

스마트 시대 빅데이터

기술과 주요 이슈

빅데이터가 최근 화제이다. 빅데이터는 저장되지 않았거나 저장되더라도 분석되지 못하고 버리게 되는 방대한 양의 데이터를 말한다. 실제로 빅데이터가 스마트 사회에서는 매우 많이 발생하고 있는데, 이러한 빅데이터를 분석하여 유의미한 지식을 발생 시킴으로써 가치창출을 이루고 자원을 절감하며 범죄율을 낮추고 우리의 미래를 예측하여 대처할 수 있는 수단을 제공할 수 있다는 것은 매우 흥미로운 일이다. 따라서 본 고에서는 스마트 사회의 변화와 빅데이터의 이슈를 설명하고, 빅데이터의 주요 기술과 활용사례를 살펴보고자 한다.

■ 조영임

(수원대학교 컴퓨터학과)

I. 서론

우리나라 전자정부가 2010, 2012년 2회연속 UN 평가에서 세계 1위를 차지하고 있으며, 2013년에도 1위를 달성하기 위해 노력하고 있다. 이로 인해 글로벌 전자정부 선도국 위상이 높아지고 있다.

전자정부 벤치마킹을 위해 해외 고위급 인사의 방안이 급증하고 있고 외국정부의 방문인원도 2011년 25개국 140여명에서 2012년 62개국 364명으로 증가하고 있다. 우리나라 전자정부 수출실적도 지속적으로 상승하여 2012년 12월 현재 3억불 이상 달성하였다. 수출지역도 동남아 등에서 중남미, 아프리카, 중앙아시아 등으로 확대되고 수출품목도 데이터센터, 주민등록시스템 등으로 다양화되고 있다. 또한 개도국 IT봉사단이 파견되고 정보접근센터가 구축되는 등 국가간 우호관계가 형성되고 전자정부수출확대 등 행정한류확산에 기여하고 있다[1]. 이러한 우리나라 전자정부의 약진은 수 년전부터 준비해온 스마트 사회로의 변화에 대한 대응과 노력의 결과이다.

최근에는 클라우드, SNS 등의 활성화로 방대한 양의 데이터가 중요한 이슈가 되고 있다[2]. 빅데이터로 인해 절감되거나 예측되는 상황들이 매우 효율적임은 여러 사례들을 통해서 입증되고 있다. 향후 우리나라가 자랑하는 전자정부도 전자정부 시

스템에서 발생하는 빅데이터를 효율적으로 분석하여 예측할 수 있는 시스템을 개발하기 위해 노력중이다.

따라서 본 고에서는 스마트 사회의 변화와 이로 인한 이슈 등을 조망해 봄으로써 정보화 사회에서 빅데이터의 중요성에 대해 살펴보고자 한다.

II. 스마트 사회의 변화

2000년대 초반부터 '언제 어디서나...'를 표방하는 '유비쿼터스'라는 개념이 우리나라에 도입되었고, 이후 지금까지 약 10여년간 유비쿼터스 응용기술들이 우리생활의 곳곳을 변화시켜왔다. 내가 살고 있는 아파트의 gas와 전기를 집 밖에도 제어할 수 있는 유비쿼터스 도시가 TV 광고의 등장으로 이제는 전 국민의 상당수가 '유비쿼터스'라는 단어를 일상생활에서 익숙하게 사용하게 되었다. '유비쿼터스'라는 단어는 2011년부터 '스마트'라는 단어로 바뀌기 시작하였는데, 이유는 기본 개념은 동일하나 최근 스마트폰 등 다양화되는 디바이스를 통해 그만큼 복잡하고 새로운 기술들이 개발되면서 정보화의 3.0 세대를 표방하면서 자연스럽게 변화하게 되었다[3].

유비쿼터스는 유비쿼터스 도시개발로 응용영역이 확대되면서 발전하고 있는데, 이것 역시 유비쿼터스의 트렌드 변화에

따라 '유비쿼터스 도시' 대신 'Smart City' 로 바뀌고 있다. Smart City는 유비쿼터스와 같은 첨단 정보통신 인프라와 도시 기반시설이 융합된 도시공간을 구축하여 도시 내에서 발생하는 각종 정보를 시민들에게 서비스를 제공함으로써 보다 편리하고, 안전하며, 안락한 생활을 시민에게 제공하자는 개념은 동일하나, 기술들이 좀더 복잡적이며 지능적이고 더구나 스마트폰 등의 디바이스가 중요한 매체로 작용하는 신개념의 도시를 의미한다.

Smart City를 논하기 전에 먼저, 우리나라 및 세계 정보화의 흐름을 살펴보면 다음과 같다[3,4]. 우리나라 정보화는 2002년부터 중점적으로 추진되어 왔고, 이명박 정부에서는 지식기반사회인 k-Korea 전략에 의해 최근 들어서는 스마트 코리아 전략에 비중을 두고 추진하고 있다. 국가정보화의 컨트롤 타워로써 2009년 11월 10일 국가정보화전략위원회가 출범되었고 산하 국가정보화실무위원회에서는 각 분과별 전문위원을 두어 국가정보화의 현안을 중점적으로 추진하고 있다. 국가정보화의 중점 추진방향은 고품격의 사회로의 진입을 위한 미래 성장잠재력 확충, 녹색기반, 안전하고 성숙한 사회구현 등과 같이 과거 기술개발만을 위한 추진방향과는 다소 방향이 달라지고 있다.

행정안전부의 연구보고서인 IT기반 한국사회 패러다임 변화 연구에 의하면[5], 향후 중요하게 부각될 핵심기술서비스에 대해 ①정보보호 및 보안기술 ②IT/바이오 융합기술 ③모바일 단말의 지능화 통합화 순으로 전망하였다. 물론 국내외 전문가의 의견 차이는 다소 있었으나 공통적으로 향후 IT기술서비스의 고도화와 일상화가 전망됨에 따라 정보보호 보안기술의 중요성에 대한 인식이 월등한 반면, 가상증강현실 기술과 IP-TV의 중요성은 상대적으로 낮게 나타났다. 구체적으로 보면, 국내 전문가는 IT/바이오 융합기술을, 해외 전문가는 정보보호/보안기술, 그린 IT, 모바일 단말의 지능화/통합화, 클라우드 컴퓨팅, 통합 커뮤니케이션을 미래에 가장 중요한 기술/서비스로 평가하였다.

또한 이 보고서에서 국내외 전문가들은 위의 16개 핵심 기술/서비스들이 본격적으로 활용되는 시기를 전망하였는데, 2010년은 정보보호 및 보안기술, 소셜 네트워크 서비스, IP-TV 분야, 2015년은 차세대 이동통신, 모바일 단말의 지능화/통합화, 클라우드 컴퓨팅, 센서기반 통합모니터링 기술, 스마트 그리드, 지능형 검색기술, 위치기반 서비스, 통합 커뮤니케이션 기술, 그린 IT, 사물의 인터넷 분야, 2020년은 IT/바이오 융합기술, 서비스 로봇, 가상/증강현실 기술 분야라고 제시하였다.

한편, 포레스터 리서치에서는 전 세계 그린 IT 시장은 2008년

에는 약 5억 달러에 불과하였지만 2013년에는 48억 달러로 연평균 60%이상 성장할 것으로 전망하였다. 선진국 정부 및 글로벌 IT기업들은 그린 IT시장 선점을 위한 기술 및 제품개발에 막대한 예산을 투자하고 있으며, 일본 경제산업성은 2008년 데이터 센터, 네트워크, 디스플레이 등 3대 분야의 전력효율성을 30% 이상 향상시키기 위한 그린 IT 기술개발에 30억 엔을 투자하였으며, 2009년에는 그린 클라우드 컴퓨팅 기술개발을 위해 68억 엔을 투자하였다. IBM은 그린 IT기술개발에 매년 10억 달러를 투자할 계획이며 시스코는 인천 송도 경제자유구역에 약 20억 달러를 투자하여 '시스코 글로벌 센터' 를 설립하고 지능형 도시 구축사업을 전개하는 등 그린 IT 테스트베드로 활용할 계획을 갖고 있다[6,7].

우리나라는 지난 10여 년의 지속적인 전자정부 추진노력의 결과 2년마다 평가하는 'UN 전자정부지수' 에서 2010, 2012년 연속해서 세계 1위를 달성하였는데, 이는 정부주도의 'Top-Down' 식 한국의 정보화 추진방식이 단순히 인프라 구축단계를 넘어서 서비스 분야에서까지 인정을 받은 사례라고 할 수 있다. 앞으로의 10년은 더 이상 정부주도가 아닌 지금까지와는 전혀 다른 방식의 정보화 추진이 요구되고 있다.

이상에서 우리나라 및 세계 정보화의 흐름을 살펴보았는데, Smart City의 핵심은 국가와 국가, 중앙정부와 지방정부, 정부와 기업, 기업과 개인, 개인과 개인을 이어주는 네트워크화 사회를 구현함으로써 언제 어디서나 편리한 삶을 국민들에게 제공하려는 노력들의 결정체라고 할 수 있다.

Smart City의 궁극적 비전은 기술적응용(computing paradigm), 기술적원리(technical principles), 응용도메인(application domain), 응용공간(application space)등 4가지 측면에서의 각기 다른 비전들이 상호작용하고 있다. 이러한 비전들은 infra technology, core technology, application fields의 3가지 계층구조에 의해 기술적으로 융합되어 응용될 수 있다. 따라서 Smart City는 정보의 최고 기술력들이 모두 응집되어 복합적으로 개발되는 응용분야이다.

국토해양부에서는 Smart City 건설을 위해 지난 2008년 9월 「유비쿼터스도시의 건설 등에 관한 법률 시행령」을 마련하여 유비쿼터스 도시건설의 법적 근거가 마련되었으며, 2010년 10월, 유비쿼터스 도시를 추진하려는 도시들의 활성화를 위해 개정안을 마련하여 유비쿼터스 도시 정보이용 활성화 및 산업 활성화 기반 마련 등 유비쿼터스 도시발전을 위한 제도적 기반을 마련함으로써 실제적인 도움이 되도록 지원하고자 하였다. 이로써 현재 대부분의 도시들에서 유비쿼터스도시를 추진 중이

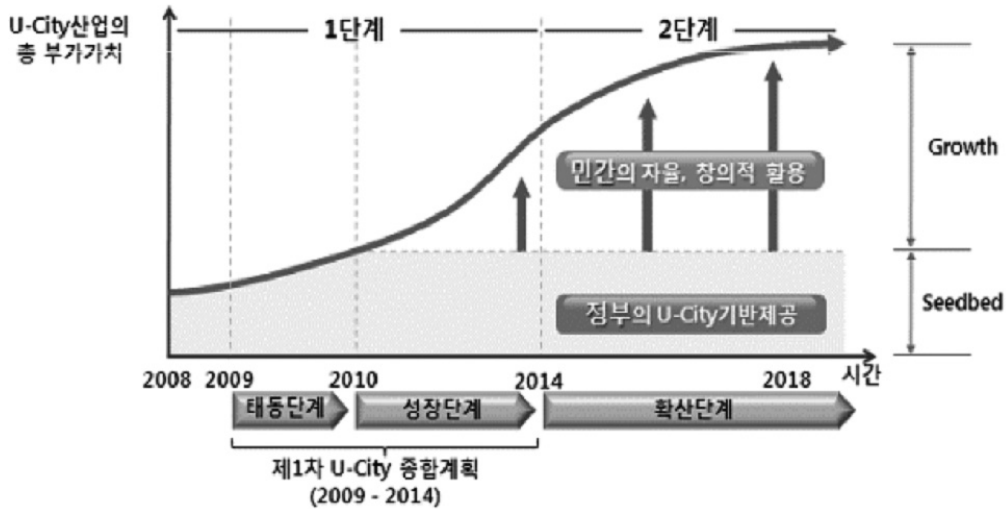


그림 1. 유비쿼터스 도시의 부가가치 전망.

며, 특히 2012년은 광명, 파주, 여수, 부산, 성남에서 유비쿼터스 도시계획을 수립하여 국토해양부에 계획을 제출하였다.

또한 국토해양부에서는 그림 1에서 유비쿼터스 도시의 부가가치를 제시하였다[8]. 즉, 2013년까지는 기반마련을 하고, 2014년부터는 기반제공을 통해 확산단계에 가겠다는 계획이다. 이를 위해서라도 앞에서 제시한 많은 정보화 기술들이 개발되고 응용되어야만 할 것이다.

이러한 국가정보화의 한 흐름에서 그림 2와 같은 Smart City의 라이프 사이클을 이해하는 것이 중요하다. 먼저, Smart City의 초반 기대치에 대해 환상을 갖게 되는데 제도적이나 기술적 지원이 뒷받침되지 않으면 회의론에 빠지게 되므로 이를 방지하기 위해 노력을 해야 한다. 또한 국가에서의 「유비쿼터스 도시의 건설 등에 관한 법률 시행령」마련 등 제도적이나 기술적 지원 노력들을 추진해야 하며, 이와 병행하여 관련 기술들의 개발이 매우 필요하다.

국내의 정보화의 흐름을 볼 때, Smart City에서 어떠한 기술들이 개발되어야 할 것인지에 대해서는 먼저 Smart City의 라이프 사이클에서의 이슈 사항들과 Smart City의 비전, 그리고 국가정보화전략위원회의 중점추진방향과 주요개념들 등을 분석해 보면 알 수 있다. 각 핵심기술 및 주요사업군에는 여러 기술과 사업들이 포함되어 있지만, 기본적으로 센서기반 서비스기술과 모바일 서비스기술, 그리고 클라우드 서비스기술로 정의할 수 있다. 이러한 기술들의 복합체로써 Smart City에 더욱 다양한 서비스와 기술들이 접목되고 개발될 수 있는 것이다. 물론 이외에도 이러한 사업군들의 실현과정에서 더 많은 기술들이 개발될 수 있을 것이다.



그림 2. Smart City의 라이프 사이클.

이와 더불어 최근 몇 년 전부터 Smart City관련 많은 R&D 연구가 진행되고 있는데, 이중 빼놓을 수 없는 중요한 이슈는 CCTV 통합관제센터의 구축에 관한 것이다. 과거에는 CCTV를 각 서비스 목적에 따라 따로따로 구축하였으나, 최근에는 CCTV를 통합적으로 구축하여 통합적인 도시 관리에 사용하려는 시도가 확산되고 있다. 따라서 CCTV 통합관제센터의 구축방식이나 운영방법 및 성과측정방법 등 CCTV를 중심으로 하는 관리방법의 중요성이 증가하고 있다[9].

각 도시들마다 범죄예방이나 교통단속 등 공공목적의 CCTV가 4년 평균 45.7%로 증가하여 2012년 현재 약 35만대가 설치되어 있다. 이중 방법이 39.6%로 가장 많고, 일반시설물 관리가 35.1%, 주정차단속 및 교통시설물관리는 14.5% 순으로 나타나는데, 정부에서는 2015년까지 전국 230개 기초 자치단체에 CCTV 통합관제센터 구축을 유도하여 Smart City를 효율적으로 운영하려고 계획 중이다. CCTV 통합관제센터에서의 이슈는 운영유지보수 미비, 고장방치, 이로 인한 CCTV 신규갯수증가,

CCTV장비들 간의 호환성이나 상호운용성 미비, CCTV의 통합 운영미비 등 근본적 문제를 갖고 있으므로 이러한 문제점들이 해결될 수 있는 기술개발들이 필요할 것이다.

이러한 기존 문제점들을 해결하고 원래의 CCTV 통합관제센터의 목적달성을 위해 국가영상정보자원 기본계획수립 등 다양한 노력들이 추진되고 있다. 정부에서는 CCTV 통합관제센터 구축 가이드라인을 2011. 2월에 배포함으로써 각 자치단체에서는 보다 효율적인 CCTV 통합관제센터 구축을 위해 노력하고 있다.

그러나 현재 CCTV 통합관제센터의 구축의 효율성 평가를 위한 방법이 개발되어 있지 않아서 사업의 중복성이 발생할 뿐 아니라 CCTV 사업의 고도화 등을 이유로 지자체의 예산이 발생하나, 그 효율성 등을 판단하기 어려워 예산수립에 있어서 어려움이 발생하고 있다. 따라서 현재 스마트 도시 관련된 많은 R&D가 진행되고 있으나, 스마트 도시에서의 CCTV 통합관제센터가 도시보안을 제대로 담당할 수 있도록 하기 위해서는 운영의 효율화를 위한 전략을 연구하고 마련하는데 또한 많은 노력을 해야 할 것이다.

Smart City 라이프 사이클에서 결론적으로, 국토해양부에서 유비쿼터스 관련 법령이나 인프라 등에 노력을 하고 있으니, 이와 더불어 스마트 관련 다양한 기술들이 개발되고 응용되어,

Smart City의 철학이 실천된다면 국민 및 기업들이 모두 다 바라는 '잘사는 사회'를 좀더 일찍 달성할 것이다.

지금까지 Smart City로의 사회변화와 주요 기술들에 대해 살펴보고자 한다.

최근에는 Smart City의 주요 기술로 빅데이터가 화두이다. CCTV에서도 여전히 많은 양의 데이터를 생산해 내고 있다. 따라서 빅데이터 이슈는 다량의 매우 많은 정보로부터 유의미한 정보를 추론하고 의미를 산출해 내야 하기 때문에 발생하는 기술이다. 앞 언급한 Smart City의 핵심은 USN에 의한 정보수집과 이를 통한 지능화의 실현이다. 즉, 주요 기술로 상황인지 (situation awareness), 자가성장(self-growing), 인간중심(human-oriented)이다. 따라서 USN으로부터 수집된 많은 정보가 빅데이터가 되고 지능화의 실현을 위해서는 이들의 분석과 유의미한 데이터의 추론이 무엇보다 중요하다. 따라서 다음 장에서는 빅데이터의 기술과 주요 이슈에 대해 살펴보고자 한다.

Ⅲ. 빅데이터의 기술과 주요 이슈

IT 기술의 일상화가 이루어지면서 하드웨어, 소프트웨어에 이어 데이터가 정보화 사회의 새로운 핵심기술로 등장하고 있다. 빅데이터로의 발전은 정보사회의 패러다임을 견인할 정도

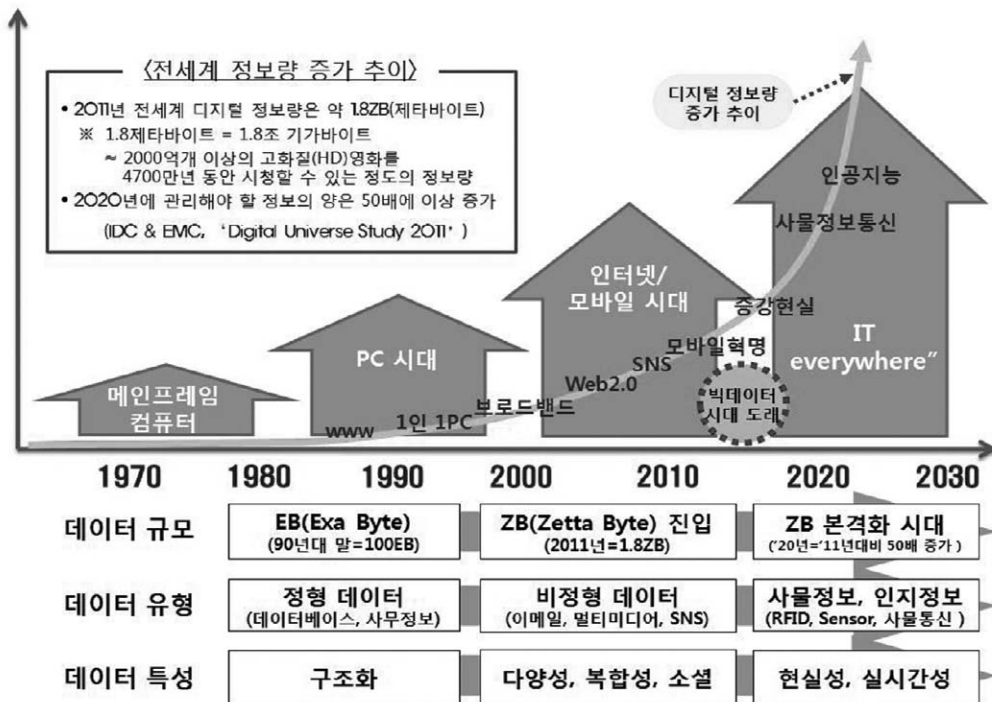


그림 3. 빅데이터의 출현.

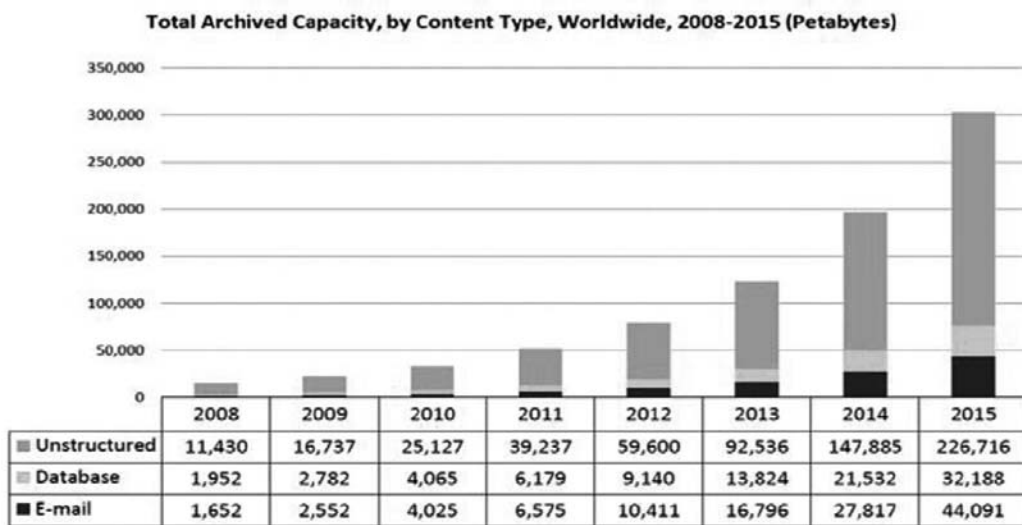
의 큰 힘을 발휘하고 있다. 스마트 단말기 확산, SNS 활성화, M2M (Machine to Machine) 네트워크의 확산으로 데이터 폭발이 더욱 가속화되어 빅데이터의 기반이 확대되고 있다. 따라서 향후 5년 이내 빅데이터를 위한 고급분석 등 관련기술은 성숙될 것으로 전망하고 있다. 다음 그림 3은 이러한 빅데이터의 특성과 함께 데이터 규모, 유형을 나타낸 그림이다[10,11].

인터넷이 일상화된 최근 10년 사이 디지털 데이터가 폭증하고 있어서 2011년 자료에 의하면 전 세계 데이터에 생성될 디지털 정보량이 1.8ZB에 달하고 있다고 한다. 정보량은 기하급수

적으로 증가하여 2020년에는 관리해야할 정보량이 50배 급증하고 10배 많은 서버가 필요할 것이라고 전망되고 있다. 기존의 데이터 관계된 기술로는 감당할 수 없는 규모로 성장하게 된다.

데이터의 유형 또한 매우 다양하게 변화하고 있는데, 그림 4에서 보는 바와 같이 소셜 데이터외에 이메일, 동영상 등과 같은 비정형 정보들이 향후 10년동안 생성되는 전체 데이터의 약 90%에 달할 것이라 전망하고 있다[12].

따라서 빅데이터는 단순히 규모가 크다는 것 외에 몇 가지 특성을 더 갖고 있다. 즉, 4V로 요약되는데, 먼저 규모(volume)로서



Source: Enterprise Strategy Group, 2010.

그림 4. 데이터의 유형의 변화.

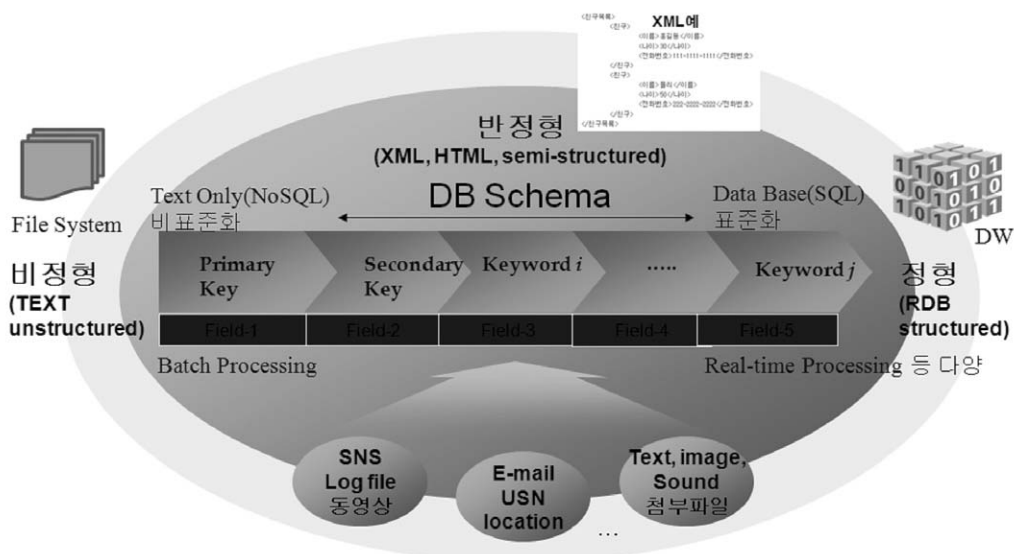


그림 5. 정형과 비정형의 차이.

이는 앞서 언급한 내용과 같다. 둘째는 다양성(velocity)으로서 로그기록, 소셜, 위치, 소비, 현실데이터 등 데이터의 종류가 증가되어 텍스트 이외의 멀티미디어 등 비정형화된 데이터의 유형이 다양화 되었다는 것을 의미한다. 셋째는 복잡성(complexity)으로서 구조화되지 않은 데이터, 데이터 저장방식의 차이, 중복성 문제 등 데이터 종류가 확대되고 외부 데이터의 활용 등으로 관리대상이 증가됨으로써 점차적으로 데이터 관리 및 처리가 복잡화되고 심화되어 새로운 처리 및 관리기법이 요구되는 상황을 의미한다. 넷째는 속도(velocity)로서 센서나 모니터링 등 사물정보, 스트리밍 정보 등 실시간성 정보가 증가하고 있고, 이러한 실시간성으로 인한 데이터 생성, 이동과 유통의 속도가 증가하고 있으며 대규모 데이터처리 및 가치있는 실시간성 정보활용을 위해 데이터 처리 및 분석 속도가 매우 중요하게 되었음을 의미한다[12].

정형과 비정형은 다음 그림 5에서 보는 바와 같이 DB로 정의하느냐 일반적인 파일시스템 형태로 유지하느냐의 차이로 보면 된다. DB와 파일의 차이는 저장과 분석을 위한 기술이 적용되느냐 단순히 저장기술만 적용하느냐의 차이로 정의된다.

최근 가트너는 그림 6에서 보는 바와 같이 2011년 이머징 기술전망에서 빅데이터를 새롭게 포함시키고 앞으로 주목해야할

기술로 소개한바 있다. 이 자료에 의하면 빅데이터는 현재는 기술발생단계(technology trigger)이나 향후 5년내에 성숙될 것이라 전망하였다[12,13]

빅데이터가 왜 이슈가 되는가를 보면, 데이터분석의 목적이 저장의 필요성 보다는 대증의 경향과악이 우선일 때인데, 저장에 따른 오버헤드를 최소화하고 분석결과를 활용하려는 경우에 적용한다. 또한 데이터자체가 저장을 체계적으로 해야 할 만큼 중요도가 높지 않거나 많을 때 이용하는데, SNS, log 파일 등의 중요도가 다른 DB 스키마 정보에 비해 높지 않으므로 굳이 정형화할 필요성이 없을 때 적용한다. 또한 데이터의 신뢰성이 그다지 높지 않을 때 필터링할 필요성 있을 때 적용한다.

빅데이터는 IT의 흐름이 하드웨어, 소프트웨어에 이어 데이터로 이동하면서 독립적인 주체로 발전하기 시작하였다. 사용되는 정보형태의 변화, 정보관리 및 기대치의 변화에 원인이 있으며, 유비쿼터스에서 스마트 사회로 변화되면서 기술적, 감성적 요구사항 증대되며, 단순 분석이나 집계보다는 분석과 추론의 중요성을 강조하는 방향으로 진화하고 있어서 데이터의 의미와 가치에 더 많은 관심을 가지기 시작하면서 발전하기 시작하였다[13,14]. 또한 스마트 사회 도래에 따른 디바이스의 다양화, 응용의 다양화로 인해 생산되는 데이터량이 기하급수적으

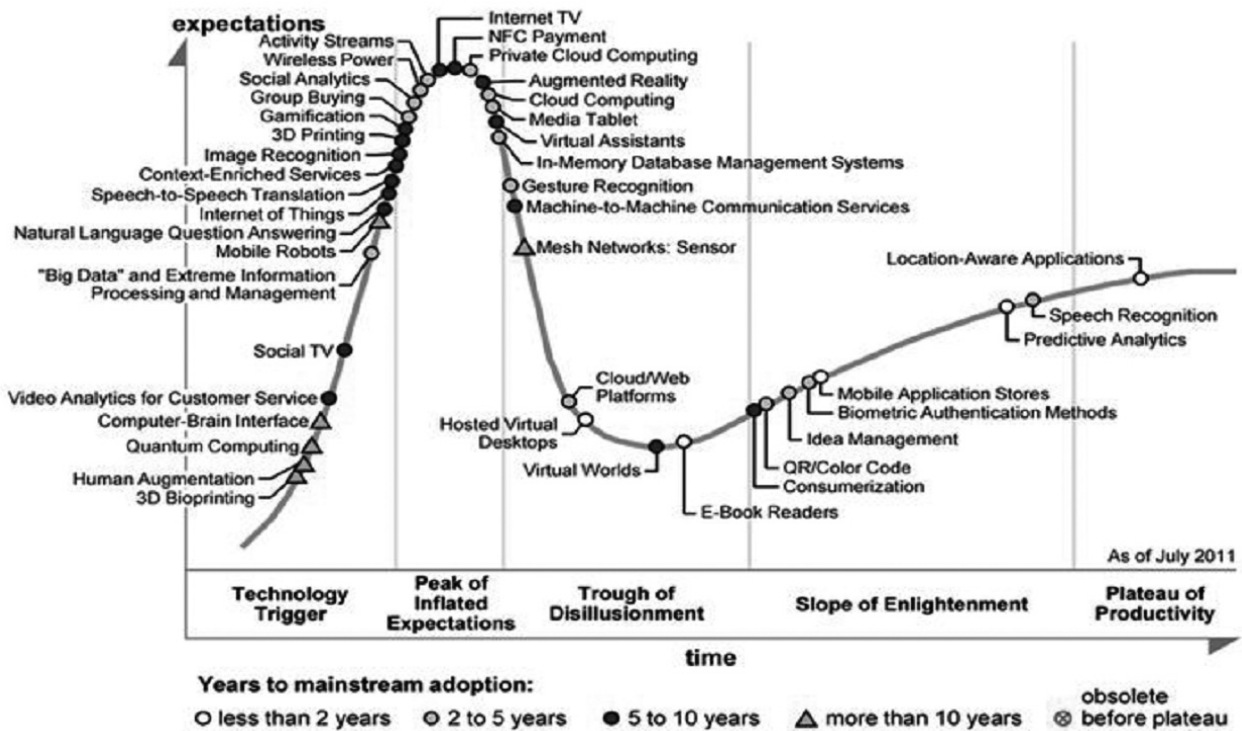


그림 6. 가트너의 이머징 기술 hype cycle.

로 증대함으로써 처리해야 할 데이터범위와 대상도 변화되는 것도 발전의 원인인데, 매일 250경 바이트 비정형 데이터발생, 매달 300억여개 페이스북 메시지와 10억개의 트윗이 생성되고 이러한 비정형 데이터의 처리가 시급하게 되었다. 또한 빅데이터의 엄청난 잠재력으로 인해 사회경제의 원동력임을 인식하게 되어 향후 공공, 민간의 성패를 좌우할 원천임을 인식하고 가치인정하기 시작하면서 발전하게 되었다. 또한 시간이 지날수록 디바이스의 수는 점점 늘어나고 디바이스의 크기는 점점 작아지기 때문에 센싱해야 할 정보의 양이 점점 많아질 수밖에 없기 때문이며, USN 등장으로 소규모 네트워크 형태가 점점 증가하고 있기 때문이다.

앞서 설명한 바와 같이 빅데이터는 많은 데이터들의 축적과 공유를 통해 유의미하게 분석할 수 있어서 상황을 분석하고 추론하고 전망할 수 있도록 진화되고 있다. 이러한 빅데이터는 사회 환경변화를 신속하게 감지하고 대응하는 역량과 스마트한 지능형 및 개인화 서비스를 창출하는 원천요소로 작용하고 있다. 즉, 과거에는 DB를 통한 데이터 축적이 중요하고, 이로부터 활용을 위한 정보검색엔진이 활발히 연구되었다. 이후 지식관리시스템(KMS, Knowledge Management System)과 웹 2.0을 통해 데이터가 확산되고 공유되기 시작하였다. 최근에는 매우 많은 데이터들의 분석을 통해 추론하고 상황인지 함으로써 새로운 가치를 창출하는 것이 중요한 이슈가 되고 있다.

이미 잘 아는 바와 같이 정보화 사회는 매우 급변하게 변화하고 있고 과거 시스템 중심에서 사람중심으로 변화되고 있으며, 소셜네트워크의 파워가 매우 증가하고 있다.

이러한 시대에 빅데이터는 미래사회의 특성인 불확실성, 리스크, 스마트, 융합 등 미래사회의 특성에 대응하는 역할을 수행하며 기회요인을 창출하는 핵심엔진으로 작용하고 있다 [11,12,13,14].

먼저 불확실성에 대한 빅데이터의 역할은 통찰력을 제공하는데 있는데, 주로 사회현상과 현실세계의 데이터를 기반으로 패턴분석과 미래전망을 제공하고 있고, 여러 가지 가능성에 대한 시나리오 시뮬레이션과 다각적인 상황이 고려된 통찰력을 제시하고 있다. 또한 다수의 시나리오로 상황변화에 유연하게 대처하게 한다.

두 번째 특징인 리스크에 대해서는 대응력을 제공하고 있다. 빅데이터는 환경, 소셜, 모니터링 정보의 패턴분석을 통한 위험징수, 이상신호를 포착하고 이슈를 사전에 인지하여 분석하고 빠른 의사결정과 실시간 대응지원을 하고 있다. 또한 기업과 국가경영의 투명성을 제고하여 낭비요소를 절감하게 한다.

세 번째 특징인 스마트에 대해서는 경쟁력을 제공하고 있다. 주로 대규모 데이터분석을 통한 상황인지, 인공지능 서비스 등이 가능하도록 하고 있다. 여기에는 개인화 및 지능화 서비스를 제공할 수 있도록 확대하고 있고, 소셜 니즈분석과 평가, 신용, 평판 분석을 통해 최적의 선택이 지원되도록 하고 있다. 또한 트렌드 변화분석을 통한 제품경쟁력을 확보할 수 있도록 한다.

네 번째 특징인 융합에 대해서는 창조력을 제공하고 있다. 타 분야와의 결합을 통한 새로운 가치를 창출할 수 있도록 즉, 의료정보, 자동차정보, 건물정보, 환경정보 등을 제공할 수 있도록 하고 있다. 또한 인과관계, 상관관계가 복잡한 컨버전스 분야의 데이터 분석으로 안정성을 향상시키고 시행착오를 최소화 할 수 있도록 한다. 또한 방대한 데이터 활용을 통한 새로운 융합시장 창출이 가능하도록 한다.

이러한 빅데이터의 역할이 성공을 거두려면 공공과 민간부문이 통합된 데이터 분석을 위한 플랫폼 개발이 필요하며, 다학제적 이해와 통합적 사고와 직관력을 갖춘 데이터 사이언티스트의 양성이 필요하고, 개인 프라이버시를 위한 기법의 도입과 데이터 자원의 결합과 협력축진을 위한 신뢰기반형성이 필요하다[15].

IV. 빅데이터 플랫폼

빅데이터에서 플랫폼 개발이 매우 중요한데, 플랫폼에 대한 기본형을 살펴보고자 한다. 최근 빅데이터 플랫폼은 특성이 복잡하기 때문에 NP-Hard 문제에 귀결될 수 있다. 따라서 이러한 NP-Hard 문제에 대한 해결책으로 인공지능 기술이 대두되고 있다[10,13]. 따라서 빅데이터 플랫폼에 인공지능기술이 접목되면, 센싱정보로부터 학습, 추론, 인지 등의 기본적 인공지능 과정을 통해 분석결과를 해석하고 의사결정 할 수 있는 지능을 가진 인간과 유사한(human-like platform)을 구축할 수 있는 것이며, 이러한 기술개발이 빅데이터 플랫폼의 기술관건이 되고 있다.

빅데이터 플랫폼은 민간보다는 공공분야에서 공공분야정보의 공유를 위해 더 필요하다[14]. 최근 스마트 사회에서 빅데이터는 이러한 플랫폼을 통한 데이터개방을 필요로 하고 있다. 미국은 데이터 생성부터 공개까지 효율적인 플랫폼과 관리체계가 구축되어 가장 활발한 데이터 공개를 하고 있다. 미국은 원천데이터를 제공하는 서비스 형태가 sheet, chart, map, apps, open API이며, 통계와 지리정보 데이터를 제공하고 있고, link와 다운로드가 가능한 데이터 파일을 공개하고 있다. 영국은 미국에 비해 상대적으로 정보공개현황의 수준이 미흡하나 원천데이터를

중심으로 데이터를 공개하고 있는 점이 차이점이나 미국과 마찬가지로 데이터를 다운로드할 수 있는 데이터 파일형태로 제공하고 있다. 호주도 다운로드 할 수 있는 데이터 파일형태로 제공하고 있다. 그러나 우리나라는 원천데이터의 공개는 이루어지지 않고 있으며 open API 형태로 가공된 프로그램만을 제공하고 있으므로 개발된 데이터의 양이 절대적으로 부족하고, 공개를 위한 자원소모가 상대적으로 크다는 점이 다른 나라들에 비해 단점으로 보인다. 이로 인해 공공정보의 활용율이 매우 낮다.

따라서 공공정보의 활용율을 높이고 빅데이터의 효율적 분석과 활용성을 위한 플랫폼의 기본구조와 기능들에 대해 설명

하고자 한다.

Human-like 기능의 빅데이터 플랫폼은 다음 그림 7에서와 같이 설명된다. 즉, 빅데이터 처리인프라에 기반하여 데이터 수집/통합에서 데이터 전처리단계를 거쳐 데이터저장/관리기술로 이어지고 데이터 분석 및 데이터분석 가시화 단계로 구성된다.

먼저 빅데이터 수집 및 통합은 새로운 데이터 생성, 네트워크에 산재해 있는 외부데이터 수집, 내외부 이종데이터 통합 등 데이터의 형태와 소재에 무관하게 데이터를 확보하는 기술을 말한다. 과거 기술적 가능성, 분석성능의 미비, 분석결과와 타당성 등을 이유로 등한시했던 데이터들을 빅데이터로 명명하고 이를 분석하면서 통계분석기법, 인공지능적 기법, machine

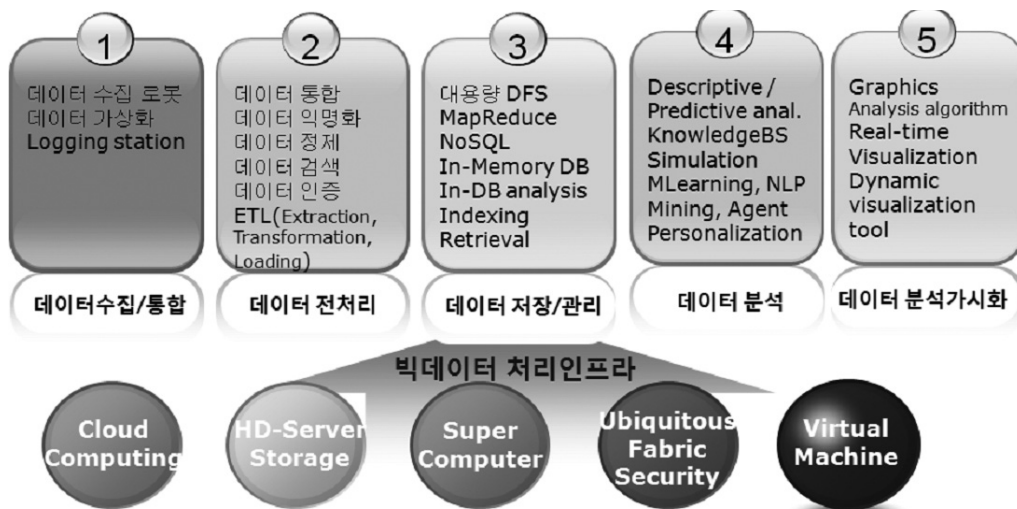


그림 7. Human-like 빅데이터 플랫폼.

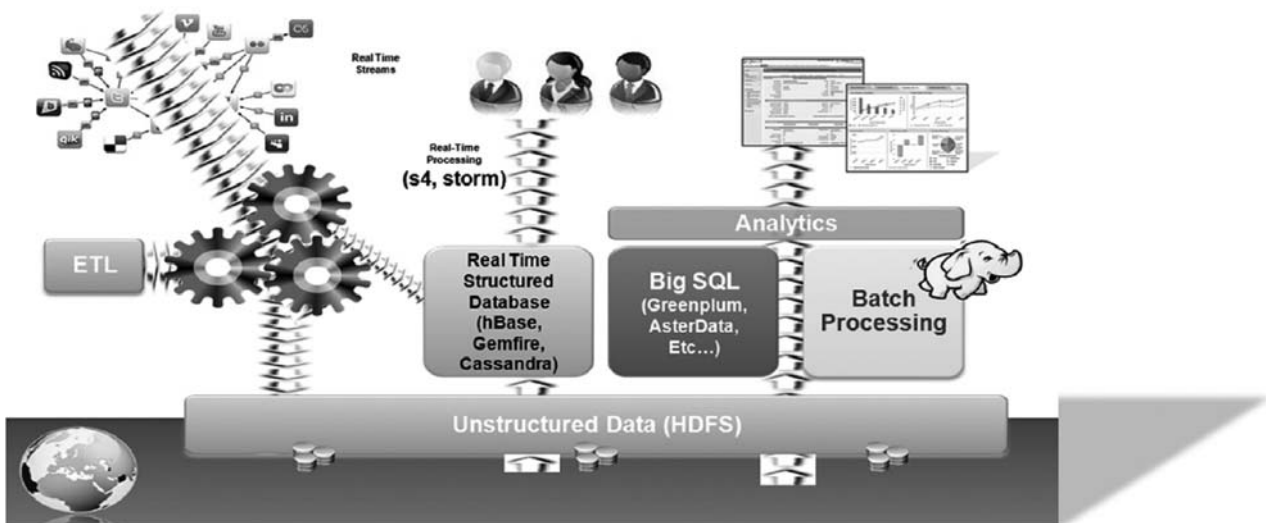


그림 8. 빅데이터 플랫폼의 흐름.

learning 기법들을 총체적으로 적용하면서 연관성을 찾아 기업의 경쟁력 강화, 기업의 수익, 소비자 성향파악에 주력하는 경향이 있다.

데이터 전처리에는 센싱정보, SNS 등 지속적으로 발생하는 비정형 스트림 데이터를 정제하여 분석 가능한 형태로 구조화하여 분석의 정확성을 높이고 심층분석을 가능하게 하는 기술을 말한다.

데이터 저장/관리 기술은 웹 데이터, 소셜미디어, 비즈니스 데이터, 센싱정보 등의 증가하는 다양한 형식의 데이터를 실시간으로 저장 및 관리할 수 있는 분산 컴퓨팅 기술을 말한다. 이는 빅데이터 플랫폼의 핵심역할을 수행하고 있다. NoSQL (Not Only SQL)은 Relational DB (RDB)를 넘은 빅데이터 저장을 위한 새로운 DB개념으로 다양한 형태의 인터페이스 제공하고 있으며 비표준화 상태이고 종류로는 MongoDB, Cassandra, Hbase 등이 있다.

데이터 분석기술은 빅데이터에 내재된 가치를 추출하기 위해 필요한 대규모 통계처리, 데이터 마이닝, 그래프 마이닝 등의 분석기술, machine learning, artificial intelligence 기술을 활용한 다양한 심층분석 기술로서, 빅데이터 플랫폼의 핵심 중 핵심 기술이다. 따라서 이러한 이유로 최근에는 데이터 사이언티스트의 양성이 필요하다.

데이터 분석 가시화 기술은 비전문가가 데이터분석을 수행할 수 있는 환경을 제공하는 분석도구기술과 분석결과를 함축

적으로 표시하고 직관적인 정보를 제공하는 인포그래픽스 기술을 말한다. '빅쿼리'는 2011년 11월 구글의 빅데이터 분석 솔루션 플랫폼으로, 오픈소스 데이터 분석 툴인 Hadoop을 활용해 신속히 분석하여 시각화할 수 있는 특징이 있다.

다음 그림 8은 이러한 기능을 갖는 빅데이터 플랫폼을 개략적인 흐름으로 나타낸 그림이다.

앞으로 빅데이터 플랫폼을 지능화하여 융통성 있게 개발함으로써 AI와 빅데이터가 상호 시너지효과를 발휘할 수 있도록 하고 결과물의 신뢰성을 향상시켜야 한다. 기존 고객관리나 자료관리기술의 단순 확장 또는 기존 검색기술의 단순확장 및 적용이 아니라 창의적 활동이 포함되는 Human-like 플랫폼 개발 필수적이다.

지금까지 많이 사용되는 Hadoop은 저장/관리위주 분산처리 tool로서 제한적 기능만 제공하는 낮은 수준의 플랫폼 역할을 수행하고 있다. 따라서 Hadoop 초기버전을 참고로 차후로는 Human-like platform 개발이 필수적으로 수행되어야 한다.

따라서 향후로는 다음 그림 9와 같은 공공정보와 민간정보의 융합을 통한 정보공유체제를 갖추어야 할 것이다.

V. 빅데이터 활용사례

빅데이터의 활용사례는 매우 많다. 미국의 오바마 대통령 선거시 SNS 분석을 통한 예측과 2011년 서울시장 선거에서 박원

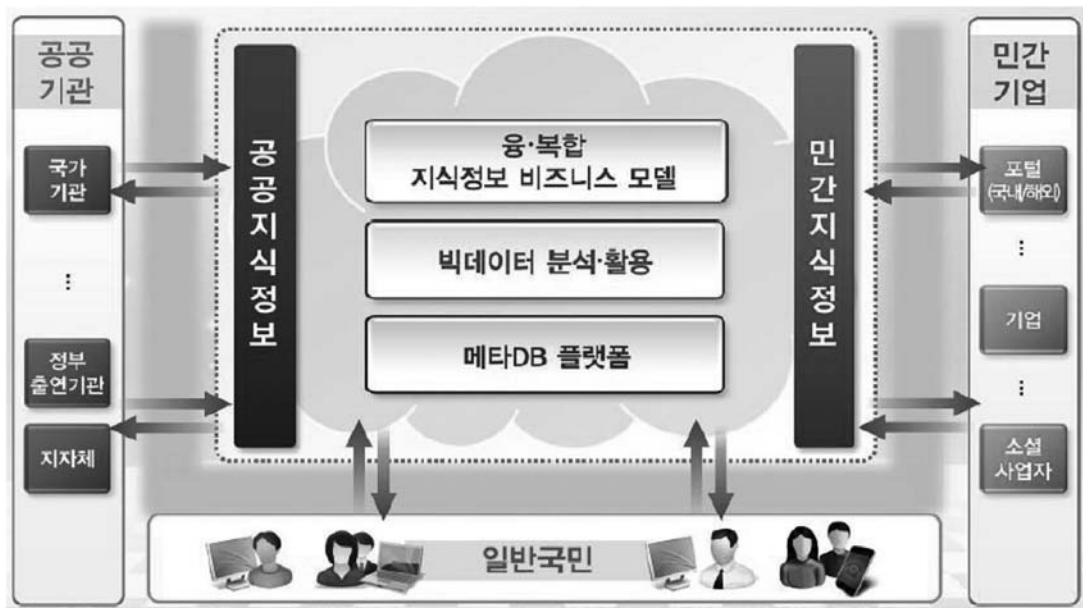


그림 9. 공공과 민간의 빅데이터 공유 형태.

순 서울시장 당선을 정확히 예측한 사례는 이미 많이 알려진 사실이다[16].

한국 SAS가 발표한 사례를 소개하면, 석유시추선은 매초 엄청난 석유를 생산하며 연간 36억 달러의 원유를 생산한다고 한다. 이 거대한 석유시추선을 구성하는 모든 부품 하나하나마다 소량의 데이터를 계속 생산해 내고 있는데, 많은 부품들에서 엄청난 양의 데이터를 만들어내고 있다. 그러나 장애발생시 대체 부품을 구하고 수리하는데 5일이 걸린다고 하고 수리하는 동안 600억원의 손실이 발생한다고 한다.

그런데 이러한 피해를 엄청난 양의 데이터를 분석함으로써 매년 7억달러를 줄였다고 한다. 즉, 하나하나의 부품마다 쏟아내는 소량의 데이터를 분석하면 장애발생 즈음하여 특정부품이 정후를 나타내며 48시간내 고장날 확률이 80%라는 것을 알 수 있었다. 따라서 이러한 빅데이터 분석을 통해 생산량을 5% 늘리고 고장으로 인한 정지일이 80%로 감소하게 되어 결국 연간 7억달러를 감소할 수 있었다고 한다.

또 다른 예로 구글의 미국내 독감발생추이 예측을 들 수 있다. 미국 질병관리국내 인플루엔자 관리국 소속 역학예방법팀은 3000여개 진료소, 3000여명의 외래 진료환자 진단의, 122개 도시의 인구통계사무소, 각 주 보건국 관계자들로부터 모은 자료로 독감환자의 수의 추이를 집계 및 추정하였다[12]. 이를 위해

미국은 많은 시간과 비용이 들었다. 그러나 구글 트렌드에 '독감'을 검색어로 입력하면 그림 10과 같은 추이가 분석되어 나오는데, 실제로 미국 질병관리국이 분석한 자료와 매우 비슷한 트렌드를 예측하였음을 알 수 있다. 따라서 굳이 많은 시간과 돈을 들여 추이를 집계할 것이 아니라 빅데이터 분석을 통한 분석이 효과적임을 알 수 있다.

또 다른 사례를 들면, 말콤 블래드웰의 '깨진 창문' 이론 또는 '상황의 힘' 이론에 의해 범죄충동은 사람의 개성보다는 환경과 상황에 따른다고 한다. 예를 들면, 어떤 집 창문이 깨진 채로 방치되면 '그 집은 사람이 없구나, 다른 창문도 깨도 되는구나, 들어가도 되는구나, 물건을 갖고 나오자' 등과 같이 범죄가 일어난다는 것이다. 도저히 범죄를 생각할 수 없을 정도로 지하철이 깨끗하면 범죄가 발생하지 않는다는 것이다. 이는 IT의 활용을 통해서도 가능한데, GIS를 활용해 살인 등 강력사건이 자주 발생하는 지역에 집중적으로 가로등, CCTV를 설치하고 깨진 창문을 모두 없애고 순찰을 강화하였더니 살인건수가 60%이상 줄었다고 한다.

이상의 몇 가지 예를 보더라도 빅데이터를 통한 절감사례나 효율성은 많은 곳에서 나타나고 있다고 볼 수 있다.

가트너(Gartner)는 2013년 전 세계 기업 IT 지출이 2012년 2조 6,030억 달러 대비 2.5% 증가한 2조6,790억 달러에 달할 것으로

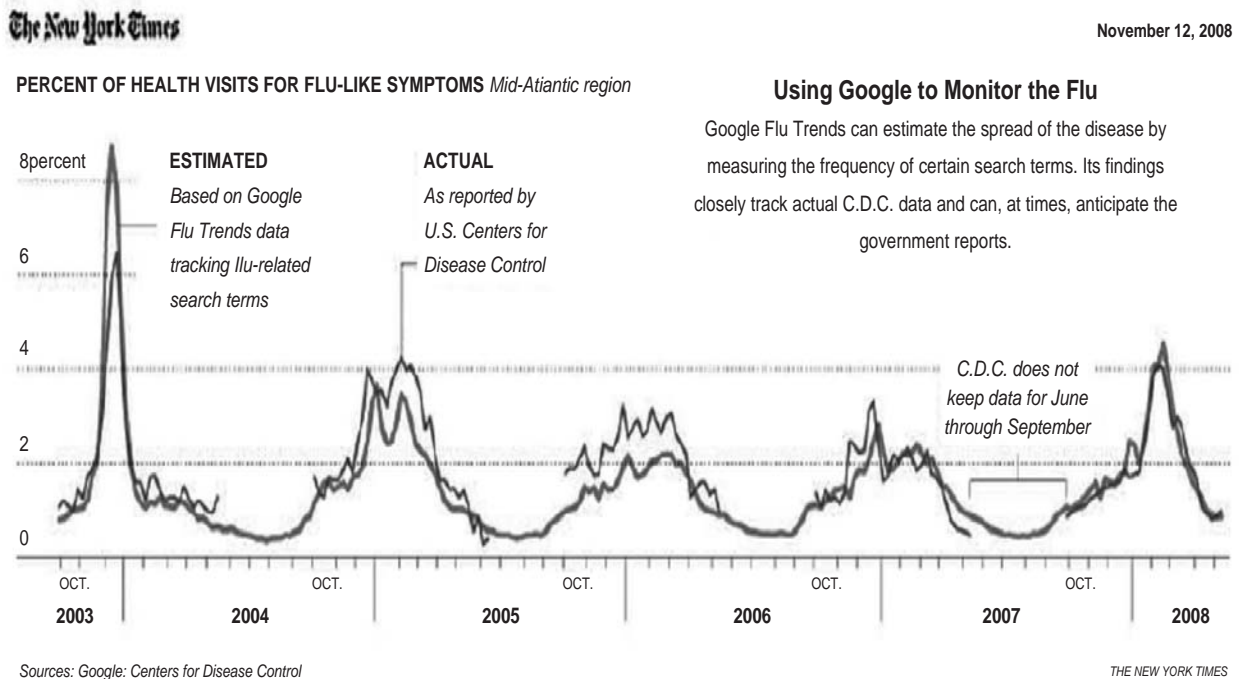


그림 10. 구글 독감동향과 실제 확산의 비교결과.

전망하였다[6,11]. 산업별로는 금융, 통신 미디어 서비스(CMS), 제조업 분야 등이 2016년까지 가장 높은 성장률을 기록할 것으로 관측하였다. 2012년에는 경기 불안으로 인해 기업 IT 지출이 전반적으로 위축되었으나, 미국과 유럽에서 심각한 금융 위기가 발생하지 않는다면 2013년 기업 IT 지출은 상당 수준 회복될 것으로 전망하였다. 제조업과 천연자원 분야의 2013년 IT 지출은 2012년 4,670억 달러 대비 2.3% 성장한 4,780억 달러에 달할 것으로 예상되고 있다. 제조업체들의 경우 특히 실적 부진에 대비, 상당 부분의 IT 비용을 매출과 연동해 관리하며, 이로 인해 전 세계 제조업체들은 2008년 글로벌 경제위기 이후 IT 구매 축소되었다. 그러나 소셜, 모바일, 빅데이터, 클라우드와 같은 새로운 기술력의 결합(a nexus of new technology forces)으로 인한 가치 창출을 기업 리더들이 적극 지원하게 되면서 제조업 IT 지출은 지속적 증가할 전망이라고 하였다.

VI. 결론

스마트 사회에서 빅데이터는 필수불가결하게 발생하는 요소이며, 이것을 어떻게 잘 활용하느냐가 스마트사회의 성과를 좌우한다고 해도 과언이 아니다.

우리나라가 자랑하는 전자정부도 빅데이터의 활용으로부터 효율성을 높일 수 있도록 패키지화 하는 것이 향후 지속적인 경쟁우위를 점하는 요소가 될 것이다.

이제는 버리는 데이터가 하나도 없이 다 모아서 분석하여 유의미한 지식을 생산함으로써 가치창출을 이루는 정보화 사회를 이루어나가야 할 때가 도래한 것이다.

참고문헌

- [1] 주요정책 과제 추진실적 보고서, 행정안전부, 2012.11.21
- [2] 신가치창출 엔진, 빅데이터의 새로운 가능성과 대응전략, 한국정보화진흥원 IT&Future Strategy, 2011.12.30
- [3] 2012 국가정보화백서, 행정안전부, 2012.8.31
- [4] 스마트시대 국가발전전략, 행정안전부, 2011.10.26

- [5] IT기반 한국사회 패러다임 변화연구, 행정안전부, 2009.12.31
- [6] 2012년 IT트렌드 전망 및 정책방향, IT정책연구시리즈, 한국정보화진흥원, 2011.12.29
- [7] 경제사회 발전을 이끄는 IT생태계 전략, IT정책연구시리즈, 한국정보화진흥원, 2011.12.2
- [8] U-City IT 인프라구축 세부 가이드라인, 한국정보화진흥원, 2009.11
- [9] 조영임, 통합관제센터 성과모형 및 성과측정지표 개발, 2012.1
- [10] 모바일 시대를 넘어 AI 시대로, 한국정보화진흥원 IT&Future Strategy, 2010.8.25
- [11] 빅데이터 확산에 따른 도전과 기회, 최신 IT동향, 정보통신 산업진흥원, 2011.5.31
- [12] 빅데이터로 진화하는 세상, 한국정보화진흥원, 빅데이터연구소센터, 2012.5.31
- [13] 빅데이터 시대, AI의 새로운 의미와 가치, 한국정보화진흥원 IT&Future Strategy, 2012.7.25
- [14] 빅데이터 시대 : 효과적인 공공정보 개방을 위한 데이터 플랫폼 구축방향, 한국정보화진흥원 IT&Future Strategy, 2012.9.28
- [15] 빅데이터 시대의 인재, 데이터 사이언티스트의 역할과 가능성, 한국정보화진흥원 IT&Future Strategy, 2012.8.8
- [16] 권대석, 클라우드와 슈퍼컴퓨팅이 이끄는 미래 : 빅데이터 혁명, 21세기북스, 2012.10.8

저자약력



조영임

- 1988년 고려대학교 컴퓨터학과 졸업.
- 1990년 동 대학원 석사. 1994년 동 대학원 박사.
- 2000년 Univ. of massachusetts post-doc.
- 1996년 삼성전자 선임연구원.
- 현재 수원대학교 컴퓨터학과 교수.
- 관심 분야 : 유비쿼터스 시스템, 전자정부, 인공지능, 정보검색, 빅데이터 등