

빅 데이터 시대의 한국 갈라파고스가 되지 않으려면

손민선 책임연구원 msson@lgeri.com
문병순 선임연구원 psmoon@lgeri.com

- I. 빅 데이터, 생각하는 기계의 시대를 열다
- II. 지능형 서비스가 온다
- III. 빅 데이터에서 지능형 플랫폼으로
- IV. 한국은 지금 데이터 갈라파고스
- V. 한국형 십년지대계가 필요하다

최근 애플의 시리나 IBM의 왓슨처럼 사람의 말을 이해하고, 그를 바탕으로 대안을 제시하는 ‘생각하는 기계’가 화제가 되고 있다. 한 때 잊혀졌던 인공 지능 기술이 빅 데이터를 기반으로 재부상하고 있는 것이다.

빅 데이터 기술은 지금까지 처리할 수 없었던 대용량의 비정형 데이터를 처리하는 기술로서, 데이터를 기반으로 한 지능형 서비스를 구현하는 기반 기술이다. 실시간으로 수집되는 데이터와 그를 분석하는 시스템을 기반으로 한 지능형 서비스는 머지 않은 장래에 서비스의 제공 체계나 그를 제공하는 전문가의 역할과 가치를 바꿔놓을 것이다.

이러한 상황에서 IT 거대 기업들은 그들의 축적된 데이터, 노하우, 대규모의 컴퓨팅 인프라가 결합된 빅 데이터 솔루션과 서비스를 내놓고 있다. IBM의 왓슨은 지능형 서비스 솔루션 시장을 열고 있으며, 구글은 자사의 클라우드 서비스를 기반으로 한 빅 데이터 서비스를 준비 중이다. 이러한 서비스의 종착점은 다양한 분야의 지능형 서비스를 구현하는 지능형 플랫폼이 될 것이며, 이 플랫폼은 지금의 OS와는 비교가 되지 않을 정도로 강력한 시장 지배력을 갖게 될 것이다.

이에 비해 한국은 데이터 기반의 사업 환경 자체가 미성숙한 상황이다. 제한된 내수 시장 규모, 언어적 한계, 글로벌 서비스 업체가 없는 시장 구조, 데이터 활용이 까다로운 규제 환경 등이 맞물리면서 데이터 산업의 갈라파고스로 전락할 가능성도 농후하다. 뛰어난 IT 인프라를 가지고도 인터넷 시대의 변방으로 머물렀던 아픔을 또 겪지 않으려면 더 이상 머뭇거릴 시간이 없다. IT를 넘어 각 분야의 서비스 경쟁력이 모두 융합된, 한국형 빅 데이터 마스터플랜이 필요하다. ■

“연산하는 기계를 넘어
생각하는 기계가
현실화되고 있다.”

“나 지금 기분 꿀꿀한데, 재밌게 좀 해줘 봐.”

스마트폰에 대고 이렇게 말하면 최신 유머나 좋아하는 연예인의 소식을 알려주면 어떨까? 아니면 평소 미워하던 사람을 때리면서 스트레스를 푸는 게임을 띄워준다면? 애플의 음성 인식 서비스인 시리가 화제가 되면서 한번쯤 상상해 보았을 만한 상황이다.

기기를 음성으로 제어하는 자연어 인터페이스는 최근 IT 업계의 가장 큰 관심거리다. 이렇게 되면 ‘심심해’, ‘우울해’와 같이 단짝 친구한테나 할 말을 스마트폰에 하는 상황도 어색하지 않아질 것이다. 인간이 기계에게 고민을 토로하고, 또한 기계가 인간을 위로하는 시대를 향하고 있는 것이다.

이러한 미래를 단순히 음성 인식 기술의 쾌거라 말하는 것은 조금 모자라다. 음성 인식은 기기 조작 방식의 변화로서 아이콘을 누르는 대신 말을 하는 것에 불과하다. 그보다 더 중요한 것은 ‘기분이 꿀꿀하다’는 것이 무엇이며, 어떤 이야기로 그 마음을 달랠 수 있는 지 ‘이해하는’ 힘이다. 연산하는 기계를 넘어 생각하는 기계의 등장이다.

I. 빅 데이터, 생각하는 기계의 시대를 열다

빅 데이터로 거듭난 인공지능

생각은 어떻게 가능한 것일까? 생각의 출발점은 정보다. 갓난 아이가 성인이 되어 복잡한 사고를 할 수 있게 될 때까지, 사람은 감각과 언어 기관을 통해 수집된 자극에 반응하면서 이 정보를 뇌 속에 저장한다. 새로운 정보를 과거의 정보와 연결하면서 개념이 생겨나고, 이러한 개념이 발전하면 다단계의 사고로 이어진다.

물론 동물도 감각을 통해 주변 정보를 받아들여 뇌를 통해 이해하고 반응한다. 하지만 동물의 뇌활동을 사고라 칭하지 않는 것은, 정보를 저장하고 그것을 시간과 인과 관계에 따라 배치하면서 의미를 찾아내는 활동의 비중이 극히 낮기 때문이다.

“ 잊혀졌던 인공지능
기술이 빅 데이터와
함께 재부상하고
있다. ”

동물에게는 그 순간의 정보와 그 순간의 반응이 중요하다. 하지만 인간에게는 시간에 따라 정보를 배열하여 저장하는 기억력이 있고, 저장된 정보를 유형화하면서 의미를 찾아내는 분석력이 있다. 이로 인해 하나의 정보에 담긴 복합적인 의미를 이해할 수 있을 뿐 아니라, 앞으로 일어날 일을 상상하고 예측하는 추상적 사고도 가능하다.

기계의 생각도 마찬가지라는 것이 인공 지능 기술의 가장 기본적인 가정이다. 인간의 사고 과정과 유사한 프로그램을 만들면 기계도 마치 사람처럼 복잡한 의미를 이해하고, 새로운 사실을 학습할 수도 있다는 것이다. 나아가 상황에 가장 적합한 대응 방안을 생각해 내는 것은 물론이고, 그것이 가져올 결과도 예측할 수 있다는 것이다. 이러한 개념은 컴퓨터의 발명과 역사를 거의 같이 하는 오래된 개념 중 하나지만, 몇 년 전까지만 하더라도 거의 잊혀졌던 것이 사실이다. 인간의 사고 과정이 너무 복잡하고, 이런 작업을 수행할 만한 성능을 가진 컴퓨터를 구현하기가 불가능했기 때문이다.

그러나 최근 등장한 빅 데이터 기술과 함께 인공 지능이 다시 부상하고 있다. 정보의 양과 처리 능력이 제한되었을 때 기계는 마치 동물처럼 그 순간 주어진 문제를 해결하는데 그친다. 어떤 생각도 거치지 않고 나오는 반사 작용처럼, 시스템에 프로그램된 방식으로만 정보를 처리하는 연산 기계일 뿐이다. 하지만 엄청나게 많은 정보를 수없이 다양한 방법으로 분석할 수 있게 된다면 어떨까?

빅 데이터는 지금까지 이해할 수 없었던 정보를 이해하고, 분석할 수 없었던 방대한 양의 정보를 분석하는 기술이다. 불가능할 것으로 생각되었던 음성 인식이나

자동 번역이 빅 데이터를 기반으로 현실화되면서 IT 업계는 다시, 생각할 수 있는 기계를 꿈꾸고 있다. 이미 현실화되기도 했다. IBM에서 개발한 지능형 컴퓨터인 왓슨이 대표적이다. 작년 제퍼디 퀴즈쇼에서 IBM 왓슨이 퀴즈 달인을 제치고 우승을 거머쥔 것은 앞으로의 미래에 많은 변화가 있을 것임을 암시한다. 그리고 이것을 구현하는 데 가장 중요한 역할을 한 것이 바로 빅 데이터다.



미국의 퀴즈쇼 제퍼디에서 우승을 차지한 IBM 왓슨

‘생각’을 만드는 빅 데이터 기술

“ 빅 데이터는 다양한 기술이 복합적으로 결합되어 있는 기술 시스템이다. ”

빅 데이터라는 이름으로 통칭되기는 하지만, 사실 빅 데이터 기술은 기계 학습이나 자연어 처리, 각종 통계 기법, 분산 병렬 컴퓨팅 기술 등의 다양한 기술들이 복합적으로 결합되어 있는 기술 시스템이다.

우선 빅 데이터 기술의 원리부터 간단히 살펴보자. 컴퓨터 명령어로 짜여진 정보가 아니라 사람들이 평상시에 쓰는 말이나 글을 컴퓨터가 이해하고, 정보화하는 것이 시작이다. 이를 통해 모아진 대용량의 정보를 분석할 수 있도록 프로그래밍하고, 여기에 여러 가지 통계 기법, 기계 학습과 같은 인공 지능 프로그램을 사용하여 이 정보가 담고 있는 복합적인 의미를 분석하고 추론하는 것이 빅 데이터 분석이다.

컴퓨터가 정보를 받아들이는 첫 단계에서 가장 중요한 기술은 자연어 처리 기술이다. 즉, 사람들이 일상적으로 쓰는 말과 글을 컴퓨터에 가르치는 기술이다. 여기에 더해 영상이나 행동, 온도, 습도 등 사람이 감각 기관으로 감지하는 정보를 기계가 받아들여 처리하는 기술도 같이 요구된다. 영상 인식이나 센서 기반의 상황 인식 기술이 여기에 해당된다.

스마트 혁명과 빅 데이터 기술

생각하는 기계의 출발점이 정보라면, 정보란 무엇일까? 그야말로 모든 것이다. 스마트 환경은 예전엔 정보가 될 수 없었던 많은 것들을 정보로 만든다. 수많은 센서들이 탑재된 기기들이 모두 네트워크에 연결되어 있고, 표준화된 소프트웨어 환경을 통해 전달되는 서비스들은 각종 사용 기록을 남기기 때문이다. 우리가 하는 말, 우리가 하는 행동, 우리의 위치, 우리가 접하는 환경, 이 모든 것이 정보가 된다. 검색엔진에서 찾을 수 있는 지식이나 온라인 사이트에서 듣는 음악이나 영화 같은 콘텐츠는 물론이고, 서비스를 이용하는 사람들의 행동 정보와 이들을 둘러싸고 있는 주변 환경의 상황도 모두 정보다. 정류장에서 버스를 기다리는 사람의 수, 이

들이 듣고 있는 음악, 이들 각자가 기다리고 있는 버스의 번호, 기다리는 버스의 도착 시간에 영향을 미치는 도로 교통 상황 등 시시각각 발생하는 모든 행동과 상황들이 모두 정보가 되어 기록될 수 있는 환경이 바로 스마트 환경이다.

엄청나게 다양한 형태의 데이터가, 엄청난 양으로, 엄청난 속도로 쏟아지는 스마트 환경에서 빅 데이터 기술의 부상은 필연적이다. 수없이 많은 정보를 감지하는 스마트 기기들과 센서들이 스마트 환경의 감각 기관이라면, 빅 데이터 기술은 스마트 환경의 뇌로서 작동한다. 감각 기관과 뇌가 유기적으로 작동하듯, 스마트 환경과 빅 데이터는 서로가 서로를 요구하는 필연적 조합이다.

“ 모든 것을 정보화하는
스마트 환경과 무수한
정보를 이해하는
빅 데이터 기술은
서로가 서로를 요구하는
필연적 조합이다. ”

하둡, 맵리듀스, NoSQL 등 최근 IT 업계에서 빅 데이터란 이름으로 빈번하게 언급되는 기술들은 앞에서 수집된 데이터를 실시간으로 분석할 수 있도록 하는 컴퓨팅 기술들이다. 대용량의 데이터를 다루기 위해서는 데이터를 여러 대의 서버에 분산 저장하고 이들을 동시에 처리하기 위한 병렬 처리 방식을 써야 하기 때문에 분산 병렬 컴퓨팅 기술이라고 한다. 또한 빅 데이터 특유의 기술들이기 때문에 빅 데이터 기술이라는 말을 협의로 쓰면 주로 이들 기술을 뜻하는 경우가 많다.

마지막 기술 요소는 이들 데이터에서 의미를 도출하고 활용하기 위한 기술이다. 데이터 속에서 패턴을 찾아내고, 새로운 정보를 계속 덧붙여서 좀 더 정교화된 패턴을 만들어 내는 것은 물론이고, 실시간으로 업데이트되는 정보와 결합하여 현재 상황을 분석하고 예측치를 만들어낼 수도 있다. 이를 위해서는 회귀 분석이나 아노바와 같은 일반 통계 기법과 기계 학습과 같은 컴퓨터 프로그래밍 기술, 분석된 결과치를 손쉽게 이해할 수 있도록 보여주는 시각화(Visualization) 기술이 필요하다. 이러한 단계를 통해 빅 데이터는 스마트 환경이 만들어낸 정보를 기억하고, 분석하는 ‘생각하는 기계’의 원동력이 되는 것이다.

Ⅱ. 지능형 서비스가 온다

빅 데이터가 만드는 지능형 서비스

빅 데이터를 사용하면 무엇이 어떻게 달라지는 것일까? 생각할 수 있는 기계가 제공하는 지능화된 서비스란, 과연 지금의 서비스와 무엇이 달라지는 것일까?

지능형 서비스의 첫번째 장점은 의사 결정의 효율화다. 고도의 지식을 가졌거나, 오랜 경험을 가진 사람에 의존할 수 밖에 없었던 의사 결정 사항들을 축적된 데이터를 기반으로 훨씬 더 빠르고 쉽게 처리할 수 있게 된다. 지금까지 IT 시스템의 역할은 주문을 처리하고, 판매량을 기록하며, 과거 정보를 조회하는 등의 보조 수단에 불과했지만, 빅 데이터를 기반으로 한 IT 시스템은 좀 더 적극적으로 현상을 파악하고 필요한 대책을 제시할 수 있다. 물류 효율화나 재무 정보의 실시간 파악, 소

“지능형 서비스의 가장 큰 기대요소는 난치병 치료와 지능형 교통망과 같은 꿈의 현실화 가능성이다.”

설 정보를 활용한 상품 기획, 마케팅 의사 결정의 실시간화 등 기업 현장에서의 빅 데이터 분석은 그 자체가 최고의 경영 컨설팅 서비스다. 맥킨지의 분석에 따르면 빅 데이터 분석을 통해 상품 개발 및 생산 비용을 50%까지 줄일 수 있고 재고 자산 등의 효율화를 통해 운전 비용도 7% 이상 절감 가능하다고 한다. 순간 순간 긴박한 의사 결정이 필요한 의료 현장에 빅 데이터가 적용되면 의사는 이를 바탕으로 즉시 의사 결정을 내릴 수 있다. 실제로 온타리오 대학 병원의 경우 IBM의 빅 데이터 솔루션을 활용하여 미국아 사망률을 혁신적으로 줄이는 데에 성공하기도 했다.

지능형 서비스의 두번째 특성은 고도의 개인화다. 축적된 개인 정보와 현재 파악되는 상황 정보를 결합하여 한 사람, 한 사람에 특화된 서비스를 제공할 수 있게 되는 것이다. 개인화된 검색 결과를 제공하는 구글이 이 개념을 가장 먼저 현실화시켰고, 넷플릭스나 아마존 등의 콘텐츠 업체에서 제공하는 추천 데이터 역시 방대한 양의 사용 정보를 분석한 결과로 넷플릭스의 경우 매출의 80% 정도가 이 추천에 의해서 발생하는 것으로 알려져 있다. 페이스북 역시 8억 명의 사용자 정보와 검색어를 실시간으로 분석하여 맞춤형 광고를 제공한다. 개인화는 서비스의 맞춤화 측면도 있지만, 기기를 사용하는 인터페이스가 좀 더 사용자 중심으로 맞춰지는 측면도 있다. 시리와 같은 음성 인식 기술, 구글의 자동 번역 기술 역시도 빅 데이터 기술의 결과물이다. 구글은 음성과 텍스트에 이어 영상과 얼굴, 표정 인식에도 공을 들이고 있는데, 이것은 단말기를 통해 파악할 수 있는 사용자 정보가 더 많아진다는 의미이기도 하다. 궁극적으로 개인화 서비스는 지금의 미디어/광고 영역을 벗어나 건강 관리나 패션 코디처럼 등 사용자의 의식주 전반에 걸친 영역으로 확장될 것으로 본다.

마지막으로, 지능형 서비스에 거는 가장 큰 기대 요소는 각 분야의 오래된 난제를 해결할 실마리를 제시할 수 있다는 점이다. 환자의 신체 특성과 가족력, 생활 습관, 업무 환경 등의 데이터를 빠짐없이 수집하여 치료에 사용하는 지능형 의료에서는 방대한 임상 데이터를 기반으로 난치병 연구가 혁신적으로 진전될 수 있다. 차량의 위치와 운전자가 네비게이션에 입력한 목적지 데이터를 실시간으로 파악하여 교통량을 예측하고, 그에 따라 최적화된 운전 경로를 제시하여 교통 정체를 막아내는 지능형 교통 서비스도 빅 데이터를 통해 가능해 질 것이다. 정해진 노선을 따라 운행하는 버스 대신, 목적지가 유사한 사람들을 실시간으로 조합하여 그 때 그 때 다

“지능형 서비스는 기존의 서비스 패러다임을 깨뜨리는 파괴적 속성을 가지고 있다.”

른 노선을 운행하는 택시 같은 버스가 생긴다고 생각해 보자. 대중교통의 불편함과 고질적인 적자 노선 문제를 동시에 해결할 수 있을 것이다. 지금까지 그저 꿈 같은 이야기로 여겨졌던 많은 시나리오들을 방대한 데이터를 기반으로 실현해내는 지능형 서비스야말로 빅 데이터의 본질적 가치이다.

지능형 서비스 시대, 삶의 변화

지능형 서비스는 전통적인 서비스 패러다임과 경쟁력의 원천을 근본부터 바꾼다는 점에서 파괴적이다. 개인, 기업, 나아가 국가의 운명까지도 바꿀 수 있다.

우선, 지능형 서비스는 기존의 서비스를 구성하는 가장 중요한 요소인 현장성을 거의 요구하지 않는다. 병원에 가지 않아도 몸 상태를 진단받을 수 있게 되면 동네 병원이 필요할까? 만약 의사와 자동 번역 시스템으로 대화할 수 있다면, 굳이 한국에 있는 병원을 이용할 필요도 없다.

전문가의 역할도 변화될 것이다. 소위 지식 노동이라고 하는 분야에서 사람이 해 오던 일을 기계가 대신할 것이기 때문이다. IBM의 왓슨은 이미 수백만 건의 의료 특허 문헌을 분석하여 250만 개의 화학 혼합물 데이터를 찾아낸 바 있다. 제약 R&D 분야에서 한번쯤 해보고 싶었던 일이지만, 그 일의 규모가 너무 방대해서 업무를 내지 못하던 일을 기계를 통해 빠르게 할 수 있게 된 것이다. 다수의 사무직 노동자들이 수행하는 시장과 경쟁 동향에 대한 리서치, 소송이나 특허 판매를 위해 검토해야 하는 다량의 자료들에 대한 검색은 빅 데이터 시대에는 기계가 할 일이다. 작업 절차가 복잡하고, 사안에 대한 지식적 이해를 요구하기에 전문가의 영역으로 구분되었던 제약, 의료, 법무, 회계, 경영 컨설팅 분야의 업무 중 상당 부분은 기계가 대신해 줄 수 있다.

혁신 메커니즘도 달라질 것이다. 각 분야의 전문가들이 부단한 연구와 노력을 통해 얻어내는 혁신 대신, 성실하고 집요하게 축적된 데이터가 어느 순간 놀라운 답을 내주는 방식으로 바뀔 수도 있다. 그리고 이 결과는 전 세계 어느 지역이나 적용될 수 있다. 사람이 아니라 시스템이 중심이 되는 지능형 서비스는 하나의 서비스가 한 국가를 넘어 글로벌로 확장되는 데 핵심적인 역할을 할 것이다.

“ 지능형 서비스는 새로운 패러다임으로 개인과 기업, 국가의 운명을 바꿀 것이다. ”

물론 지능형 서비스가 아무리 발전해도 인간이 하는 일을 기계가 완전히 대체하지는 못한다. 인간 의식과 직관의 메커니즘은 아직도 밝혀지지 않은 미지의 영역이기 때문이다. 하지만 가장 방대한 의료 데이터를 가지고 있는 병원, 가장 다양한 금융 데이터를 보유한 은행, 가장 폭넓은 학업 성취 데이터를 보유한 교육 기관이 경쟁자가 쉽게 따라올 수 없는 노하우를 확보하고, 이것을 세계 곳곳에 이식할 수는 있다. 같은 의료나 금융 서비스 안에서도 데이터를 잘 다루는 보통의 의사나 은행가가 지금의 숙련된 의사와 노련한 은행가를 능가하는 영역이 분명히 생겨날 것이다.

지능형 서비스는 개인과 기업, 그리고 국가의 운명을 가르는 새로운 패러다임으로 세상을 바꿔놓을 것이다. 개인 측면에서 보면 출세를 담보하는 일의 종류가 달라질 것이고, 기업 측면에서 보면 인재의 정의와 경쟁력의 원천이 달라지는 것이다. 국가의 경쟁력도 마찬가지다. 글로벌 시장을 주도하는 지능형 서비스 경쟁력이 그 나라의 경쟁력을 좌우할 것이다. 높은 교육 수준, 근면한 사회 문화와 같은 일반적인 인적 자본이 아니라, 끊임없이 정보를 수집하여 전송하는 정보 인프라, 좀 더 정확하고 빠르게 데이터를 분석하는 기술력, 각각의 서비스에 맞는 데이터 수집과 분석 모델을 설계하여 서비스 지능화를 주도하는 인적 인프라가 한 나라의 경쟁력을 결정하는 요소로 부각될 날이 다가오고 있는 것이다.

Ⅲ. 빅 데이터에서 지능형 플랫폼으로

빅 데이터 솔루션 vs. 빅 데이터 서비스

빅 데이터를 제품화하는 데 가장 앞장 서고 있는 기업들은 역시 오라클, IBM, HP, EMC처럼 기업용 솔루션 사업을 하고 있는 기업들이다. 이들은 마치 약속이라도 한 듯, 빅 데이터 분야에서 두각을 나타내고 있는 신생 벤처 기업들을 인수하고, 그들의 솔루션을 자신들의 기존 제품과 결합하여 빅 데이터 분석 솔루션을 내놓고 있다. 이러한 제품은 기업 내부 정보나 SNS 정보를 분석하여 상품 기획, 마케팅, R&D, ERP 등에 활용하고자 하는 기업 고객들을 대상으로 한다.

“ 빅 데이터를
솔루션으로 상품화하는
대신 클라우드
컴퓨팅과 결합하여
서비스로 제공하려는
움직임에 주목할
필요가 있다. ”

여기에는 오픈 소스를 활용한 신생 기업들도 가세하고 있다. 하둡은 빅 데이터 분야의 리눅스라고 할 수 있을 만큼 광범위하게 사용되고 있는 오픈소스 기술이다. 값비싼 상용 제품과 달리 기존 서버 인프라 위에서 혁신적으로 저렴한 가격으로 빅 데이터 분석 시스템을 구축할 수 있다는 장점으로 각광을 받고 있다. 그러나 오픈소스 특유의 불안정성이나 시스템 구축에 따르는 난이도 등의 문제가 있어, 하둡을 기

반으로 하되 좀 더 안정적인 빅 데이터 분석 솔루션을 제공하는 벤처 기업들도 등장하고 있다. 클라우드라나 호튼웍스 등이 대표적인 기업이다.

그러나 빅 데이터를 솔루션으로 상품화하는 대신 클라우드 컴퓨팅과 결합하여 서비스로 제공하려는 움직임에 주목할 필요가 있다. 바로 구글의 빅 퀴리라는 서비스다. 얼마 전까지 비공개 시험판이었던 것이 현재 프리뷰 단계로 공개되어 있어, 얼마지 않아 상용화에 돌입할 것으로 전망되고 있다. 빅 퀴리는 빅 데이터 분석에 필요한 인프라 투자 비용이 부담스러운 기업들에게 빅 데이터 저장 공간과 분석 솔루션을 함께 제공한다. 고객들은 자체 서버와 솔루션을 구축하는 대신 빅 퀴리 인프라를 사용하여 데이터를 저장하고, 그를 분석하는 프로그램 역시 빅 퀴리를 통해 개발하여 서비스를 운영할 수 있다. 상용 솔

루션에 비해서는 가격이 저렴하고, 오픈소스인 하둡보다는 기술적 안정성이 높다는 점이 장점일 것이다.

빅 데이터 서비스에서 지능형 플랫폼으로

빅 데이터 솔루션과 빅 데이터 서비스의 차이는 무엇일까? 데이터가 모이는 위치다. 빅 데이터 솔루션은 그 제품을 구매한 기업, 병원, 학교가 보유한 서버로 정보가 모인다. 반면 빅 데이터 서비스의 경우에는 서비스 제공자의 시스템으로 모이게 된다. 결과적으로 빅 데이터 서비스 제공자는 엄청난 데이터 기반을 갖게 될 것이다.

빅 데이터 시장 규모

위키본에 따르면 2012년 빅 데이터 시장 규모는 약 50억 불이며 2017년에는 500억 불 규모로 성장할 것이라고 한다. 그러나 이것은 협의의 빅 데이터 제품, 즉 빅 데이터 처리를 위한 컴퓨팅 인프라(서버)와 솔루션 시장을 말한다. 빅 데이터가 적용된, 즉 지능형 서비스 시장 규모는 반영되어 있지 않은 수치라는 것이다.

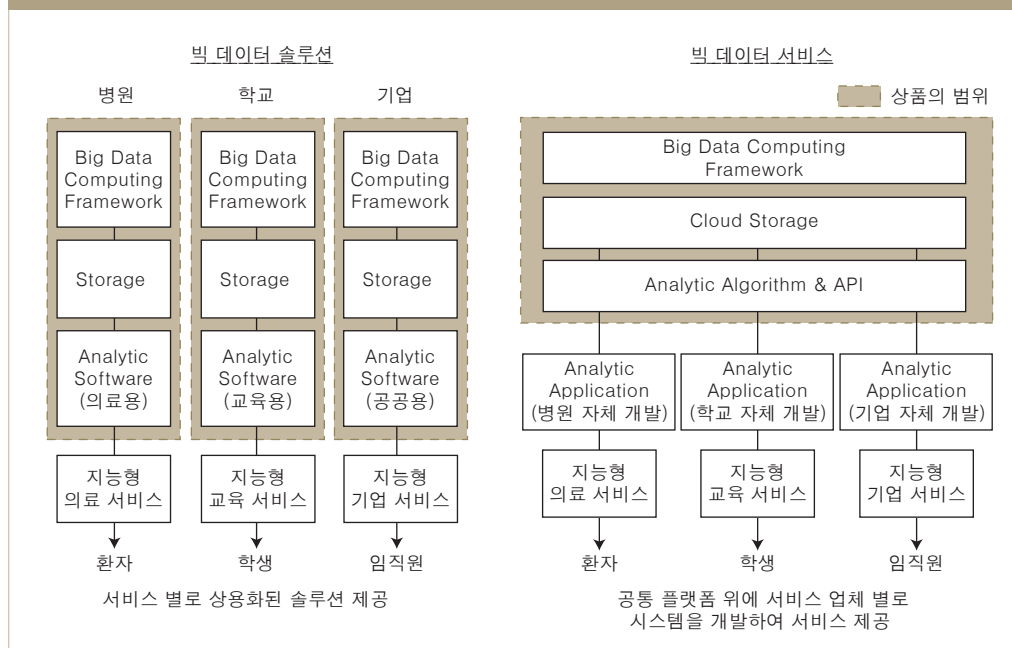
빅 데이터가 실제 서비스 현장에 적용되었을 때 어느 정도의 가치가 발생할 지에 대해서는 아직 명확한 추정치가 없다. 빅 데이터의 활용이 미국 의료 분야에서 3,300억불, 유럽 공공 부문에서 1,500억 유로에서 3,000억 유로의 잠재 가치가 있다는, 컨설팅 회사인 맥킨지의 부분적 추정치를 통해 대략적인 시장 가치를 가늠하는 수준이다.

“ 빅 데이터 서비스는
지능형 서비스
플랫폼으로 발전할 수
있다. ”

양 쪽의 데이터 분석 능력이 유사하다면 빅 데이터 서비스는 또 다른 형태의 솔루션에 불과하다. 하지만 수만 건의 문헌으로 자동 번역 기술을 개발하려던 IBM은 실패한 반면, 동일한 프로젝트를 수억 건의 문헌으로 수행한 구글이 성공했던 것을 생각해 보자. 데이터의 규모는 결국 시스템의 지능 지수와 직결된다. 시간이 지날수록 빅 데이터 서비스의 데이터 이해, 처리, 분석 능력은 진화하고 결국 개별 솔루션을 능가하는 수준으로 발전할 수 있다. 또한 분산 컴퓨팅 인프라를 구축하는 관점에서 보아도 솔루션보다는 서비스 쪽이 유리하다. 개별 업체가 각자의 컴퓨팅 시스템에 투자하는 것보다 다수의 고객을 보유한 전문 서비스 업체가 대규모로 투자하는 쪽이 비용 효율이 높기 때문이다.

이렇게 빅 데이터 시스템의 성능 우수성과 비용 효율성이 솔루션을 능가하는 수준으로 발전하면, 빅 데이터 서비스는 하나의 플랫폼이 될 수 있다. 모든 형태의 데이터를 이해하고, 이것을 실시간으로 분석하는 빅 데이터 서비스가 컴퓨팅 플랫폼이 되고, 업체들은 이것을 기반으로 자신들의 서비스 시스템을 구축하게 되는 것이다. 빅 데이터 서비스는 지능형 의료, 지능형 교육과 같은 개별 서비스를 에코시스

〈그림〉 빅 데이터 솔루션과 빅 데이터 서비스의 사업 적용 모델



“지능형 플랫폼을 기반으로 한 에코시스템은 플랫폼에 대한 의존성이 대단히 강하다.”

템을 거느린 지능형 플랫폼이 된다.

이러한 에코시스템은 OS를 기반으로 형성된 에코시스템과는 비교할 수 없을 정도로 플랫폼에 대한 의존성이 강하다. OS가 서비스 어플리케이션을 개발할 수 있는 도구와 어플리케이션 유통망을 제공해 주긴 하지만 OS와 서비스의 관계는 사실, 근본적으로는 독립적이다. 예컨대, 안드로이드 OS로 도로교통정보 서비스를 제공할 때 안드로이드 OS의 역할은 도로교통정보 어플리케이션을 만들 수 있는 개발 도구와 이 어플리케이션을 배포할 수 있는 마켓을 제공하는 것일 뿐이다. 안드로이드 OS가 없다고 해서 도로교통정보 서비스가 없어지지는 않는다. 다른 OS용의 어플리케이션이나 PC를 기반으로 해서도 서비스를 할 수 있기 때문이다. 하지만, 지능형 플랫폼은 다르다. 플랫폼 자체가 서비스의 정수다. 서비스에 필요한 모든 정보와 운영 노하우가 플랫폼에 축적된다. 이 플랫폼이 작동하지 않으면 서비스가 중단될 것이며, 혹여 다른 플랫폼으로 교체하게 될 때는 막대한 전환 비용이 발생하게 될 것이다.

때문에, 지능형 플랫폼은 오픈 OS와 달리 분명한 수익 모델을 갖는다. 구글에게 안드로이드 OS는 구글 서비스를 전파하는 도구고, 애플에게 iOS는 아이폰을 판매하는 도구일 뿐이다. 하지만 클라우드와 결합한 지능형 플랫폼은 사용량과 사용 강도에 따라 명확히 과금할 수 있다. OS 사업은 실제 수익원은 다른 곳에 두고 마케팅 수단으로서 플랫폼을 활용했지만, 지능형 플랫폼은 그 자체로 판매 가능한 독립형 사업이 된다. GPT(General Purpose Technology)로서 빅 데이터가 가질 수 있는 가장 명확한 사업 모델은 바로 이것이 될 것이며, 빅 쿼리를 상용화하는 구글의 궁극적 목표도 여기 있을 것이다. 지능형 서비스를 에코시스템으로 거느린 플랫폼이 되는 것이다.

IT 공룡의 거대한 꿈

웹이나 OS 가상화로 OS 중심의 에코시스템의 경계가 점차 허물어지면 지금의 플랫폼 사업자들은 다음 대안을 찾아야 할 것이다. 지금은 에코시스템을 주도하지 못하지만, 새로운 기회를 엿보고 있는 기업들에게도 지능형 플랫폼은 중요한 전략적 고지다.

지능형 서비스 분야로 가장 빠른 움직임을 보이는 곳은 IBM이다. 지능형 컴퓨터인 IBM 왓슨은 2억 페이지의 자연어 DB를 바탕으로 사용자의 질문을 이해하고 필

빅 데이터 거품론

최근의 빅 데이터에 대한 관심은 거의 열풍 수준이다. 빅 데이터가 모든 비즈니스의 필수적인 경쟁력으로 인식되는 분위기다. 그러나 실제로 빅 데이터 분석을 시행해 본 서비스 현장에서는 실망감을 표출하는 목소리가 적지 않다. 페이스북이나 넷플릭스의 성공 모델은 매력적이지만, 실제 자신들의 비즈니스에 적용해 본 결과는 과거의 데이터마이닝 모델과 크게 다르지 않다. 투자비에 비해 얻을 수 있는 인사이트가 적다라는 의견도 나온다. 가트너의 경우 Fortune 500 기업 중 85%가 빅 데이터 활용에 실패할 것이라는 비관적인 전망을 내놓기도 했다. 높은 관심에도 불구하고, 빅 데이터에 대한 관심과 전망에 거품이 끼어 있다는 것이다.

때문에 우리는 빅 데이터 기술과 그 의미를 좀 더 깊이 고민해 볼 필요가 있다. 기술이 아무리 좋아도 그것을 딱히 활용할 곳이 없을 때 오는 실망감을 우리는 거품이라 부른다. 빅 데이터 거품론은 바로 이런 문제를 지적하는 것이다. 지금까지 해오던 방식과 유사한 고객 조사나 비즈니스 데이터 분석은 빅 데이터가 없이도 가능하다. 여론을 이해하고 고객의 반응을 이해하는 방법이 SNS 데이터에 기반한 소셜 분석만 있는 것도 아니며, 그것이 가장 정확한 것도 아니다. 무엇을 위해 빅 데이터가 필요한지, 왜 굳이 빅 데이터여야 하는지 생각하지 않고, 막연히 유행만 좇아서 빅 데이터를 제대로 활용하기 힘들다.

빅 데이터의 기준 역시 상대적이다. 통상 페타바이트 급의 데이터를 빅 데이터라 하지만, 컴퓨팅 성능이 제한

적이던 과거에는 기가바이트의 데이터도 빅 데이터였다. 결국 빅 데이터의 특성은 단순히 용량이 크다는 것이 아니다. 지금까지 정보가 될 수 없었던 상황과 행동들을 정보화할 수 있다는 점이고, 이것을 실시간으로 처리해서 판단할 수 있다는 점이다. 때문에 빅 데이터는 그에 어울리는 서비스에 사용되어야 한다. 사람이 일일이 처리하기에는 너무 광범위하거나 시간이 오래 걸리는 일, 서비스 이용자 개개인에게 모두 맞춤형된 대안을 주어야 하는 일, 지금까지 풀 수 없었던 어려운 일에 사용될 때만이 비로소 그 의미를 가질 수 있다.

중요한 것은 빅 데이터 자체가 아니라 이것을 사용하는 서비스 모델을 만들어 내는 것이다. 전기가 인터넷처럼 특정한 신제품이나 신시장을 위한 기술이 아니라, 이미 존재하고 있는 시장에 적용되어 가치를 창출하는 기술을 경제학에서는 GTP(General Purpose Technology), 혹은 일반목적기술이라고 부른다. 90년대 인터넷 비즈니스의 경우에서 보듯, GPT는 기술의 도입보다 기술을 활용하는 목적과 방법에 대한 정의, 즉 사업 모델의 설정이 중요하다. 아마존, 이베이, 구글 등 인터넷과 함께 성장한 기업 모두 인터넷을 잘해서가 아니라, 혁신적인 서비스 모델이 먼저 있었고 그를 구현하는 수단으로 인터넷을 잘 활용했기에 지금의 자리에 오른 것이다.

빅 데이터도 마찬가지다. 목적은 혁신적인 서비스이고 빅 데이터는 수단이다. 목적과 수단의 선후가 분명해져야만 거품이 아닌 빅 데이터의 진정한 가치가 빛날 것이다.

요한 정보를 검색하여 답을 주는 솔루션인데, 얼마 전 미국의 병원인 세톤 헬스케어에서 이를 도입하면서 빅 데이터 기반의 의료 서비스의 포문을 열었다. 세톤 헬스케어는 의료용 콘텐츠 및 예측 분석 기술을 사용하여 환자의 재입원과 병원 방문 횟수를 줄이는 서비스를 제공할 예정이라고 한다.

의료용 왓슨은 온타리오 공과 대학, 콜럼비아 의과 대학, 메릴랜드 의과 대학 등과 IBM이 진행한 협력 프로젝트에 기반해 있다. IBM은 다른 분야와의 협력 프로젝트도 지속하면서 왓슨을 활용할 수 있는 서비스 범위를 확대할 것이다. 의료에 이어

“지능형 플랫폼은
IT 거대 기업의
다음 목적지다.”

금융과 교육 분야가 거론되고 있다.

왓슨을 만들어 내기 위해 IBM의 왓슨 연구소는 4년의 시간이 걸렸다고 하지만, IBM의 인공 지능 컴퓨터 프로젝트는 인간과 체스 게임을 한 것으로 잘 알려진 인공 지능 컴퓨터 딥 블루까지 거슬러 올라간다. IBM이 딥 블루 개발을 시작한 시점은 1989년이다. 빅 데이터를 기반으로 왓슨이 완성되기까지, IBM은 무려 20여 년의 시간을 쏟은 것이다. 게다가 지금도 왓슨은 진화 중이다. 빅 데이터를 기반으로 구현되어 있는 자연어 처리 엔진을 바탕으로, 의료, 금융, 교육 등의 현장 데이터가 왓슨에 다시 쌓인다. 이를 통해 왓슨은 좀 더 정교한 분석과 사고 능력을 갖게 될 것이다.

구글의 빅 쿼리 서비스는 아직 상용화 전이기 때문에, 구글이 어떤 서비스 전략을 보일 지 단언하기는 어렵다. 그러나 구글이 사용자 정보는 물론이고, 다양한 산업 분야의 정보를 모두 모으고 있다는 점은 눈여겨 볼 필요가 있다. 의료 분야가 대표적인데, 구글은 구글 헬스라는 서비스를 통해 개인 의료 기록을 수집하고, DNA 분석 및 관리 업체에 투자하여 자사의 클라우드 서버에 데이터를 저장하도록 하는 등 의료 정보 수집에도 열심이다. 구글 맵을 기반으로 한 교통 서비스는 물론이고, 항공권 예약, 게임, 쇼핑, 사진 등 구글이 인수와 제휴를 통해 확보하는 데이터의 종류는 매우 다양하다. 안드로이드 기기를 통해 확보되는 정보는 물론이고, 스피커나 헬스 기기 등과 안드로이드 기기를 연결하는 플랫폼인 안드로이드웨어도 구글의 정보 수집에 큰 역할을 할 것이다. 구글은 사용자의 운동 기록이나 건강 정보는 물론이고, 집 안에 있는 기기의 종류와 교체 연한 정보까지, 누구도 상상하지 못했던 정보들을 바탕으로 자신의 플랫폼을 진화시킬 수 있다.

이렇게 보면 구글이나 아마존 같은 서비스 업체가 왜 하드웨어 사업에 공을 들이는 지도 다시 한 번 생각해 볼 필요가 있다. 하드웨어 단말은 정보를 수집하는데 가장 중요한 축수다. 서비스를 유통하기에도 좋지만, 단말을 통해 얻어지는 사용자 정보는 무엇보다 귀중하다. 아마존이 킨들을 출시하면서, 단말에서 얻어지는 GPS 정보를 바탕으로 사용자의 위치를 예측하는 기술을 개발하고 이에 대한 특허를 출원한 것이 대표적이다.

또한, 빅 쿼리와 같은 빅 데이터 서비스를 제공할 수 있는 업체가 비단 구글만은 아니다. 클라우드 컴퓨팅의 강자인 아마존도 빅 쿼리와 유사한 서비스를 내놓을 것

“ 거대한 꿈을 꾸는
IT 공룡들의 레이스는
이미 시작되었는지도
모른다. ”

이라는 전망이 있으며, 최근 기업용 서비스를 내놓은 페이스북의 행보도 주목해 볼 만하다. 이들 업체는 사업을 하는 과정에서 이미 빅 데이터 분석 기술을 축적해 놓은 상태이기 때문에 이 기술을 다른 기업들이 이를 활용할 수 있도록 개방할 가능성은 충분하다. 특히 페이스북의 경우 학교와 학생, 지역 정부와 거주민을 실시간으로 연결하면서 생겨나는 데이터를 서비스에 활용하는 플랫폼을 제공할 수도 있다. 만약 페이스북이 단말 산업에 진출한다면 이러한 연결과 그를 통해 확보할 수 있는 정보를 좀 더 광범위하게 모으기 위해서가 아닐까?

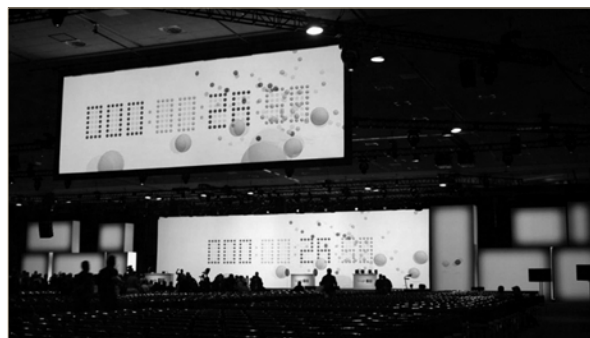
보이지 않는 물밑에서는 이미, 거대한 꿈을 꾸는 IT 공룡들의 상상 불가능한 수준의 레이스가 시작되었는 지도 모른다.

V. 한국은 지금 데이터 갈라파고스

태생적 한계

한국의 IT 인프라는 최고 수준이다. 세계 최고 속도를 자랑하는 네트워크와 스마트폰 보급률 50%에 육박하는 사용자 기반은 전 세계 어디에서도 찾을 수 없는 환경이다. 시스코에 따르면 국내 2015년 국내 모바일 가입자는 미국의 15%, 일본의 32%, 중국의 8%에 불과하나, 모바일 데이터 트래픽 발생량은 미국의 43%, 일본의 70%, 중국의 112%에 해당한다고 한다. 인당으로 보면 미국의 3배, 일본의 2배, 중국에는 무려 14배에 해당하는 사용량이다.

그러나 데이터 경쟁력은 사용 강도도 중요하지만 규모는 더욱 중요하다. 아무리 인프라가 좋아도 한국의 내수 시장 규모로는 한계가 있다. 전 세계 데이터 트래픽에서 한국이 차지하는 비중은 8~9% 수준에 불과하다. 국내의 대형 포털과 통신 업체가 페타바이트 급 데이터를 보유하고는 있으나, 전 세계를 대상으로 서비스를 제공하면서 데이터를 수집하는 글로벌 서비스 업체에 비하기는 어렵다. 언어의 한계도 있다. 빅



작년 5월 구글의 I/O 컨퍼런스에서 소개된 빅 퀴리는 구글의 클라우드 서비스를 기반으로 빅 데이터 저장 공간과 분석 알고리즘을 제공하는 서비스다.

“ 한국의 IT 인프라는
훌륭하지만 데이터
기반은 대단히
취약하다. ”

데이터 기술의 첫 단계인 자연어 이해 분야에서 가장 필수적인 언어 기반이 글로벌 공용어가 아니라는 점도 한계다. 기업용 빅 데이터 솔루션이나 국내 시장을 대상으로 한 소셜 분석 서비스는 나올 수 있겠지만, 지능형 플랫폼과 같은 메가트렌드를 선도하기는 힘든 환경이다.

기술적 인프라도 부족하다. 빅 데이터는 원래 구글이나 야후 같은 검색 서비스 업체에서 시작된 기술이다. 데이터를 본업으로 하는 업체가 그들이 보유한 다양한 형태의 데이터를 분석하는 과정에서 필요한 여러 분야의 엔지니어들을 모은 것이 출발점이다. 그리고 이들이 협력하는 과정에서 생겨난 종합적 결과가 바로 빅 데이터인 것이다. 어느 한 분야만 잘해서 되는 것이 아니다. 하드웨어부터 소프트웨어까지, 컴퓨터 공학에서 인간공학, 심지어는 뇌과학과 언어학까지 망라한 기술이 모두 적용되는 분야가 빅 데이터다. 이것이 지능형 서비스로 이어지기 위해서는 빅 데이터 기술 뿐만 아니라, 각 서비스 현장의 노하우를 기반으로 빅 데이터 분석 및 적용 모델을 설계할 수 있는 데이터 사이언티스트의 역할도 중요하다.

이처럼, 빅 데이터는 지금까지와 차원이 다른 기술과 인적 인프라를 요구하는 융합 기술이지만, 기술과 사람의 인프라는 역지로 만들어지는 것이 아니다. 시장이 이를 필요하다고 느낄 때 비로소 자연스럽게 육성된다. 그러나 데이터가 곧 본업이며 소프트웨어 원천 기술을 가진 선도 기업이 드문 한국 시장의 경우, 빅 데이터에 대한 관심은 대부분 상용 솔루션을 기반으로 한 기업용 시장 중심으로 논의가 전개되고 있을 뿐이다. 원천 기술은 해외에 두고, 현장에서의 시스템 구축이나 관리만을 주로 하는 한국의 왜곡된 소프트웨어 산업 구조가 빅 데이터 분야에서도 그대로 재현되고 있는 것이다. 개인 맞춤형 서비스, 지능형 의료나 교육, 지능형 정부와 같은 차세대 혁신 서비스에 대한 비전이 없다면, 결국 외국에서 적용된 빅 데이터 사례를 한국에 들여오는 일만 남는다. 빅 데이터나 지능형 서비스 분야의 전문가는 없고, 소프트웨어 기능공만 남게 될 것이다. 세계 최고의 인터넷 인프라를 가지고도 인터넷 비즈니스 혁명에는 동참하지 못한 채 갈라파고스로 남았던 아쉬운 경험이 빅 데이터 시대에 와서도 또 재현될 가능성이 농후하다. 빅 데이터 시장만의 문제가 아니다. 궁극적으로는 우리의 의료, 교육, 공공 시스템 등 사회 전반의 경쟁력을 약화시키는 결과로 이어질 것이다.

까다로운 규제 환경

데이터 기반의 서비스 사업이 미국에서 유독 활성화되어있는 것은 개인 정보의 자유로운 이용을 옹호하는 미국의 사회, 문화적 분위기와도 깊이 관련되어 있다. 미국은 프라이버시 규제가 거의 없다고 할 정도로 자유로운 정보 이용이 가능한 나라다. 사생활을 헌법 상의 기본권으로 보장하는 우리나라와 달리 미국의 연방 헌법은 프라이버시를 명문으로 인정하고 있지 않으며, 구체적인 개별 법령에 한해 프라이버시 보호 규정이 적용된다. 때문에 개인 정보의 공개 그 자체만으로는 불법 행위가 성립하지 않으며, 명예훼손죄는 아예 없다. 금융계좌정보는 프라이버시로 보지 않아서 법원의 영장 없이도 자유롭게 검색할 수 있다. 인터넷은 자택과 같이 내밀한 사생활 영역이 아니므로 온라인 상의 프라이버시는 엄격하게 보호하지 않는 것이 사회적인 후생을 높인다는 주장이 지지를 받고 있기도 하다.

반면, 우리나라의 프라이버시 규제는 매우 엄격해서 데이터를 활용한 상당수의 서비스들이 위법이다. 예를 들어 소비자들의 쿠키 정보 등을 수집해서 특정 계층을 대상으로 하는 타게팅 광고나 SNS에 공개된 개인정보를 기반으로 한 다양한 산업들이 미국에서는 허용되지만 한국에서는 법적으로 금지되거나 매우 제한적으로만 허용되고 있다. 이로 인해 사용자의 정보를 축적하고, 이것을 활용하거나 공개하는 활동이 초기 단계에서 차단되는 경우도 많다.

개인 정보에 대한 보호 규정은 까다롭지만, 주민등록번호로 인해 한 번 정보 유출이 발생하면 그 피해나 파장이 상당히 큰 것도 문제다. 정부가 주민등록번호의 사용을 금지하기로 법률을 개정하였지만 인터넷 실명제와 같이 주민등록번호의 사용이 강제되는 경우가 있어서 공백은 여전히 남아 있다.

정보 활용에 대한 부정적 인식

정보 보호에 대한 생각은 비단 법률적인 문제뿐만 아니라, 사람들의 인식에도 영향을 미친다. 보이스피싱이나 광고전화에 시달려온 한국인들은 개인정보유출에 극히 민감하다. 그래서인지, 빅 데이터를 이야기할 때 빠지지 않고 등장하는 말이 바로

“ 한국의 프라이버시 규제는 매우 엄격해서 데이터를 활용한 상당수의 서비스들이 위법이다. ”

“ 중요한 것은 정보를 활용하는 기업의 활동을 판단하고 감시할 잣대를 세우는 일이다. ”

빅 브라더이다. 구글이나 페이스북이 엄청난 정보를 바탕으로 개인의 사생활을 엿담하고, 막강한 정보력을 행사한다는 것이다. 빅 브라더로부터 사생활을 지키기 위해 인터넷에 함부로 사진을 올리지 말아야 된다고, 서비스 자체를 사용하는 것을 재고하라는 내용의 주장도 종종 등장한다. 정보의 수집과 활용에 기반한 서비스에 대해 소비자들도 거부감을 가지고 있는 것이다. 정보를 공개해서 적극적으로 활용하기보다, 가능한 공개를 최소화하고 이것을 이용하는 것을 의심스러운 눈초리로 바라보는 분위기에서는 어떤 기업도 사용자 정보를 활용하여 비즈니스에 활용하겠다는 생각을 하기 어렵다.

폭발적인 정보 공개 시대에 개인 정보 관리에 유의해야 한다는 문제 의식은 중요하다. 하지만 빅 브라더 문제는 좀 더 냉정하게 생각해 볼 필요가 있다. 감찰을 목적으로 하는 정부 기관이 아닌 바에야, 일반 기업이 개인 단위의 정보를 엿볼 필요는 사실상 없다. 수억 명의 고객을 가진 기업들이 한 사람, 한 사람의 사생활을 엿맬한다는 것은 불가능한 일이거나 이 정보를 사용자 모르게 다른 기업에 판매한다는 발상도 현실성이 떨어진다. 구글의 경우 얼마 전까지만 하더라도 사용자 정보를 절대 공개하지 않았기 때문에 조사 목적으로 정보 공개를 요청한 각국 정부와 마찰을

빚었던 것을 기억해 보자.

구글의 개인 정보 수집 논란

최근 구글이 MS와 애플 소비자들의 개인정보를 수집한 것이 문제가 되고 있다. MS의 인터넷 익스플로러와 애플의 사파리에는 별도의 보안 프로그램이 작동하는데, 구글이 이 기술의 허점을 이용하여 쿠키와 같은 사용자 개인 정보를 수집한 것이다.

그러나 구글의 정보 수집은 우리나라에서는 위법이지만 미국에서는 합법이다. 문제가 되는 것은 개인 정보 수집이 아니라 구글이 과거에 애플의 보안정책을 존중하겠다는 공지에 위배되는 행위, 즉 허위 공지를 했다는 점이다. 구글은 작년, 미국 FTC와 허위 공지를 할 경우 매일 16,000불의 과징금을 지불할 것을 합의하였기 때문에 이 부분에 대해서는 과징금이 부과될 가능성이 있다.

또한 이들이 정보를 독점한다는 발상도 개연성이 떨어진다. 만약 이런 기업이 정말 있다면 경쟁자들이 가만 있지 않을 것이며, 하루아침에 소비자들의 신뢰를 잃고 추락하게 될 일이다. 실제로 빅 데이터 분석에 필요한 정보는 굳이 개인화될 필요성이 낮으며, 분석 알고리즘이 형성된 이후에는 기반 데이터를 폐기할 수도 있다. 기억이 사라져도 사고 활동이 마비되는 것이 아닌 것과 유사하다.

중요한 것은 정보를 활용하는 기업의 활동을 판단하고 감시할 잣대를 세우는 일이다. 빅 브라더인지 그렇지 않은지는 이 잣대를 기준으로 판단하고 감시할 문제다. 미국은 정보 공개에 대한 사회적 인식도 비교적 관대하

고 연구 목적으로 자신의 의료 정보를 공개하는 사람들도 많지만, 지나친 개인 정보 활용에 대해서는 제동 장치를 적용한다. 페이스북이 사용자의 주소와 전화번호를 공개하는 것에 대해 미 의회가 중지 요청을 한 것이 대표적이다. 페이스북은 난감해하는 모습이지만, 일단 공개 방침을 보류하겠다는 입장을 발표했다. 이러한 잣대가 명확하면 기업들 스스로가 정보의 활용과 소비자 보호 사이에서 현명한 판단을 할 수 있다. 막연한 거부감은 이미 거스를 수 없어진 정보 공개 시대에 역행할 뿐이다.

“지능형 서비스 트렌드는
지금 막 시작되었을 뿐,
현재 선두업체와의
격차는 극복 못할
수준은 아니다.”

IV. 빅 데이터 십년지대계가 필요하다

아직 늦지 않았다

새로운 기술이 떠오르면, 그 기술 자체의 놀라움에 압도되기 쉽다. 그러나 정말 중요한 것은 기술 그 자체가 아니라 그것이 향하는 미래에 있다는 것을 IT 산업의 역사는 이미 여러 번 증명해 주었다. 그런 의미에서 보면 최근 IT 업계의 기술적 화두인 빅 데이터, 클라우드 컴퓨팅, 스마트화, 지능형 인터페이스는 모두 한 몸이다. 이 모든 기술이 그리는 미래는 지능형 서비스를 기반으로 개개인의 삶이 좀 더 편리하고 윤택해지는 것이다. 그리고 이 트렌드는, 다행스럽게도 이제 막 시작되었을 뿐이다.

빅 데이터 기반의 지능형 서비스 시장은 기술적으로도, 사업적으로도 아직은 시작 단계이다. 선두 업체와 격차는 분명하지만 극복할 수 없을 만큼은 아니다. 일단 IBM의 왓슨은 경이로운 분석 능력을 가지고 있으나, 빅 데이터를 분석하고 서비스에 활용하는 단위가 개별 병원, 은행 등이기 때문에 서비스 적용 범위나 확산 속도가 얼마나 빠른 지는 미지수다. 반면 구글의 빅 데이터 서비스는 아직 미완의 그림이다. 아무리 구글이라 해도 지능형 서비스에 필요한 광범위한 인프라 투자를 혼자서 할 수는 없고, 안드로이드앳홈과 같은 프로젝트는 이제 겨우 시작되었을 뿐이다. 실제 서비스 에코시스템이 형성되기까지는 상당한 시간과 시행 착오가 있을 가능성이 높다. 출발은 다소 늦었지만, 아직도 시간은 남아있다.

“지능형 서비스 시스템을
전 세계 어느 나라보다
빨리 구현하고
이 시스템을 해외로
확대해야 한다.”

한국형 데이터 십년지계

한국의 환경은 테스트베드로서는 분명한 장점이 있다. IT 기기에 대한 높은 관심, 도심에 집중되어 있는 인구, 신기술을 적용하기 좋은 대도시의 건물과 주거 환경 등은 제대로 활용한다면, 시장 초기 단계에서 서비스 경쟁력을 확보하기에 유리한 조건이다. 이 조건을 십분 활용하면 지능형 서비스를 전 세계에서 가장 빨리 실현할 수 있다. 스마트 기기를 기반으로 한 원격 지능형 의료, 전 국민의 의료 기록을 기반으로 한 난치병 연구, 지능형 교통 시스템, 수준 별 진도 학습과 다양한 학습 수행 자료를 기반으로 한 적성 지도를 제공하는 스마트 교육 등의 지능형 서비스 시스템을 전 세계 어느 나라보다 먼저 구현하고, 이것을 해외로 가져가면서 시장을 확대할 수 있어야 한다.

앞서 언급한 바와 같이, 한국의 경우 IBM이나 애플, 구글과 같은 단일 기업이 지능형 서비스나 플랫폼 사업을 주도하기는 쉽지 않다. 글로벌 서비스 시장 기반이 취약하고, 글로벌 업체가 이미 가지고 있는 물적, 인적 인프라를 따르기 쉽지 않기 때문에 단일 기업 단위의 경쟁에는 어려움이 따르는 현실을 직시해야 한다. 플랫폼을 만들어 서비스를 육성하는 방식보다, 서비스를 먼저 만들고 이것이 시장에 확산되면서 플랫폼이 성장하게 되는 전략이 한국에는 좀 더 적합하다.

한국어를 비롯한 비영어권 언어에 대한 자연어 처리 기술이나 서비스 별로 정교화된 분석 기법의 개발, 빅 데이터 처리에 필요한 컴퓨팅 인프라의 집종화된 육성은 플랫폼 차원에서 이루어질 일이다. 하지만 이 일에는 IT 기업은 물론이고 의료, 교육, 공공, 복지 등의 서비스 주체들의 노하우와 협력이 함께 어우러져야 한다. 지능형 서비스에 필요한 하드웨어 및 장비의 개발이나 세부적인 서비스 솔루션의 개발, 실제 서비스 현장에서의 지능화 서비스 기술 적용 등은 개별 기업들이 수행할 일이지만 이러한 일들도, 공통적인 기술 기반과 인프라를 중심으로 수렴되고 조율되어야 한다. 여기에 더해, 서비스 정책과 규제 환경의 개선, 서비스의 해외 확대 등에 대한 국가 차원의 투자 지원도 필요할 것이다. 1~2년 안에 될 일이 아니고, 10년을 내다본 마스터 플랜이 필요한 일이다.

흩어진 데이터의 통합도 필요하다. 의료, 교육, 금융 분야의 정보 통합과 표준화된 정보 축적 체계가 있어야만 한다. 글로벌 서비스 기반을 가지고 블랙홀처럼 정보

“ 10년을 내다 본
지능화 시대 마스터
플랜이 필요하다. ”

를 빨아들이는 거대 IT 기업과 경쟁하려면 이것이 무엇보다 중요하다. 데이터센터 구축이나 빅 데이터 분석 기술은 일종의 장치 산업처럼 진행되고 있다. 글로벌 업체들은 조 단위의 투자를 통해 엄청난 규모의 데이터 센터를 짓고 누구도 따라올 수 없는 비용 경쟁력을 확보하고 있다. 이 분야에 전문적인 노하우를 가지고 있고, 글로벌 영업이 가능한 선도 기업도 필요하다. 이러한 기업이 있어야 방대한 클라우드 시스템을 설계하는 소프트웨어 아키텍처가 육성되고, 대접받을 수 있는 환경이 생겨날 것이다. 이것은 시급하다 못해 절박한 과제다.

지능형 서비스를 위한 정보 인프라망에 대한 투자는 한국만큼 빠르고 효율적으로 진행할 수 있는 나라가 없을 것이다. 의사가 멀리서도 환자의 상태를 진단하기 위해서는 환자가 보유한 기기가 환자의 상태를 실시간으로 알려줘야 하고, 어떤 의의 식생활을 진단하려면 그 사람이 무엇을 먹는지, 냉장고에는 어떤 음식이 들어있는지, 어떤 요리를 하는 지를 알아야 한다. 구제역과 같은 전염병의 발생 가능성을 사전에 경고하려면 바이러스를 감지할 수 있는 센서가 설치된 환경이 필요하다. 지금의 기기와 인프라로는 이것을 정확히 파악할 길이 없다. 스마트 가전과 스마트 의료 기기의 확산, 스마트 건물과 도로에 설치되는 각종 센서와 이를 기반으로 한 M2M 네트워크는 차세대 성장 동력이다. 이를 기반으로 한다면, 전 세계 어디에서도 볼 수 없는 지능형 서비스를 구현할 수 있고, 또한 서비스와 인프라가 결합된 시스템을 해외로 확산하기도 유리하다.

새로운 방식의 인재 육성도 필수적이다. 의료나 법무, 회계 등의 전문 지식과 통계 지식을 함께 가졌거나, 컴퓨터 공학과 전문적인 서비스 지식을 함께 가진 융합형 인재를 키워내기 위한 학제가 준비되어야 한다. 서서히 윤곽을 드러내기 위한 지능화 시대를 예비하고, 그에 준비된 인재를 키워내는 일이야말로 지금 상황에서 가장 필요한 십년지대계다.

제도 환경이 곧 경쟁력

지금 우리는 정보가 자원이 되는 새로운 시대를 눈 앞에 두고 있다. 지금까지 한번도 경험하지 못한 것이며, 어떤 모습일 지도 쉽게 예상하기 어렵다. 자유로운 정보

“ 선진적인 제도 환경이
향후 시장의 성장과
기술 발전에 중심축
역할을 할 것이다. ”

공개와 현명한 활용을 통해 모든 이들이 지금껏 경험하지 못한 편리함 속에서 살게 될 지, 무분별한 정보 공개와 사생활 유출에서 누구도 자유롭지 못한 혼란 속에 놓이게 될 지는 아직 아무도 알 수 없다. 다만 확실한 것은 정보 활용을 통한 가치 창출과 개인의 권익 보호라는 두 가지 토끼를 잡는 선진적인 제도 환경이 향후 시장의 성장과 기술 발전에 중심축 역할을 할 것이라는 점이다. 미국의 자유로운 분위기 속에서 구글이나 페이스북 같은 기업이 생겨날 수 있었다고는 하지만, 이것이 무조건

국가별 온라인 프라이버시 규제의 차이

프라이버시라는 것이 어느 나라나 비슷하게 생각하는 개념으로 알기 쉽지만, 의외로 각국 간에 프라이버시 개념과 규제가 크게 다르다. 올해 발표한 구글의 개인정보 통합 관리 정책이 미국에서는 합법이지만 유럽에서는 위법일 가능성이 높다는 점은 이를 잘 보여주고 있다. 먼저 우리와 비슷한 유럽의 접근 방식은 프라이버시, 즉 사생활을 헌법상 기본권으로 보장하고, 프라이버시를 침해하는 일체의 행위를 금지한다. 헌법뿐만 아니라 프라이버시를 보호하는 일반 법률을 제정한 나라들이 많으며, 결과적으로 소비자의 프라이버시를 침해하는 모든 행위를 금지하고 있다.

반면 미국은 법적으로만 보면 프라이버시를 유럽이나 우리만큼 존중하지 않는다. 미국의 연방 헌법은 프라이버시를 명문으로 인정하고 있지 않다. 프라이버시는 헌법상 부당한 압수수색 금지 조항의 반사적인 효과로 인정되며, 프라이버시보다 수정헌법 1조가 보장하는 알 권리와 보도할 권리를 중시하고 있다. 그 결과 구체적인 법령이 특정 개인정보의 수집과 공개를 금지하는 경우 이외에는 프라이버시가 보호되지 않는다. 프라이버시를 보호하기 위해서는 반드시 구체적인 법령이 필요한 것이다. 이에 따라 개인정보의 공개 그 자체만으로는 불법 행위가 아니며, 개별 법령이 금지하는 개인정보 공개만이 위법이다. 민간뿐만 아니라 정부도 정보를 자유롭게

이용할 수 있다. 금융계좌정보는 아예 프라이버시로 보지 않아서 수사기관은 법원의 영장 없이도 자유롭게 검색할 수 있다. 알 권리를 보장하고 프라이버시를 존중하지 않는 미국의 법제도는 구글이나 페이스북과 같은 IT 혁신 기업들의 모태가 되기도 하며, 수사기관의 수사력을 제고해 주어 미국을 세계에서 가장 사회적 투명성이 높은 나라로 만들었다.

그러나 자유로운 정보의 이용을 중시하는 현 법제도가 개인의 프라이버시를 너무 경시한다는 비판도 제기되고 있다. 프라이버시 침해에 따른 사회적 비용이 지나치게 커질 수 있다는 점을 우려하고 있는 것이다. 복잡한 프라이버시 규제 체계도 문제라고 지적된다. 미국은 프라이버시를 보호하는 법률이 부족해서 자율적 규제에 해결하는 경우가 많다. 그 때문에 상당히 많은 자율적 규제와 정부와 기업 간의 프라이버시 관련 합의가 있어서 이를 단순화하자는 움직임이 있다.

이에 백악관은 2012년 2월 24일 Consumer Privacy Bill of Rights를 제안하였지만, 강제성이 없다는 문제가 있다. 또한 미 의회 일각에서는 우리나라와 유럽처럼 프라이버시를 포괄적으로 보호하는 법률을 제정하려는 움직임이 있다. 그러나 미국은 워낙 수정헌법 1조의 알 권리를 강조하고 있기 때문에 프라이버시 규제 개혁이 실제 입법화될 지는 미지수다.

“ 정보를 보는
패러다임을 다시
설정해야 한다. ”

옳다고도 할 수 없다. 빅 데이터 시대에 맞는 사생활 보호와 개인 정보의 활용 범위에 관한 논의가 공론화되어야 하고, 사회적인 중지도 모아져야 한다. 정보를 보는 패러다임을 다시 설정해야 한다.

첫 번째 할 일은 개인 정보의 종류와 활용 범위를 좀 더 명확하게 규정하는 것이다. 명백한 개인 정보와 개인 정보가 아닌 일반 정보의 종류는 물론이고, 쿠키나 구매 기록, IP와 같이 개인을 특정할 수 없지만 다른 정보를 함께 이용하면 개인을 식별할 수 있는 비식별 개인 정보도 개인 정보의 범주에 두어 엄격하게 보호해야 하는지와 같은 논의가 진행되어야 한다. SNS에 소비자들이 공개한 개인정보도 마찬가지다. 제도가 명확해져야 개인은 자신의 정보가 중요함을 인식할 것이고, 기업은 최선의 이용 범위를 정할 수 있다.

개인의 정보 권리에 대한 논의도 필요하다. 개인 정보가 중요하다지만, 정작 개인이 자신의 정보를 통합하여 열람하거나, 외부에 노출된 자신의 개인 정보를 관리하는 것은 쉽지 않다. 의료 정보, 교육 정보, 금융 정보 등 생활에 필요한 모든 정보를 개인이 통합적으로 열람하고 관리할 권리를 보장해야 한다. 또한 정보를 축적한 기업이 개인의 요구에 따라 정보를 공개하고, 제 3자에 제공하거나 또는 완전히 폐기하는 의무가 정착되어야, 사용자들도 자신의 개인 정보 공개에 안심할 수 있게 될 것이다.

빅 데이터의 일차적 목적은 개인을 추적하는데 있는 것이 아니라 광범위한 자료 속에서 인과 관계를 찾아내는 데에 있으므로, 서비스 기업은 개인을 추적할 수 없는 비식별 정보만을 축적하고 개인 식별이 가능한 정보는 개인이 지정한 별도의 서버에 저장하는 방식도 생각해 볼 수 있다. 이른바 개인 정보 금고다. 개인 맞춤형 서비스를 제공하려는 서비스 업체는 개인 정보 금고에 접속하여 사용자의 상황이나 특성 정보를 이용하고, 사용자는 누가 어떤 목적으로 자신의 정보를 열람했는지 한 눈에 알 수 있다. 이처럼 제도가 잘 정의되고 정보 이용 가이드라인이 명확해지면, 그에 맞춘 새로운 사업 모델도 육성할 수 있다.

정보 유출에 대한 처벌과 수사는 지금보다 훨씬 강화되어야 한다. 개인의 권리와 사회적 경쟁력에 직결되는 문제인만큼, 사용자의 주의를 환기하는 수준을 넘어 철저한 수사와 엄격한 처벌 규정이 마련되어야 한다. 새로운 시대에는 그에 맞는 새로운 제도가 곧 경쟁력이다. 더 이상 머뭇거릴 시간이 없다. www.lgeni.com