

빅데이터의 국내·외 활용 고찰 및 시사점

이성훈*, 이동우**

Current Status of Big Data Utilization

Seong-Hoon Lee*, Dong-Woo Lee**

요약 정보기술 및 통신과 관련된 기술들을 융합하고자 하는 노력들이 지속적으로 이루어지면서 우리주변에는 다양하면서도 수많은 데이터들이 만들어지고 있다. 스마트폰이 일반화 되고 있으며, 태블릿PC와 카메라, 게임기등을 통하여 다양한 비 정형 데이터들이 생성되고 있으며 이러한 데이터들로 인한 데이터 트래픽 또한 급증하는 추세이다. 또한 데이터의 크기와 형태가 다양하고 데이터의 증가 속도가 가파른 이른바 ‘빅데이터 시대’가 도래하고 있는 것이다. 현재 다양한 분야에서는 이러한 빅데이터를 활용하여 새로운 가치 창출을 이루고자 하고 있다. 본 연구에서는 이러한 빅데이터의 국내·외 활용에 대한 고찰 및 시사점등을 기술하였다.

주제어 : 빅데이터, 융합화, 스마트폰, 활용, 정형, 비정형.

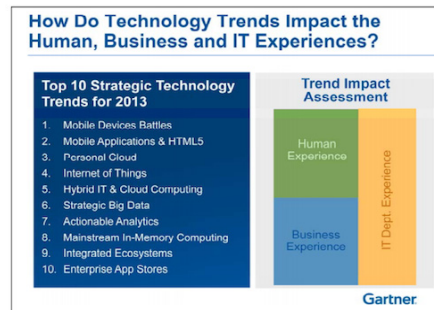
Abstract The technologies related with information communication regions are progressing continuously. These technologies in today are converged with different industries in rapidly. Because of these properties, A number of data are made in our life. Through many devices such as smart phone, camera, game machine, tablet pc, various data types are produced and the traffic is increased. We called it Big Data. There are many efforts to create new worth creation through Big Data utilization. Therefore, we described current trends and future of Big Data in this paper.

Key Words : Big Data, Convergence, Smart phone. Utilization, Structured, Unstructured.

1. 서론

오늘날 정보통신 분야에서의 화두는 스마트, 융복합과 더불어 빅데이터를 이야기할 수 있을 것이다. 리서치 자문기업인 가트너는 최근에 모바일 기기 전쟁, 전략적 빅 데이터 등 2013년 기업들이 전략적으로 대응해야 하는 10대 기술 및 트렌드를 발표하였다. 발표된 10대 기술 및 트렌드는 그림 1과 같다[1].

·모바일 기기 전쟁 ·모바일 애플리케이션과 HTML5
·퍼스널 클라우드 ·기업용 앱 스토어 ·IoT(Internet of Things) ·하이브리드 IT 및 클라우드 컴퓨팅 ·전략적 빅 데이터 ·실행 가능한 분석 ·인메모리 컴퓨팅 ·통합 예코 시스템.



[그림 1] 2013년 10대 기술 트렌드

가트너는 2013년에는 모바일폰이 전세계에서 가장 널리 사용되는 웹 액세스 기기로서 PC를 추월하게 될 것이며, 2015년에 이르면 선진국 시장에서 판매된 휴대폰의

*백석대학교 정보통신학부 교수.

**우송대학교 컴퓨터정보학과 교수, 교신저자.

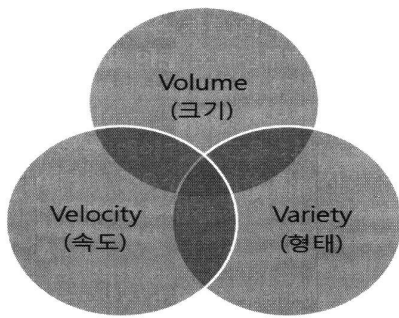
논문접수: 2013년 1월 8일, 1차 수정을 거쳐, 심사완료: 2013년 1월 25일, 확정일: 2013년 2월 20일

80% 이상을 스마트폰이 차지하게 될 것으로 예측하고 있다[3].

또 개인이 자신의 개인적인 콘텐츠를 보관하고, 자신의 서비스와 선호하는 대상에 접근하며 자신의 디지털 생활을 집중시키는 장소는 PC에서 퍼스널 클라우드로 점차 대체될 것으로 예측했다.

또 2014년에 많은 조직이 전용 앱 스토어를 통해 직원들에게 모바일 앱을 제공할 것으로 예상했으며, 빅데이터는 개별 프로젝트에서 기업의 전략적 정보 아키텍처로 초점이 옮겨가고 있다고 분석했다.

이러한 빅데이터는 기존 데이터에 비해 너무 방대하여 기존의 방법이나 도구로 수집, 저장, 분석등이 어려운 정형 및 비정형 데이터들을 의미한다. 세계적인 컨설팅 기관인 Mckinsey지는 2011년 한 보고서에서[5] 빅데이터의 정의는 “기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석하는 역량을 넘어서는 데이터 셋 규모로서 그 정의는 주관적이며 앞으로도 계속 변화될 것이다”라고 언급하고 있다. 빅데이터가 갖추어야 하는 요소기술로서 미디어관련 데이터 크기(Volume), 데이터 입/출력 속도(Velocity), 데이터 형태(Variety)가 있으며 그림 2와 같다.



[그림 2] 빅데이터 요소

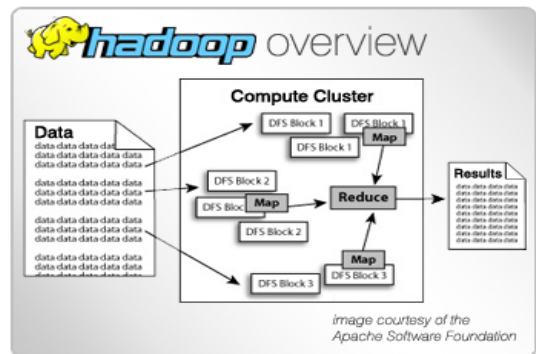
크기는 일반적으로 수십 테라 혹은 수십 페타 바이트 이상 규모의 데이터 속성을 의미한다. 속도는 대용량의 데이터를 빠르게 처리하고 분석할 수 있는 속성이다. 융복합 환경에서 디지털 데이터는 매우 빠른 속도로 생산되므로 이를 실시간으로 저장, 유통, 수집, 분석처리가 가능한 성능을 의미한다. 형태는 다양한 종류의 데이터를 의미하며, 정형화의 종류에 따라 정형(Structured), 반정형(Semi-Structured), 비정형(Unstructured)으로 표 1과 같이 분류될 수 있다[2].

<표 1> 빅데이터 유형

정형	· 고정된 필드에 저장된 데이터 · 관계형 데이터베이스, 스프레드시트 등
반정형	· 고정된 필드에 저장되어 있지 않지만, 메타데이터나 스키마 등을 포함하는 데이터로 XML이나 HTML 텍스트 등
비정형	· 고정된 필드에 저장되어 있지 않는 데이터로 텍스트 분석이 가능한 텍스트 문서 및 이미지/동영상/음성 데이터 등

오늘날 전 세계에서 다루어지는 디지털 정보량은 2년에 2배씩 증가하고 있다고 한다[4][6][7]. 정보통신 기술이 다른 산업들과 융복합되면서 방대한 량의 데이터들이 생성되고 있는 가운데 사회변화에 따른 삶의 질에 대한 욕구 및 현안 해결에 빅데이터들의 활용이 매우 중요한 과제로 떠오르고 있는 것이다.

빅 데이터 처리에서 가장 중요한 것은 위에서 기술된 다양한 유형의 거대한 데이터를 수집할 수 있는 저장기술과 저장된 기술을 이용하여 이를 의미 있는 데이터로 분석해낼 수 있는 데이터 분석기술이라고 할 수 있다. 빅 데이터 시대의 도래로 인하여 하둡과 같은 새로운 기술이 대두되었으며 기존기술로는 처리/분석할 수 없었던 데이터들을 처리/분석할 수 있는 기능을 제공한다. 다음 그림 3은 하둡의 개요를 보여주고 있다.



[그림 3] 하둡 구조

빅데이터의 활용이 기업을 중심으로 한 '빅데이터 경영'을 넘어 대국민 공공 서비스의 영역으로까지 확대되고 있다[3]. 빅데이터가 단순히 기업 경쟁력 향상의 수단에 머무르지 않고 국가 경쟁력 제고에도 이용되고 있는 것이다.

그동안 빅데이터는 경영 혁신의 카테고리에서 주로

논의돼 왔었다. 다양한 모바일 기기와 소셜미디어로 부터 생성되는 방대한 데이터를 수집하고 분석해 고객의 요구를 파악하고 기업 경영에 시장의 요구를 도입하려는 움직임이었다. 또한 생산라인에서 발생하는 수많은 센서 데이터를 활용해 제품의 불량률을 줄이고, 체계적인 물류 작업을 진행할 수 있도록 하는 데에도 빅데이터가 고려 대상이 되었다. 대규모 글로벌 IT 기업들은 이에 따라 한국 시장에 빅데이터 솔루션을 출시하면서 기업 경영에 초점을 맞춰 제품 홍보와 교육을 진행해 왔다.

하지만 최근에는 빅데이터가 공공 서비스 부문에 도입되면서 기업 경영 뿐 아니라 전 사회적으로 활용될 수 있는 광의의 개념으로 인식되는 분위기다.

해외에서는 이미 빅데이터와 시스템통합(SI)이 결합된 새로운 공공 서비스 모델이 제시되며 시민들에게 양질의 서비스를 제공하고 있다. 빅데이터 솔루션이 공공기관의 정보화 시스템의 두뇌역할을 하며 행정서비스의 질을 개선하는데 중요한 역할을 하고 있는 것이다.

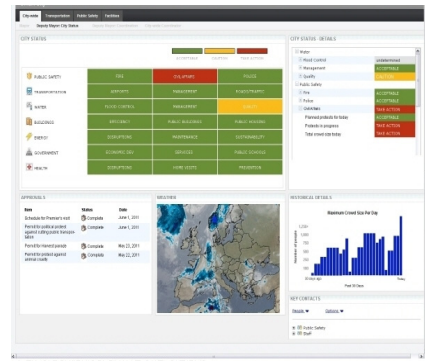
2. 빅데이터 활용

빅 데이터는 우리가 살아가는 모든 영역에서 발생하는 데이터들을 기반으로 한다. 따라서 이러한 데이터는 공격적인 서비스의 영역에 속할 수도 있으며 또한 기업 및 일상생활에 필요한 데이터일 수 있다. 따라서 본장에서는 빅데이터의 활용을 공공서비스 측면과 일반적인 측면으로 나누어 기술하고자 한다.

2.1 공공서비스부문 활용

브라질 리오데자네이로시가 빅데이터로 각종 도시재해에 대비하는 것은 기존의 빅데이터가 기업 경영의 혁신 수단으로 이용되던 수준을 넘어 국가경영 활동에 적용된 대표적인 사례라 할 수 있다.

2014년 월드컵과 2016년 올림픽을 준비하는 리오데자네이로는 지능형운영센터(IOC)를 통해 도시 관리와 긴급 대응 시스템을 갖췄다. 30여개에 이르는 여러 기관의 데이터와 프로세스를 지능형운영센터에 통합해 도시의 총체적인 움직임을 24시간 365일 모니터링할 수 있다.



[그림 4] 지능형 운영센터

그림 4와 같은 지능형운영센터에는 교통, 전력, 홍수, 산사태 등의 자연재해와 수자원 등을 통합 관리할 수 있는 체계가 갖춰져 있는데, 여기에는 IBM의 분석 솔루션이 적용돼 비상사태를 효과적으로 예측하고 대응할 수 있도록 한다. IBM이 제공한 고해상도 날씨 예측 시스템과 수문학적 모델링 시스템은 날씨 및 수문 관련 방대한 데이터를 분석해 폭우를 48시간 이전에 예측한다.

강 유역의 지형측량 자료와 강수량 통계, 레이더 사진 등의 데이터에서 추출한 통합 수학적 모델에 기초해 강수량과 갑작스런 홍수를 예측한다. 뿐만 아니라 강수량과 교통체증, 정전 사태 등 도시에 영향을 미치는 상황들도 평가한다.

또한 새로운 자동경보 시스템은 리오데자네이로의 홍수와 산사태 예측에 변동이 있을 시에 시청 공무원과 긴급 대응팀에 통보된다. 이같은 경보시스템은 통보를 수작업으로 전달하던 과거 시스템과 달리 비상사태 발생 시 자동 이메일 통보나 문자 메시지 등의 즉각적인 커뮤니케이션 툴을 이용해 긴급 대응팀과 시민에게 알려 대응시간을 크게 줄일 수 있도록 한다.

싱가폴에서는 차량의 기하급수적인 증가로 많은 교통체증을 겪고 있으며, 빅데이터는 이러한 싱가포르 행정기관에 새로운 해법을 제시하였다.

싱가포르의 빅데이터 분석을 통해 실시간 교통정보에서 한 단계 더 나아간 '교통량 예측 시스템(TPT)'을 운영하고 있다. 싱가포르 교통국(LTA)은 아이 트랜스포트 시스템(i-Transport system)을 활용한 도시 교통 정보와 교통 예측 툴로 교통의 흐름을 예측하고 있는 것이다. 이같은 예측 시스템은 교통흐름 분석과 예측 시스템으로 구현되는데, LTA의 교통 관제관은 교통 통제관이 센서를 통해 실시간 교통 데이터를 보내주면 이를 교통 시나

리오로 모델링 해 1시간 뒤의 통행량을 예측할 수 있도록 하였다.

솔루션을 제공한 IBM에 따르면 전체적인 예측 결과는 85% 이상의 정확성을 보이고 있고 특히 교통량이 가장 많은 비즈니스 중심가에서는 정확도가 85% 이상으로 측정됐다.

미국 워싱턴DC의 상하수도를 관리하고 있는 디씨 워터(DC Water)는 배수 및 수집 시스템의 효율적인 관리를 위해 빅데이터 시스템을 도입했다. 예측 분석 시스템을 통해 배수관과 밸브, 공공 수도전, 수집관, 맨홀 및 계량기 등의 자산을 효율적으로 관리할 수 있게 된 것이다. 이를 통해 직원들은 상세한 지도에서 회사 자산의 위치와 상태를 확인하고, 자산 내역과 총 자산 비용, 각 지역의 문제점, 문제 유형 및 지역별 수질 문제에 신속하게 접근하고 있다.

특히 디씨 워터는 예측 분석을 통해 서비스 중단을 예방하고 서비스 수요를 기반으로 새로운 비용 모델을 구축할 수 있었다. 또한 향상된 예방 관리와 자동 점검으로 고객 전화가 36%나 줄었고, 프로세스를 간소화시켜 10분 이내에 신속하게 처리해야 하는 업무를 기존 49%에서 93%까지 늘렸다.

2.2 일반생활부문 활용

소비자가 전기 요금 '폭탄'을 맞았다면? '전기를 아껴 써야겠다'고 마음을 먹는 것 외에는 별다른 수가 없을 지 모른다. 특히 전기나 수도요금은 카드요금처럼 사용내역이 세부적으로 나오지 않아 따져 보지도 않고 매달 청구되는대로 지불할 수밖에 없다. 하지만 앞으로는 빅데이터가 이러한 문제들을 해결하는 역할을 할 수도 있을 것이다.

'맥아더 펠로우(MacArthur Fellow) 상을 받은 워싱턴 대학 컴퓨터 공학과 조교수인 슈택 페이텔은 이러한 공과금 산출에 가장 합리적인 방법을 발견했다. 외부에서 집으로 유입되는 모든 전기와 수도, 가스 관련 기기마다 특이한 디지털 신호를 갖고 있다는 사실을 기반으로 하는 것이다.

이에 따라 간단한 알고리즘을 통해 신호를 감지하는 센서를 제작했으며 이 센서는 가스 및 전기 배선, 배수관, 환기구 등에 설치돼 디지털 신호로 만들고 무선으로 태블릿PC에 전송해 실시간으로 수치를 확인하게끔 한다.



[그림 5] 전기요금에의 활용

예를 들어 그림 5와 같이, 어떤 전자기기가 많은 양의 전기를 사용했는지, 얼마만큼의 수도나 가스가 사용되고 있는지 확인할 수 있게 되는 것이다. 페이텔은 사촌이 사는 집의 수도와 전기 사용량 측정 센서를 아이패드에 설치해 사용량을 확인한 결과, 총 사용하는 전력의 11%가 수영장 전기 펌프에서 사용된다는 것도 파악했다.

슈택 페이텔이 발명한 이러한 획기적인 에너지 감지 센서는 머지 않아 우리 일상에서 접할 수도 있다. 지난 2010년 벨킨이 슈택 페이텔의 에너지 감지 센서에 대한 특허권을 사들이고 상용화 준비에 돌입했기 때문이다. 이 센서를 통해 어떤 기기가 전력을 많이 사용하는지 파악할 수 있어 예방적인 조치가 가능할 것이다. 또한 앞으로 스마트시티내 수십만 가정이나 사무실에 전력을 공급하는데 있어서도 많은 정보를 얻을 수 있을 것이다.

싱가포르에 사는 사람이라면 누구나 비가 오는 날에 택시를 잡기 힘들다는 사실을 알고 있다. 싱가포르-MIT 기술 연구소에서는 두 달여 간의 날씨 위성 데이터와 택시 운행 기록에서 제공된 8억3천만 개의 GPS 기록을 비교하는 과제를 진행하였다. 지난 2011년 이 프로젝트를 진행하던 중에 특이한 패턴을 발견했다. 1만6천대의 택시에서 나온 데이터를 분석한 결과, 많은 택시들이 폭풍우 속에서 움직이지 않는다는 점이였다. 실제 GPS 기록 결과에서도 비가 내릴 때 수많은 기사들이 차를 멈추고 손님을 더 이상 태우지 않았다.

이같은 분석결과를 토대로 택시기사의 하루 일과를 조사한 결과 원인을 알 수 있었다. 싱가포르 택시회사의 경우, 택시 사고 발생시 원인규명이 끝나기 전까지는 무조건 택시 기사의 월급에서 1천 달러를 강제로 보관하게끔 하는 게 문제였다. 이 때문에 택시 기사들은 비가 내

리먼 팬한 위험을 무릅쓰고 운전을 하는 것보다 차라리 차를 세우고 날씨가 좋아지길 기다리는 것이 낫다고 판단했던 것이다. 결국 이같은 원인 파악 후 택시 회사와 택시 기사, 고객들에게 더 좋은 방향으로 회사의 규정은 수정됐다. 빅데이터가 어떤 식으로 시민들의 삶의 질을 향상 시킬 수 있었는지를 보여주는 사례라 할 수 있다.

이처럼 데이터를 분석하는 일은 단지 데이터를 읽는 것으로 그치는 것이 아니라 데이터를 분석하는 통찰력을 통해 전혀 예상치 못한 원인과 결과를 찾아내는 것이 중요하다.

빅데이터 분석을 통하여 새로운 유통 트렌드를 읽는 중요한 도구가 되고 있는 또 다른 사례를 들어보자. 미국에서는 11월 넷째 주 목요일인 추수감사절 바로 다음 날인 '블랙 프라이데이(Black Friday)'에 연중 최대 규모의 세일과 쇼핑이 이뤄진다. 하지만 빅데이터 분석 결과 최근에는 블랙 프라이데이를 능가하는 새로운 쇼핑 날짜로 '사이버 먼데이(Cyber Monday)'가 주목받고 있음을 밝혀냈다.

사이버 먼데이는 추수감사절과 주말을 지낸 첫 월요일로, 이날 직장에 복귀한 사람들이 온라인 쇼핑에 몰두하는 트렌드가 형성되면서 최근 유통업체가 새롭게 주목하고 있다는 것이다. 보고서에 따르면 지난 해의 경우 사이버 먼데이 당일 온라인 매출은 블랙 프라이데이보다 29.3%나 높았다. 모바일 트래픽과 모바일 매출 비중도 각각 10.8%, 6.6%로 블랙 프라이데이의 14.3%, 9.8%에 근접했다.

이같은 결과는 IBM이 미국 전역의 500개 주요 유통기업에서 발생하는 하루 백만건 이상의 거래량과 테라바이트급(TB)의 빅데이터를 분석해 발견해 낸 수치다.

3. 결론

현재 우리 사회는 스마트폰이 대중화되고, 태블릿PC와 카메라, 게임기 등 데이터를 생산하는 다양한 기기들이 출현하면서 트래픽도 급증하고 있다. 또한 데이터의 크기와 형태가 다양하고 데이터의 증가 속도가 가파른 이른바 '빅데이터 시대'에 놓여 있는 것이다. 본 연구에서는 최근 화두가 되고 있는 빅 데이터와 관련된 내용을 활용성 측면에서 기술하였다. 공공서비스 영역에서의 활용 및 일상생활에서 접할 수 있는 가능성등에 대하여 다

루었다.

최근 기업들이 빅데이터 분석을 통해 기업경영 개선과 마케팅 효율화를 도모하고 있다. 반면에 공공분야인 교통시스템, 수자원시스템, 방범시스템, 탈세방지시스템, 의료시스템 등 다양한 분야에서 빅데이터를 통한 시스템 효율화가 전개될 것으로 보인다.

이러한 빅데이터에 대한 관심 및 활용성은 지속적으로 확대될 것이며 영역 또한 순수한 정보기술 영역을 넘어 모든 영역으로 광범위하게 적용될 것이다. 이 같은 노력이 지속적으로 확대, 발전된다면 향후 다가오는 미래사회는 우리에게 무한한 가능성을 제시할 수 있을 것이다.

참 고 문 헌

- [1] 각 언론사 뉴스 자료 (2012).
- [2] 이성훈 (2012). 빅 데이터 활용과 통신산업에 대한 시사점.
- [3] Gartner. (2011). Big Data Analytics. Gartner Group.
- [4] IDG Korea. (2012). 빅데이터를 클라우드에서.
- [5] Mckinsey Global Institute. (2011). Big Data: The next frontier for innovation, competition, and productivity.
- [6] Warden, P. (2011). Big Data Glossary.
- [7] www.itkorea.co.kr/ news.

이 성 훈



- 1998년 2월 : 고려대학교 컴퓨터학과(이학박사)
- 1998년 3월~현재 : 백석대학교 정보통신학부 교수.
- 관심분야 : 분산 시스템, 무선 통신, 유전 정보, 웹서비스
- E-Mail : shlee@bu.ac.kr

이 동 우



- 2005년 2월 : 고려대학교 전산학과(이학박사)
- 1995년 3월~현재 : 우송대학교 컴퓨터정보학과 교수
- 관심분야 : 웹기반 분산시스템, 능동시스템, 데이터베이스.
- E-Mail : dwlee@wsu.ac.kr