	1,774	1,689
- 사망	248	268	277	211	258
- 부상	1,685	1,623	1,594	1,563	1,431
재산피해(억원)	2,541	3,983	3,723	3,970	3,868



[그림 5-57] 전국 건축물 화재 피해 발생 추세

□ 건축물 화재 발생 특성 분석

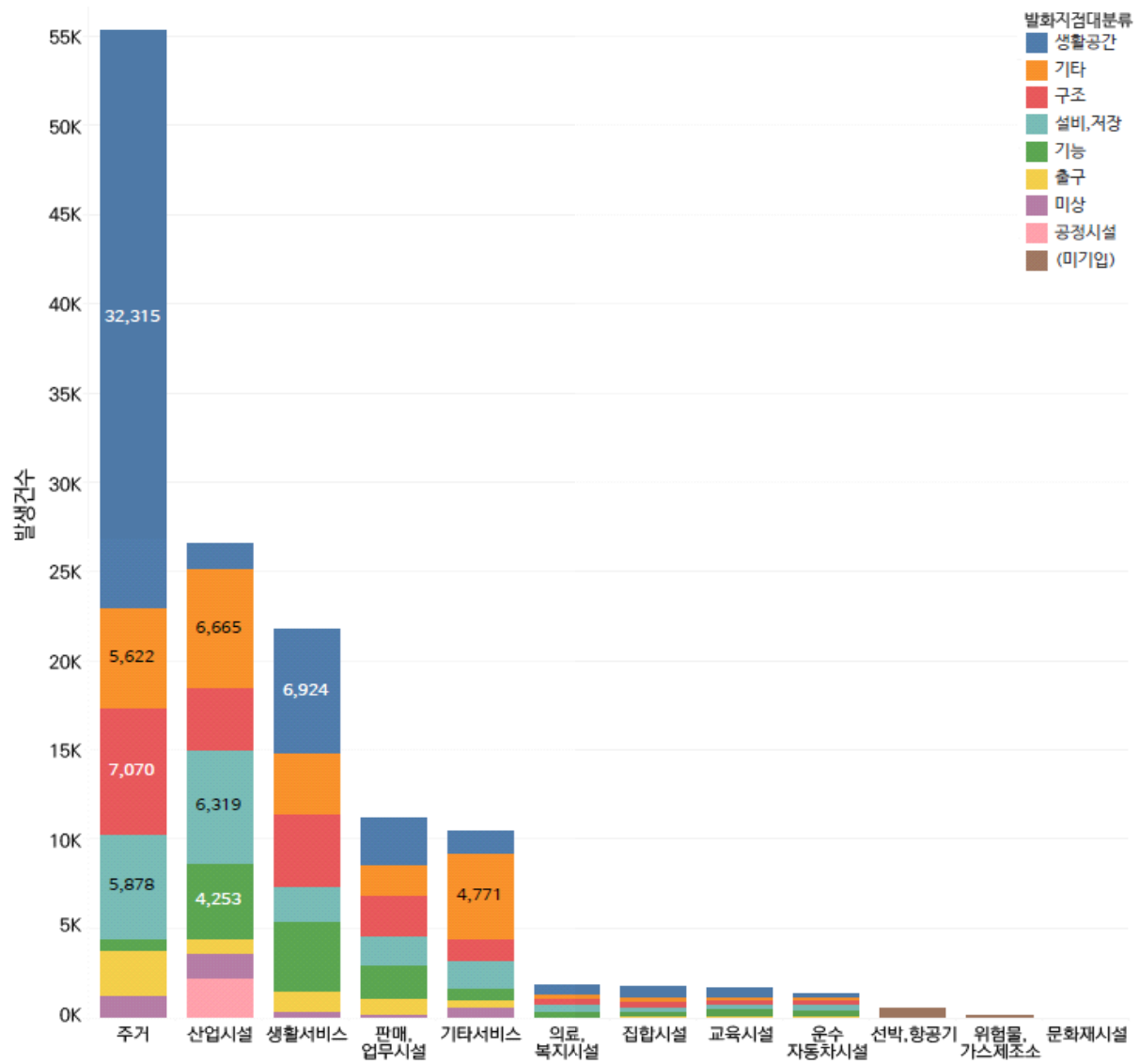
건축물 화재의 발생 특성을 분석하기 위하여, 먼저 건축물 용도별 화재 발생 및 피해 통계를 분석하였다. 건축물 화재 발생건수를 건축물 용도별로 나누어 보면 주거 55,268건 (41.8%), 산업시설 26,499건 (20.0%), 생활서비스 21,741건 (16.4%) 순으로 나타난다. 인명피해는 주로 주거에 집중하여 발생하였으며, 재산피해 및 소실면적은 산업시설에서 높게 나타났다.

[표 5-7] 건축물 용도별 및 발화지점별 화재 발생 및 피해 통계

건축물 용도별	발생건수	인명피해(명)	재산피해(억원)	소실면적(㎡)
주거	55,268	4,981	2,457	816,014
산업시설	26,499	1,168	11,481	3,436,389
생활서비스	21,741	1,109	1,117	302,522
판매,업무시설	11,137	851	1,388	202,141
기타서비스	10,412	330	536	388,368

건축물 용도별	발생건수	인명피해(명)	재산피해(억원)	소실면적(㎡)
의료,복지시설	1,800	172	79	15,555
집합시설	1,742	102	209	52,561
교육시설	1,686	75	134	24,030
운수자동차시설	1,350	212	416	119,332
선박,항공기	544	69	203	20
위험물,가스제조소	129	89	55	8,280
문화재시설	16	0	10	432
발화지점별	발생건수	인명피해(명)	재산피해(억원)	소실면적(㎡)
생활공간	45,853	4,764	2,063	657,341
기타	23,359	839	2,717	1,223,669
구조	19,230	619	1,432	550,004
설비,저장	18,390	902	3,933	988,516
기능	12,543	858	2,299	638,962
출구	5,919	332	1,359	174,050
미상	3,895	487	2,981	897,118
공정시설	2,462	199	1,041	227,684
(미가입)	673	158	258	8,300

건축물 발화지점별로 화재 발생을 분석하였을 때도 생활공간이 45,853건(34.7%)로 높게 나타나나, 대부분은 주거 화재 발생이 높은 것에 기인하였다. 생활공간을 제외하면 기타, 구조, 설비 및 저장 지점 순으로 화재 발생이 많았으며, 이들 발화지점 특성은 용도별로 큰 편차를 보이지 않았다. 결론적으로 주거공간 화재가 발생 건수가 가장 많고 인명 피해도 집중되어 있다는 점을 확인할 수 있다.



[그림 5-58] 건축물 용도별 발화지점별 화재 발생 통계

□ 발화요인 및 열원별 건축물 화재 발생

건축물 화재의 발생원인별 통계를 분석한 결과, 발화요인은 부주의가 60,228건 (45.5%)으로 가장 많으며, 그 뒤에 전기적 요인 36,649건 (27.7%), 알 수 없음 14,685건 (11.1%) 순이었다. 건축물 화재를 일으킨 열원도 작동기기가 56,647건 (42.8%), 그 다음으로 담배 및 라이터, 불꽃, 발화원인 미상 등이 비슷한 규모로 나타났다.

[표 5-8] 건축물 발화요인 상위 5개 (발생건수 기준)

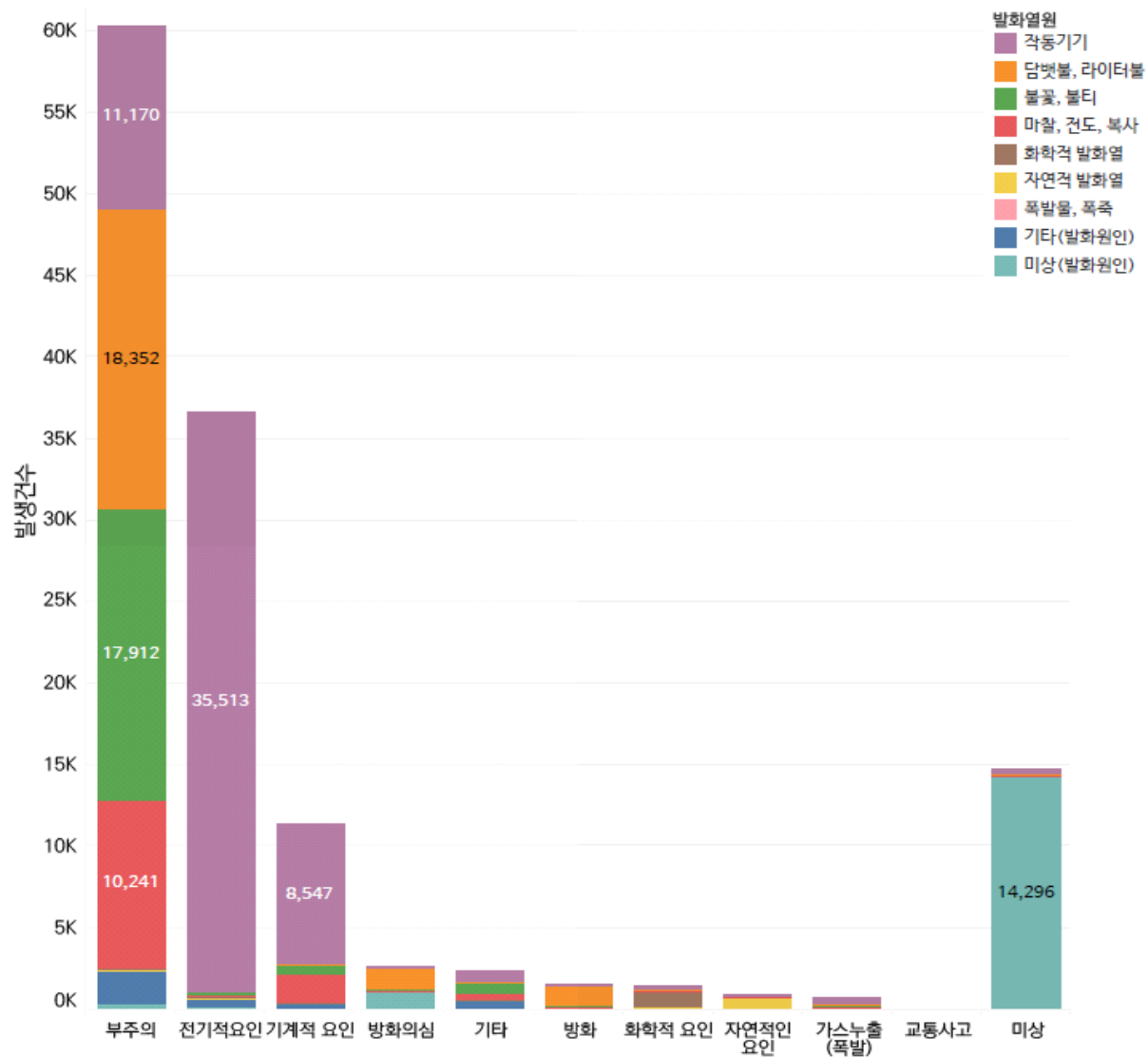
발화요인별	발생건수	인명피해(명)	재산피해(억원)	소실면적(㎡)
부주의	60,228	3,022	2,774	1,025,815
전기적 요인	36,649	1,666	3,488	1,441,732
미상	14,685	2,048	8,668	2,061,863
기계적 요인	11,339	385	1,137	360,946
방화의심	2,607	486	168	53,520

[표 5-9] 건축물 발화열원 상위 5개 (발생건수 기준)

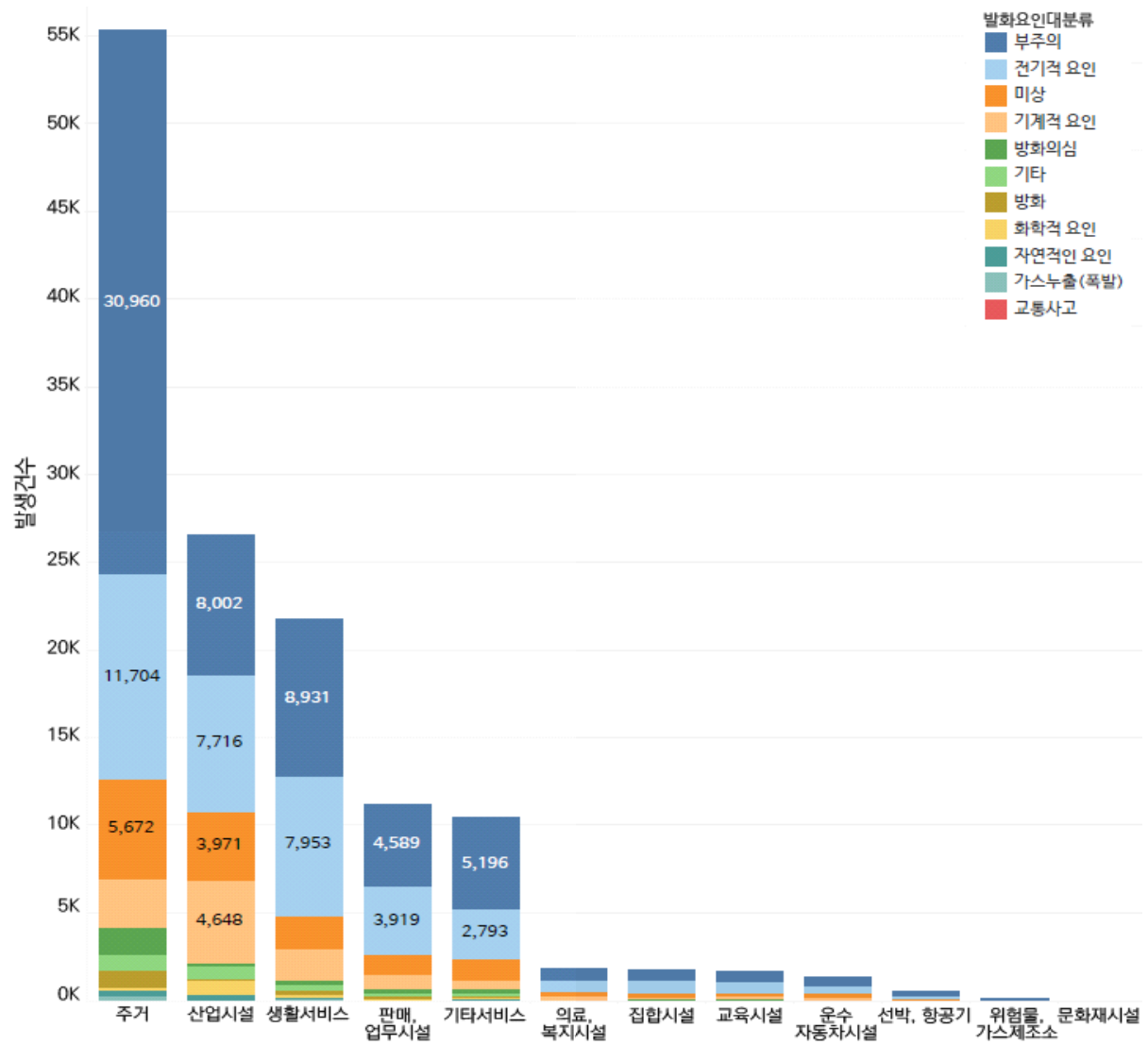
발화열원별	발생건수	인명피해(명)	재산피해(억원)	소실면적(㎡)
작동기기	56,647	2,680	5,016	1,827,929
담뱃불, 라이터불	21,158	1,770	1,180	269,646
불꽃, 불티	19,693	1,250	1,722	753,183
미상(발화원인)	16,327	2,389	8,837	2,099,386
마찰, 전도, 복사	12,767	641	559	178,332

두 요인을 함께 살펴보면, 크게 다양한 형태의 불 취급시 부주의와 작동기기의 전기·기계적 요인으로 나누어 설명할 수 있다. 다양한 불씨와 열원을 부주의하게 다루지 않도록 하는 것은 건축물 화재에 특화된 문제라기보다는 일반적인 화재예방 정책 차원의 문제일 것이다. 전기·기계적 요인으로 인한 화재의 경우, 각종 기기의 결함을 예방하는 것은 건축물 화재안전 정책의 범위에서 벗어나나, 화재 발생 가능성이 있는 기기를 파악하고 화재 피해를 줄일 수 있는 설계기준 등을 세울 수 있다.

이러한 발화요인별 화재 발생을 건축물 용도별로 나누어 분석해보면, 부주의로 인한 화재는 대부분 주거에서 나타나면서 (30,960건), 주거 화재 발생의 과반을 차지한다. 전기·기계적 요인은 다양한 건축물 용도에 고르게 분포하는 모습을 보인다.



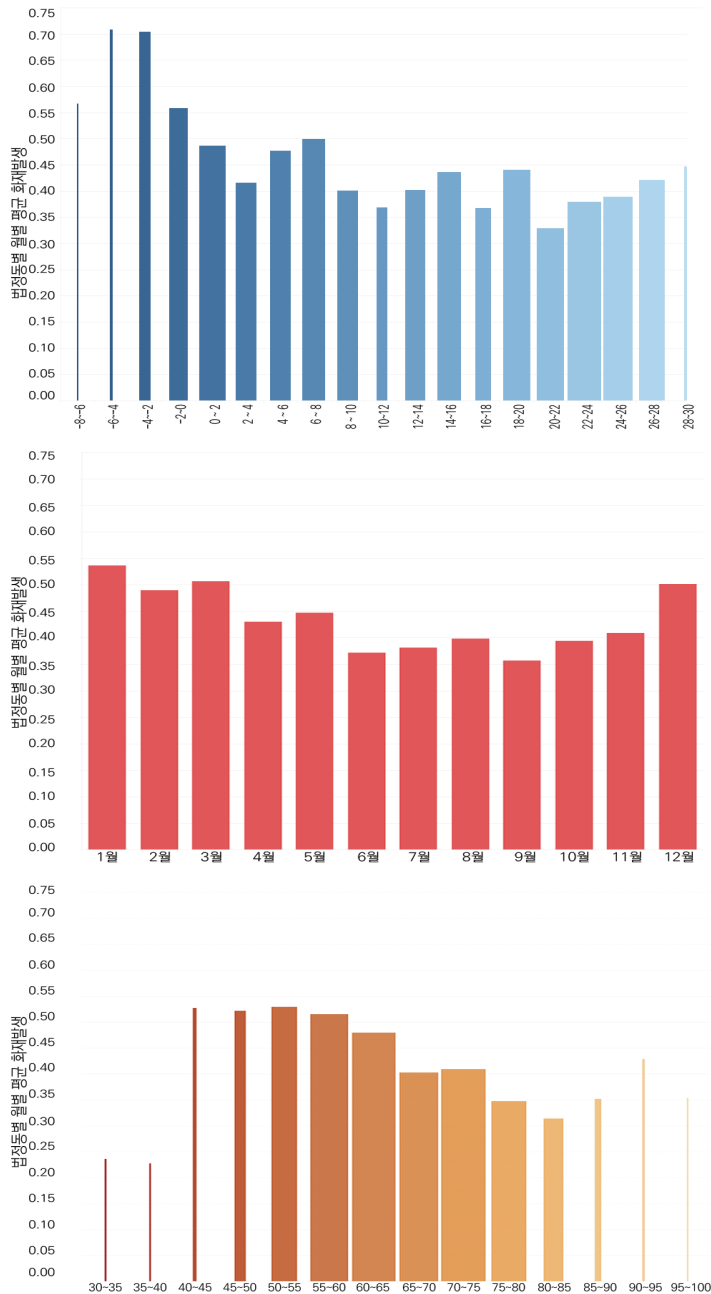
[그림 5-59] 건축물 발화요인 및 열원별 화재발생 통계



[그림 5-60] 건축물 용도별 발화요인별 화재 발생 통계

앞서 건축물 용도별 분석 결과와 마찬가지로, 주거 용도 건축물에 각종 불 및 열원을 부주의하게 다루어 일어나는 화재가 전체 건축물 화재의 상당 부분을 차지하는 것을 확인할 수 있다.

□ 날씨에 따른 건축물 화재 발생

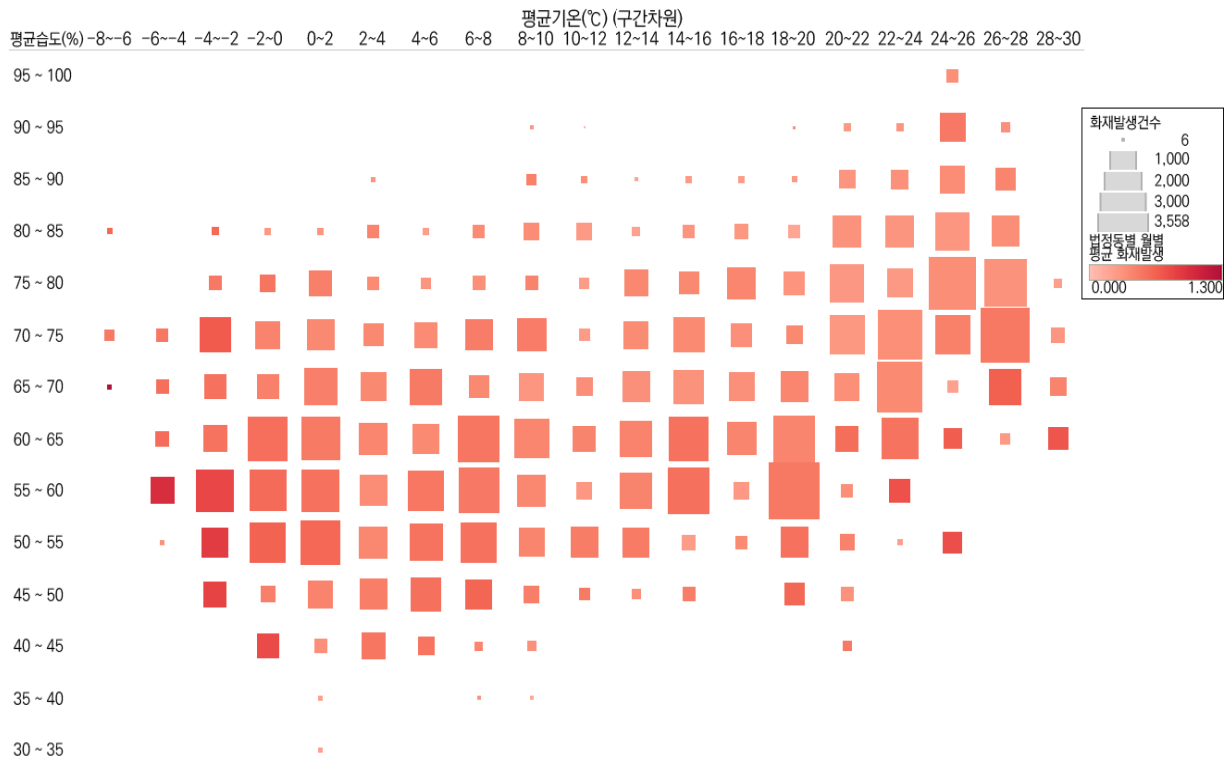


[그림 5-61] 월별 및 월평균 온·습도에 따른 화재 발생 통계
(가로축 단위: 해당 온·습도 구간 빈도, 세로축 단위: 법정동별 평균 화재발생건수)

월별 화재 발생건수를 분석한 결과 12월부터 3월까지 겨울~봄 기간 동안 건축물 화재 발생이 증가하는 경향이 있음을 확인하였다. 이러한 계절적 변동이 날씨와 어떠한 연관이 있는지를 분석하기 위하여, 기상청 지상기상관측 통계⁸⁾에서 전국 45개 지점의 월평균 기상 데이터와 인접 법정동의 월별 화재를 연계하여 분석하였다.

월 평균기온이 0도보다 높을 경우 기온과 화재발생 사이에 유의한 관계는 나타나지 않는다. 한편 월 평균기온이 영하로 떨어지면 화재 발생은 점차 증가하여 최대 1.5배까지 늘어난다. 그러나 실제로 월 평균기온이 영하가 되는 것은 드문 일로, 계절에 따른 전반적인 화재 증감을 설명하기 어렵다.

습도의 경우 화재발생과 매우 강한 관계를 보여준다. 이러한 관계는 특정 구간에서만 나타나는 것이 아니라 전 구간에서 습도가 낮으면 화재발생이 증가하고 반대로 습도가 높으면 감소하는 것을 확인할 수 있다. 구체적으로 살펴보면 월평균 습도 50%~85% 구간에서 10% 습도 하락 시 화재발생이 약 19% 증가하는 것으로 나타났다.



[그림 5-62] 월평균 온·습도에 따른 화재 발생 통계
(크기: 해당 온·습도 구간 화재 발생건수, 채도: 법정동별 평균 화재 발생건수)

8) 기상청, “지상기상관측”, <http://sts.kma.go.kr/jsp/home/contents/statistics>. (검색일: 2017.08.03)

이러한 경향성은 온습도를 결합하여 분석할 때 가장 잘 나타난다. 영하의 기온에 습도가 60% 이하일 경우 화재 발생 확률(해당 구간에 속한 법정동당 화재 발생 건수)이 가장 높으나, 그 경우에 해당하는 화재는 전체 건축물 화재에 비해 매우 적다. 반대로 모든 온도 구간에서 습도가 낮아질 때 화재 발생건수와 화재 발생 확률이 모두 높아지는 것을 볼 수 있다.

습도가 낮은 건조한 날에 실화로 인한 화재가 늘어나는 것은 이미 알려진 사실이다. 그러나 최근 5년간 건축물 화재 전수 데이터에 대한 분석을 통해 이러한 사실을 다시 구체적인 수치로 검증하였고, 주거 용도 건축물에 부주의로 인한 화재가 집중되어 있다는 점과 연계하여 계절별 특화 정책 개발 필요성을 확인할 수 있었다.

□ 시도별 화재 발생 및 피해

시도별로 화재 발생건수 및 피해 규모를 분석한 결과, 발생건수, 인명피해, 재산피해가 모두 가장 많은 곳은 경기도이며, 그 다음으로는 서울시, 경상남도, 경상북도, 부산시 순으로 화재가 많이 발생하였다. 이들 시도는 인구가 가장 집중된 곳으로, 시도 내 인구, 대지 면적, 건축물 연면적 등으로 화재 발생건수를 통제하였을 때는 화재 발생이 다른 시도에 비해 특별히 높게 나타나지 않았다.

인명피해는 발생건수와 유사한 순위를 보였으나, 부산이 3위로 올라서는 등 인구 순위와 더 일치하는 분포를 보였다. 재산피해 규모는 경기도, 경상남도, 충청남도, 경상북도 순으로 크게 나타났는데, 이는 인구보다는 공업지역이 많이 위치하는 지역 순위를 따르는 것으로 해석할 수 있다.

이러한 경향성은 화재 발생 및 피해 규모를 인구 등의 변수와 상관분석을 수행했을 때 더 잘 드러났다.⁹⁾ 시도별 화재 발생건수는 시도별 인구($R^2=0.978$) 및 시도별 건축물 연면적($R^2=0.990$)과 매우 강한 상관관계를 보인다. 인명피해 또한 완전한 선형적 관계는 아니지만 강한 상관성을 확인할 수 있다.

시도별 재산피해의 경우, 인구, 건축물 동수, 대지 면적, 건축물 연면적 모든 변수에서 특정 시도를 제외하지 않고 상관성을 설명하지 못하는 것으로 나타났다. 이는 재산피해 규모가 단순 인구 및 건축·도시 개발 규모가 아닌 숨겨진 변수를 따른다는 앞선 해석과 일치하는 결과이다.

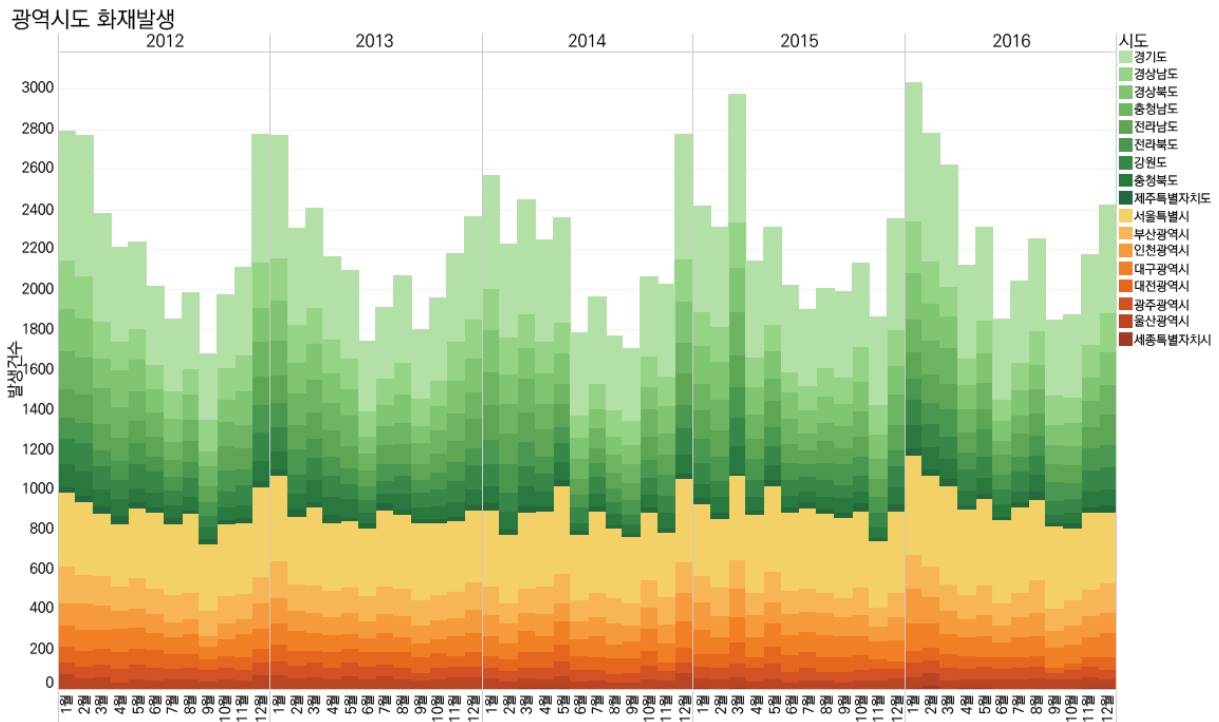
9) 인구는 2015년 인구총조사 '인구, 가구 및 주택' 통계, 대지 면적은 국토지목별현황, 건축물 동수 및 연면적은 건축물대장 건축물 현황 통계를 적용하였다.

□ 시도별 월별 건축물 화재 발생

앞서 분석한 바와 같이 화재 발생은 계절에 따라 변동하는데, 이러한 변동이 시도별로 어떻게 나타나는지 분석을 수행하였다.

화재 발생건수 1, 2위를 차지한 서울시 및 경기도는 연중 화재 발생 비율이 대체로 일정하게 나타났다. 광역시(세종특별자치시 포함) 지역에서도 연중 일정하게 화재가 발생하여, 전국적 변동과 다른 모습을 보였다.

한편 광역시를 제외한 도 지역은 여름·가을에 비해 겨울철에 최대 2배까지 화재 발생이 높아져, 계절적 변동이 도 지역에 집중되어 나타나는 것을 확인할 수 있었다. 이렇게 광역시 지역의 화재 발생은 연중 일정하고, 도 지역에서만 겨울·봄철 건축물 화재 발생이 늘어나는 원인에 대해서는 추가적인 분석이 요구되나, 이러한 패턴을 확인한 것만으로도 맞춤형 정책을 개발하기 위한 근거가 될 수 있다.



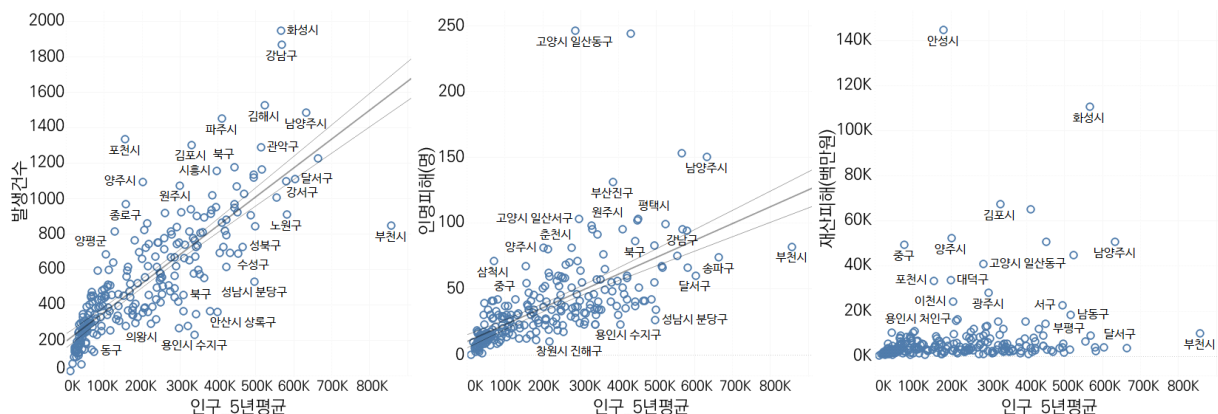
[그림 5-63] 시도별 월별 건축물 화재 발생건수 통계

□ 시군구별 화재 발생 및 피해

시군구별 화재 발생 및 피해와 인구에 대한 상관분석을 실시한 결과, 화재 발생건수와 인명피해 모두 시군구 인구와 통계적으로 매우 유의한 상관성이 있었으나, ($p < 0.0001$) 결정계수(R^2)는 각각 0.639 및 0.410으로 감소하였다. 인구와 건축물 연면적 각각에 대해 화재 발생과 인명피해가 거의 비례하였던 시도별 통계와 달리, 10만명당 화재 발생건수 및 인명피해가 시군구별로 적지 않은 차이를 보였기 때문이다. 재산피해는 인구 또는 대지 면적과 상관성이 없는 것으로 나타났는데, 시도별 분석 결과와 연계하여 해석하면 시군구의 규모와 관련 없는 다른 변수를 따르는 것을 확인한 것으로 볼 수 있다.

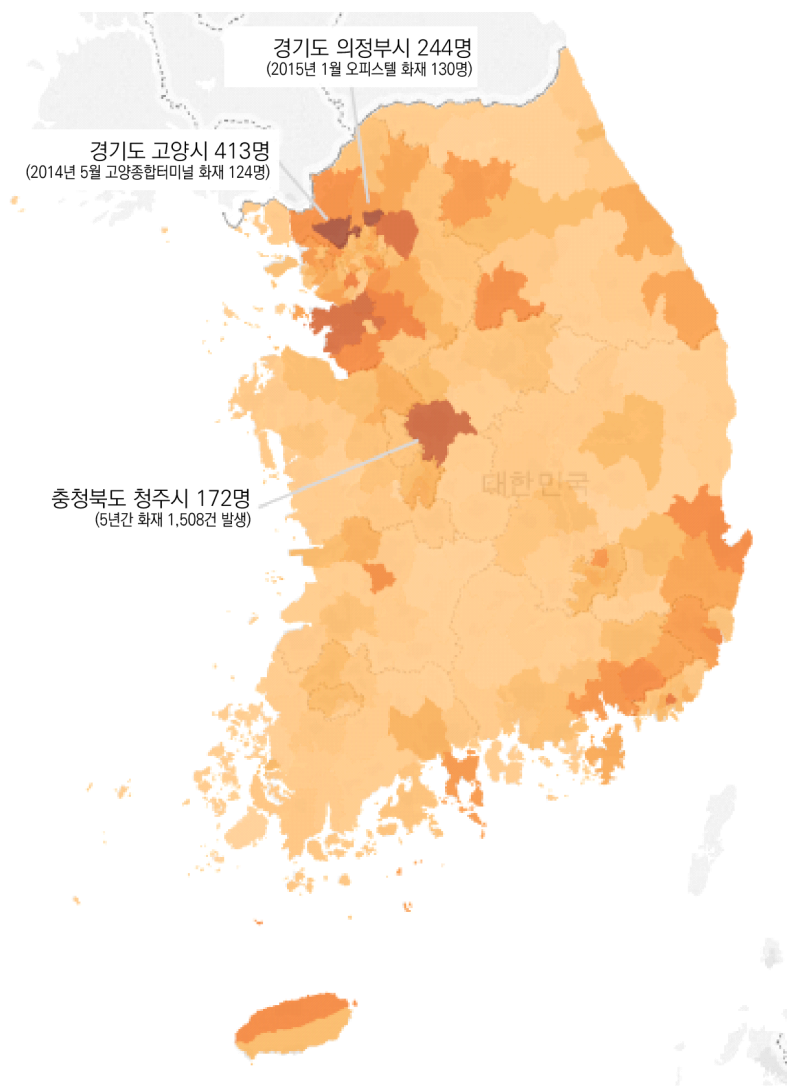
[표 5-10] 건축물 화재 발생건수 상위 시군구 통계

시도명	시군구명	발생건수	발생건수 10만명당	발생건수 대지 1km ² 당	인명피해 (명)	인명피해 10만명당(명)	재산피해 (백만원)	재산피해대지 1km ² 당(백만원)
경기도	화성시	1,943	343	50	153	27	110,606	2,868
서울특별시	강남구	1,866	327	117	95	17	8,946	559
경상남도	김해시	1,523	291	63	99	19	44,589	1,851
경기도	남양주시	1,482	234	66	150	24	50,550	2,243
경기도	파주시	1,447	351	48	95	23	65,076	2,145
경기도	포천시	1,333	855	70	67	43	33,408	1,764
경기도	김포시	1,301	393	65	95	29	67,264	3,367
서울특별시	관악구	1,285	250	157	67	13	2,765	337
서울특별시	송파구	1,226	184	90	74	11	3,391	248
대구광역시	북구	1,175	265	87	86	19	9,162	675
중앙값		446	300	37	29	19	3,134	286



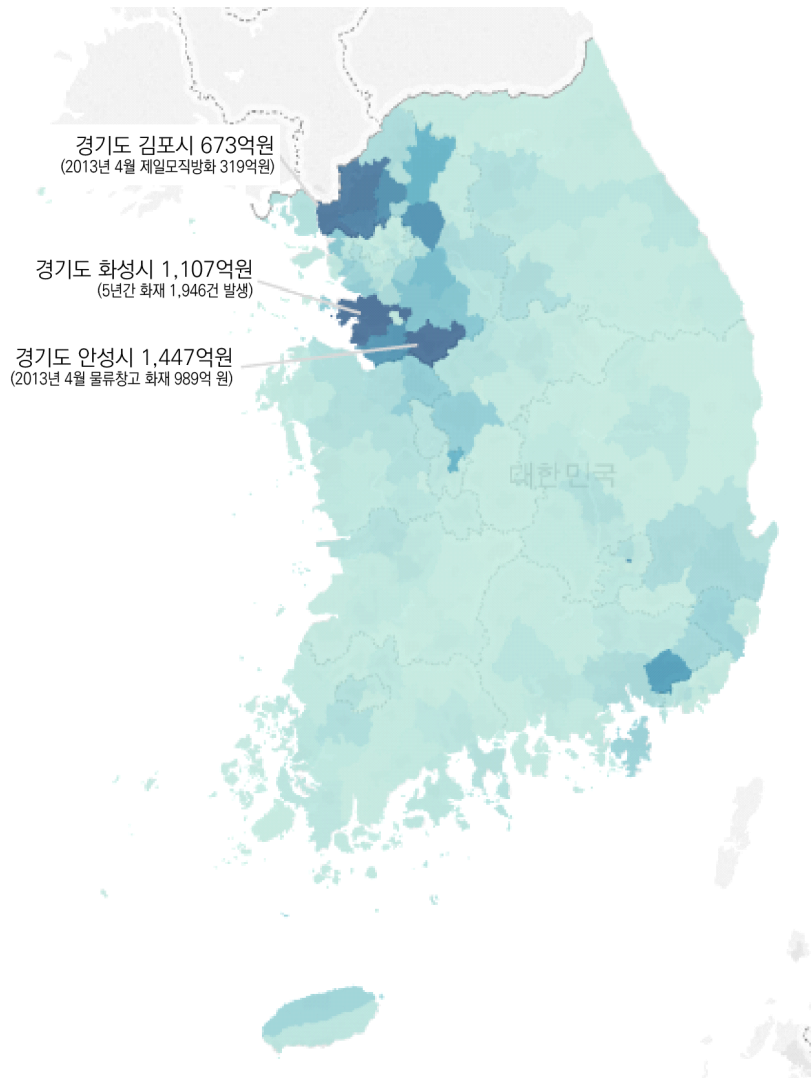
[그림 5-64] 시도별 화재 발생 및 피해와 인구·건축 밀도 상관분석

화재 발생 상위 10개 시군구는 크게 두 가지 부류로 나타났다. 공업지역이 밀집한 시군구는 거주 인구 10만명당 화재 발생건수가 많고 재산피해가 크며, 반대로 주거지역이 많은 경우 화재 발생에 비해 재산피해가 적게 나타난다. 따라서 주거가 집중된 시군구에서는 작은 규모의 화재가 자주 발생하며, 산업시설에서는 대형 화재사고로 인한 큰 규모의 재산피해가 발생함을 확인하였다.



[그림 5-65] 시군구별 건축물 화재로 인한 인명피해
(상자 안은 상위 시군구 피해 규모 및 특이사항)

인명피해 및 재산피해 상위 시군구를 각각 살펴보았을 때, 인명피해 상위는 주로 인구밀집 지역에, 재산피해 상위는 산업밀집 지역에 위치하며, 대형 화재 발생 장소도 동일한 패턴을 보이고 있다. 또한 시군구 단위에서는 대형 화재사고의 영향이 화재 피해 순위에 영향을 강하게 미치고 있음을 알 수 있다. 인명피해 상위 3개 시군구 중 2곳, 재산피해 상위 3곳 모두 분석 대상 기간 중 대형 화재로 인해 큰 피해가 발생한 곳으로 나타났다. 따라서 시군구별 화재 피해 경감을 위한 정책의 목표는 대형 화재사고를 예방하고 화재 발생시 대형 피해 발생을 억제하는 것이 될 것이다.



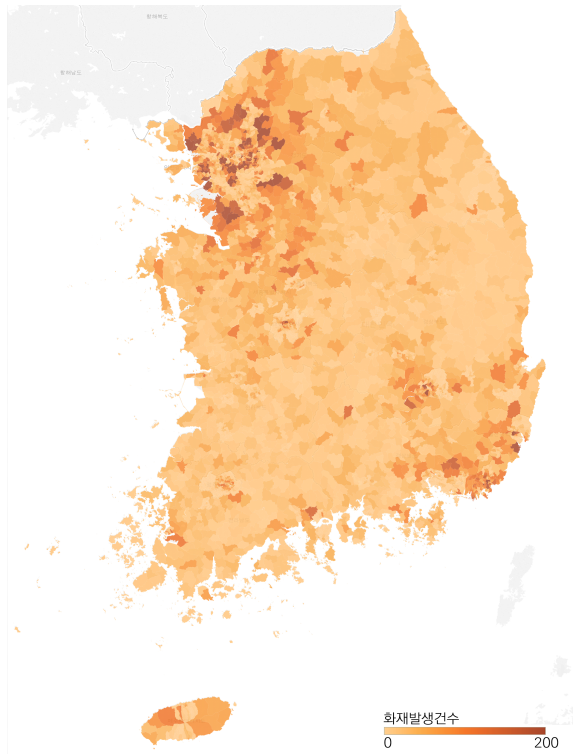
[그림 5-66] 시군구별 건축물 화재로 인한 재산피해
(상자 안은 상위 시군구 피해 규모 및 특이사항)

□ 법정동별 화재 발생 및 피해

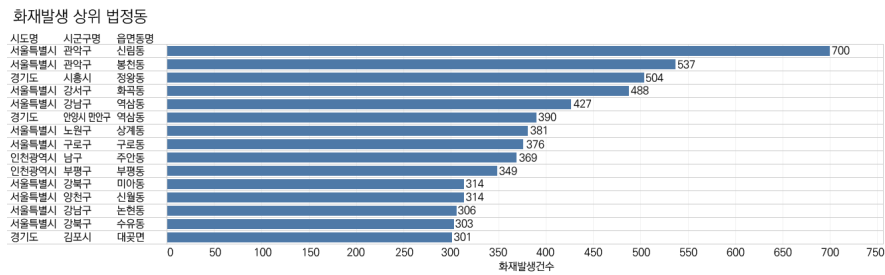
최근 5년간 화재 발생건수 구간별로 해당 법정동 수의 분포를 분석한 결과, 대부분의 법정동은 화재 발생이 매우 적고 (중앙값 5년간 14건) 극소수의 법정동에서 매우 많은 화재가 발생하는 거듭제곱 분포(power law)를 따르는 것을 알 수 있었다. 이러한 특징은 지도에서도 잘 표현되는데, 인구분포와 선형적으로 비례하지 않고 몇몇 법정동에 매우 집중되는 모습을 보인다.

[표 5-11] 건축물 화재 발생건수 상위 법정동 통계

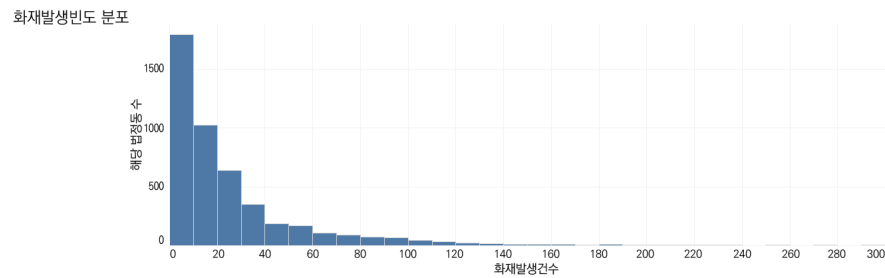
시도	시군구	읍면동	발생건수	발생건수 대지 km2당	인명피해 (명)	재산피해 (백만원)	재산피해 대지1km2당 (백만원)
서울특별시	관악구	신림동	700	162	41	1,420	328
서울특별시	관악구	봉천동	537	152	22	1,185	335
경기도	시흥시	정왕동	504	105	19	7,053	1,471
서울특별시	강서구	화곡동	488	125	56	1,088	279
서울특별시	강남구	역삼동	427	173	18	2,513	1,016
경기도	안양시 만안구	안양동	390	122	34	1,210	379
서울특별시	노원구	상계동	381	102	16	820	219
서울특별시	구로구	구로동	376	151	19	1,067	430
인천광역시	남구	주안동	369	98	31	1,230	328
인천광역시	부평구	부평동	349	137	19	1,466	574
서울특별시	강북구	미아동	314	117	19	780	290
서울특별시	양천구	신월동	314	131	6	2,311	963
서울특별시	강남구	논현동	306	159	19	1,041	541
서울특별시	강북구	수유동	303	123	20	946	384
경기도	김포시	대곶면	301	144	17	10,872	5,194
서울특별시	양천구	신정동	296	91	22	407	125
서울특별시	양천구	목동	293	99	14	790	266
서울특별시	서초구	서초동	292	83	13	663	188
대구광역시	서구	비산동	291	169	11	789	459
부산광역시	연제구	연산동	282	75	15	263	70
중앙값			14	37.9	1	82	190.6



[그림 5-67] 법정동별 건축물 화재 발생



[그림 5-68] 법정동별 건축물 화재 발생



[그림 5-69] 법정동별 건축물 화재 발생

공간적으로도 인접한 법정동의 화재 발생건수는 큰 편차를 보인다. 수도권의 많은 법정동은 전국적으로 화재 발생이 가장 낮은 편에 속하는 등, 화재 발생이 인구와 비례하여 나타나지 않고 복합적인 양상을 나타내고 있다. 그러나 이러한 건축물 화재 발생 특성이 발생하는 원인을 분석하기 위해 필요한 인구나 건축물 규모 등의 통계자료는 행정동 단위까지만 작성되어, 법정동 단위 분석에는 적용할 수 없는 상황이다. 월별 화재 발생 변동에서 나타난 시도별 차이, 시군구별 상관분석에서 나타난 인구와의 관계를 동시에 고려하여, 수도권 및 광역시 지역에서 전반적으로 화재 발생이 감소하는 경향과 인구가 집중된 지점에 국지적으로 화재 발생이 증가하는 경향이 동시에 작용한 결과일 것이라는 추측만이 가능하다.

화재 발생건수 상위 20개 법정동의 특성을 살펴보면, 화재 발생이 중위 법정동에 비해 20배~50배에 달하며, 대부분은 주거 밀집지역이나 김포시 대곶면 등 공장 밀집지역도 일부 포함되어있다. 시도별, 시군구별 분석에서 경기도가 화재 발생 상위권을 대부분 차지했던 것과 달리 20개 법정동 중 13곳이 서울시 내 법정동으로, 주거지역 화재 빈도에 대한 인구밀도의 영향을 보여주고 있다.

⑦ 건축물 대장 및 화재 정보 연계 빅데이터 분석

연관규칙의 평가 척도는 지지도(Support)와 신뢰도(Confidence)가 주어진다. 지지도는 ‘전체 건축물 중 규칙에 해당하며 동시에 화재가 발생한 건축물의 비율’을 나타내는 데, 전체 건축물 중 화재가 발생한 건축물의 비율이 강남구(6.4%)와 김해시(1.7%)가 크게 다르기 때문에 비교하기 어렵다. 따라서 지지도를 전체 화재 발생 비율로 나누어준 ‘화재 발생 건축물 중 규칙에 해당하는 건축물의 비율($\text{sup} / \text{sup}(\text{all})$)’을 해당 규칙의 설명력으로 삼아 비교하였다.

신뢰도는 ‘규칙에 해당되는 건축물 중 화재가 발생한 비율’을 나타내는데, 이를 전체 화재 발생 비율로 나누어주면 ‘전체 건축물 대비 화재 발생 비율의 배수($\text{conf} / \text{conf}(\text{all})$)’를 얻을 수 있다. 이 지표가 1보다 큰 값을 가질 때, 규칙에 해당되는 건축물에서 화재가 전체 대비 몇 배로 발생한다는 의미이다.

이러한 두 지표는 서로 상호보완적으로 사용될 수 있다. 전체의 45%에 해당하는 건축물에서 화재 발생이 1.57배가 되는 유형(김해, 철근콘크리트 구조)과 전체 건축물 중 16% 밖에 해당되지 않지만 화재 발생은 5.67배로 늘어나는 유형(김해, 높이 23m 이상)은 서로 다른 차원의 정책적 목표를 제공하며, 복합적인 정책 개발을 뒷받침하는 근거가 될 수 있다.

분석 결과로 도출된 규칙 중 주건축물, 콘크리트 지붕, 철근콘크리트 구조, 단독주택 제외는 대부분의 화재 발생 건축물을 포함하면서 일부 건축물 유형을 배제하는 형식의 규칙으로, 주로 화재 발생이 많은 유형을 간접적으로 지칭하는 경우이다. 반대로 비주거 집합건축물의 경우 전체 화재 발생의 일부분만을 설명하지만, 전체 건축물 대비 최대 7.8배나 되는 화재 발생률을 보이면서 화재 발생과 깊은 상관관계가 있는 유형을 드러낸다. 높이, 건축면적, 연면적, 용적률 등 규칙들은 서로 독립적인 조건이 아니라 건축물 규모라는 공통된 대상을 지칭하고 있으며, 실제 분석 결과에서 도출된 건축물 유형에서는 가장 효율적인 한 가지 규칙만 사용된다.

[표 5-12] 건축물 화재에 대한 연관규칙 및 분석 지표

규칙 (설명)	지역	해당 건축물 수	화재 발생 건축물 수	Support	sup / sup(all)	Confidence	conf / conf(all)
MAIN_ATCH_GB_CD=0	강남	23,849	1,557	6.4%	99.8%	6.5%	1.03
주건축물 (주로 김해 공장 부속건축물 제외)	김해	58,467	1,134	1.7%	99.6%	1.9%	1.14
ROOF_CD=10	강남	22,097	1,461	6.0%	93.7%	6.6%	1.04
콘크리트 지붕	김해	28,375	569	0.9%	50.0%	2.0%	1.18
STRCT_CD=21	강남	18,371	1,338	5.5%	85.8%	7.3%	1.14
철근콘크리트 구조	김해	19,257	515	0.8%	45.3%	2.7%	1.57
FMLY_CNT=0	강남	14,900	1,155	4.7%	74.0%	7.8%	1.22
가족 0 (주로 단독주택 및 주상복합 제외)	김해	34,961	848	1.3%	74.5%	2.4%	1.42
MAIN_PURPS_CD!=1000	강남	17,323	1,336	5.5%	85.6%	7.7%	1.21
단독주택 제외	김해	35,601	907	1.4%	79.7%	2.5%	1.49
REGSTR_GB_CD=2	강남	7,164	559	2.3%	35.8%	7.8%	1.23
집합건축물	김해	4,697	248	0.4%	21.8%	5.3%	3.09
MAIN_PURPS_CD=2000	강남	6,290	407	1.7%	26.1%	6.5%	1.02
공동주택	김해	4,054	160	0.2%	14.1%	3.9%	2.31
REGSTR_GB_CD=2 MAIN_PURPS_CD!=(1000 2000)	강남	880	152	0.6%	9.7%	17.3%	2.71
비주거 집합건축물 (오피스텔, 상가, 업무시설 등)	김해	658	87	0.1%	7.6%	13.2%	7.75
HEIT='(23-inf)'	강남	2,595	344	1.4%	22.1%	13.3%	2.08
높이 23m 이상	김해	1,923	186	0.3%	16.3%	9.7%	5.67
ARCH_AREA='(499-inf)'	강남	2,537	348	1.4%	22.3%	13.7%	2.15
건축면적 499m ² 초과	김해	8,554	504	0.8%	44.3%	5.9%	3.45
TOTAREA='(5748-inf)'	강남	1,901	302	1.2%	19.4%	15.9%	2.50
연면적 5,748m ² 초과	김해	1,642	165	0.2%	14.5%	10.0%	5.89
VL_RAT_ESTM_TOTAREA='(1653-inf)'	강남	3,973	528	2.2%	33.8%	13.3%	2.09
용적률 산정 연면적 1,653m ² 초과 (지상층)	김해	4,109	341	0.5%	30.0%	8.3%	4.86

이러한 분석 과정을 통해 최종적으로 도출된 화재 발생 집중 건축물 유형은 4가지이다. 첫 번째 유형은 주건축물이면서 단독주택 이외의 철근콘크리트 구조인 건축물이며, 두 번째 유형은 주건축물이면서 단독주택 이외의 콘크리트 지붕인 건축물이다. 이 두 유형은 전체 건축물을 절반 가까이 제외하면서도 강남구 화재의 대다수를 설명한다. 김해시에서는 화재 발생률이 더 높게 나타나나, 많은 건축물이 콘크리트가 아닌 기타 구조/지붕으로 분류되어 있어 화재의 1/3를 설명하는 선에 그친다. 이러한 현상은 조적조 등 다른 구조로 지어진 경우가 많은 단독주택의 화재 발생이 낮은 것을 지칭하는 것으로 설명될 수 있다.

세 번째 유형인 집합건축물의 주건축물이면서 용적률 산정 연면적 1,653m² 초과하는 건축물은 전체 화재 발생 건축물 중 해당되는 비율이 적지만, 발생률은 전체 건축물에 비해 2.5배~5.9배에 달한다. 마지막 유형인 주거 용도 집합건축물(공동주택)을 제외하면 그 비율은 최대 11.9배까지 높아진다. 비주거 집합건축물에는 오피스텔, 상가, 업무시설 등이 포함되는데, 대규모 비주거 집합건축물 중 1/5~1/4 정도에서 최근 5년 사이 화재가 발생했다는 것은 주목할 필요가 있는 수치이다. 특히 대규모 건축물의 화재를 조기에 대처하지 못한다면 대형 화재로 번질 수 있으므로 비주거 집합건축물의 화재안전을 위한 집중적인 관리가 요구된다.

[표 5-13] 화재 발생 비율이 높은 건축물 유형

규칙 (설명)	지역	해당 건축물 수	화재 발생 건축물 수	Support	sup / sup(all)	Confidence	conf / conf(all)
MAIN_ATCH_GB_CD=0 STRCT_CD=21 MAIN_PURPS_CD!=1000	강남	13,915	1,169	4.8%	74.9%	8.4%	1.32
주건축물/철근콘크리트 구조/단독주택 제외	김해	9,363	408	0.6%	35.9%	4.4%	2.55
MAIN_ATCH_GB_CD=0 ROOF_CD=10 MAIN_PURPS_CD!=1000	강남	11,719	1,247	5.1%	79.9%	10.6%	1.67
주건축물/콘크리트 지붕/단독주택 제외	김해	7,960	415	0.6%	36.5%	5.2%	3.06
REGSTR_GB_CD=2 MAIN_ATCH_GB_CD=0 VL_RAT_ESTM_TOTAREA=(1653-inf)	강남	2,251	359	1.5%	23.0%	15.9%	2.51
집합건축물/주건축물/용적률 산정 연면적 1,653m ² 초과	김해	1,993	200	0.3%	17.6%	10.0%	5.88
REGSTR_GB_CD=2 MAIN_PURPS_CD!=(1000 2000) MAIN_ATCH_GB_CD=0 VL_RAT_ESTM_TOTAREA=(1653-inf)	강남	407	96	0.4%	6.2%	23.6%	3.71
비주거/집합건축물/주건축물/용적률 산정 연면적 1,653m ² 초과	김해	364	74	0.1%	6.5%	20.3%	11.91

⑧ 결과 해석 및 정책 옵션 개발

□ 비주거 집합건축물 화재안전 강화 및 소방 관련 법제도 개선

비주거 집합건축물은 전체 건축물 수에 비하면 극소수에 불과하나 화재 발생률은 몇 배 이상 높아 화재안전을 위한 집중 관리가 필요한 상황이다.

준주택인 오피스텔과 실제 비주거 용도로 이용되는 집합건축물을 건축물대장 상에서 구분할 수는 없으나, 공동주택 뿐만 아니라 집합건축물이 아닌 비주거 건축물에 비해서도 화재 발생률이 크게 높은 점을 볼 때 비주거 집합건축물의 상당 부분을 차지하는 오피스텔이 화재 발생에 미치는 영향이 적지 않을 것이라는 추측이 가능하다.

비주거 집합건축물을 준주택과 실제 비주거 용도로 나누어 화재안전을 관리한다는 대원칙 아래, 신축되는 준주택에 대하여 공동주택에 준하는 설계기준을 적용하고, 기존 준주택에 대하여 부주의로 인한 화재예방 홍보를 강화하고, 준주택 밀집지역의 소방 출동 시간을 단축하는 등의 종합적인 화재안전 정책 패키지를 도입한다면 화재 취약 건축물의 안전을 크게 향상시킬 수 있다.

특히 모든 건축물에서 연면적 1,653m² 초과하거나, 건축면적 499m² 초과, 높이 23m 이상 건축물의 경우 상대적으로 화재발생 위험이 높은 것으로 나타나 건축물의 소방관련 법제도와 지침에 관련 기준 수립 시 해당 결과를 반영할 필요가 있다.

□ 산업시설 대형 화재 예방

산업시설의 화재 발생은 많지 않으나 한번 대형 화재가 일어나면 재산피해가 크게 발생하는 것이 특징이다. 산업시설의 화재 발생 원인은 작동 기기의 전기·기계적 요인이 과반을 차지하는데, 이러한 화재 발생을 모두 예방하지 못하더라도 대형 화재로 이어지지 않도록 할 필요가 있다. 공장의 소방시설 기준을 강화하고, 산업 특성에 맞추어 인명과 재산을 보호할 수 있는 조기 대응 훈련을 실시하는 등의 정책을 통해 화재로 인한 재산피해를 줄일 수 있다.

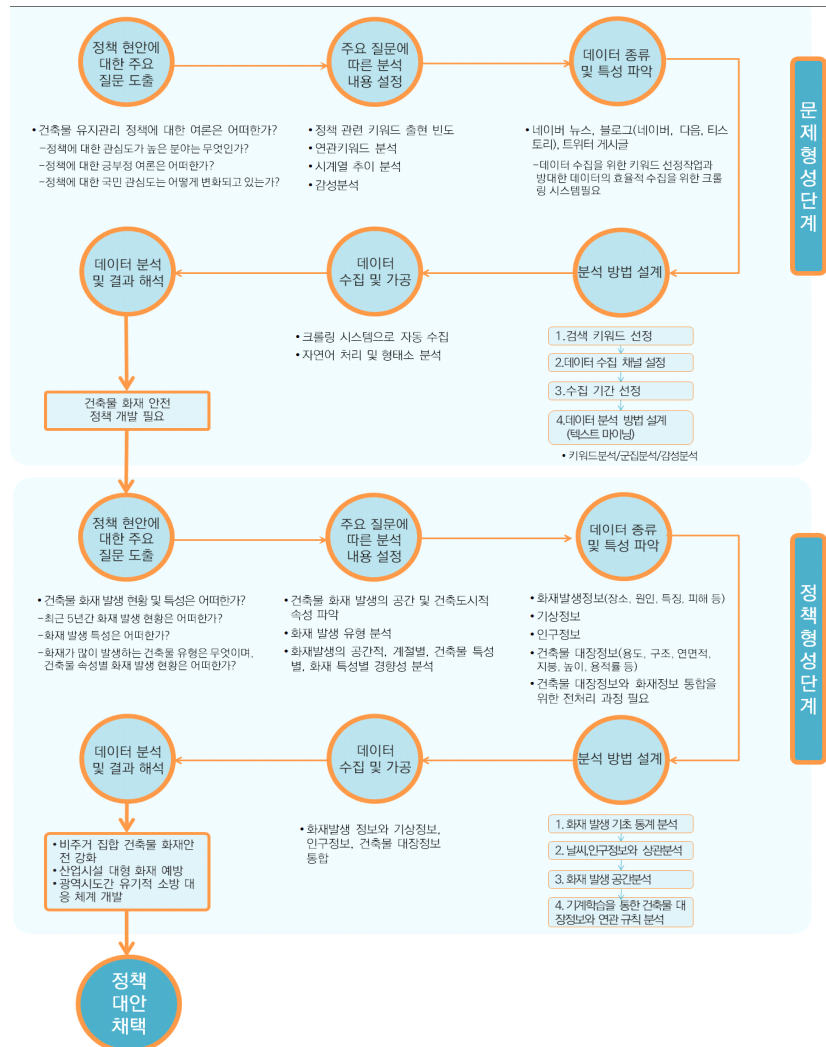
□ 광역시·도간 유기적 소방 대응 체계

겨울철 도 지역의 화재 발생은 크게 늘어나는 반면, 광역시 지역은 연중 화재 발생이 일정하게 나타난다. 이러한 차이가 나타나는 원인을 분석하고 화재를 예방하는 노력에 더하여, 각 지역별 소방 인력 및 자원을 탄력적으로 운용하여 겨울철 화재 증가에 대응할 필요가 있다. 현재와 같은 지자체별 대응으로는 어려운 일이므로, 이를 달성하기 위해서는 국가 차원의 종합적 조율 체계가 필요하다.

4. 평가 및 환류

□ 빅데이터 기반 정책개발 체계를 통한 정책대안 도출 실증

정부가 보유하고 있는 공간 빅데이터를 활용하여, 건축물 화재안전을 위해 건축·도시·소방·방재 분야의 효율성을 높일 수 있는 정책 개선 아이디어를 도출하였다. 특히 서로 다른 부처에서 관리하고 있는 빅데이터를 연계하여 새로운 가치를 창출할 수 있었다는 점에서, 다양한 출처의 빅데이터에 대한 연계 분석을 통하여 건축·도시 관련 다양한 분야 정책 개발이 가능함을 확인하였다.



[그림 5-70] 빅데이터 기반 정책개발 체계 실증 과정

□ 건축·도시 정책을 위한 빅데이터 연동 체계 구축

건축·도시 정책은 건축물과 길을 기본 단위로 한다. 일반적인 행정에 필요한 통계는 행정구역(행정동) 단위로 충분하나, 건축·도시 정책은 그보다 더 정밀한 공간정보가 뒷받침되어야 효과적인 수행이 가능하다. 건축물대장과 공간 빅데이터를 연계할 수 있도록 개별 건축물 단위로 통계정보를 산출하고 교환할 수 있도록 하는 공간 빅데이터 활용 가이드라인을 구축할 필요가 있다.

□ 빅데이터의 호환성 향상

행정구역 단위로 산출되는 통계는 시계열적 분석에서 해결할 수 없는 문제가 발생되는데, 행정구역의 경계가 변화함에 따라 공간코드와 통계치도 변동하기 때문이다. 서로 다른 시점에 생산된 데이터를 완벽하게 일치시키는 것은 단순히 노력이 많이 드는 것이 아니라 불가능하다. 또한 한번 일치시킨 데이터도 영원히 유지된다는 보장이 없으며, 시간이 지나면 모든 자료는 재작업이 필요한 기술부채(Technical Debt)가 된다.

다양한 시점의 공간 빅데이터를 쉽게 연동할 수 있도록, 행정구역이나 도로 등 시간에 따라 달라지는 대상이 아닌 공간좌표를 기준으로 집계하고 제공하도록 하여야, 공간 빅데이터에 기반을 둔 정책 개발 체계를 안정적으로 구축할 수 있을 것이다.

□ 건축물 단위 빅데이터에서의 개인정보 보호

빅데이터 분석시 개인정보를 보호하는 것은 필수적인 조치이다. 그러나 이를 위해 지나치게 데이터를 단순화시켜 가공할 경우 빅데이터 분석의 효율성이 감소한다. 특히 개별 건축물 단위에서 위치와 인접 관계가 중요한 건축·도시 공간 빅데이터에 이러한 제약을 건다면 분석 자체의 효용을 크게 떨어뜨릴 수 있다.

정부는 2016년 6월 30일 행자부, 방통위, 미래부, 금융위 등 정부부처 합동으로 개인정보 처리를 위한 비식별화 조치 가이드라인¹⁰⁾을 발표하였는데, 빅데이터의 유용성을 최대한 유지하면서 개인정보를 보호할 수 있는 17종의 비식별화 기술을 소개하였다. 이러한 기법을 활용하면 개인 단위 데이터의 정밀도를 희생하지 않고도 개인을 식별할 수 없도록 처리한 데이터를 활용할 수 있게 된다. 여러 기관의 빅데이터를 결합하여 사용하는 경우에도 제3의 공공기관에서 결합 처리 후 비식별화하여 제공하도록 하였다.

그러나 이 가이드라인은 법적 효력이 없어 활용 주체를 보호하지 못한다는 지적과, 개인정보 재식별을 막기 위한 보완이 필요하다는 비판을 동시에 받았다.¹¹⁾ 이러한 비판을 수용하면서, 정부 가이드라인에 맞추어 빅데이터를 활용하기 위한 체계를 구축할 필요가 있다.

10) 정부부처 합동(국무조정실 외)(2016), 「개인정보 비식별 조치 가이드라인」, 국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부

11) 한겨레 (2016), “개인정보 비식별 정부 가이드라인 따랐다간 큰코다친다”, <http://www.hani.co.kr/arti/economy/it/754640.html>.(검색일: 2017.08.11)

□ 빅데이터의 비식별화 전략 필요

개인정보 비식별화 가이드라인은 개인 단위 데이터에 초점이 맞추어져 있어, 공간 빅데이터에서 중요한 건물 혹은 위치에 대한 고려는 부족하다. 위치 및 세대, 건물, 길, 동네 등 다양한 정밀도에서의 개인정보 보호 필요성을 검토하고 그에 맞는 비식별화 전략을 수립할 필요가 있다.

서울시 심야버스 노선정책 수립 사례의 경우, 1km 직경 블록 안의 빅데이터를 범주화하여 활용하였으며,¹²⁾ 비씨카드 및 한국감정원의 점포 평가 서비스의 경우, 공간가중회귀 및 보간법을 적용하여 입지 특성을 범주화하는 방식으로 비식별화를 하였다.¹³⁾

이와 같은 선행 성공사례를 반영하여 광역 정책 개발에는 1km 정도의 블록 범주화 처리를 통해 빅데이터를 활용하도록 하는 가이드라인을 구축하고, 그보다 정밀한 데이터를 활용하여야 하는 마을만들기 등 소규모 지역에 대한 데이터 분석에도 적용할 수 있는 비식별화 전략을 수립할 필요가 있다.

□ 빅데이터를 활용한 정책개발체계와 의견기반 정책개발체계 비교

개발 체계의 실증을 위하여 문제형성에서 대안채택에 이르기까지의 일련의 정책개발 과정을 4장에서 개발한 빅데이터를 활용한 정책 개발 체계에 기초하여 수행하였다. 대상정책은 국토교통부 건축정책관실의 2017년 업무계획 분석을 통하여 ‘건축물 유지관리 정책’을 선정하였다.

정책의제 선정을 위하여 건축물 유지관리 정책 관련 키워드의 소셜빅데이터와 미디어 빅데이터를 활용한 여론 빅데이터 분석을 통해 건축물 유지관리에 관한 여러 주제 중 ‘건축물 화재안전’이 정책의제로 선정되었다. ‘건축물 화재 안전’에 관한 정책 대안 개발을 위하여 여러 정형데이터를 결합하여 머신러닝을 포함한 다양한 빅데이터 분석 방법론을 적용하여 다 수의 대안을 도출하였다. 도출된 대안 중 시행이 필요한 최종대안을 정량적 분석결과에 기초하여 제시하였다.

정책개발을 위한 과정은 앞서 논의한 것처럼 문제형성, 정책형성, 대안채택 단계로 구성되며, 문제형성 단계에서는 정책의제 선정이 정책형성 단계에서는 다수 정책대안 개발이 되며 대안채택 단계에서는 최종 대안이 선정되는 일련의 절차이다. 정책의 과정으로 볼 때 의견기반 정책개발과 본 연구에서 제안한 빅데이터 활용 정책개발은 과정상 차이가 없으나, 일련의 정책개발 과정에서 의사결정에 활용하는 근거에 차이가 있다. 본 연구에서 제안한 개발체계는 모든 의사결정을 빅데이터 분석 결과에 기초한 증거기반 정책개발 체계이며, 의견기반 정책개발은 정책의제는 정책개발자 또는 정책결정자의 주

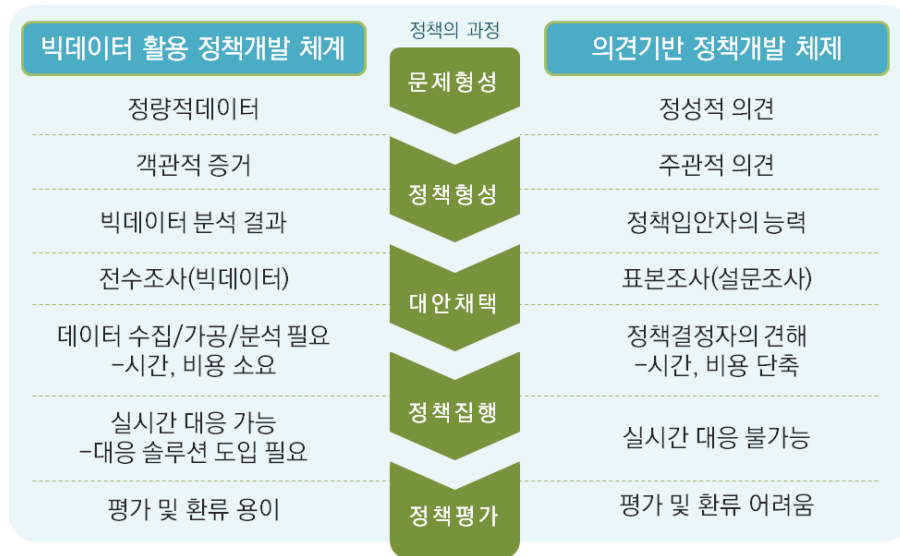
12) 한국정보화진흥원(2014), 「빅데이터 활용 위한 개인정보 비식별화 사례집」, 서울:한국정보화진흥원, p.45

13) 상동, p.41

관이나 언론보도 등의 사회적 이슈에 따라 선정되고, 정책대안 개발은 전문가 의견, 설문조사, 정책개발자의 주관에 따라 이루어지며, 최종 대안채택은 정책결정자의 주관적 판단에 의하여 결정되는 경우가 많다.

의견기반 정책개발은 정책을 개발하는 실무자와 정책을 결정하는 결정자가 해당 분야의 전문성이 높을수록 유용한 수단이다. 그러나 공무원 직무순환제를 운영하고 있는 우리나라에서는 정책을 개발하고 결정하는 공무원의 전문성을 기대하기 어렵고, 건축·도시 분야는 전문분야가 광범위하여 정책개발에 분야별 전문성이 요구되기 때문에 의견기반 정책개발은 바람직한 방향이 아니다.

빅데이터를 활용한 증거기반 정책개발은 본 연구에서 검증한 것과 같이 정책의 과정에서 의사결정이 과학적이고 객관적으로 이루어져 신뢰도가 높은 정책을 개발할 수 있다. 또한 정책의제와 대안개발이 정량적으로 이루어진 만큼 해당 정책에 대한 평가와 환류가 용이한 장점이 있다. 그러나 데이터의 수집, 가공, 분석에 상대적으로 많은 시간이 소요되고, 또한 분석과정에 대한 설계가 정교하지 못하면 결과가 나오지 않는 경우도 있어 준비과정에서 상당한 시간이 필요하다. 따라서 이러한 시간 단축을 위해서는 정책분야별로 관련 빅데이터를 활용할 수 있는 플랫폼의 개발이 필요하며, 본 연구에서 제안한 개발체계에 따라 분야별 빅데이터 플랫폼을 선제적으로 개발하여 운영한다면 증거기반 정책개발의 시간단축은 물론이거니와 실시간으로 대상 정책에 대한 모니터링과 피드백이 가능하다.



[그림 5-71] 빅데이터 정책개발 체계와 의견기반 정책개발 체계 비교

제6장 결론

1. 연구의 결과
 2. 건축·도시 빅데이터 활용을 위한 정책제언
 3. 연구의 한계와 기대효과
-

1. 연구의 결과

빅데이터의 정의는 2001년 Gartner에서 데이터의 급성장에 따른 시대 변화를 데이터의 규모(Volume), 속도(Velocity), 다양성(Variety)의 3가지 개념으로 정의한 것(3Vs)을 그 시작으로 볼 수 있다. 그 후 전문가들에 따라 추가적으로 복잡성(Complexity), 정확성(Veracity), 가치(Value) 등 개념들을 추가하여 빅데이터의 개념에 대하여 정의하고 있지만, 여전히 3Vs는 빅데이터의 기본적인 개념이라 할 수 있다. 따라서 빅데이터란 기존 데이터에 비해 규모가 크고, 복잡 다양하며, 생성·처리 속도가 빨라 새로운 관리 및 분석 방법이 요구되는 데이터이며, 이에 더하여 데이터의 활용·분석 결과가 유의미하고 유용한 데이터를 말한다.

빅데이터의 등장은 과거에는 크기가 크고, 형태가 복잡하여 다루기 어렵다고 생각했던 데이터를 비교적 손쉽게 다룰 수 있도록 하는 컴퓨터 하드웨어, 소프트웨어 등 관련 정보통신기술의 혁신적인 발전에 기인한다고 할 수 있다. 빅데이터 종류는 크게 비정형·반정형 데이터와 정형데이터로 구분할 수 있으며, 전자는 소셜 데이터, 로그데이터, 미디어 데이터 등이 속하며, 후자는 각 종 통계 데이터, 건축물 대장 정보 등 일정한 항목에 저장한 데이터를 의미한다.

본 연구에서 활용한 정책과정 모형은 정책의 과정을 문제형성, 정책형성, 대안채택, 정책집행, 정책평가가 다섯 단계로 구분한 Anderson(1975)의 모형이다.

국내 빅데이터 사례 분석 결과 다양한 유형의 빅데이터가 활용되고 있음을 확인할 수 있었으며, 여론 수렴과 동향 분석을 위해서는 소셜 데이터, 지역적/공간적 문제도출과 대

안개발을 위해서는 로그데이터(휴대전화, 신용카드)가 많이 사용되었다. 또한 정책과정 단계별로 구분하면 문제형성, 정책형성, 대안채택 과정에 빅데이터가 중점적으로 활용되고 있음을 확인할 수 있었다.

해외 사례 분석결과 다양한 데이터의 활용과 대안채택 과정 이전 단계인 정책개발단계에 빅데이터 활용이 많다는 점은 국내사례와 같다. 다만 국외에서는 범죄데이터, 의료데이터를 활용한 사례를 다 수 볼 수 있으나, 국내 사례와 같이 휴대전화와 신용카드 정보를 활용한 사례는 부족하였으며, 이는 개인정보에 대한 국가별 인식의 차이가 있음을 알 수 있다. 또한 국외사례의 경우 해당 주제에 따라 자체적인 분석 알고리즘을 개발하여 활용하는 사례가 다 수 있으며, 통상 자체 알고리즘을 개발하기 위해서는 인력과 비용이 필요한 만큼 프로젝트에 보다 많은 자원을 투자를 하고 있음을 보여준다.

문제형성 단계에서는 여론모니터링을 위한 소셜 빅데이터, 지역적인 문제를 파악하기 위한 휴대전화, 신용카드 빅데이터, 대중의 주요관심사 파악을 위해서는 검색 빅데이터를 문제발굴(정책의제선정)에 활용할 수 있다. 정책형성 단계에서는 소셜빅데이터의 연관 키워드 등 분석을 고도화하여 다수의 정책 대안을 만들 수 있으며, 휴대전화·신용카드 데이터의 위치정보(로그데이터)를 통하여 다양한 정책 옵션을 개발할 수 있다. 또한 선거공약 빅데이터는 지역의 정책대안을 집적해 놓은 자료로 정책대안 개발에 유익하다. 정책형성 단계에서 가장 많이 활용되는 빅데이터는 정형 빅데이터로 기존에 보유하고 있는 자료를 적극적으로 활용하는 것이 정책대안 개발에 가장 필요하다. 대안채택 단계에서는 다수의 대안 중 최적의 대안을 탐색하는 것으로 알고리즘 개발을 통해서 정책형성 단계에서 발굴한 대안을 정량적으로 검증한다. 정책집행 단계에서 빅데이터 활용은 실시간 분석도구와 시스템이 요구되어 대규모 플랫폼 구축이 필요하며 대표적인 예로 스마트시티 플랫폼을 들 수 있다. 정책평가 단계에서는 사실상 문제형성단계에서 활용한 빅데이터를 다시 활용하여 해당 정책이 정책의제를 해결했는지를 검토한다.

건축정책관실의 2017년 업무계획 내용 분석 결과 28개의 세부내용 중 안전강화가 9건으로 가장 높게 나타났으며, 정책과정으로는 정책형성 및 대안채택 단계가 13건으로 가장 높게 나타났다. 도시정책관실의 경우 53개의 세부내용 중 24건이 경제살리기 관련 정책으로 가장 높게 나타났으며, 정책형성 및 대안채택 단계가 27건으로 가장 높게 나타났다. 따라서 건축도시 분야의 빅데이터 활용체계는 정책옵션도출에 초점을 맞출 필요가 있으며, 특히 건축분야는 안전, 도시 분야는 경제에 중점을 두는 것이 바람직한 것으로 나타났다.

정책개발은 정책과정 5단계 중 정책집행 이전 단계인 문제형성, 정책형성, 대안채택 단계에 해당된다. 문제형성 단계에서 정책의제를 선정하고, 정책형성 단계에서 정책대안을 개발하며, 대안채택 단계에서 최적 대안을 선정한다.

각 정책 단계별 빅데이터 활용 체계는 ‘질의개발’, ‘분석내용 결정’, ‘데이터 종류 및 특성 파악’, ‘분석방법 설계’, ‘데이터 수집 및 가공’, ‘분석 및 결과해석’의 6단계로 구성하였다. 이렇게 개발한 빅데이터 활용 체계 실증을 위하여 건축분야에서 정책개발이 시급한 안전 분야를 포괄하는 ‘건축물 유지관리’를 대상으로 파일럿 테스트를 진행하였다. 파일럿 테스트는 비정형 데이터와 정형 데이터를 두루 활용할 수 있도록 구성하였다.

건축물 유지관리에 관한 정책의제를 발굴하기 위하여 질의는 “건축물 유지관리 정책에 대한 관심도가 높은 분야는 무엇인가?”, “건축물 유지관리 정책에 대한 공부정 여론은 어떠한가?”, “건축물 유지관리에 대한 국민 관심도는 어떻게 변화되고 있는가?”로 도출하였으며, 질의에 따른 분석내용은 국민여론에 대한 여론분석으로 설정하였다. 분석대상 데이터는 소셜 데이터(네이버/다음/티스토리 블로그, 트위터)와 미디어 데이터(네이버 뉴스)로 선정하고, 수집·가공·분석방법 등을 설계하였다. 웹 크롤링을 통하여 데이터를 수집하였으며, 수집기간은 2014년~2017년 까지 약 4년간으로 하였다. 분석은 키워드별 빈도수와 시계열 추이 분석, 키워드에 대한 공부정 빈도수와 시계열 추이 분석, 연관 키워드 분석, 데이터 수집 채널별 키워드 분포 분석을 수행하였다. 분석결과 건축물 유지관리 중 화재에 대하여 여론의 관심이 높게 나타났으며, 감성 분석결과 긍정보다는 부정 감성이 높았다. 건축물 관련한 연관 키워드가 많이 등장하여, 설비, 구조, 탐지, 경보, 설치 등 화재 발생 예방을 위한 유지관리 방안에 대한 관심이 높은 것으로 나타났다.

정책형성 단계의 실증을 위하여 정책의제 중 건축물 화재를 선정하여 정책 대안을 개발하였다. 질의는 “최근 5년간 화재 발생 건수, 인명피해, 재산피해 등 화재 발생 현황은 어떠한가?”, “공간적 특성, 지역, 계절, 화재 특성별(발화효인, 열원 등) 화재 발생 경향은 어떠한가?”, “화재가 많이 발생하는 건축물 유형은 무엇이며, 건축물 속성별 화재 발생 현황은 어떠한가?”로 도출하였다. 분석내용은 화재 발생의 공간 및 건축·도시적 속성으로 설정하였으며, 데이터는 정형 빅데이터인 화재발생 통계, 건축물대장, 기상청 기상 데이터, 인구총조사 자료를 대상으로 하였다. 분석 방법은 기술통계, 상관분석, 기계학습 알고리즘 분석을 수행하였다. 데이터수집은 공개된 공공데이터와 해당 자료생산 기관으로부터 협조를 받았으며, 자료 가공은 법정동을 기준으로 수행하였다. 분석 결과 비주거 집합건축물 화재안전 강화, 산업시설 대형 화재 예방, 광역시·도간 유기적 소방 대응 체계 수립 등의 정책대안을 발굴할 수 있었다.

본 연구에서 개발한 빅데이터 활용 정책 개발 체계는 증거기반 정책개발을 시행할 수 있는 방법론으로 객관적이고 과학적인 데이터 기반의 정책을 개발할 수 있음을 연구를 통하여 실증하였다. 이러한 증거기반 정책개발 체계는 의견기반 정책개발 체계와 비교할 때 정책의 결정 과정이 투명하고, 특정 전문가나 집단의 의견이 아닌 데이터 기초하여 정책의 의제와 대안 등을 개발할 수 있으며, 데이터를 통하여 정책을 개발한 만큼 평가 및 환류과정에도 유익하게 활용될 수 있다.

2. 건축·도시 빅데이터 활용을 위한 정책제언

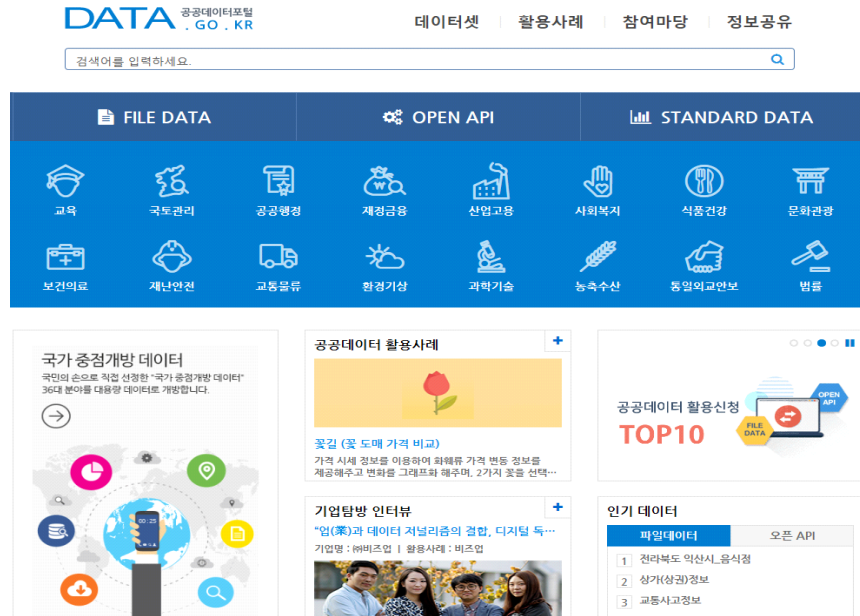
빅데이터의 종류는 일일이 열거할 수 없을 정도로 많다. 휴대전화 기록, 신용카드 사용 내역, 각종 교통기록, 에너지사용실태, 건축물의 규모, 용도, 가격 등의 속성정보, 기상청의 날씨정보, 통계청에서 제공하는 수많은 통계정보 등 공간좌표(통상 위경도)위에 구현할 수 있는 모든 데이터가 빅데이터에 해당한다고 할 수 있다. 이러한 도시 빅데이터를 공간빅데이터라는 용어로 따로 구분한다.

다양한 분야에서 빅데이터 활용으로 정부와 지자체가 제공하는 공공데이터의 가치가 더욱 높아지고 있다. 우리나라에서도 매년 공개되는 공공데이터의 양이 지속적으로 증가하고 있고, 공개되는 데이터의 품질 또한 좋아지고 있다. 그러나 공간빅데이터의 경우는 아직까지 해상도가 낮은 자료(읍·면·동 또는 시·군·구 단위)를 제공하고 있어 위치기반의 세밀한 분석은 불가능한 실정이다.

빅데이터 분석은 서로 다른 데이터를 연계 또는 결합하여 새로운 시사점을 찾는 것이다. 이를 위해서는 데이터를 상호 연결할 수 있는 고유한 식별정보가 필요하다. 공간빅데이터 분석을 위해서는 고유식별정보로 주소 또는 공간좌표가 필수적이나, 관계 기관에서 ‘개인정보보호법’에 근거하여 제공 및 열람을 금하는 경우가 많다. 그러나 ‘개인정보보호법’은 특정 개인의 주소정보를 보호하는 것인지 건축물의 주소정보가 개인 정보에 포함된다고 해석하기에는 다소 무리가 있다. 오히려 ‘공공기관 정보공개에 관한 법률’에 의거 적극적으로 제공하는 것이 올바른 해석일 것이다.

공간빅데이터는 건축·도시 분야의 정책수립 뿐만 아니라 민간에서도 활용가치가 매우 높지만, 휴대전화 로그, 신용카드 사용내역, 건강보험 관련기록 등 개인정보 유출로 인한 피해가 우려가 높은 것도 사실이다. 따라서 개인정보 등 민감정보에대한 익명화가 가

능하고 빅데이터로 활용 가능한 분석단위를 개발하고 공공에서부터 분석단위로 제공하는 것이 좋은 방안이다.

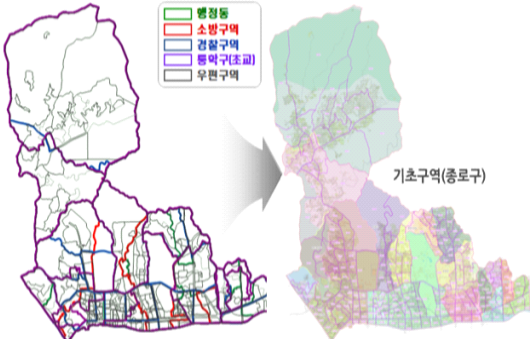


[그림 6-1] 공공데이터 포털
출처 : 공공데이터 포털 www.data.go.kr

우리나라는 2014년부터 국가기초구역을 설정하여 운영하고 있다. 국가기초구역제도는 국민생활의 기초단위지역으로 소방, 경찰, 학교, 선거, 우편, 통계기관 등이 구획하는 각종구역의 기본단위가 되는 지역단위 경계를 지칭하며, 통상 행정경계의 1/10 규모로 기존의 행정읍면동 보다 규모가 작으면서 동질적이고 경계변화가 작은 국가의 최소공간단위를 목표로 설정되어 있다.¹⁾ 전국 3만여 개의 국가 최소단위 구역을 기반으로 지형지물, 인구, 사업장 수 등을 고려하여 도로, 하천, 철도 등 잘 변하지 않는 지형지물을 기준으로 격자 형태로 구분되어 있다. 국가기초구역의 핵심가치는 각종구역 설정의 최소단위로서 뿐 아니라 다양하게 산출되는 국가기초구역기반 공간데이터를 이용하여 공공 및 민간의 다양한 공간의사결정지원이 이루어질 수 있도록 하는 것이다.²⁾

1) 강영욱, 강애피, 국가기초구역기반 의사결정지원시스템 구축전략, 한국공간정보학회 21권 2호, 2013, 85-97
2) 강영욱, 강애피, 국가기초구역기반 의사결정지원시스템 구축전략, 한국공간정보학회 21권 2호, 2013, 85-97

[표 6-1] 공간빅데이터

기관별 관할구역	현 황 도 면	국가기초구역
〈관할구역별 설정 기준〉 • 행정동 : 지번 • 소방 : 법정동, 지번 • 경찰 : 법정동, 지번 • 학군 : 행정동, 통 • 우편 : 법정·행정동, 대형빌딩·아파트 • 통계 : 행정동, 대구역 등		〈국가기초구역 설정 기준〉 • 가급적 불변하는 지형 지물들을 이용하여 인구 및 동일생활권을 고려해 설정 • 하나의 읍면동 면적보다 5~10개 정도로 작게 분할 ※ 하천 : 국가하천, 지방하천 도로 : 대로, 로, 길급능선(일부 행정경계) 등

출처 : 서울특별시(2014), 국가기초구역 관리·활용 계획

공간빅데이터의 효용을 높이기 위해서는 공공데이터부터 선행적으로 국가기초구역 단위로 제공되어야 하며, 특히 거의 모든 공공데이터가 공간빅데이터범주에 포함되는 건축·도시 분야의 관련 통계 역시 이를 포함하여 제공할 수 있는 시스템 기반을 우선적으로 구축하여야 한다. 공공데이터의 국가기초구역 단위 제공 완료 후 민간이 보유하고 있는 휴대전화, 신용카드 등의 데이터 역시 해당 단위로는 제공할 수 있도록 제도적 기반을 마련하고 이를 시행할 필요가 있다. 이렇게 구축된 빅데이터는 단순히 데이터의 활용을 넘어 제4차 산업혁명시대의 기본적인 동력원이자 가장 중요한 역할을 하게 될 것이다.

공공빅데이터는 생산하는 주체가 정부, 지자체, 공공기관 등 매우 다양하며, 한 기관이 구축한 데이터를 다른 기관이 이를 가져다가 해당 기관의 정보를 결합하여 새로운 데이터를 생산하는 경우도 많다. 빅데이터의 활용을 위해서는 한 곳에서 생산된 데이터를 해석하는 것도 중요하지만, 각기 다른 기관의 데이터를 연계 활용하여 새로운 시사점을 찾는 것이 보다 더 중요하다. 이를 위해서는 해당 데이터를 연계할 수 있는 연계플랫폼 개발이 필수적이라 할 수 있다.

뉴욕시의 MODA(New York Mayor's Office of Data Analytics)에서는 기관들 간의 정보의 조정 및 공유를 위하여 기관들이 기관의 데이터에 접근하고 추출 할 수 있는 연계플랫폼인 DataBridge를 개발하여 운영 중에 있다. 현재 약 20개 기관 및 외부 조직(기상청(National Weather Service), 연방 비상 관리국(Federal Emergency Management Administration), ConEdison, Long Island Power Authority, 뉴욕 주정부 포함)에 속한 50개 이상의 자료 시스템을 자료를 연계하고 있다.³⁾

뉴욕시에서는 DataBridge를 통해 불법적으로 개조 된 아파트를 찾는데 건물 검사관이 소요한 시간의 5배를 절약하고, 검사관이 밀수담배를 파는 상점을 발견하는 비율이 2배가 향상(30%에서 82%)되는 등 다양한 효과를 거두고 있다.⁴⁾ 또한 DataBridge를 활용하는 사용자가 오류를 수정하면 데이터 원본을 제공하는 담당자에게 이를 자동으로 전달되도록 하여 데이터의 무결성 유지에도 효과를 거두고 있다.

공공데이터의 개방과 활용을 적극적으로 시행하고 있는 우리나라에서도 뉴욕의 DataBridge의 사례가 시사하는 바가 크다. 건축·도시 분야의 대표적인 빅데이터인 건축물대장데이터의 경우 재산세납입, 에너지통계 등 다양한 데이터의 1차 데이터로 활용되고 있으나 피드백 채널이 없어 원본데이터에서 오류가 있는 자료가 재배포 될 위험성이 상존한다. 또한 데이터를 갱신할 때 역시 매번 정제와 병합작업을 다시 하는 등 비용과 시간이 중복되는 경우도 있어 건축·도시 분야 공공빅데이터 연계플랫폼 개발이 시급하다.

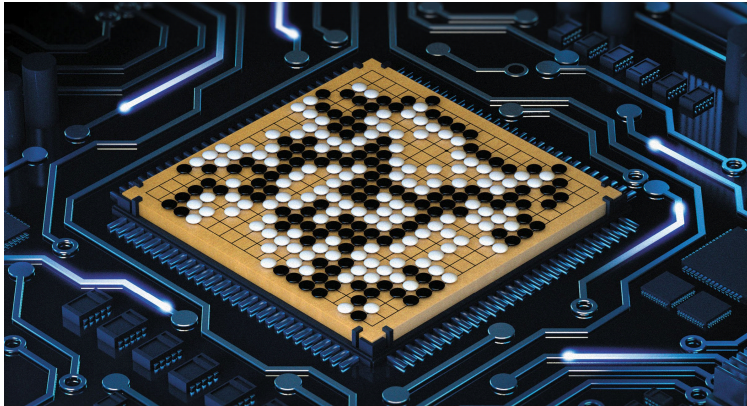
3. 연구의 한계와 기대효과

본 연구에서 개발한 빅데이터 활용한 정책 개발 체계는 구체적인 설계도가 아니라 건축·도시 분야의 정책개발에 빅데이터를 활용할 수 있는 방법론이다. 이는 수립대상 정책에 따라 활용해야 하는 빅데이터의 종류가 다르고, 하나의 빅데이터를 분석해서 정책을 개발할 수도 있지만 여러 종류의 빅데이터를 연계 분석해야 하는 경우가 많이 있다. 데이터의 중요성이 나날이 강조되고 있는 요즘 새로운 빅데이터와 분석방법이 등장하고 있어, 현실적으로 모든 건축·도시 분야 주제에 적용 가능한 통합 플랫폼을 만드는 것은 어렵다. 향후 본 연구를 토대로 건축·도시 분야의 주제별 빅데이터 연구를 통하여 구체적인 개발 체계 수립이 이루어 져야 할 것이다. 그리고 본 연구는 정책의 과정 중 정책 개발에 해당하는 문제형성과 정책형성을 중심으로 수행되었으며, 추후 정책집행과 정책모니터링에 관한 빅데이터 활용 연구도 필요하다. 또한 제 4차 산업혁명 시대의 도래로 인공지능(Artificial Intelligence)에 대한 연구가 관련 분야에서 많이 수행되고 있다. 본

3) NYCAalytics, <http://www1.nyc.gov/site/analytics/index.page> [latest accessed 2017.09.13]

4) Accenture, City of New York: Using Data Analytics to Achieve Greater Efficiency and Cost Savings, 2013

연구에서도 일부 빅데이터의 분석과정에서 머신러닝을 수행하였지만, 건축·도시 분야가 복잡하고 다양한 만큼 인공지능 분야와의 접목을 통한 시너지효과가 클 것으로 기대된다. 이에 관련한 융합 연구의 개발이 필요한 시점이다.



[그림 6-2] AlphaGo

출처: NATURE 표지, Vol.529, No.7587, 1/28/2016

본 연구에서 정책개발체계의 실증을 위하여 데이터의 종류도 다양화 하고, 데이터의 수집, 가공, 분석 또한 자체적으로 진행하였다. 따라서 건축·도시 정책개발을 위해서 빅데이터를 활용하기 위해서는 본 연구와 같이 데이터 수집, 가공, 분석 기술을 갖추고 있어야 할 수 있다고 판단할 수도 있다. 본 연구의 제2장2절에서 소개한 민간과 공공의 빅데이터 분석 서비스 플랫폼을 활용한다면 복잡한 많은 과정을 생략할 수 있다. 따라서 본 연구에서 제시한 빅데이터 활용체계에 따라 어떤 정책과정에서 빅데이터를 활용할 것 인지를 정하고, 해당단계의 질의개발부터 분석결과해석까지 계획을 수립하고, 빅데이터 서비스 플랫폼을 활용한다면 간단하게 빅데이터를 활용하여 여론을 수립하고 정책대안을 개발할 수 있다.

본 연구에서는 빅데이터를 활용한 건축·도시 분야 미래정책 개발체계를 개발하고 이를 실증하였다. 이를 통하여 건축·도시 분야의 정책 개발에 다양한 빅데이터를 활용할 수 있음을 확인할 수 있다. 빅데이터에 기초한 정책개발은 증거기반 정책 수립에 기반이라 할 수 있으며, 데이터를 통한 정량적 분석을 통한 정책개발은 정책의 신뢰성과 합리성을 높일 수 있을 것이라 기대된다.

- 강영욱, 강애띠(2013), “국가기초구역기반 의사결정지원시스템 구축전략”, 「한국공간정보학회지」 V.21(2), 한국공간정보학회
- 김대중·윤서연(2013), 「빅데이터를 활용한 국토정책반응 모니터링 및 정책수요 예측 방안」, 국토연구원
- 김창수·송민정(2014), 「빅데이터 경영론」, 학현사
- 남광우·신강원·김대중·신동빈·안종욱·김승범·최선화·장동인(2014), “[특집] 빅데이터와 스마트도시”, 「도시정보(대한국토도시계획학회 정보지)」, v.390, 대한국토도시계획학회
- 남기범(2006), 「현대 정책학 강의」, 한국학술정보
- 노규성(2016), 「빅데이터와 공공 혁신 10대 사례」, 커뮤니케이션북스(주)
- 박시룡·안순권·장선화·이효정(2013), 「빅데이터를 활용한 과학적·효율적 행정 구현방안」, (주)서울경제신문 백상경제연구원
- 송태민·송주영(2015), 「빅데이터 연구 한 권으로 끝내기」, 한나레아카데미
- 서울특별시(2014), 「국가기초구역 관리·활용 계획」, 서울특별시
- 양현철·김진철·신신애·김배현(2014), 「빅데이터 활용 단계별 업무절차 및 기술 활용 매뉴얼」, 미래창조과학부, 한국정보화진흥원
- 윤미영(2013), 「더 나은 미래를 위한 데이터 분석」, 한국정보화진흥원
- 윤미영·권정은(2012), 「빅데이터로 진화하는 세상」, 한국정보화진흥원
- 이각범(2011), 「빅데이터를 활용한 스마트 정부 구현(안)」, 국가정보화전략위원회
- 이미숙·이창훈·김지연(2014), 「빅데이터를 활용한 환경분야 정책수요 분석」, 한국환경정책·평가연구원
- 이수진·전유나(2015), 「소셜미디어 분석을 통한 경기도 관광 트렌드 연구」, 경기연구원
- 이승환·김인철·권성수·윤익준·조희진(2015), 「지방자치단체 정책지원을 위한 데이터 분석 운영체계 연구」, 대구대학교 산학협력단
- 이영주(2015), 「빅데이터 시대, 미래전략의 새로운 접근법」, 한국정보화진흥원,

- 이원규, 오동하(2013), 「빅데이터(Big Data)를 활용한 부산시 도시관리 전략」, 부산발전연구원
- 이진욱·유국현·문병민·배석주(2017), “감성분석과 Word2vec을 이용한 비정형 품질 데이터 분석”, 「품질경영학회지」, v.45(1), 한국품질경영학회
- 전도홍·김형래·김정래(2015), 「데이터 분석을 중심으로 한 빅데이터 입문」, 정익사
- 정부부처 합동(국무조정실 외)(2016), 「개인정보 비식별 조치 가이드라인」, 국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부
- 정상현(2001), 「정책학의 이해」, 신아출판사
- 정정길·최종원·이시원·정준금·정광호(2014), 「정책학원론」, 서울: 대명출판사
- 정지선(2011), 「新가치창출 엔진, 빅데이터의 새로운 가능성과 대응전략」, 한국정보화진흥원
- 조영진·이은석·이경신(2016), 「건축도시정책동향」, v.39, 건축도시공간연구소
- 차재필(2013), 「빅데이터 정책동향」, 한국교통연구원. 세미나 자료
- 한국소프트웨어기술인협회 빅데이터전략연구소(2016), 「빅데이터 개론」, 광문각
- 한국정보화진흥원(2014), 「빅데이터 활용을 위한 개인정보 비식별화 사례집」, 서울:한국정보화진흥원
- 행정학용어 표준화연구회(2010), 「이해하기 쉽게 쓴 행정학용어사전」, 새정보미디어
- 황명화(2014), 「여론 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구」, 국토연구원
- IDG Korea(2012), 「빅 데이터의 이해」, IDG Tech Report
- Accenture(2013), “City of New York: Using Data Analytics to Achieve Greater Efficiency and Cost Savings”, High Performance Delivered, Accenture
- Agrawal, Rakesh and Ramakrishnan Srikant (1994), “Fast algorithms for mining association rules”, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB. pp.487-499. Santiago, Chile.(1994.9.12.-15.)
- Beyer, Mark A. and Douglas Laney (2012), The Importance of ‘Big Data’ : A Definition, Gartner
- Blei, David M. (2012), “Probabilistic Topic Models”, COMMUNICATIONS OF THE ACM, v.55(4), ACM
- Laney Doug (2001), 3D Data Management: Controlling Data Volume, Velocity, and Variety, META Group
- Manyika, James, Michael Chul, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh and Angela Hung Byers (2011), Big Data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute
- Olofson, Steve Conway and Natalya Yezhkova (2012), Worldwide Big Data Technology and Services 2012-2015 Forecast, idc
- Russell, M. G. et al. (2011), Semantic Analysis of Energy-Related Conversations In Social Media: A Twitter Case Study, International Conference on Persuasive Technology

Turner, Margery Austin (2013), Evidence-Based Policymaking · Requires A Portfolio Of Tools, THE URBAN INSTITUTE

Vesset, Dan, Henry D. morris, Gard Little, Lucinda Borovick, Susan Feldman, Matthew Eastwood, Benjamin Woo, Richard L. Villars, Jean S. Boman, Carl W. Witten, Ian H., Eibe Frank, Mark A. Hall and Christopher Pal (2016), Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann

국토교통부 홈페이지, http://www.molit.go.kr/USR/WPGE0201/m_35312/DTL.jsp

기상청 홈페이지, <http://sts.kma.go.kr>

블로터(2013), “앱리뷰, 스트리밍앱, 타임머신...타맵 4.0”, <https://www.bloter.net/archives/140459>

블로터(2016), “뉴스 빅데이터 분석 시스템 ‘빅가인즈’ 공식출범”, <https://www.bloter.net/archives/254773>

비즈워치(2016), “날씨경영, 빗줄기에 수백억 오락가락”, <http://www.bizwatch.co.kr/pages/view.php?uid=23379>

서울시 공식 관광정보 웹사이트, http://korean.visitseoul.net/tours/K-Pop-EXO-KR_/10457

서울관광마케팅(2014), “[소식] 서울관광마케팅, SK텔레콤, 신한카드와 함께 관광 빅데이터 사업 추진”, <http://www.seoulwelcome.com/kr/?p=5091>

소방청, 소방청국가화재정보센터, <http://www.nfds.go.kr>

스마트인사이트, <http://www.smartinsight.co.kr/pr/industryReport/industryReport.do>

연합뉴스(2016), “어린이날 어디갈까 SNS 빅데이터 분석 결과”, <http://www.yonhapnews.co.kr>

에스리 코리아, <https://www.esrikr.com>

전자신문(2016), “교통카드 빅데이터로 대중교통 개선하고 광고 수익 올린다”, <http://www.etnews.com/20160525000173>

제로노바(2013), “서울시 심야버스 노선 최적화 빅데이터 활용 사례”, <https://zeronova.kr/2013/08/07/seoul-bus-route-optimization/>

조선일보(2015) “남자는 강남 술집, 여자는 강북 맛집...약속장소 남남북녀”, http://news.chosun.com/site/data/html_dir/2015/02/10/2015021000456.html

한겨레(2016), “개인정보 비식별 정부 가이드라인 따랐다간 큰코다친다”, <http://www.hani.co.kr/arti/economy/it/754640.html>

한국정보통신기술협회(2013), “네이버 지식백과: 소셜 계층”, ICT 시사용어 300. <http://terms.naver.com/entry.nhn?docId=3587142&cid=59277&categoryId=59280>

한국행정학회(2013), “행정학전자사전: 정책과정론”, 한국행정학회, http://www.kapa21.or.kr/data/kapa_dictionary_view.php?num=957

한종수(2013), “서울시, 빅데이터로 심야버스 노선 구축”, 뉴스1. <http://news1.kr/articles/1219217>.

Amsmarterdam city, “A tool that stimulates the use of renewable energy”, <https://amsterdamsmartcity.com/projects/energy-atlas>

Amsmarterdam Energy Map, https://maps.amsterdam.nl/energie_gaselektra/

Butler, Declan (2013), "When Google got flu wrong", Nature,
<http://www.nature.com/news/when-google-got-flu-wrong-1.12413>

Garrahan M.(2012), "Hollywood turns to tweet analysis", Financial Times,
<https://www.ft.com/content/9c810f04-907e-11e1-9e2e-00144feab49a>

Google, "Google Flu Trends", <https://www.google.org/flutrends/about/>

Heaton, Brian(2015), "New York City Fights Fire with Data",
<http://www.govtech.com/em/safety/New-York-City-Fights-Fire-Data.html>

Komoran : www.shineware.co.kr / twitter : <https://github.com/kkma> : kkma.snu.ac.kr

National Public Radio, "Google's Flu Tracker Suffers From Sniffles", <https://www.npr.org/>

NTT Docomo, <https://www.nttdocomo.co.jp>

NYC Crime map, <https://maps.nyc.gov/crime/>

NYC Analytics, <http://www1.nyc.gov/site/analytics/index.page>

Redflagnews, <http://www.redflagnews.com/headlines/senators-warn-911-terror-network-still-active-in-us>

Stamen, <https://stamen.com/work/crimespotting/>

Urban Omnibus(2014), "The Analytics of Fire", <http://urbanomnibus.net/2014/06/the-analytics-of-fire/>

WEKA, "Weka 3: Data Mining Software in Java", <http://www.cs.waikato.ac.nz/ml/weka/>

Policy Development Framework of Architecture and Urbanism using Big Data

SUMMARY

Cho, Young-Jin
Yoo, Kwang-Heum
Kim, Shin-Sung

Evidence-based policy making is an approach of policy development that is effective in the field of architecture and urbanism where public interest is high, civil affairs are often generated, and policy demand is changing quickly. However, to date, there have been very few researches on this issue. In order to improve the quality of the architecture and urban policy, it is urgent to study the evidence-based policy making formulation method.

Advanced countries are actively adopting big data as a technology innovation and a new problem solving method for policy development. In Korea, there is a need to introduce big data such as public data and social data into policy development. In particular, it is necessary to develop a methodology that can diagnose and predict the demand and change of rapidly changing architecture and urbanism policies by using big data.

The first objective of this study is to find out how to use big data in the field of architecture and urbanism. The second purpose is to derive a policy development method that uses big data in the field of architecture and urbanism. The final objective is to present a policy development framework for evidence-based policies in the field of architecture and urbanism.

The characteristics of big data are described according to data volume, velocity, variety, veracity, and value. Depending on the type of data, it can be classified into formal, informal, and semi-formal data. Formal data refer to data stored in a certain field, and can be managed by a database such as building information, while informal data refer to data such as blogs, social data, news media, and the like, which are not defined in an item.

Policy process theory is a theory that deals with the process from the recognition of policy problems to the finalization of policy through the process of goal setting, alternative analysis, decision making, legalization, enforcement, and evaluation. The representative policy process model that is used in policy studies is Anderson's model (1975). It is composed of five steps: problem identification and agenda formation, formulation, alternative adoption, implementation, and evaluation. The scope of policy development corresponds to problem identification and agenda formation, formulation, alternative adoption in the five stages of the policy process model.

The problem identification and agenda formation stage is necessary to monitor public opinion on specific topics and issues. The formulation stage is focused on creating policy alternatives by combining and linking various information. The alternative adoption stage is to find the best alternative among the various alternatives derived from the formulation stage.

The big data based architecture and urbanism policy development framework proposed in this study is composed of 'development of query', 'determination of analysis contents', 'comprehension of data types and characteristics', 'design analysis method', 'data collection and processing', 'analysis and interpretation of results'. Pilot test was conducted for building maintenance policy to verify the framework.

In order to verify the problem identification and agenda formation stage, main questions of the building maintenance policy were derived. Based on the extracted key questions, public opinion analysis on building maintenance policy was conducted. The analysis model used is keyword frequency analysis, related keyword analysis, time series trend analysis, and emotional analysis. For the purpose of publicity analysis, social and news data from April 1, 2014 to April 30, 2017 were collected using a web crawler.

As a result of the keyword analysis of the maintenance of the building, the urgent keyword to solve the policy problem was 'safety'. Next, the result of detailed analysis of 'safety' keyword revealed that the keyword with the highest number of

buzz was 'fire', which showed that public opinion about 'fire' was the most interested. Finally, based on the results of public opinion analysis, 'fire safety of buildings' was chosen as the policy agenda.

In order to verify the formulation stage, main questions of fire safety of buildings were derived. The analysis model was designed based on the derived query and analyzed. The data used for the analysis were fire statistics, building register information, weather information, and census information from 2012 to 2016.

As a result of the analysis, it was confirmed that the building fire occurred most in the residential space and the casualties were also concentrated there. Combined analysis of fire statistics and meteorological data revealed that a 10% drop in monthly average humidity could increase the incidence of fire by about 19%. Combined analysis of fire statistics and census showed that most of the villages had very little fire, but very few villages had very frequent fires, which revealed the power law between village and fire frequency. In addition, according to the combined analysis of fire statistics and building property information through machine learning, the specific types of buildings where fire is frequent were found.

Through case studies, three policy alternatives related to fire safety were developed in the maintenance management policies of buildings: 1) improvement of the fire safety of non-residential buildings, 2) prevention of large-scale fire in industrial facilities, 3) establishment of flexible fire fighting system among local governments.

The big data based policy development framework developed in this study demonstrated through the research that it can develop objective and scientific data based policies as a methodology to implement evidence-based policy. This evidence-based policy development framework could be transparent to the decision-making process, and policy agenda and policy alternatives could also be implemented based on the objective data. In addition, evaluations and feedback could be easier because they are developed based on data. Therefore, it is expected that policy development through quantitative analysis of data would enhance the credibility and rationality of policy.

Keywords :

Big Data, Policy Development Framework, Evidence-based Policy, Building Maintenance, Building Fire