

빅데이터와 환경

02. 빅데이터 분석 방법론과 활용 사례



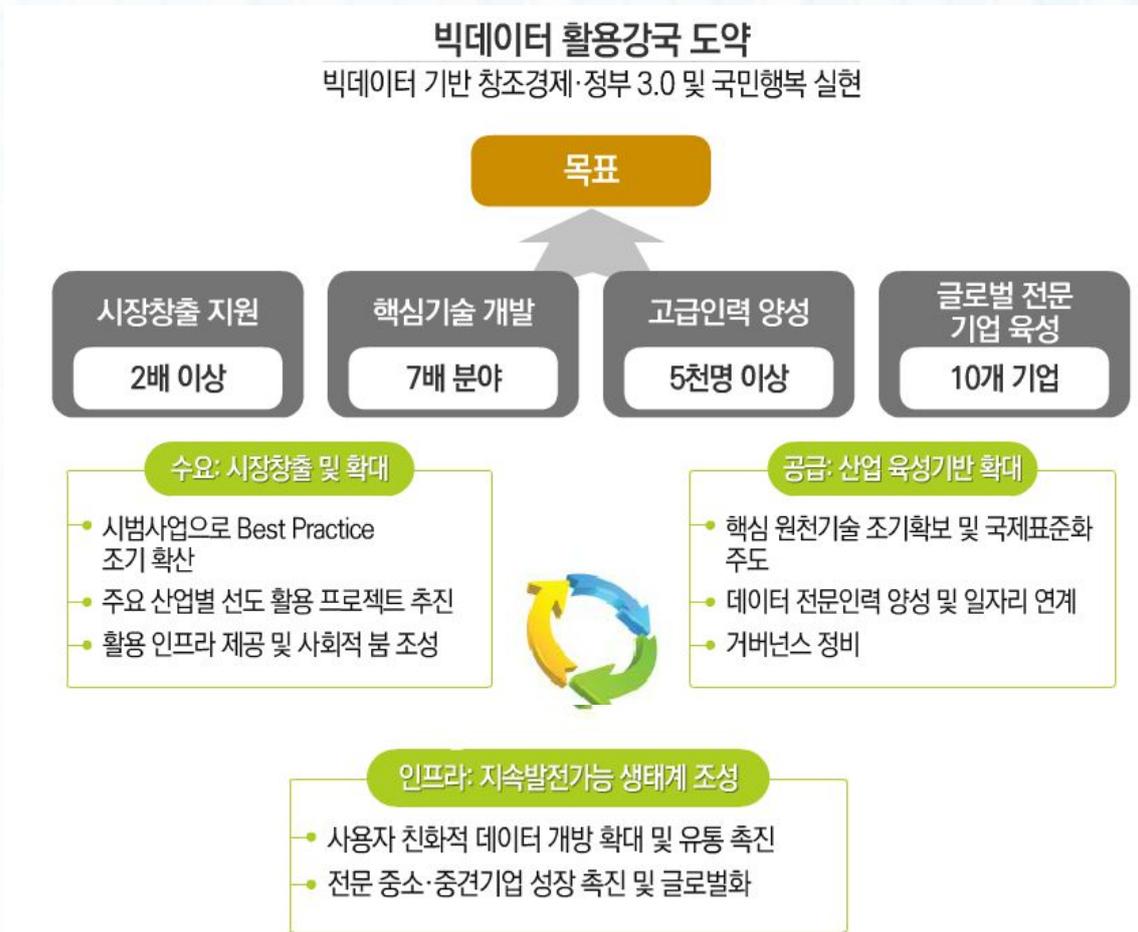
1. 빅데이터 정책 현황

(1) 우리나라

지난 강의에서 살펴본 바와 같이 빅데이터는 기업의 생산성 제고에서부터 보건, 복지, 교육 등 공공서비스 개선에 이르기까지 다양한 분야에 활용되고 있다. 빅데이터 활용의 가장 큰 장점은 직관에 의존하던 기존의 의사결정 방식에서 벗어나 정보에 기반한(evidence-based) 과학적 정책 수립이 가능해졌다는 점이다. 특히 공공부문의 경우 하향식(top-down) 의사결정 방식에서 벗어나 일반 국민들의 의견을 직접 반영한 상향식(bottom-up) 정책을 수립할 수 있다는 점에서 빅데이터의 활용 가능성이 크다고 할 수 있다. 실제로 트위터, 페이스북 등의 소셜미디어는 정부 정책에 대한 국민들의 인식을 직접적으로 파악하는 데 용이하게 활용될 뿐만 아니라, 의견을 개진할 수 있는 의사소통의 장으로도 역할을 수행하고 있다.

우리나라에서는 2013년 12월 미래창조과학부가 관계부처 합동으로 빅데이터의 활용을 촉진하고 관련 산업을 육성하여 새로운 국가 원동력으로서의 토대를 마련하기 위한 「빅데이터 산업 발전전략」을 발표하였다. 국내 시장은 잠재력에 비해 불확실성 등의 리스크 때문에 빅데이터의 활용이 저조한 상태이며, 이를 해결하기 위해 수요, 공급, 인프라 측면의 중점전략을 바탕으로 빅데이터 활용을 촉진하고 관련 산업 육성을 전략적으로 지원한다는 내용이다. 빅데이터 산업 발전전략의 주요 목표로는 국내 빅데이터 시장 2배 이상 확대, 7개 분야에서 핵심 원천기술 개발을 통한 기술경쟁력 제고, 5,000명 이상의 고급인력 양성, 그리고 10개 이상의 글로벌 전문기업 육성 등이 제시되었다.

[그림 1] 빅데이터 산업 발전전략의 목표 및 과제



출처: 관계부처 합동(2013). 창조경제 및 정부 3.0 지원을 위한 빅데이터 산업 발전전략. 경제장관회의.

또한 정부는 공공기관의 정보공개 기준을 완화하고 공공데이터포털(<https://www.data.go.kr>)을 통해 공공정보를 개방하는 등 빅데이터 이용 활성화에 앞장서고 있다. 2013년 기준으로 공공데이터의 약 16%를 일반에 공개하였고, 2017년까지 60% 개방을 목표로 하고 있다.

(2) 미국

해외 주요국들도 자연재해, 테러, 탈세, 전염병 확산 등 더욱 복잡해지는 국가 현안들을 해결하기 위한 방안으로 빅데이터를 적극 활용하고 있으며, 미국은 사회·경제·정치 등의 분야에서 빅데이터를 가장 활발하게 적용하고 있는 국가 중 하나이다. 오바마 정부는 범정부 차원에서 빅데이터 거버넌스 체계를 구축하는 등 빅데이터 기술 개발을 지원·활용하기 위한 방안을 다각적으로 모색해 왔다. 일찍이 2010년 대통령 과학자문위원회가 IT기술 관련 보고서를 통해 연방정부 차원에서 빅데이터 전략의 필요성을 강조하였고, 2012년 3월 대통령실 내 과학기술정책실은

‘빅데이터 연구개발 이니셔티브(Big Data R&D Initiative)’를 발표하고 약 2억 달러 규모의 투자계획을 수립하여 지원하고 있다.

이처럼 적극적인 정부의 지원을 바탕으로 미국 내에서는 이미 의료, 국방, 교통, 기후, 범죄예방 등 다양한 분야에서 빅데이터 기술이 활용되고 있다(표 1). 대표적으로 미국 국세청은 정부 내외의 빅데이터를 통해 세금 납세자들의 행동을 분석하고 탈세 증후를 예측해 관련 범죄를 사전에 차단하는 시스템을 구축한 바 있다. 2010년 기준 미국 내 탈세 규모는 저소득층 의료보장 총액을 초과하여 약 3,373억 달러에 달하였으며, 탈세로 인한 국가재정 손실 문제가 크게 대두되었다. 이에 미국 국세청은 정부기관의 데이터뿐만 아니라 소셜미디어 데이터 등을 통합적으로 활용해 납세자의 계좌정보 일치 여부를 확인하고 탈세징후를 포착·감시하고 있다.

[표 1] 미국의 빅데이터 활용 사례

기관명	프로젝트/시스템명	활용
미국 국립 문서기록 관리처(NARA)	전자기록 아카이브 (Electronic Records Archive)	정부 문서들을 디지털화하여 정부·민간에서 편리하게 이용하도록 제공
미국 에너지부(DOE) & TerraEchos	외곽침입 감지 시스템 (Perimeter Intrusion Detection)	국경지대 주요 기반시설들을 보호 및 모니터링하고 위험 조기 감지
온타리오 대학	메디컬 모니터링 (Medical Monitoring)	생후 24시간 이내 영아의 감염 여부 예측 진단
미국 항공우주국(NASA)	유인 우주비행 이미저리 (Human Space Flight Imagery)	기업 및 일반 대중에 우주와 관련된 이미저리 제공
AM Biotechnologies	DNA 염기서열 분석 (DNA Sequence Analysis for Creating Aptamers)	방대한 DAN 염기서열 정보를 분석하여 질병 치료 목적으로 특수한 앵타머 화합물 합성
미국 국립 해양대기청(NOAA)	일기예보 서비스 (National Weather Service)	날씨, 기후 데이터를 분석하여 일기예보 및 자연재해 경보 정보 제공
미국 국세청(IRS)	납세정보 저장소 (Compliance Data Warehouse)	납세자들이 올바르게 세금을 납부하도록 관리
미국 메디케어 & 메디케이드 센터(CMS)	의료기록 분석 (Medical Records Analytics)	의료기록을 통합적으로 관리하여 미국인들의 건강 관리

자료: TechAmerica Foundation(2012), Demystifying Big Data: A Practical Guide to Transforming the Business of Government.

(3) 영국

유럽연합의 공공부문 데이터 개방 정책과 함께 영국 정부는 2010년부터 세출, 범죄율, 질병 등 중앙정부와 지방자치단체의 공공데이터를 일반에 공개해 왔다. 2012년 3월에는 ‘데이터전략위원회(Data Strategy Board)’를 설립하고 16개 부처별로

각각의 특성에 맞는 오픈데이터 전략을 수립하도록 하였으며, 이에 따라 각 부처는 데이터 접근성을 강화하고 데이터 공개 서비스를 활성화시킬 수 있는 방안을 모색하였다.

또한 영국은 불확실한 미래에 대응하고 기회를 포착하기 위하여 일찍이 미래이슈 분석에 관심을 가졌다. 미래이슈 분석은 정부, 민간, 학계에 축적된 수많은 정보들을 검토하여 신뢰할 수 있는 정보만을 추출하고 특정 분야에 대한 미래를 예측하는 활동이다. 이를 위해서는 수많은 정보의 홍수 속에서 신뢰할 수 있는 정보를 추출하여 분석하는 기술이 중요하다. 영국 정부는 1990년대 초부터 정부부처 간 협력을 통해 과학분야의 주요 선진기술을 예측 및 연구하는 방법론을 개발하기 시작하였으며, 1994년에 영국 과학기술국(Office of Science and Technology)하에 '미래이슈 분석 프로그램(Foresight Programme)'을 설치하였다. 초기에는 과학기술 분야의 잠재된 기회를 포착하는 데 집중하였으나, 점차 분석대상을 확대하여 현재는 기후변화 대응, 비만대책 수립, 식량안보 대응, 사이버 범죄 방지 등 다양한 분야에 대한 분석을 수행하여 정부정책 수립에 적극적으로 활용하고 있다.

(4) 프랑스

프랑스도 정부 차원에서 정부와 공공기관의 데이터를 일반에 공개하고 있으며, 공공데이터의 활용도 또한 유럽 국가 중에서 비교적 높은 편이다. 2011년에 '국무총리 산하 공공정보 개방 및 공공데이터 플랫폼 발전을 위한 서비스(Etalab)'와 '국가정보사이트(data.gouv.fr)'를 통해 공공데이터의 폭넓은 활용을 지원하였으며, 2013년 초에는 빅데이터 분석 관련 프로젝트가 1,150만 유로 규모의 국가미래투자계획(Cadre des Investissements d'avenir) 지원 대상으로 선정되어 빅데이터에 대한 정부 차원의 관심이 높은 것을 확인할 수 있다.

특히 환경 관련 정책에서 정부, 지방자치단체, 기업, 시민 등 이해당사자들이 협력하여 빅데이터를 활용한 사례를 찾아볼 수 있다. 예를 들어, 프랑스 내에서는 연간 약 23만 톤 이상의 플라스틱 생수통이 버려지고 이 중 절반 정도만 재활용되고 있다. 프랑스 정부는 이러한 문제를 해결하기 위하여 빅데이터를 활용한 수원(水源)검색 시스템(Eaupen)을 구축하고 국민들이 언제 어디서든 수원에 대한 정보를 얻을 수 있도록 제공하였다. 즉, 지방자치단체와 민간업체(상점, 레스토랑) 등의 도움을 받아 프랑스 전역의 모든 분수, 수도 등 수원에 대한 정보를 집계하며, 시민들은 인터넷과 휴대전화를 통해 편리하게 정보를 확인할 수 있고, 언제든지 수원 위치에 대한 정보를 추가할 수 있어 양방향 정보 공유가 가능하다. 이러한 정부의 서비스는 결국 플라스틱 통에 담긴 생수의 구매를 절감시키고 환경보호에 기여할 것으로 기대된다.

(5) 스웨덴

스웨덴 역시 유럽연합 차원의 공공데이터 개방정책 기조를 따르며, 정부의 빅데이터 분석 역량을 증대하기 위한 정책들을 수립하고 있다. 2013년에는 스웨덴 성장정책분석청(Tillväxtanalys)이 국가혁신과 경제발전에 빅데이터를 활용하기 위한 사전작업으로 미국·영국·일본·한국·인도 등의 사례연구를 수행한 바 있다. 스웨덴 기술혁신청(MINNOVA)은 빅데이터 분석을 국가전략적 혁신의제로 선정하고 산업계, 국가기관, 연구기관, 대학으로 구성된 국가 빅데이터 분석 네트워크를 구축하였다. 또한 공공데이터 개방을 위한 오픈데이터 포털을 개설하였으며, 환경분야의 데이터로는 스웨덴 환경청(Naturvårdsverket)이 산업별 온실가스 배출량, 이산화질소 배출량 등의 정보를 제공한다.

환경분야의 대표적인 빅데이터 활용 사례로는 스웨덴 왕립공과대학(KTH)에서 수행한 스톡홀름 시 교통개선방안 연구가 있다. 스웨덴의 수도이자 인구밀집 지역인 스톡홀름의 교통 흐름을 실시간 데이터 분석으로 관리하고, 교통혼잡제 시스템을 보완해 수송부문 온실가스 배출량 감축을 달성하기 위한 목적이다. 이를 위해 도시 출입로에 설치된 카메라, GPS, 대중교통망, 날씨정보 등을 종합적으로 분석하여, 운전자들에게 실시간 교통상황 및 최적운행경로 정보를 제공하고 교통량을 분산시킨다. 또한 주변 국가인 덴마크의 풍력발전 설비기업 Vestas가 지리정보, 날씨정보 등을 분석해 풍력발전 터빈을 설치할 최적의 장소를 선정하는 사례가 있어, 향후 스웨덴의 풍력발전 설비 및 에너지 효율 향상에 활용될 것으로 예상된다.

(6) 호주

호주 정부는 2012년 10월 정부의 공공서비스에 정보통신 기술을 적극적으로 활용하겠다는 내용이 포함된 국가전략 'Australian Public Service ICT Strategy 2012-2015'를 발표하고 빅데이터 전략의 필요성을 강조하였다. 또한 2013년 2월에는 빅데이터 관련 기관들이 참여한 빅데이터 워킹그룹을 구성하여 빅데이터 전략을 수립·이행토록 하였으며, 빅데이터 관련 국가역량 강화를 위하여 데이터분석센터(DACoE)를 설치하기도 하였다. 이러한 정부의 지원에 힘입어 호주 연방과학원에서는 자체적으로 소셜미디어 모니터링 소프트웨어인 Vizie를 개발하였다. Vizie는 소셜미디어 데이터를 모니터링하고 분석하는 소프트웨어로, 소셜미디어상의 주요 이슈를 포착하고, 여러 온라인 콘텐츠를 분류하며, 자체적으로 분석 내용을 시각화하는 기능이 있어 사용자가 손쉽게 분석 결과를 파악할 수 있다.

호주 이민국(Department of Immigration and Citizenship, DIAC)에서는 빅데이터를 활용한 국경경보시스템(Border Risk Identification System)을 구축하여 국제공항 등

국경 지대에서 발생할 수 있는 위험에 대비하고 있다. 국가간의 자유로운 방문이 빈번해지면서 테러 등의 위험에 대응하기 위하여 공항·항구 등에서 출입국 절차를 강화하였는데, 이는 출입국 심사에 소요되는 시간을 크게 증가시켜 업무의 비효율성이 문제로 지적되었다. 호주 정부는 이러한 문제를 해결하기 위하여 국경경보시스템을 구축하고 정부부처 및 기관에 축적되어 있는 데이터들을 종합적으로 분석하는 방식을 통해 실시간으로 잠재적 위험 인물(입국거부 대상자 등)을 선별하고 있다. 이러한 시스템은 출입국 심사 대기시간을 단축시키고 업무 효율성을 높인 것으로 평가받고 있다.

[그림 2] 호주의 국경경보시스템(Border Risk Identification System)



자료: Felsche(2013), Future of Business Intelligence & Analytics in Financial Services.

(7) 아시아

아시아 지역에서는 대표적으로 싱가포르와 일본이 공공정책 수립에 빅데이터를 활용하고 있다. 싱가포르는 테러, 전염병 확산 등 국가안보 문제에 초점을 맞추어 잠재된 위험에 체계적으로 대응하기 위한 방안으로서 'Risk Assessment and Horizon Scanning(RAHS)' 시스템을 구축하였다. RAHS 시스템은 정부 및 국가기관에 산재되어 있는 방대한 데이터 시스템간의 연계를 통해 정보를 공유하고 안보 문제에 신속하게 대응할 수 있도록 한다. 특히 싱가포르는 지리적으로 해상안보의 비중이 크기 때문에, RAHS 시스템을 통해 해상 교통량, 선로 패턴 등을 수집·분석하여 전략적 요충지를 파악하고 해안지역에서의 안보를 구축하고 있다. 또한 전 세계적으로 조류독감이 유행하던 2007년에는 조류독감의 전염경로와 전염 시 발생할 수 있는 여러 가지 시나리오를 예측하고 대응방안을 마련하는 데 이 시스템을 활용하였다.

일본의 경우 2012년 7월 정보통신 기술의 활용을 활성화하기 위해 'Active Japan ICT' 전략을 발표하였으며, 국가경쟁력 제고를 위한 5대 주요 추진전략으로 빅데이터를 활용한 사회·경제 성장을 포함시켰다. 이에 따라 일본 총무성은 빅데이터 및 오픈데이터 신규시장 창출을 위해 2013년도 예산으로 89.3억 엔을

배정하였으며, 온라인 포털(e-gov.go.jp)을 통해 공공데이터를 공개하는 등 빅데이터 활용을 위해 다양한 정부 전략을 모색하고 있다.

2. 빅데이터 활용 사례

(1) 공공부문 빅데이터 활용 사례

빅데이터는 데이터 생성주체와 유형에 따라 정부 출처의 공공데이터부터 실시간으로 생성되는 소셜데이터까지 다양한 특성의 데이터를 포함한다. 데이터의 특성뿐만 아니라 활용주체에 따라 다양한 분야에서 활용될 수 있는데, 일반적으로 활용 사례는 크게 민간부문과 공공부문으로 구분할 수 있다. 현재 사용되고 있는 대부분의 빅데이터 분석 플랫폼들은 민간부문에 초점이 맞춰져 있다고 볼 수 있다.

민간부문에서의 빅데이터 활용 사례로, 멕시코의 통신사 Telcel은 소셜미디어 분석을 통해 고객행동 패턴을 분석하고 고객이탈 가능성을 예측함으로써 예상 해지율을 절반 이상 감축하는 성과를 도출하였다. 또한 자동차 제조회사인 BMW는 보증수리 분석시스템을 통해 다양한 유형의 원시 데이터를 통합하여 자동차 품질 문제의 원인을 신속하게 식별하고 대응한 결과, 차량당 보증수리 청구율이 1.1%에서 0.85%로 감소했으며, 보증수리 및 반복 보증수리 건수가 감소하여 연간 약 3,000만 유로의 비용을 절감할 수 있었다.

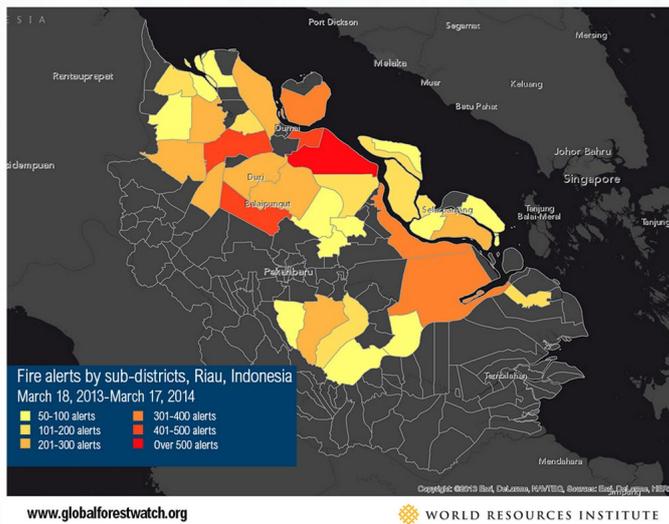
반면 공공부문에서는 맞춤형 정책을 통해 공공서비스의 질을 제고하거나 국가 정책을 수립하기 위한 근거자료로 빅데이터가 활용된다. 앞서 살펴본 바와 같이, 이미 많은 국가가 다양한 공공서비스 부문에 빅데이터 기술을 적용하고 있다. 국가안보 차원에서 범정부적인 데이터 수집·분석체계를 마련하여 세금탈루 방지, 테러 사전 탐지, 국경 모니터링에 빅데이터를 이용하기도 하며, 전국적인 의료정보를 통합적으로 관리하여 개인 맞춤형 의료 서비스를 제공하기도 한다. 우리나라에서도 영유아 보육정책, 청소년 자살예방정책, 재난관리 및 방지정책을 위해 소셜 빅데이터 분석을 적용한 사례가 있다.

(2) 환경부문 빅데이터 활용 사례

공공부문에서의 이러한 노력에도 불구하고 아직 환경분야에서는 빅데이터를 활용한 사례가 충분하지 않은 상황이다. 하지만 소셜미디어는 현실 세계의 물리적인 장벽을 허물어 장소와 시간에 구애받지 않고 상호작용을 가능하게 한다는 점에서 산림보호, 기후변화 등 국제적인 환경문제 대응에 더욱 유용하게 활용될 수 있는 잠재력이 있다. 일례로, 다양한 환경문제를 연구하는 국제 민간연구소 World Resource Institute(WRI)는 2014년 2월에 전 세계 산림자원의 손실 및 조림 현황을 실시간으로 파악할 수 있는 온라인 모니터링 시스템을 구축하였다. Global Forest

Watch(GFW)라고 불리는 이 시스템은 미국항공우주국(NASA)과 구글의 인공위성 데이터와 함께 일반인의 소셜미디어 데이터를 적극 활용한다. 크라우드소싱(crowdsourcing) 방식을 통해 트위터를 사용하는 사람은 누구나 산림과 관련된 정보나 사진, 비디오 등을 실시간으로 업로드할 수 있으며, 이는 지리정보 데이터 및 정부 데이터와 함께 전 세계 산림 현황을 모니터링하기 위해 중요하게 활용된다. 이러한 정보수집 방식은 기존에 파악하기 어려웠던 산림의 작은 변화도 실시간으로 확인할 수 있도록 하여, 산림에 대한 이슈가 발생할 경우 신속하게 대응할 수 있도록 하고 있다(그림 3).

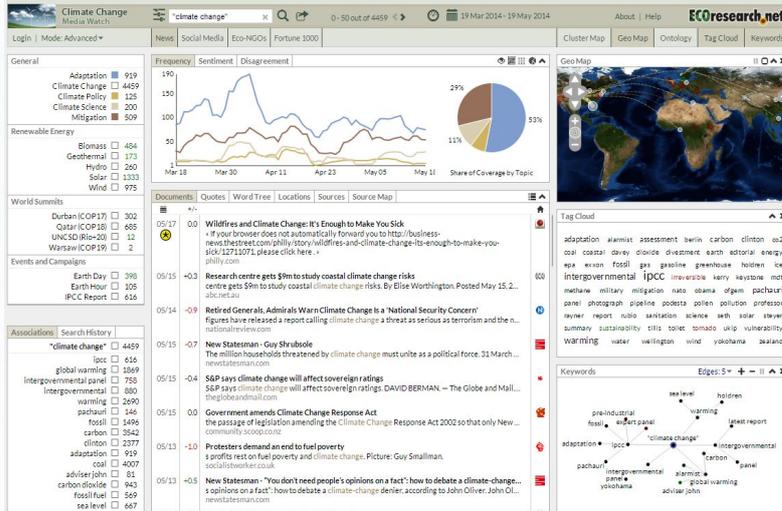
[그림 3] Global Forest Watch의 산림 화재위험 모니터링



자료: Global Forest Watch 홈페이지 참조(<http://blog.globalforestwatch.org/>).

환경분야에 소셜 빅데이터 기술을 활용한 또 다른 사례로 기후변화 모니터링 시스템을 들 수 있다. 오스트리아의 모듈 대학교는 2011년 기후변화 대응을 위한 미디어 모니터링 시스템(Media Watch on Climate Change)을 구축하였다. 이 시스템은 과학 논문에서부터 뉴스, 트위터, 유튜브 등 소셜미디어에 이르기까지 다양한 출처의 정형/비정형 데이터를 분석하여 기후변화 문제를 실시간으로 모니터링하고 있다. 사용자가 기후변화와 관련된 키워드를 입력하면 그와 연관된 최신 자료 및 지리적 정보를 확인할 수 있으며, 연관어 분석, 감성 분석 등의 서비스도 제공하여 다양한 분석이 가능하다(연관어 분석, 감성 분석 등에 대해서는 이후 3절에서 제시할 예정). 영어 이외의 언어는 아직 데이터에 포함되지 않아 한계는 있지만, 전 세계적인 관심 추이를 한눈에 볼 수 있어 기후변화와 관련된 동향을 파악하고 정책을 수립하는 데 유용하게 활용될 것으로 기대된다(그림 4).

[그림 4] Media Watch on Climate Change 홈페이지 정보



3. 빅데이터 분석 방법

지금까지 공공부문, 특히 환경 분야에서 빅데이터 분석을 활용한 사례에 대해 살펴보았다. 이러한 빅데이터 분석을 수행하기 위해서는 전문적인 분석 방법을 필요로 하는데, 본 강의에서는 비정형 데이터인 소셜 미디어 데이터를 분석하는 기법에 대해 집중적으로 살펴보도록 한다.

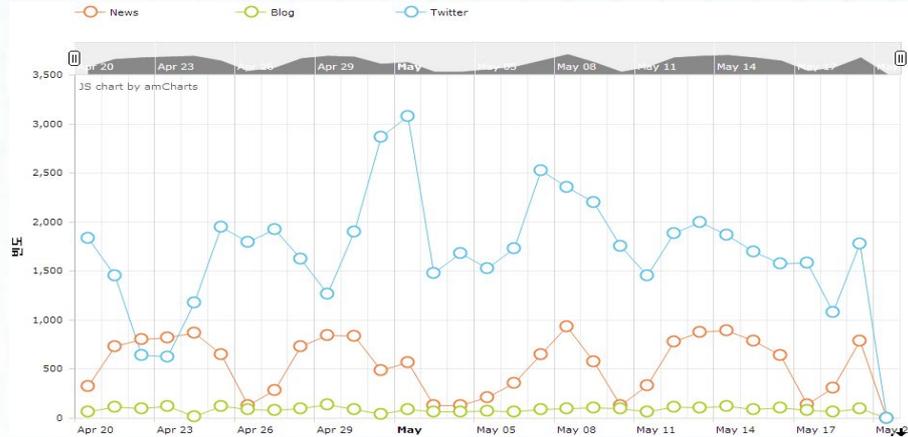
소셜 빅데이터 분석을 수행하기 위해서는 데이터 수집, 텍스트 마이닝, 데이터 시각화 등 다양한 기술과 역량을 필요로 한다. 특히 텍스트 마이닝(text mining)은 사람이 사용하는 언어로 표현된 텍스트 데이터를 자연어 처리(natural language processing) 방식으로 가공·처리하여 유용한 정보를 추출하는 기술로 빅데이터 분석의 핵심이라고 할 수 있다. 소셜 빅데이터 분석은 목적에 따라 다양한 분석 기법이 사용되며, 대표적으로 빈도 분석, 감성 분석, 연관어 분석, 이슈 반응 분석, 네트워크 분석 등이 있다.

(1) 빈도 분석

빈도 분석(frequency analysis)은 가장 기본적으로 활용되는 소셜 빅데이터 분석 기법 중 하나로, 관심 키워드를 포함한 채널별 문서의 발현빈도 총량을 통해 해당 주제에 대한 국민의 관심도를 분석한다. 특정 기간 동안 관심 키워드에 대한 발현빈도의 변화를 통해 해당 주제에 대한 국민의 관심도 변화를 살펴보는 추이 분석(trend analysis)도 함께 수행할 수 있으며, 서로 다른 키워드에 대한 빈도 총량을 비교하여 관심도의 차이를 확인할 수 있다. 예를 들어, [그림 5]는 '환경'이라는 키워드에 대한 트위터, 뉴스, 블로그 채널의 문서 발현빈도를 나타낸 결과이며, 지정된 기간 동안에 일별로 환경에 대한 국민의 관심이 어떻게

변화하였는지 확인할 수 있다.

[그림 5] 환경 키워드에 대한 빈도 분석 사례



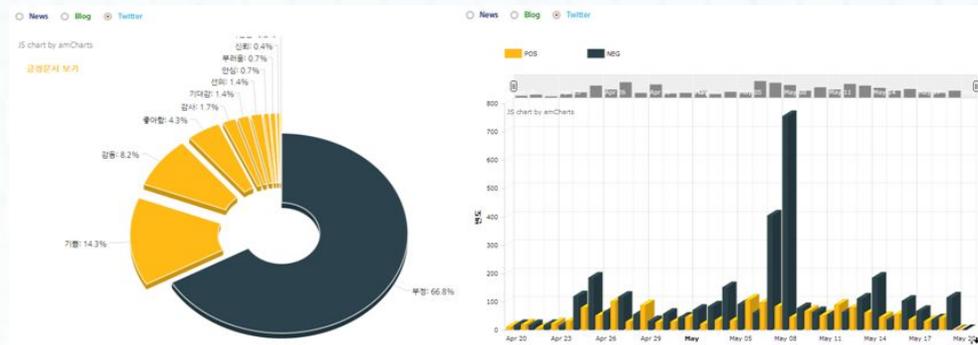
자료: 마인즈인사이트 홈페이지(<http://www.mindsinsight.co.kr>).

또한 빈도 분석을 통해 특정 이슈 또는 사건에 대한 국민의 반응 정도나 변화를 확인할 수도 있다. 예를 들어, 채널별 문서의 평균 발현빈도에 대한 특정 이슈의 최대 발현빈도의 비율을 분석하여 국민들이 해당 주제에 얼마나 민감하게 반응하는지 볼 수 있으며(민감도 분석, sensitivity analysis), 특정 사건을 전후로 발현된 문서 빈도를 비교하여 사건에 대한 반응도를 확인할 수도 있다(이슈반응 분석, issue response analysis).

(2) 감성 분석

감성 분석(sentiment analysis)은 소셜 미디어의 특성을 잘 보여주는 분석 기법이다. 일반적으로 사람들은 트위터, 페이스북, 블로그 등 사적인 매체에서는 좀 더 솔직한 의견을 자유롭게 개진하는 성향이 있다. 감성 분석은 이러한 특성을 이용하여 관심 키워드가 포함된 문서에서 긍정 또는 부정 감성이 발현되는 빈도를 통해 해당 주제에 대한 국민들의 여론을 파악한다. 예를 들어, [그림 6]은 트위터 채널에서의 '환경' 키워드에 대한 감성 분석 결과를 나타내며, 지정된 기간 동안의 일별 감성이 어떻게 변화했는지 확인할 수 있다.

[그림 6] 환경 키워드에 대한 감성 분석 사례

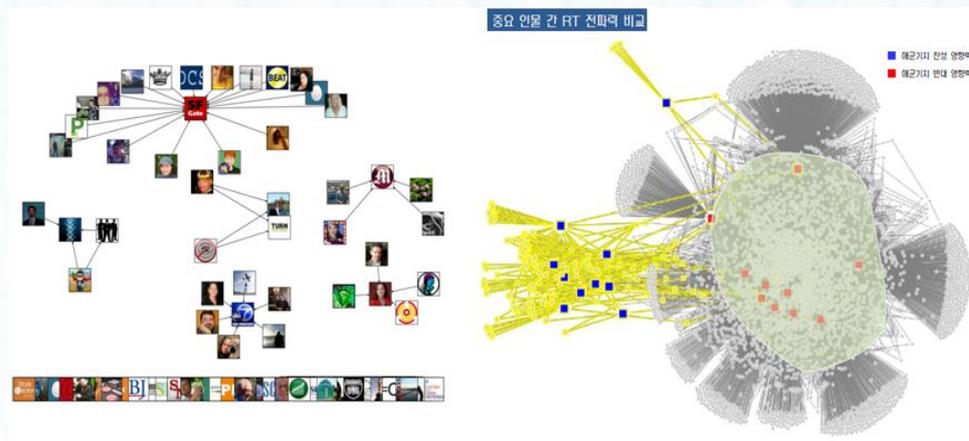


자료: 마인즈인사이트 홈페이지(<http://www.mindsinsight.co.kr>).

(3) 연관어 분석

연관어 분석(association keyword analysis) 역시 유용하게 활용되는 빅데이터 분석 기법으로, 관심 키워드를 포함한 대상 문서 내에서 함께 발현되는 키워드를 추출하여 국민들이 해당 키워드와 어떤 주제를 연결시켜 생각하는지 분석한다. 분석 목적과 기술에 따라 연관어를 일반명사, 인물, 장소, 서술어 등으로 그룹화 하기도 하며, 이를 시각화하는 방법 역시 다양하다. 예를 들어, [그림 7]의 (A)는 '환경' 키워드에 대한 연관어를 발현빈도가 높은 순서에 따라 원형으로 나열하여 제시하며, (B)는 'Smart Grid'에 대한 연관어를 워드 클라우드(word cloud) 이미지를 통해 시각화하여 제시한 사례이다. 또한 (C)는 'Energy' 키워드에 대한 해시태그(hashtag)를 분석하고 연관어 간의 상관관계를 네트워크 형태로 제시한 결과를 나타낸다.

[그림 8] 네트워크 분석 사례



[정리하기]

1. 빅데이터 정책 현황

- 미국, 영국, 호주 등 주요 국가들은 대부분 공공데이터를 일반에 공개하여 활용도를 높이고 빅데이터 분석 역량을 키우기 위한 정책을 추진하고 있으며, 실제로 의료·안보·교통 등 다양한 분야에서 빅데이터 기술을 적용하고 있다.
- 전 세계적인 추세에 맞추어 우리나라 역시 “빅데이터 산업 발전전략”을 수립하여 빅데이터 활용을 촉진하고 관련 산업을 육성하기 위한 정책을 추진하고 있다.

2. 빅데이터 활용 사례

- 빅데이터의 활용분야는 민간부문과 공공부문으로 구분되며, 현재까지의 빅데이터 분석은 주로 민간부문에서 기업의 상품 판매 및 마케팅 전략 수립을 위해 활용되는 사례가 대부분을 차지한다.
- 한편 공공부문에서는 정부 정책을 수립하거나 실효성을 제고하기 위한 목적으로 빅데이터가 활용될 수 있으며, 이미 많은 국가가 다양한 공공서비스 부문에 빅데이터 기술을 적용하고 있다.
- 현재 환경분야에서 소셜 빅데이터를 활용한 사례는 부족하지만, 산림보호나 기후변화 등의 환경문제에 빅데이터 분석을 유용하게 활용할 수 있는 잠재력이 있다.

3. 빅데이터 분석 방법

- 소셜 빅데이터 분석을 수행하기 위해서는 전문적인 기술을 필요로 하며, 특히 텍스트 데이터를 처리하고 정보를 추출하는 텍스트 마이닝 기술은 빅데이터 분석의 핵심이라 할 수 있다.
- 소셜 빅데이터 분석 목적과 대상에 따라 빈도 분석, 감성 분석, 연관어 분석,

네트워크 분석 등 다양한 분석 기법을 활용할 수 있다.

[참고문헌]

- 관계부처 합동(2013), 「창조경제 및 정부 3.0 지원을 위한 빅데이터 산업 발전전략」. 경제장관회의.
- 데이터 솔루션(2014), 「데이터사이언티스트가 말하는 빅데이터 분석 사례 세미나」. 세미나 발표 자료(2014년 3월 25일).
- 박대현, 송동현(2014), 비정형 데이터 활성화의 정치, 경제, 문화적 함의, 「Internet & Security Focus」, 한국인터넷진흥원.
- 사이람(2014), “소셜 미디어 분석 사례: 공공/연구기관에서의 활용을 중심으로”. KEI 전문가 세미나 발표 자료(2014년 2월 24일).
- 삼성경제연구소(2012), “빅데이터: 산업 지각 변동의 진원”. 「CEO Information」 제851호.
- 이미숙 외(2014), 빅데이터를 활용한 환경분야 정책수요 분석, 한국환경정책평가연구원.
- 한국정보화진흥원(2012), “소셜 분석으로 살펴본 청소년 자살예방정책의 시사점”. 「IT & Future Strategy」 제17호(2012년 12월 30일).
- 한국정보화진흥원(2012), “신가치창출을 위한 주요국의 빅데이터 추진전략 분석”. 「IT & Future Strategy」 제11호(2012년 11월 21일).
- 한국정보화진흥원(2012), 「영유아 보육정책에 대한 국민정서분석: 빅데이터 분석 따라하기-사례①」.
- 한국정보화진흥원(2012), 「Big Data 글로벌 10대 선진사례」.
- Felsche(2013), Future of Business Intelligence & Analytics in Financial Services.
- Russell et al.(2011), “Semantic Analysis of Energy-Related Conversations I Social Media: A Twitter Case Study”. International Conference on Persuasive Technology.
- TechAmerica Foundation(2012), Demystifying Big Data: A Practical Guide to Transforming the Business of Government.