



글 · 박원준 전임연구원
정책연구본부 융합정책연구부
(02)2142-2117,
sumboy2@kca.kr

주제어: 빅데이터, 데이터 패러다임 전환, 개인정보, 저작권

전 세계적으로 빅데이터가 가장 뜨거운 이슈가 되고 있다. 국내 역시 빅데이터에 대한 활용을 놓고 정부와 기업이 기술개발과 인력양성에 투자하고 있다. 빅데이터는 분명 새로운 비즈니스 전략을 수립하는 중요한 정보를 제공할 것이다. 하지만 개인정보, 사생활침해, 저작권 등과 관련된 문제 역시 빅데이터 시대 중요한 화두가 될 것이다.

‘빅데이터(Big Data)’ 활용에 대한 기대와 우려



I. 서론

최근 빅데이터(big data)가 우리사회의 핵심 키워드로 등장하고 있다. 빅데이터는 새로운 개념이 아니라 1990년 이후 인터넷이 확산되면서 정형화된 정보와 비정형 형태의 정보가 무수히 발생하게 되면서 정보홍수(Information overload)나 정보폭발(information explosion)이라는 개념으로 논의되었고, 오늘날 ‘빅데이터’라는 개념으로 이어지게 된다. 그 동안 인터넷에서 발생한 수많은 정보는 인터넷 서비스 기업이 보관하거나 일부 상업적으로 이용되기도 하였다. 더불어 모바일 스마트 기기의 확산으로 개인과 관련된 비정형 데이터가 축적되면서 데이터는 더욱 증가하게 된다. 특히 소셜 미디어의 증가는 공적인 정보뿐만 아니라 사적인 정보까지 교류함으로써 빅데이터의 서막을 알리는 계기가 되었다.

빅데이터에 대한 분석은 2008년 미국의 인디애나 주립대학의 조한 볼렌(Johan Bollen)교수에 의해서 시작되었다고 볼 수 있다. 그는 트위터에 올라온 글에서 개인과 관련된 정보가 포함되어 있음을 인지하고, 이 모든 데이터를 종합하여 분석하면 어떠한 거대한 흐름을 발견할 수 있을 것으로 기대하였다. 그는 2008년 상반기 트위터에 올라온 모든 데이터를 분석해 이용자들의 집단적 기분 변화가 전국적인 행사와 일치한다는 사실을 발견하게 된다. 또한 추수감사절이 다가올수록 트위터에 올라온 기분변화는 행복지수를 나타내고, 반대로 사람들이 불안감을 느끼면 며칠 뒤 주가지수가 하락한다는 사실도 발견하였다(유지연, 2012). 이러한 분석은 과학적 데이터 즉, 통계나 실험을 통해

서 얻은 것이 아니라 인간의 정서나 심리 정보에 해당하는 기분이나 감정을 분석한 것이다. 이와 같은 비정형화된 데이터는 인터넷을 비롯해 스마트 기기 이용을 통해서 생산되는데, 전자상거래 서비스에서 발생하는 거래내역이나 소비자 의견은 물론 게시판, 소셜 네트워크 서비스 이용자들의 모든 메시지가 여기에 해당된다.

최근 비정형화된 데이터는 숫자에 해당하는 기존 데이터의 양적인 측면과 질적인 측면을 넘어 비교할 수 없을 정도의 많은 데이터가 생산되고 있다. 2012년 7월 통계법 제18조(또는 제20조)에 의거 승인받아 통계청이 매년 발표하는 공식적인 통계 데이터는 총 860종으로서 지정통계 93종, 일반통계 767종이다. 작성형태별로는 조사통계는 353종, 보고통계는 444종, 가공통계는 63종에 달한다. 작성기관을 정부기관과 지정기관으로 구분하여 볼 때 정부기관에 의하여 작성되고 있는 통계는 705종(통계청은 56종)이며, 지정기관에서 작성하고 있는 통계는 155(157)종에 달한다. 이와 같은 데이터는 정형화된 것으로 바로 활용이 가능하다.

반면 비정형화된 데이터는 우리 사회 곳곳에서 만들어지고 있는데, 특히 페이스북, 트위터, 카카오톡과 같은 소셜 네트워크 서비스 이용자들이 많은 데이터를 생산하고 있다. 카카오톡의 가입자는 4,200만 명으로 하루에 작성되는 총 메시지의 작성건수도 지난해 비해 1억 7,000만 건 이었지만 올해는 작년에 비해 8배나 증가해 13억 건으로 나타났다. 가입자들이 하루 수신하는 메시지는 총 30억 건에 달한다(윤태구, 2012).

트위터는 전세계 1억 명의 이용자들이 하루 평균 2억 개의 트윗이 발생한다. 오늘날 11억 인구가

소셜 네트워크를 이용하고 있는데 2억 5000만 명이 매일 페이스북에 사진을 업로드하고 있다. 이와 같이 모바일 스마트 기기의 보편화와 무선 인터넷의 안정적인 사용이 빅데이터의 확산을 가속화시키고 있다(박현욱, 2012).

향후 모바일 기기를 통해서 소통이 더욱 활발해지고, 스마트 기기 이용자의 증가로 인해 생산되는 데이터의 용량은 더욱 커질 것이다. 따라서 빅데이터는 대용량의 데이터를 저장하는 기업이 저장된 데이터를 얼마나 효율성 있게 분석하는가에 따라서 기업 운영전략에 새로운 기회를 만들어갈 수 있을 것이다. 최근 생산되는 데이터의 크기는 테라바이트(Terabyte TB)에서 제타바이트(Zettabyte ZT)에 이르고 있는데, 데이터의 양이나 크기는 빅데이터의 특징 중 하나이다. 수십 년간 데이터를 축적해 온 Data warehouse와 Data mart는 그동안 수집한 데이터를 설계 및 분석함으로써 새로운 가치를 창출하고 있다.

하지만 2010년 10월 TechTarget이 발표한 ‘플랫폼 분석: 전통적 개념의 데이터 웨어하우스, BI 리서치 및 IT 시장 전략(Analytic Platforms: Beyond the Traditional Data Warehouse, BI Research and IT Market Strategy)’에 따르면, 조사기업 중 53%가 관련된 데이터베이스 관리 시스템(RDBMS)에 저장하지 않는 데이터를 일상적인 오프라인 방법으로 분석하고 있다고 응답하였다. 즉, 웹 활동, 고객 우대 프로그램, 원격 센서 기술, 콜 센터 레코드 및 소셜 미디어에서 쏟아져 나오는 데이터 홍수로 인해 기업은 데이터의 비즈니스 가치를 어떻게 극대화해야 할지 혼란을 겪고 있음을 보여준다.

현재 빅데이터는 산업혁명과 IT 혁명에 버금가

는 혁신이라 불리며, 기업의 경쟁력 강화, 생산성 향상을 위한 중요한 원천으로 떠오르고 있다(McKinsey, 2011). 이에 본 글은 빅데이터의 의미를 기반으로 해외에서 빅데이터가 어떻게 활용되고 있는지 살펴보고, 이를 바탕으로 빅데이터의 국내 활용가능성을 살펴보았다. 아울러 빅데이터 시대에 나타날 수 있는 문제점에 대해서 논의하였다.

II. 빅데이터의 의미와 활용현황

1. 빅데이터의 의미

전통적 개념의 데이터는 Google이나 Apple 등과 같은 대기업이나 NASA의 연구과학 프로젝트에서 분석하는 대용량의 정형화된 데이터를 일컫는다. 하지만 빅데이터는 전통적인 데이터의 개념으로 설명할 수 없다. 아울러 빅데이터는 상대적인 개념으로 받아들이고 있는데, 이유는 기업 규모에 따라 데이터 저장과 분석 용량이 다르기 때문이다. 최근 논의되는 빅데이터는 ‘대용량의 데이터 외에 앱, SNS 등에서 생산되는 데이터’를 포함하는 개념이다(채승병, 2011). 이러한 광범위한 데이터는 기존 민간 기업이나 정부의 데이터 분석 범위를 넘어선 것이다. 따라서 빅데이터를 ‘기존의 시스템, 서비스, 기업 등에서 주어진 비용이나 시간 내에 처리 가능한 데이터 범위를 넘어서는 데이터’(김형준, 2012)라고 보는 시각이 대부분이다. McKinsey(2012)는 빅데이터를 ‘전통적인 데이터베이스 S/W를 통해 저장, 관리, 분석할 수 있는 규모를 초과하는 데이터’라고 정의하였다. 또한 정용찬(2012)은 ‘데이터의 양, 생성주기, 형식(수치대

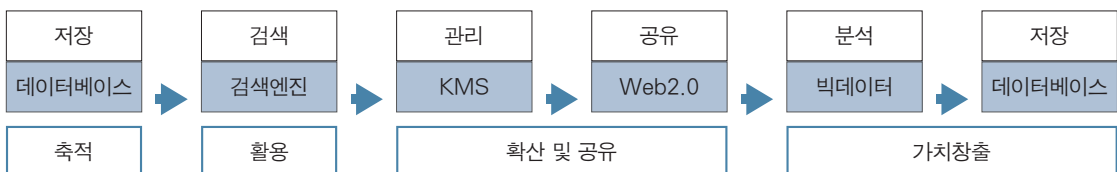
이터 뿐만 아니라 문자와 같은 비정형 데이터 포함) 등에서 과거 데이터에 비해 규모가 크고, 형태가 다양하여 기존의 방법으로는 수집, 저장, 검색, 분석이 어려운 방대한 크기의 데이터'라고 빅데이터를 정의하고 있다. 이러한 빅데이터의 정의는 기존 분석 시스템의 한계를 기반으로 빅데이터를 바라보고 있다. 그렇다보니 빅데이터의 정의에서도조차 빅데이터를 매우 모호하게 규정하고 있다.

이에 본 글에서는 빅데이터를 '숫자, 그림, 문자, 언어 등을 포함하는 다양한 형태의 대용량 데이터를 S/W를 통해 분석하여 새로운 가치해석이 가능한 데이터'라고 정의하였다. 이러한 정의에 의해서 유무선 네트워크 사용에 따라 생산 또는 저장되는 정형화된 데이터와 SNS, 블로그, 홈페이지, 전자상거래 등에 의해서 생산되고 있는 비정형화된 데이터를 포함한다. 따라서 빅데이터는 정형화(Structured)된 데이터, 반정형화(Semi Structured) 데이터, 비정형(Unstructured)데이터로 구분할 수 있다. 정형화된 데이터는 일정한 규칙을 갖고 체계적으로 정리된 데이터를 의미한다. 예를 들어 매년 통계청에서 발표하는 통계자료, 방송통신 실태조사, 각종 과학적 데이터 등이 이에 해당한다. 정형화된 데이터는 그 자체로 의미 해석이 가능하며, 바로 활용할 수 있는 정보를 내포하고 있다. 반정

형화된 데이터는 한글이나 MS Word 등으로 작성된 데이터를 의미한다. 대표적인 예가 인쇄매체의 텍스트라 할 수 있다. 반정형화된 데이터는 표나 그림이 될 수도 있지만 일반적으로 문자로 서술된 정보를 담지하고 있다. 비정형화된 데이터는 스마트 기기 등을 통해서 형성되는 데이터로 페이스북, 트위터, 카카오톡 등으로 상호 교류되는 정보가 이에 해당한다. 비정형화된 데이터는 개인, 집단, 사회, 국가 등과 관련된 주제를 스마트 미디어 이용자들이 상호 의견을 교류함으로써 생산되는 정보들이다. 특히 오늘날 빅데이터는 비정형화된 데이터에 관심을 두고 있다.

이러한 빅데이터는 해당 데이터를 분석하고 처리함으로써 기존의 데이터에서 볼 수 없었던 새로운 의미를 산출할 수 있다. 즉, 수많은 데이터를 분석하여 사용자에게 유용한 정보를 제공할 수 있어야 빅데이터는 효용성을 갖는다. 따라서 빅데이터에서 중요한 것은 형식적인 데이터 소스 내에서 외부로 새로운 가치를 창출할 수 있느냐 하는 것이다. 결국 새로운 가치와 의미를 산출하기 위해서는 축적된 데이터를 갖고 무엇을 분석할 것인가에 대한 고민이 필요하다. 이러한 고민은 데이터 마이닝과 연결되는데 빅데이터에서 데이터 마이닝은 텍스트 마이닝(Text Mining)¹⁾과 웹 마이닝(Web

● 그림 1 데이터의 분석과정

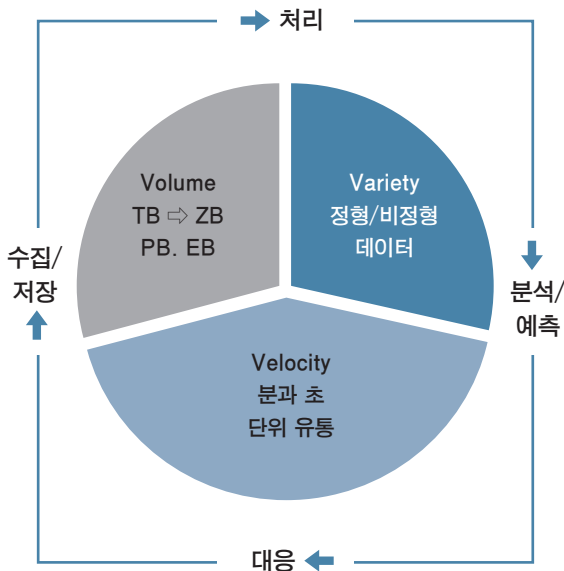


1) 텍스트 마이닝(Text Mining)은 자연어로 구성된 비정형 텍스트 데이터에서 의미있는 패턴이나 관계를 추출하여 가치 있는 정보는 찾아내는 마이닝기법. 이는 사람들이 말하는 언어를 이해할 수 있는 자연언어처리(Natural Language Processing)기술에 기반.

Mining)²⁾ 그리고 소셜 마이닝(Social Mining)³⁾을 통해서 현실 마이닝(Reality Mining)⁴⁾에 도달해야 한다. 빅데이터의 현실 마이닝은 우리가 영화에서나 볼 수 있었던 미래를 예측할 수 있는 데이터들이 산출되어 사후대책이 아니라 사전 방지 시스템을 작동시킨다.

이러한 효과를 거두기 위해서 빅데이터는 크기(Volume), 다양한 형태(Variety), 빠른 속도(Velocity) 등과 같은 조건을 갖추어야 한다. 빅데이터의 크기는 데이터의 엄청난 용량과 크기를 의미하는데, 새로운 가치를 창출할 수 있는 데이터의 양을 100terabyte 이상으로 보고 있다(McKinsey,

● 그림 2 빅데이터의 3가지 조건



※ 출처 : STRABASE (2012. 7. 11) 자료 수정

2011). 데이터의 다양한 형태는 앞서 언급한 바와 같이 정형, 반정형, 비정형 등 다양한 형태의 데이터를 의미한다. 이러한 형태의 데이터가 혼재되어 있는 복잡한 데이터를 어떻게 분류하여 분석할 것인가에 대한 데이터 분석 체계가 필요하다. 아울러 속도는 저장되어 있는 대용량의 데이터를 빠른 속도로 분석하고, 응용 가능한 정보로 전환하여 서비스 할 수 있어야 한다.

아울러 빅데이터를 활용하기 위해서는 기존의 데이터 분석방법이 아니라 새로운 기법의 분석방법을 도입해야 한다. 빅데이터 활용에 따른 분석방법의 변화는 활용의 변화와 분석방법의 변화로 구분할 수 있다. 빅데이터에 대한 활용의 변화는 앞서 언급한 바와 같이 비정형화된 데이터를 기반으로 실시간 분석이 가능해야 하며, 특정기관이 아니라 민간, 기업, 정부 등 다양한 주체들이 활용할 수 있도록 유도해야 한다. 분석의 변화는 데이터의 크기에 대한 변화를 기반으로 다중의 저장 공간과 분석틀로 데이터의 흐름을 분석할 수 있어야 한다. 최근 빅 테이블(Big Table), 카산드라(Cassandra), 데이터웨어하우스(Data ware house) 및 분석 어플라이언스(appliance), 분산시스템(Decentralized system), 구글 파일시스템(Google file system), 하둡(Hadoop), H베이스(Hbase), 맵리듀스(MapReduce), 비관계형 데이터베이스(non-relational database) 등과 같이 빅데이터와 관련된 기술이 등장하고 있다.

2) 웹 마이닝(Web Mining)은 인터넷상에서 수집된 정보를 데이터 마이닝 방법으로 분석하는 기법.

3) 소셜 마이닝(Social Mining)은 소셜 미디어에 올라오는 글과 사용자를 분석해 소비자의 성향과 패턴 등을 분석함으로써 판매 및 홍보에 주로 이용하며, 여론변화나 사회적 흐름을 파악하기도 함.

4) 현실 마이닝(Reality Mining)은 사람들의 행동패턴을 예측하기 위해서 사회적 행동과 관련된 정보를 스마트 기기를 통해 얻어 분석하는 기법으로 현실생활에서 발생하는 정보를 기반으로 인간관계나 행동의 양태를 추론.

빅데이터의 분석은 일반 자연언어를 페타바이트(Peta Byte)에서 제타바이트(Zetta byte) 용량을 실시간으로 분석함으로써 새로운 흐름을 찾아내는 것이다. 분석결과를 갖고 국가기관, 민간 등의 분석기획 담당자들이 자유롭게 이용할 수 있어야 한다. 특히 빅데이터 분석을 통해서 나타난 복잡한 정보를 한눈에 볼 수 있도록 도표나 3D 형태의 정보의 시각화(Infographics)가 이루어져야 한다.

빅데이터가 이슈화되는 이유는 크게 3가지 요인으로 볼 수 있다. 첫째 요인은 스마트 폰을 비롯하여 모바일 스마트 기기 보급의 활성화이다. 모바일 스마트 기기에 탑재된 센서, 원격감지기술, 소프트웨어, 카메라, RFID 리더 등을 통해서 비정형

화된 데이터를 수집할 수 있게 된 것이 데이터 증가의 원인이라고 할 수 있다. 둘째, 클라우드 서비스이다. 클라우드 서비스를 통해서 개인과 조직의 데이터가 한곳으로 축적되고, 저장된 데이터를 분석하여 활용하고자 하는 요구가 증가하고 있다. 셋째, 소셜 미디어의 활용이 일상화되면서 정보유통 구조가 새롭게 재편되고 있다. 소셜 미디어의 특성상 쌍방향 커뮤니케이션이 활발하게 이루어지면서 상호작용 데이터의 증가를 가져왔다. 이러한 상호작용 데이터를 비즈니스 측면에 활용하려는 관심이 높아졌기 때문이다(McKinsey, 2011).

현재 빅데이터는 다양한 산업적 측면에서 활용되고 있으며, 그 가능성이 무한대로 확산될 전망이

● 표 1 빅데이터와 연결된 기술

용어	설명
Big Table	<ul style="list-style-type: none"> 구글 파일 시스템 상에 구축된 상용 분산 데이터베이스 시스템 H베이스에 영향을 미침
Cassandra	<ul style="list-style-type: none"> 분산 시스템에 방대한 분량의 데이터를 처리할 수 있도록 디자인된 오픈소스 데이터베이스 관리시스템 이 시스템은 페이스북에서 개발했으며, 지금은 아파치 소프트웨어 재단의 한 프로젝트로 관리
Data ware house & Appliance	<ul style="list-style-type: none"> 데이터웨어하우징을 위해 서버, 스토리지, 운영체제, 데이터베이스, BI, 데이터마이닝 등 기타 여러 가지 소프트웨어가 최적화되어 설치된 통합제품
Decentralized system	<ul style="list-style-type: none"> 동시에 일을 처리하기 위해 네트워크로 연결된 컴퓨터들의 집합 단일 또는 다수의 컴퓨터 리소스를 부분적으로 활용함으로써 시스템의 가격 대비 성능비, 안정성, 확장성을 향상시킬 수 있음
Google file system	<ul style="list-style-type: none"> 구글에서 개발한 분산파일 시스템으로 하둡(Hadoop)과 관련
Hadoop	<ul style="list-style-type: none"> 분산시스템 상에서 대용량 데이터 처리분석을 지원하는 오픈 소스의 소프트웨어 프레임워크 구글이 개발한 맵리듀스를 오픈소스로 구현한 결과물 아후에서 최초 개발되었으며, 지금은 아파치 소프트웨어 재단의 한 프로젝트로 관리되고 있음
Hbase	<ul style="list-style-type: none"> 구글의 빅테이블을 참고로 개발된 오픈소스 분산 비관계형 데이터베이스 파워셋에서 개발했으며 현재는 아파치소프트웨어 재단에서 하둡의 일련인 프로젝트로 관리되고 있음
MapReduce	<ul style="list-style-type: none"> 분산시스템 상에서 대용량 데이터 세트를 처리하기 위해서 구글이 제안한 소프트웨어 프레임워크. 하둡에도 구현 가능
non-relational database	<ul style="list-style-type: none"> 비관계형 데이터베이스는 데이터를 테이블에 저장하지 않는 데이터베이스이며 관계형 데이터베이스와는 대조적인 개념 이를 사용하면 스키마 없는 엔티티(NoSQL)를 관리할 수 있음.

※ 출처: 김성태 (2012). 新가치창출 엔진, 빅데이터의 새로운 가능성과 대응 전략, IT & Future Strategy, 한국정보화진흥원.

다. 특히 미국을 비롯한 일부 국가에서는 빅데이터 활용을 위해서 전담 부서를 개설하고, 의료, 교육, 환경 등에 적용하여 경제적 가치는 물론 위험예방 차원에서도 효과를 거두고 있다.

2. 빅데이터 활용현황

가. 해외 활용 현황

미국의 시장 조사기관인 Wikibon은 빅데이터와 관련된 컨설팅 및 IT 시스템 분야의 시장규모가 올해 50억 달러에서 2013년 102억 달러, 2017년에는 530억 달러 규모로 성장할 것이라고 예측하였다(ITWorld, 2012). 이미 Oracle, EMC, IBM, SAP, MS 등 IT 기업들이 국내 기업 및 학계 등과 제휴해 국내 빅데이터 시장에 진출하고 있다. 미국은 2012년 3월 빅데이터 기술개발과 인력양성을 위해서 2억 달러 이상을 투자한다는 ‘빅데이터 이니셔티브(Big Data Initiative)’를 발표하였다. 또한 Forbes는 2012년 7월 빅데이터에 대한 지도를 만들었다. Forbes의 빅데이터 지도에는 Oracle, Microsoft, IBM, Tableau, Amazon 등을 포함하고 있다.

Forbes가 제시한 지도에는 Log Data Apps, Vertical Apps, Business Intelligence, Analytics and Visualization, Data Providers, Analytics Infrastructure, Operational Infrastructure, Infrastructure As A Service, Structured Databases, Technologies 등으로 업종과 분야를 구분하고 있다. 기술을 기반으로 하여 데이터베이스, 서비스, 서버운영 및 설치, 분석에 관한 기반을 제공하는 사업자들을 분류하고, 데이터 기록과 앱 사용에 대한 데이터를 보관하거나 소프트웨어를 제공하는 업체 그리고 소프트웨어 비즈니스 업체, 데이터 제공자를 비롯하여 데이터를 시각적으로 분석해주는 사업자 등으로 포함되어 있다.

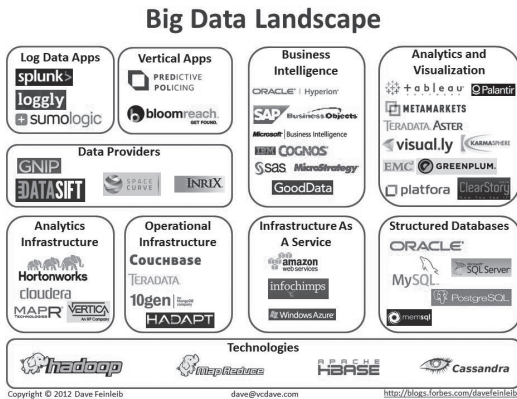
특히 미국의 경우 빅데이터 활성화를 위해서 의료 부분에 많은 투자를 하고 있다. 미국은 헬스케어 데이터를 활용할 수 있는 애플리케이션 개발 촉진을 위해 ‘Health 2.0 Boston Code-a-thon’를 개최하였다. 그 결과 수면무호흡증과 교통사고 관계에 주목한 ‘No Sleep Kills’이라는 주제의 앱 개발자들이 우승을 차지하였다. 우승자들은 Amazon

● 표 2 빅데이터 활용에 따른 분석방법의 변화

분류		현재	융합지식기반
활용변화	데이터개방	웹 기반 인터페이스	원본/분석/가시화 3계층
	이슈접근	후집계/원인파악: 사후대책	실시간 이슈탐지: 선 대응 기획
	활용형태	부처별 수직적 활용	범부처/민간 수평적 분석
	주체	업무 운영담당자	분석, 기획 담당자
분석변화	분석대상	정형화된 DB 데이터	정형데이터+SNS, 게시판 등 비정형 데이터
	규모	기가~테라바이트 급	페타~제타 바이트급
	분석범위	단일저장소	다중저장소
	적용시간	일괄(batch) 처리	인타임 처리
	데이터	저장 후 처리	흐름(on the fly) 분석 정보의 시각화(Infographics)

※ 출처 : 국가정보화전략위원회(2011). 빅데이터를 활용한 스마트 정부 구현(안)

● 그림 3 포브스의 빅데이터 지도



※ 출처: Dave Feinleib (2012). The big data landscape, Forbes, (2012. 6.19).

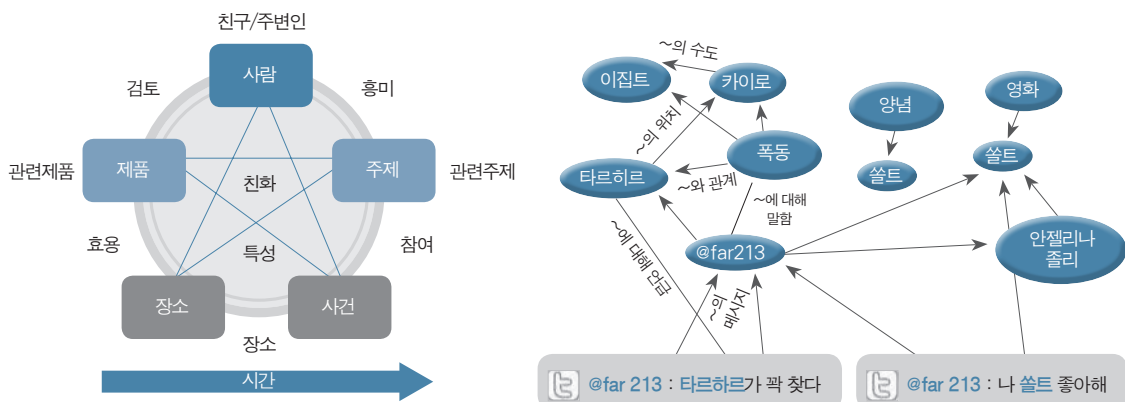
의 'Elastic MapReduce'를 활용해 데이터를 분석하였다. 그들은 사이트 방문자의 나이, 몸무게, 수면 장애일수 등 개인정보를 기입해 수면 문제를 진단하고, 수면장애가 실제 교통사고율을 높인다는 연구결과를 시각적으로 보여주었다.

월마트는 소셜미디어 회사인 Kosmix를 인수해 소셜 네트워크와 콘텐츠를 관리 및 분석함으로써 소비자 소비패턴을 조사하여 마트 운영에 반영하

였다. Walmart는 'Social Genome'라는 프로젝트를 통해서 소셜 미디어의 대규모 데이터를 수집하여 리얼타임으로 분석하고 있다. Walmart는 트위터의 트윗, 페이스북의 Feed 등의 데이터를 분석하여 인물, 사건, 장소, 제품, 조직 등의 관계를 파악함으로써 상품판매 촉진에 활용하고 있다. 그 결과 필요한 물품을 빠르게 제공하여 불필요한 재고 낭비를 방지하고, 고객이 원하는 상품을 공급함으로써 고객 만족도를 극대화하여 기업 발전에 기여하는 선 순환적 효과를 거두고 있다.

Google은 지난 2011년 11월 베타 서비스로 공개된 클라우드 기반 대규모 데이터 분석툴 'Big Query' 서비스를 제공함에 따라 업계 관심이 높아지고 있다. Google은 최근 IT 업계의 화두로 떠오른 빅데이터에 주목하고, 누구나 비용 부담 없이 클라우드 상에서 빅데이터를 활용할 수 있도록 하는 데 주력하고 있다. 기존 빅데이터 분석 시스템은 기업 내부에서 이루어지는 것이 일반적이었으나, 'Big Query'는 클라우드 기반의 분석 소프트웨어를 제공함으로써 기업이 별도의 인프라 투자 없

● 그림 4 월마트 랩의 Social Genome을 트위터에 응용한 체계도



※ 출처 : <http://www.walmartlabs.com/social-genome/>

이도 빅데이터 분석 업무를 수행할 수 있다는 장점을 갖고 있다.

일본의 경우 지진이나 해일 피해지역의 도로 교통상황을 안내할 수 있도록 일본 전역의 택시 약 11,000여 대와 데이터 제공에 동의한 사용자로부터 실시간으로 교통정보를 수집하여 분석하고 있다. 지진이나 해일 상황에서 교통 통제나 교통체증 해결 및 차량 구조를 위해 GPS 데이터에서 자동차의 주행스피드를 계산할 수 있도록 교통정보를 예측한 후 사용자의 스마트폰으로 송신함으로써 가장 빠른 길을 탐색하여 안내한다(히다찌, 2012). 일본의 경우 자연재해가 끊임없이 발생하기 때문에 인명피해나 차량피해를 최소화 할 수 있는 방안으로 빅데이터 분석에 큰 기대를 갖고 있다.

미국의 경우 주로 의료와 보건에 빅데이터 활용에 초점을 두고 있으며, 일본의 경우 자연재해나 사고에 대비한 피해의 최소화 방안에 빅데이터 활

용에 큰 기대를 걸고 있다.

그 밖에 독일의 경우 연방노동청(Federal Employment Agency)은 대규모 실업데이터 분석을 통해서 실직자 각 개인에게 맞는 맞춤형 일자리를 제공하여 100억 유로의 비용을 절감하였다. EU의 경우 빅데이터의 활용을 통해서 15~20%의 공공관리 비용감소와 2~4천 달러의 가치 창출, 향후 10년간 0.5% 생산성 증가 효과를 기대하고 있다. 특히 제조업에 빅데이터를 적용하면 상품개발과 조립 비용의 50% 절감이 가능하고 운용자본도 7% 이상 절감할 수 있을 것으로 예측하고 있다. 소매업은 이윤 60%이상 증가와 매년 0.5%~1%의 생산력 향상을 기대하고 있다. 의료와 건강분야는 매년 3,300억 가치와 년 0.7%의 생산성을 증가시킬 것이며, 개인 위치정보 서비스는 1,000억 이상의 공급자 매출과 7,000억의 사용자 혜택을 줄 것으로 전망해 각 국가들이 서둘러 빅데이터에 투자하

● 표 3 미국과 일본의 빅데이터 활성화 정책 비교

구분	미국	일본
특징	<ul style="list-style-type: none"> • 범정부 차원 이니셔티브 가동 • 기업 활동에 실질적 활용을 위한 원천 기술 및 서비스 개발 병행 추진 	<ul style="list-style-type: none"> • 산학연의 참여를 활성화 • 산업적 활용보다는 빅데이터 기초와 첨단 기술 연구개발에 초점 • 재해 대비 데이터 활용방안
추진 주체	<ul style="list-style-type: none"> • 과학기술정책실 (Office of Science and Technology Policy) 	<ul style="list-style-type: none"> • 총무성, 문부과학성, 경제산업성
주요 정책	<ul style="list-style-type: none"> • Big Data R&D Initiative 	<ul style="list-style-type: none"> • 정보폭발(Info-Plosion) 프로젝트 • 정보 대 항해 프로젝트
정책 내용	<ul style="list-style-type: none"> • 2억 달러 예산 투자 • 국립과학재단, 국방부, 에너지부 등 6개 연방 기관 주도 • 빅데이터 핵심기술향상, IT, 과학, 안보, 교육 등 산업전반으로 적용, 전문인력양성 등이 핵심 	<ul style="list-style-type: none"> • 총무성이 '빅데이터 활용'을 2020년까지의 ICT 전략 테마 선정 • 2013년 빅데이터 R&D 예산 편성 • 빅데이터 정보 추출, 분석, 활용을 위한 첨단 기술개발 • 대학과 민간연구소 및 기업 참여 유도

고 있다(Mckinsey, 2011).

이와 같이 미국을 비롯한 영국, 독일, 일본 등 주요 선진국은 이미 빅 데이터를 미래정보통신기술(ICT) 전략의 중요한 전환점으로 간주하여 전문 인력을 양성하고 주요 정책 영역에 적극 배치하고 있다. 우리나라 역시 빅데이터에 큰 관심을 보이면서 국민권익위원회는 연간 약 300만 건에 달하는 민원을 분석하는 ‘민원동향분석시스템’을 도입하여, 민원제기 이전에 원인을 파악하여 사전에 해결하기 위해 빅데이터 분석을 실시하고 있다. 또한

국가정보화위원회에서는 향후 공공부문에 빅데이터 활용 계획을 통해서 과거 재해기록과 사전 피해 예측, 유관기관의 실시간 정보 분석, 소셜 데이터 분석 등을 기반으로 대형재난의 전조현상을 감지하여 사전에 차단하는 시스템을 구축하기 위해 노력하고 있다.

나. 국내 빅데이터 동향

지난해 국가정보화전략위원회는 ‘빅데이터를 활용한 스마트 정부 구현(안)’을 통해서 지식경제

● 표 4 주요국가의 빅데이터 활용 현황

국가	활용분야	내 용
미국	국토보안	<ul style="list-style-type: none"> ○ 9.11 이후 미국은 국토안보부를 중심으로 테러·범죄 방지를 위한 범정부적인 빅데이터 수집 및 분석 • 부시 행정부 시절 국토안보부 장관 Michael Chertoff는 국토보안을 위해서 빅데이터 분석 추진 • 국내외 금융거래 감시로 테러자금 조달 색출 강화
	치안	<ul style="list-style-type: none"> ○ FBI의 종합 DNA 색인 시스템(CODIS) • DNA 포렌식, 클라우드 DNA분석 등 ‘빅DNA 데이터’의 활용을 통해 2007년 약 45천건의 범인 DNA Hit rate 달성 • 1시간안에 범인 DNA 분석을 위한 주정부 데이터연계 및 빅데이터 실시간 분석 솔루션 확보
	의료	<ul style="list-style-type: none"> ○ 오바마 Health 2.0 -필박스 프로젝트(Pillbox) • 국립보건원의 사이트로 약(drug) 검색 서비스 제공 • Pillbox를 통해 수집된 빅데이터를 통해 후천성면역결핍증(HIV) 등 관리대상 주요 질병의 분포, 연도별 증가 등에 대한 통계치 확보 가능
영국	정보공개	<ul style="list-style-type: none"> ○ 영국은 정부 사이트(data.gov.uk)를 통해 공공부문의 정보 공유 및 활용을 위한 데이터 원스톱 서비스 제공 • 정부의 투명성 제고, 국민의 알권리 향상, 사회적 가치 증대, • 일반인들의 참여를 장려하고 아이디어 수렴, 앱개발, 데이터 공개 등의 주제에 대한 커뮤니티 제공
싱가포르	국가위협관리	<ul style="list-style-type: none"> ○ 빈번히 발생하는 테러 및 전염병으로 인한 불확실한 미래 대비를 위하여 2004년부터 빅데이터 기반 위험관리 계획을 추진 • RAHS(Risk Assessment & Horizon Scanning) 시스템을 통해 질병, 금융위기 등 모든 국가적 위험을 수집 및 분석하여 위험을 선제적으로 관리 • 수집된 위험 정보는 시뮬레이션, 시나리오 기법 등을 통해 분석되어 사전에 위험을 예측하고 대응 방안 모색
호주	정보공개	<ul style="list-style-type: none"> ○ 호주 정보관리청은 정부 2.0을 통한 정보개방 • 방대한 양의 정보를 검색하고 분석 및 재사용할 수 있도록 자동화된 툴을 활용하여 시간과 자원을 절감 • AGIMO 산하 정부 2.0 전략/ 서비스 팀에서는 정부 데이터에 대한 Repository 및 검색 툴을 서비스하는 data.gov.au 웹사이트 운영

※ 출처 : 국가정보화전략위원회 (2011), 빅데이터를 활용한 스마트 정부구현

부와 행정안전부 그리고 방송통신위원회 등 국가 부처 간 협력을 통해서 공공데이터와 소셜 데이터를 분석해 대내·외 이슈를 파악하여 빠르게 대책을 수립해야 한다는 공동의 목표를 설정하였다. 국가정보화전략위원회는 정부기관이 빅데이터를 활용하면 예산절감, 주변상황에 대한 신속한 대처로 인해 삶의 질이 향상되면서 정부 신뢰도를 높일 수 있다고 기대하고 있다. 아울러 공공분야에 빅데이터를 활용할 경우 경제적 효과는 약 10조 7천억원 이상이 될 것으로 분석하였다(국가정보화전략위원회, 2011).

방송통신위원회 역시 2012년 6월 빅데이터와 관련해 시장 확대, 인력양성, 제도개선 등의 정책적 지원 과제를 담은 ‘빅데이터 서비스 활성화 방안’을 발표하였다. 하지만 우리나라는 세계 최고수준의 IT 인프라를 보유하고 있음에도 불구하고, 빅데이터 분야에서 기술력 향상, 인력이 매우 부족, 관련 데이터 공유, 법제도 및 규제 마련 등에 대한 문제를 해결한다는 방침이다(방송통신위원회, 2012). 지식경제부는 기업, 학계, 정부 관계들이 모인 컨퍼런스를 마련해 빅데이터 소프트웨어(SW) 사업 육성전략을 논의하였다.

기존 데이터 분석은 구조화된 분석을 통해서

● 표 5 빅데이터 서비스 활성화에 따른 주요 과제

분야	과제 내용
시장확대	<ul style="list-style-type: none"> • 신규 서비스 발굴, 확산을 위한 시범 서비스 추진 • 빅데이터 기술개발 및 플랫폼 경쟁력 강화 • 빅데이터 지원센터 설치 운영 및 정보공유체계 마련 • 빅데이터 산업실태 조사
인력양성	<ul style="list-style-type: none"> • 전문인력양성
제도개선	<ul style="list-style-type: none"> • 개인정보보호 관련 법제도 개선 • 빅데이터 산업 진흥을 위한 법제도 개선

논리적 해석이 중요하게 간주되었지만, 빅데이터는 창조적이며, 통합적인 사고로 직관적인 해석이 요구된다. 즉 수많은 데이터를 수집하고 축적하는 것보다 왜 분석하는지, 무엇을 분석할 것인지, 분석을 통해서 얻고자 하는 것이 무엇인지에 대한 정확한 이해가 필요하다.

현재 SK 텔레콤은 소셜 네트워크에서의 여론분석을 위한 ‘스마트 인사이트(smart insight)’ 시스템을 도입하였다. SK가 제공하는 스마트 인사이트 시스템은 기업들이 원하는 키워드를 중심으로 온라인 여론을 분석하여 실시간으로 제공한다. 여론에 대한 분석뿐만 아니라 분석결과와 관련된 기사, 블로그 댓글, 트윗 내용 등을 실시간으로 분석이 가능하다.

이러한 서비스를 통해서 새로운 제품 출시 정보가 어떤 과정으로 고객에게 전달되는지 실시간 모니터링이 가능하고, 잘못된 정보나 기업의 위기에 대해서 즉시 대처 할 수 있는 기반을 마련할 수 있다. 아울러 최신 트렌드 이슈 및 소비자 반응 조사가 가능하다.

효성 인포메이션은 ‘SAP HANA’를 공개하며 SAP HANA 어플라이언스인 유니파이드 컴퓨터 플랫폼 SAP HANA(UCP for SPA HANA)를 앞세워 실시간 데이터 분석 시장에 진출하였다. SAP HANA는 인메모리 기술을 활용해 데이터를 데이터베이스(DB)가 아닌 메모리에 저장해 처리 속도를 극대화하는 소프트웨어로 소개되고 있다. 이 기술은 Lenovo, Colgate-Palmolive 등 다양한 기업에 적용되고 있으며, HP와 IBM, Cisco 등이 HANA를 탑재한 어플라이언스를 제공하고 있다(안호천, 2012). SAP HANA는 오랜 시간과 비용이 소모되는 데이터

수집 및 분석 작업을 쉽고 빠르게 수행할 수 있도록 돕는다. 이러한 분석을 통해서 의사결정구조에 획기적인 정보를 제공할 수 있을 것으로 기대하고 있다.

그밖에 LG CNS는 최근 빅 데이터 사업조직을 신설하는 등 본격적인 시장 진출에 나섰다. 아울러 삼성 SDS, SK C&C 등 경쟁사들도 빅 데이터 분야 사업 전략을 추진 중이다.

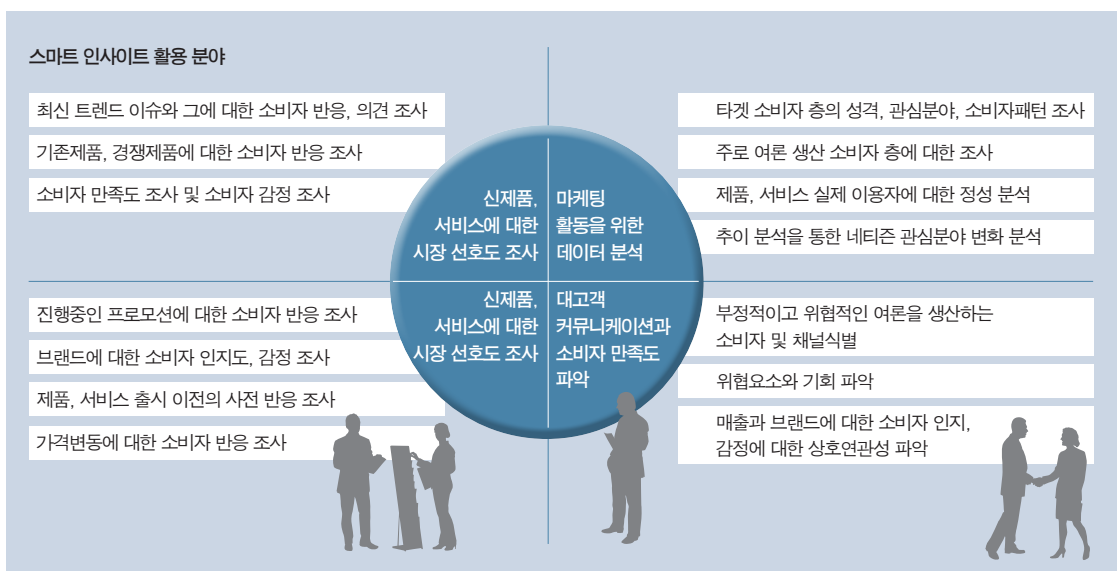
다. 국내 빅데이터 활용범위와 가능성

빅데이터를 이용하여 국내 호텔이나 기업은 고객 맞춤형 서비스 제공을 통해서 고객 만족도를 극대화할 수 있을 것이다. RitzCarlton Hotel은 체인 전체의 이용객 데이터분석 결과를 공유하여 맞춤형 서비스를 제공하고 있다. 호텔의 모든 직원은 고객 관리 인지 프로그램(Customer Recognition Program)에 고객에 대한 상세 정보를 기록하여 전

세계 약 100만 명의 고객 정보의 DB를 보유하고 있다(Barsky, 1995). 고객 정보를 통해서 투숙 고객의 특성을 기록하여 취합함으로써 고객이 투숙하기 전에 고객의 취향을 파악해 청소부를 비롯한 모든 직원이 맞춤형 서비스를 제공하고 있다. 이에 따라 비용절감효과와 고객 맞춤형 서비스로 고객 만족도를 극대화시키고 있다(윤미영·권정은, 2012).

또한 제조업분야에서 생산-유통-판매-재고관리 등에 빅데이터 활용성이 기대된다. 상품생산의 경우 빅데이터를 이용해 국내 소비자층을 설정하고, 원하는 상품을 개발하여 디자인할 수 있다. 생산된 상품은 원하는 고객에게 바로 판매됨으로써 재고관리가 더욱 수월해 질 것이다. 제조분야에서 생산된 제품을 규모가 큰 전자상거래(e-commerce) 사이트 고객에게 판매할 때 역시 빅데이터가 활용될 수 있다. 전자거래 이용자들은 사이트 검색을 이용하여 자신이 구매할 상품을 선택한다. 카테고리

● 그림 5 SK 텔레콤 '스마트 인사이트' 활용 분야



※ 출처 : 임원기 (2011). SKT, 인터넷·SNS 분석.... 기업 부정적 여론 미리 알려준다. 한국경제(2011. 8. 29).

리 기반의 검색 접근방식은 탐색이 어렵고, 특정상품에 대해 검색하면 수많은 변수가 등장하여 고객은 쉽게 관심을 잃게 된다. 하지만 프로파일 및 고객데이터를 바탕으로 고객 취향을 분석할 수 있다면 고객이 원하는 상품을 쉽게 제공할 수 있을 것이다. 이러한 방식은 MapReduce와 Data Warehouse를 이용해 소매업체가 고객 행동을 이해하고, 고객 개인에게 맞는 정보를 제공하는 것과 같다. 고객은 몇 번 클릭하지 않고도 자신의 원하는 상품을 선택하여 구매할 수 있다.

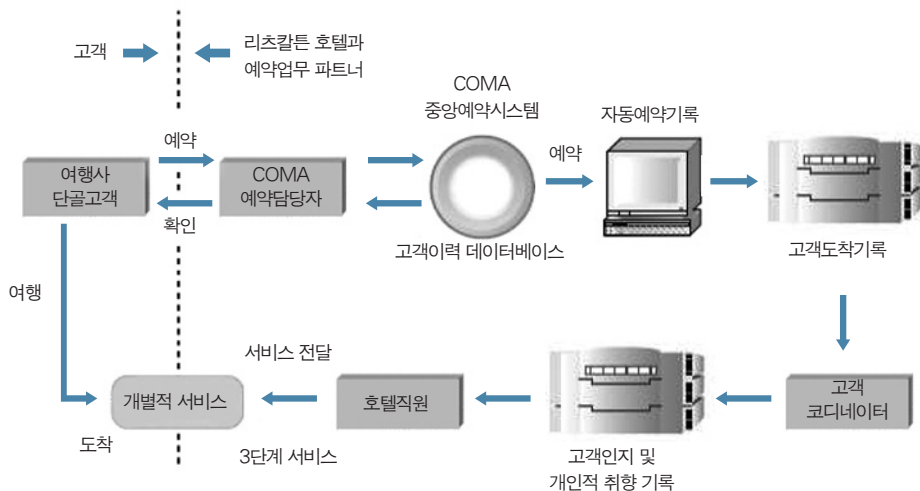
광고 분야에서 이용자 단말기의 카메라에 잡힌 실제 영상에 가상의 요소를 추가해 부가정보를 제공하는 증강현실(AR: Augmented Reality) 기술은 소비자의 구매 의사결정 과정에 매우 효과적일 수 있다. 증강현실 가상화 서비스는 소비자의 구매의사결정에 필요한 상품에 대한 구체적인 정보를 제공함으로써 편리하고 정확한 구매행위를 유도하여

기업의 매출 증대에 기여할 수 있다. 페이스북은 현재 자사 서비스 플랫폼에 올라오는 글이나 그림, 동영상 등을 분석해 이용자 관심 사항을 파악함으로써 맞춤형 광고에 활용하고 있다. Netflix 역시 서비스 사용자의 영화감상 성향을 분석하여 고전 영화 추천 서비스를 제공하고 있다.

특히 공공부분에서 의료와 헬스는 빅데이터 활용 가능성이 가장 높은 분야로 손꼽히고 있다. 미국의 의료보험사인 Wellpoint는 IBM의 왓슨 솔루션을 도입해 환자의 증상, 면담결과 등을 분석해 3초 안에 진단 또는 치료 가이드라인을 제시하고 있다(류한석, 2011).

우리나라 역시 빅데이터 분석을 통해서 의료분야에 대한 대국민 서비스가 가능할 것이다. 스마트 미디어나 스마트 폰을 통해서 자신의 건강이나 질병과 관련된 정보를 탐색할 때 현재 인터넷 수준의 검색결과가 아니라 증상의 시각화-발병원인-치료

● 그림 6 리즈칼튼 호텔의 Customer Recognition Program



※ 출처 : Barsky, J. D. (1995), World-Class Customer Satisfaction, p.143. Irwin, 윤미영, 권정은 (2012), 빅데이터로 진화하는 세상: Big Data 글로벌 선진사례, 36쪽 재인용.

가이드라인-예방 등과 관련된 다각적인 정보를 제공할 수 있을 것이다. 아울러 질병 치료를 위한 전문병원 추천과 의료진 소개 등도 포함될 수 있다. 이러한 의료 분야의 빅데이터 분석은 전국의 병원과 환자 그리고 의료진이 갖고 있는 정보를 공유할 수 있어야 한다.

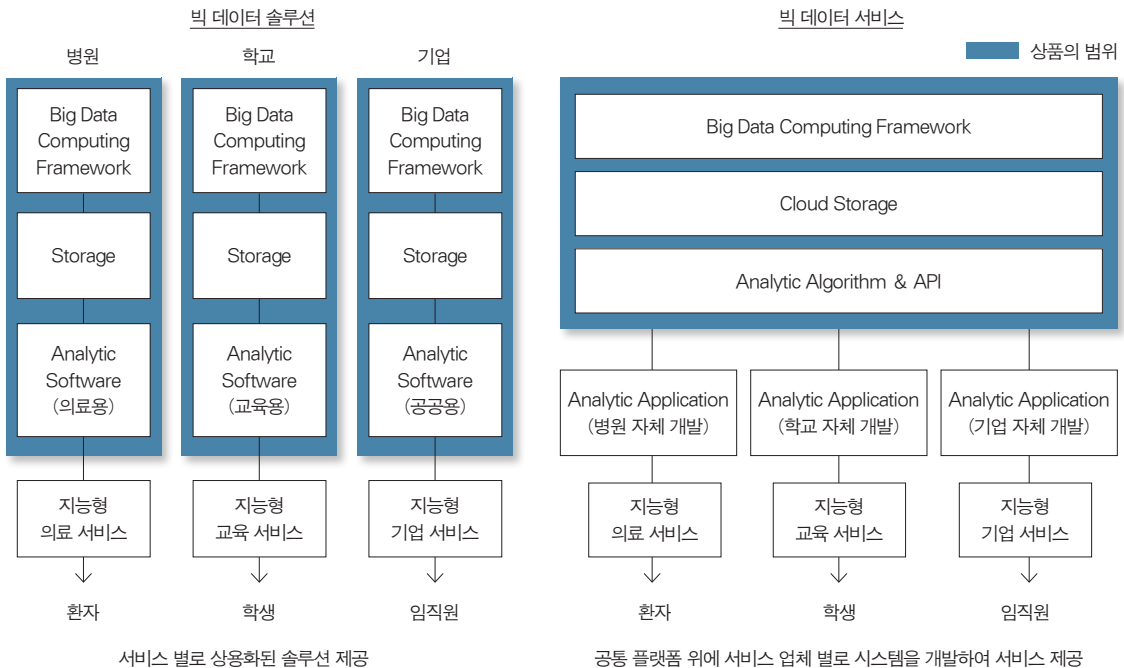
우리나라의 경우 특히 주택에 대한 관심이 매우 높기 때문에 부동산 빅데이터를 분석하여 서비스할 필요가 있다. 부동산 대책이 정권의 변화에 따라 변경되기 때문에 많은 사람들이 혼란을 겪고 있다. 이에 따라 주택(매매 및 전·월세)과 관련된 빅데이터를 분석하여 제공한다면 장기적인 부동산 대책 마련은 물론 개별 소비자들에게 유용한 정보로 활용될 수 있을 것이다. 부동산 정보와 관련해

지난 2011년 12월 국토해양부는 ‘국토정보’라는 모바일 앱을 개발하여 보급하였다. 앱을 이용하여 부동산 정보를 조회할 수 있으며 토지대장과 건축물대장 정보를 열람할 수 있다. 앞으로 주택과 관련된 빅데이터 분석은 지역, 대중교통정보, 가격, 주차, 교육환경 등에 대한 정보를 산출하여 제공해야 할 수 있을 것이다. 주택관련 빅데이터 분석은 국가에서 인정한 기관이 신뢰할 수 있는 정보를 제공해야 한다.

빅데이터 분석에 이용되는 비정형화된 데이터는 스마트폰 이용자에게 의해서 발생한다. 스마트폰 이용은 단지 이용자 편리성 뿐만 아니라 국가적 차원에서도 긍정적인 효과를 유발한다.

스마트폰 이용자들이 많거나 스마트 기기에 접

● 그림 7 공공부문 빅데이터 활용 솔루션



※ 출처 : 손민선, 문병순 (2012). 빅데이터 시대의 한국 갈라파고스가 되지 않으려면. LGRI 리포트, 3월호, 11쪽

근할 수 있는 다양한 방법이 제시되는 국가는 더 민주적이고 투명한 국가로 알려져 있다. 아울러 특정 지역의 휴대전화시간의 감소는 그 지역 인구의 소득 손실을 의미하기도 한다(손민선·문병순, 2012). 따라서 일반 시민들이 자유롭게 스마트 기기에 접근할 수 있는 기반을 확충하고, 많은 데이터를 생산할 수 있도록 국가적 차원의 배려가 필요하다.

III. 빅데이터 시대의 문제점

빅데이터는 분명 새로운 비즈니스 모델을 제시하고 있으며, 생산-유통-소비에 걸쳐 혁신적인 변화를 가져올 것이다. 하지만 빅데이터라는 개념정 의에서 알 수 있듯이 데이터 규모에 대한 규정과 데이터 종류, 데이터 보관과 활용 차원에서 사회적 문제를 유발할 수 있다. 따라서 빅데이터 효용에 대한 기대 못지않게 가외적인 문제를 어떻게 처리하는가에 따라 빅데이터의 평가가 달라질 수 있다.

1. 새로운 빅 브라더(Big brother)의 탄생

데이터 집중화 현상에 따라서 데이터 관리 및 데이터 분석 서비스 제공 업체가 나타난다. 이는 곧 빅 브라더의 탄생을 의미한다. 빅데이터를 분석하기 위해서는 데이터 보관 및 관리에 대한 책임 있는 기관이나 기업이 등장할 수밖에 없으며, 책임 있는 기관은 자신들의 역량을 강화하기 더 많은 데이터와 정보를 원한다. 결국 빅데이터는 빅 브라더라는 일종의 사회적 괴물을 만들어낼 수밖에 없다. 특히 빅데이터는 개인과 관련된 정보가 축적되기 때문에 개인정보의 유출과 사생활 침해라는 문제

가 더욱 확산될 것이다.

제로미 벤담(Jeremy Bentham)은 원형 감옥을 예로 들어 감시사회를 언급하였다. 우리는 현재 우리가 살고 있는 사회를 ‘관옵티콘’의 사회라 부르고 있다. 관옵티콘(Panopticon)은 원형감옥의 형태로 감시자의 시선을 볼 수 없는 죄수들은 감옥에서 제시하는 규율을 철저하게 따른다. 감시자들이 보이지 않는 어두운 곳에서 끊임없이 죄수들을 감시하기 때문이다. 국가인권위원회는 일반인들이 하루에 평균 80회 이상 카메라에 촬영된다고 발표하였다. 즉, 언제, 어디서나, 어디에 머물렀는지 이동 경로와 머무는 장소를 확인할 수 있다. 최근 범죄자들을 구속하는데 감시카메라의 역할이 절대적인 기여를 하고 있다.

이와 같은 감시 시대를 관옵티콘의 시대라 할 수 있다. 우리가 느끼지 못하는 사이에 우리를 감시하는 시대적 환경이 조성된 것이다. 빅데이터는 관옵티콘 시대를 넘어 슈퍼 관옵티콘을 조성할 우려가 있다. 빅데이터로 쌓이는 대량의 데이터는 상당수가 개인과 관련된 것이기 때문에 ‘프라이버시’ 침해문제가 유발될 것이다. 따라서 사생활 노출은 물론 감시와 처벌이 끊임없이 재생되는 시대의 서막을 알리고 있다.

일례로 2011년 10월 국가인권위원회는 카카오 특의 강제적이고 고압적인 개인정보수집 행위가 개인정보자기결정권을 침해하는 것으로 헌법재판소가 인정한 바 있다. 또한 헌법재판소는 정보통신망법에 대한 위반의 소지가 있다고 판단하여 방송통신위원회에게 모바일 개인정보에 대한 적절한 조치를 취하고 개인정보 처리에 대한 가이드라인을 제정할 것을 권고 하였다. 특히 카카오톡은 ‘플

리스 친구' 서비스를 통하여 수집된 개인정보를 모바일 광고를 위한 제3자에게 제공하면서 '선 동의 후 거부방식'을 적용하였다. 이는 개인정보보호법 제22조 제2항에 규정된 선택적 동의의 획득 과정에서 옵트인(opt-in) 방식이 아닌 옵트 아웃(opt-out) 방식을 적용하여 개인정보자기결정권을 침해할 소지를 야기하였다.

또한 2011년 4월 모바일 광고를 목적으로 위치 정보를 무단으로 수집하여 2억 건 넘는 개인정보를 빼돌린 개인정보 유출사태가 발생하였다. 이 사건은 인기 앱인 '대리운전' 등에 위치추적 프로그램을 탑재하여 배포하는 수법을 이용하였다. 유출된 개인정보는 모바일 광고에 이용되었는데 사용자가 심야에 유흥가에 있으면 대리운전이나 술집을 광고하고, 점심시간에는 사무실 밀집 지역에 있는 식당 등의 위치와 연락처를 앱 화면 하단에 표출하였다. 이를 통해서 모바일 광고업체는 6억 원 이상의 부당이익을 얻었다. 이러한 불법 광고가 가능했던 이유는 휴대전화 번호만 알면 날짜와 시간 대별로 단말기의 좌표가 나타나기 때문에 이 좌표를 구글, 네이버, 다음 등에서 무료로 제공하는 API(Application Programming Interface)에 입력하면 오차범위 1m 이내의 위치를 지도상에 표시해 준다. 이는 기존 휴대전화 위치정보의 오차범위(50m)와 비교할 수 없을 정도로 정밀해 사용자가 방문한 상점이나 숙박업소 등의 이름까지 추적할 수 있다. 게다가 GPS 기능을 꺼도 스마트폰 전원을 차단하지 않으면 위치가 실시간으로 수집된다.

개인정보나 사생활 침해와 관련된 문제를 해결하기 위해서 '빅데이터 관리 및 데이터 활용에 관

한 법률'(가)을 제정하여 기업이나 기관이 저장하고 있는 다양한 데이터를 활용할 때 적절한 기준을 마련해야 한다. 개인정보와 사생활은 민주주의 핵심 가치이기 때문에 법률적 보호가 필요하다. 빅데이터와 관련된 법률을 제정하여 규제함으로써 빅브라더의 탄생을 방지하고, 부정적 결과를 최소화해야 한다. 이러한 역할은 국가적 차원의 규제가 필요한 부분이며, 국가 역시 빅데이터 활용에서 나타나는 사회적 문제를 최소화하기 위해서 데이터 저장, 활용, 교류에 대한 원칙이 필요하다.

빅데이터의 대부분은 SNS를 통해서 제공되는 정보로 지식정보와 함께 정서적인 공감에 바탕을 둔 감성적인 정보가 큰 비중을 차지하고 있다. 감성적인 정보에 대한 축적은 소비자들의 라이프스타일을 파악할 수 있다. 인간의 감성적인 정보와 이성적인 정보는 사람들의 생각을 의미하며, 빅데이터는 인간의 생각을 읽어내는 작업이 된다. 따라서 인간의 생각을 읽는다는 것은 행동을 규제할 수 있음을 의미하고, 행동에 대한 규제는 결국 자유에 대한 억압 가능성을 나타낸다. 이와 같은 개인정보 유출과 사생활 침해로 인한 판옵티콘의 확장은 시대의 산물이지만, 인간의 자유를 어떻게 보장할 것인지 보다 구체적이고 적극적인 고민이 필요하다.

2. 데이터 저작권의 문제

저작권법 제2조 19항과 20항에서 데이터베이스에서 대해서 저작권법 적용을 인정하고 있다. 19항은 '데이터베이스'는 소재를 체계적으로 배열 또는 구성한 편집물로서 개별적으로 그 소재에 접근하거나 그 소재를 검색할 수 있도록 한 것을 말한다. 아울러 20항은 '데이터베이스제작자'는 데이터베이스의

제작 또는 그 소재의 갱신·검증 또는 보충(이하 ‘갱신등’이라 한다)에 인적 또는 물적으로 상당한 투자를 한 자를 말한다고 명시하고 있다. 저작권은 크게 저작인격권과 저작재산권으로 분류되는데, 데이터의 경우 저작재산권으로 볼 때 복제권, 공연권, 공중송신권, 전시권, 배포권, 대여권, 2차적저작물작성권 등에 대한 권리가 원 저작자에게 있다. 이러한 저작권 이용은 제35조 3에 의해서 저작자의 정당한 이익을 부당하게 해치지 않는 보도·비평·교육·연구 등을 위하여 저작물을 무료로 이용할 수 있다. 하지만 그렇지 않을 경우 영리성 또는 비영리성 등 이용의 목적 및 성격에 따라서, 저작물의 종류 및 용도에 따라서, 이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성에 따라서, 저작물의 이용이 그 저작물의 현재의 시장가치나 잠재적인 시장 또는 가치에 미치는 영향에 따라서 저작권자의 경제적 권리를 인정해야 한다.

미국, 독일 등을 중심으로 온라인 데이터베이스(DB) 저작권 침해 논쟁이 활발하게 전개되고 있는 가운데 국내에서도 PC통신, 인터넷 등 온라인 서비스를 통해 제공되는 데이터를 저작권 대상에 포함시켜야 한다는 주장이 제기되고 있다. 저작권심의조정위원회(2012)는 온라인 서비스상의 데이터도 저작권 보호를 받아야 하며, 이를 위해서 입법화 한다는 방침이다. 이에 따라 올해 들어 처음 PC통신 사업자와 관련된 저작권 침해소송이 발생한 국내에서도 온라인데이터의 저작권에 대한 논의가 본격화될 것이다.

빅데이터는 저작권 법 적용 대상이 될 수밖에 없으며, 현재 방송이나 책은 물론 모든 창작물은 저작권이나 특허권에서 자유롭지 않다. 빅데이터

역시 최초 데이터 생산자, 유통사, 2차 정보 생산자 등과 관련해 저작권 논란에 휩싸일 수 있다. 최초 데이터 생산자는 고유의 데이터를 비상업적으로 제공하였지만, 데이터를 분석하여 시각적인 정보를 제공하는 2차 데이터 생산자가 상업적으로 판매할 경우 저작권 분쟁이 발생할 수 있다. 따라서 학술적 용도나 개인적 용도가 아니라, 상업적 목적을 갖고 데이터를 이용할 경우를 대비하여 저작권 문제를 해결해야 한다.

빅데이터와 관련된 저작권 문제는 사업의 활성화, 빅데이터 공유, 글로벌 경쟁력 확보, 사회기여도 등을 명분으로 저작권 집중관리 시스템을 고려해 필요가 있다. 저작권 집중관리 시스템의 사례로 음원을 들 수 있다. 음원의 경우 한국음악저작권협회, 한국음악실연자협회 등에서 저작권을 집중관리하여 저작권료를 징수하고 분배한다. 음원사례에서 보듯이 데이터 집중관리 기구를 만들어 유통되는 데이터의 저작권을 관리하고, 계약에 의해서 수익의 일부를 사회에 환원할 수 있는 공익적 성격의 기구를 설립해야 한다. 따라서 데이터 활용과 이용에 관한 자율 규정 및 동의서를 작성하고, 2차 저작물에 따른 저작권 침해에 대한 예외 사항을 최대한 보장해야 한다. 아울러 기업이 모든 정보를 저장할 수 없기 때문에 필요한 정보에 대해서 상호 교환이나 이용에 대한 협조가 필요하다. 특정기업이나 집단이 정보폐쇄(Information Silo) 방침을 갖고 있을 경우 빅데이터의 활용영역은 매우 위축될 수밖에 없다. 이러한 경우 역시 교류의 조건은 저작권료나 이용에 따른 비용을 지출해야 한다.

3. 네트워크 블랙아웃(network blackout)

Gartner(2011)는 데이터가 21세기 원유와 같은 역할을 할 것이며, 빅데이터는 향후 주목해야 할 이머징 기술이라고 하였다. 또한 PWC(2010)는 빅데이터가 새로운 비즈니스 가치창출의 핵심 키워드가 될 것으로 예견하였다. 빅데이터에 대한 이러한 청사진은 데이터 저장 기술, 데이터 분석 소프트웨어, 데이터 해석 인력, 적용 분야 개발이라는 단계적 발전이 이루어져야 한다. 그렇지 않을 경우 빅데이터는 일종의 사회적 유행에 머물 수도 있다. 뿐만 아니라 기술적 차원에서 최근 나타나는 데이터 트래픽 현상과 DDoS나 해킹에 철저히 대비해야 한다. 특히 DDoS나 해킹으로 '네트워크 블랙아웃'현상이 발생할 가능성이 있다. 네트워크 블랙아웃 현상이 발생하면 네트워크 기반으로 연결돼 있는 금융·교통·의료·국방 등에서 심각한 문제가 발생할 뿐만 아니라 의료, 산업, 국방, 경제 등 전 분야에서 혼란이 예상된다.

한국인터넷진흥원 분석 자료에 따르면 DDoS 공격 피해 건수는 2009년 42건, 2010년 53건, 2011년 63건으로 계속 증가세를 보이고 있다. 얼마전 KT 이용자 870만 명의 개인정보가 해킹으로 유출되었다. 이와 같이 DDoS나 해킹으로 네트워크 블랙아웃현상이 발생할 가능성은 점점 높아지고 있다. 최근 발생한 디도스 공격의 35%가 정치적 혹은 이념적 동기, '허무주의 혹은 파괴주의에 근거(31%)', '온라인 게임 타깃(29%)', '실력 과시(25%)' 등이 DDoS 공격 동기로 나타났다(호애진, 2012). 즉, 누구나 쉽게 적은 비용으로 개인 PC는 물론 서버에 DDoS 공격이 가능하다.

트래픽으로 인해 네트워크 블랙아웃 현상이 발생할 가능성은 매우 낮지만, 지속적인 망 관리가

이루어지지 않는다면 영국과 일본의 사례에서 볼 수 있듯이 사회적 혼란을 경험하게 될 것이다. 트래픽으로 인해 2012년 7월 14일 영국 통신사 O2의 네트워크가 24시간 동안 다운되었고, 프랑스 오렌지가 9시간 동안 블랙아웃되는 현상이 발생하였다. 일본에서도 올해 초 NTT도코모가 수차례 다운되는 등 스마트폰 사용률이 높은 세계 곳곳의 통신사 네트워크가 멈추는 사고가 이어지고 있다(심나영, 2012). 이러한 현상은 결국 스마트폰 사용 급증으로 인해 문자, 동영상, 그림 등과 같은 비정형화된 데이터 사용이 급증하면서 나타나는 현상이라 할 수 있다.

빅데이터 시대 기업이나 정부가 축적한 데이터를 분석할 때 DDoS 공격으로 인해 네트워크 혼란이나 마비가 초래될 수 있으며, 해킹을 통해서 잘못된 데이터를 제공하거나 데이터를 변경할 경우 엄청난 경제적 손실이 나타날 수 있음을 염두해야 한다.

Ⅳ. 결론

빅데이터는 단순히 많은 데이터를 분석하는 것이 아니다. 분석뿐만 아니라 시스템, 서비스 자체가 이미 빅데이터에 대한 적응능력이 되고 있다. 즉, 시스템 개발, 서비스 기획 및 운영하는 능력도 빅데이터에 포함된다. 빅데이터는 하나의 솔루션으로 해결할 수 없으며 요구사항, 데이터의 성격 등에 따라 다양한 솔루션이 조합되어야 한다. 또한 오픈소스 중심의 소프트웨어 스택(Software Stack)을 구축하고 운영하기 위해서는 내부 기술력을 갖추어야 한다. 외부시스템 구축 회사나 벤더에 의존

할 경우 경쟁력 상실이 우려된다. 한번 구축하고 관리만 하면 되는 시스템이 아니라 지속적으로 진화시켜 나가야 하는 시스템이다.

따라서 단기간(6개월~1년 이내)에 전체 시스템을 구축하고자 하는 욕심을 버려야 한다. 처음의 실패를 두려워하지 말고 지속적으로 기술을 내재화하고 시스템을 진화시켜야 한다. 오픈 소스 검증에 충분한 시간을 두고 작게라도 실행에 옮기는 것이 중요하다. 정보기술(IT) 강국을 자부하는 우리나라 역시 빅데이터 분야에 정책적 관심과 지원을 통해서 새로운 시장을 형성하여 빅데이터 산업을 활성화 시켜야 한다.

빅데이터의 활용은 민간 기업뿐만 아니라 공공의 목적을 달성하기 위해서 정부의 역할이 매우 중요한 부분을 차지한다. 기업이 보유하고 있는 데이터는 시민들로부터 발생한 것으로 데이터의 주체가 시민이기 때문에 공공의 이익에 합당한 용도로 사용할 수 있도록 데이터 공유가 필요하다. 이에 정부의 데이터 공유와 활용에 대한 기본 규칙을 마련할 필요가 있다. 뿐만 아니라 빅데이터를 분석하여 시각화 할 수 있는 기술개발과 빅데이터를 분석, 적용할 수 있는 인력양성이 필요하다. 빅데이터는 기존의 정형화된 데이터 분석이 아니기 때문에 새로운 가치를 창조하는 창의적인 인재가 필요하다. 현재 ‘방송통신발전 기본법’이나 ‘국가정보화 기본법’ 안에서 빅데이터 지원에 대한 근거를 마련할 수 있을 것이다.

빅데이터는 데이터 패러다임의 전환을 가져왔다. 정형화된 데이터뿐만 아니라 비정형화된 데이터를 분석함으로써 소비자 중심의 정보를 산출할 수 있으며, 기업은 물론 공공 영역에서도 광범위하게 사용

될 수 있다. 하지만 빅데이터가 갖는 태생적 문제로 인해 개인정보유출에 따른 사생활침해, 정보 집중화 현상, 네트워크 장애 등과 같은 사회적 혼란을 사전에 예방할 수 있는 대책이 마련되어야 한다. 이러한 문제로 인해 오히려 사회적 비용이 증가한다면 빅데이터에 대한 청사진은 이루어지지 않을 것이다.

참 고 문 헌

- (1) 국가정보화전략위원회 (2011). 「빅데이터를 활용한 스마트 정부구현」.
- (2) 김성태 (2012). 「新가치창출 엔진, 빅데이터의 새로운 가능성과 대응 전략」. 『IT & Future Strategy』, 한국정보화진흥원.
- (3) 김형준 (2012). 「빅데이터란 무엇이고, 어떻게 해야 할까?」.
- (4) 류한석 (2011). 빅데이터 비즈니스의 이해와 전망. DIGIECO, ISSUE & TREND, www.digieco.co.kr
- (5) 박현욱 (2012). 「터미널 · 도서관에서도 와이파이 맘대로 쓴다」. 서울경제, A14면, (2012.07.12)
- (6) 방송통신위원회 (2012). 「방통위, 빅데이터 서비스 활성화 적극 나선다」. 방송통신위원회 보도자료, (2012. 6.21)
- (7) 손민선, 문병순 (2012). 빅데이터 시대의 한국 갈라파고스가 되지 않으려면. LGERI 리포트. 3월호, 11쪽
- (8) 심나영 (2012). 네트워크 블랙아웃, 한국도 예외 아니다. 아시아 경제 (2012. 07.19).
- (9) 세레테크놀로지 (2012). 「아시아권 7월 페이스북 이용자 수」.
- (10) 안호천 (2012). 「빅데이터 분석 · 통계 HANA 하나면 충분」. 전자신문(2012. 7.17). 10면.
- (11) 유지연 (2012). 「세계경제포럼(WEF)을 통해 본 빅데이터 논의 동향과 함의」. 『정보통신정책 동향』, 24권 4호, 통권 526, 55~62.
- (12) 윤미영, 권정은 (2012). 「빅데이터로 진화하는 세상: Big Data 글로벌 선진사례」. 36쪽
- (13) 윤태구 (2012). 「카카오톡 가입자만 4200만... 하루 수신메시지 26억건」. 아주경제, (2012. 3. 12).
- (14) 임원기 (2011). 「SKT, 인터넷 · SNS 분석.... 기업 부정적 여론 미리 알려준다」. 한국경제(2011. 8. 29).
- (15) 채승병 (2011). 「정보홍수 속에서 金脈 찾기: 빅데이터(Big Data) 분석과 활용」. SERI 경영노트, 2011, 2, 10.
- (16) 히다찌 (2012). 「비즈니스를 똑똑하게 만드는 히다찌의 빅데이터 전략 및 인사이트」. 「글로벌 빅데이터 추진동향과 활용 사례」, 한-일 IT 세미나 2012(제23회) 발표문.
- (17) Barsky, J. D. (1995). *World-Class Customer Satisfaction*, p. 143. Irwin.
- (18) Cisco (2011). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010-2015*.
- (19) Dave Feinleib (2012). *The big data landscape*. Forbes, (2012. 6.19).
- (20) Gartner (2011). *Hype Cycle for Analytic Applications*, 2011.
- (21) McKinsey Global Institute (2011). *Big data: the next frontier for innovation, competition and productivity*. McKinsey & Company.
- (22) Mckinsey (2011). *Global institute*.
- (23) STRABASS (2012. 7. 11).
- (24) ZDNet Japan(2011. 8. 30): Economist, 2010. 2. 25.
- (25) <http://www.itworld.co.kr/news/74362>
- (26) <http://finance.yahoo.com/news/reminder-health-2-0-boston-115500670.html>
- (27) <http://www.walmartlabs.com/social-genome/>