

빅데이터 핵심 기술 및 표준화 동향

Standard Trends for the BigData Technologies

김정태 (J.T. Kim) 소셜컴퓨팅연구팀 연구원
오봉진 (B.J. Oh) 소셜컴퓨팅연구팀 책임연구원
박종열 (J.Y. Park) 소셜컴퓨팅연구팀 팀장

* 본 연구는 지식경제부 및 한국산업기술평가관리원의 IT산업원천기술개발사업의 일환으로 수행하였음(KI001877, 장소/사회적 관계 인지형 Social 미디어 서비스 기술).

최근 ICT 기술의 발전으로 저장 시스템이 저렴해지고, 많은 데이터를 빠르게 처리하는 것이 가능해지면서 빅데이터(BigData)에 대한 많은 관심이 집중되고 있다. 특히 스마트폰과 모바일 인터넷 서비스 활성화에 따라 사용자의 데이터 이용이 폭발적으로 증가하고 있다. 과거 초고속 인터넷 확산이 정보화에 크게 일조한 것과 같이 거대한 분량의 데이터는 사람과 비슷한 수준의 지식을 만들어 내는 빅데이터 처리 기술 발전의 근간이 되었고, 선진국을 중심으로 다양한 사례가 보고되고 있다. 본고는 확산 일로에 있는 빅데이터 관련 이슈를 정리하고 최근의 표준화 활동에 대해서 살펴본다.

빅데이터 처리 및
분석 기술 특집

- I. 개요
- II. 빅데이터와 집단지성
- III. 표준화 기구별 동향
- IV. 결론

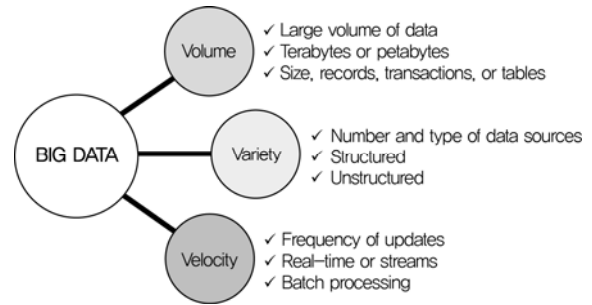
1. 개요

빅데이터는 큰 데이터를 말한다. 단순히 큰 데이터가 아니라 다음의 3가지 특징을 가지는 큰 데이터를 일반적인 빅데이터로 정의하고 있는데, 부피가 크고, 변화의 속도가 빠르며, 데이터의 속성이 너무도 다양한 데이터를 지칭한다.

산업계에서 빅데이터를 바라보는 시각은 크게 두 가지로 분류할 수 있다. 하나는 비정형 데이터를 중심으로 효과적으로 데이터를 처리하는 기술과 정형 데이터베이스에서 대규모 저장 시스템을 연구하는 분야이다. 지금까지 업계에서 이야기되고 있는 빅데이터는 전자에 속하는 분야이며, 후자의 빅데이터는 기존의 대용량 데이터베이스와 장비 시장을 장악하고 있는 글로벌 기업들을 중심으로 부각되고 있는 기술로 광의의 빅데이터를 의미한다.

또한 빅데이터를 데이터베이스 기술과 혼동하는 경우가 많으나 데이터베이스 기술은 이미 목적을 정하고 여기에 맞도록 데이터를 어떻게 하면 효과적으로 저장하고 분석할 수 있을까 하는 기술이다. 최근에는 이렇게 잘 가공된 데이터가 아닌 버려지고 있는 데이터들에 대한 관심이 부각되면서 기술에 변화를 가져왔다. 특히 대상이 되는 데이터를 전부 처리하는 것이 아니라 클러스터링과 필터링 방식을 통해 데이터를 효과적으로 걸러내는 기술이 등장하면서 더욱 활발해 졌다. 기존의 버려지고 있는 데이터들은 분석할 가치가 있는지를 알지 못하기 때문에 많은 비용을 들여서 처리하는 것이 비효율적이었다. 하지만 원시 데이터를 클러스터링하고 필터링하여 활용 가능성이 있는 데이터를 취사 선택하는 기술은 적은 비용으로 새로운 가치를 발굴할 수 있게 했다. 이는 데이터 처리 기술의 발전과 컴퓨팅 속도의 발전으로 가능하게 된 것이다.

(그림 1)은 빅데이터를 정의하는 3가지(3V) 특징에 대해서 설명하며, 데이터가 갖는 속성보다는 더 정확한



(그림 1) 빅데이터를 정의하는 3V 특징[1]

지식을 발굴하기 위해 필요한 구성 요소로 보는 것이 타당한 것으로 생각된다.

• 데이터 분량(Volume)

통계에서 표본 데이터가 많아야 정확도가 올라가는 것과 같이 분석하는 데이터의 크기가 일정 수준 이상이어야 의미 있는 데이터를 얻을 수 있다. 특히 전체 데이터를 처리하지 않고 데이터를 취사 선택하여 사용하는 빅데이터 기술에서 데이터의 분량은 분석 결과에 대한 신뢰성으로 귀결되는 문제이다. 따라서 데이터의 크기는 데이터가 갖고 있는 내재적 가치에 따라 다르기는 하지만 일반적으로 100테라바이트 이상의 데이터를 빅데이터라고 칭한다.

• 데이터의 변화 속도(Velocity)

빅데이터는 잘 가공된 데이터가 아닌 가공되지 않은 원시 데이터에서 가치를 찾는 것이다. 만약 데이터가 변하지 않는 정보라고 하면 다양한 분석 방법으로 수차례 분석을 시도했을 것이기 때문에 새로운 정보를 얻는 것은 쉽지 않다. 하지만 데이터가 계속 변하는 경우에는 변화에 따라 새로운 분석 방법, 새로운 가치 부여가 가능하기 때문에 빅데이터 분석의 좋은 예라 할 수 있다. 이는 빅데이터에서 소셜 네트워크 서비스가 항상 같이 등장하는 배경이기도 하다. 소셜 네트워크 서비스는 사람과 사람이 소통하는 것으로 끊임없이 새로운 데이터를 생성하고 소비하는 곳이다. 또한 데이터의 변화 속도가 빠르기 때문에 현재의 데이터를 포함할 가능성이 가

장 높은 곳이다.

- 데이터의 다양성(Variety)

다양성은 빅데이터에서 조금은 다른 의미를 가지고 있다. 생태계에서 종의 다양성이 생태계를 건강하게 하는 것과 같이 데이터의 다양성은 데이터가 만들어 내는 정보의 가치를 건강하게 만드는 것이다. 여기서 건강하다는 것은 사실에 가깝거나, 사람들이 체감적으로 공감하는 내용에 가깝다는 것을 의미한다. 이는 선거 여론 조사에서 표본의 다양성에 대해서 이야기하는 것과 같은 이치이다. 즉 데이터의 다양성이 떨어지면 한쪽에서 제공하는 일방적인 데이터를 반영할 확률이 높다. 또한 데이터의 출처가 한정되어 있다는 것은 데이터 정형화를 통한 가공/저장/분석이 용이하기 때문에 빅데이터보다는 정통적인 데이터베이스 기술의 적용이 더 효과적인 결과를 가져 온다.

II. 빅데이터와 집단지성

빅데이터와 함께 떠오르는 키워드에는 생각하는 기계, 즉 “기계가 사람 대신 기사를 쓴다[2]”, “기계와 사람의 대결[3]”과 같은 기계와 사람을 비교하는 글이 많이 있다. 빅데이터가 지식이라는 키워드를 널리 사용하게 하는 역할을 하기는 했지만, 아직 빅데이터가 지식을 생산하는 시스템으로 보기는 어렵다. 그럼 왜 사람들이 생각하는 기계에 관심을 갖는 것인가 의문이 생긴다. 이 답은 사람과 기계가 갖는 차이점에 대해서 이해해야 한다.

사람은 자신의 경험을 바탕으로 정신적으로나 육체적으로 성장한다. TV 방송 프로그램 중에 ‘생활의 달인[4]’이라는 프로그램이 있는데 한 분야에서 오랜 경험을 가지고 달인의 경지에 오른 사람들을 소개한다. 같은 일을 반복하면서 나름의 방식을 터득하고 남들보다 더 잘하게 되는 것인데, 사람은 같은 일을 반복할 때 항상 같지

않고 조금씩 다르게 수행한다. 이것이 사람들에게 변화를 주고 이렇게 하면 더 쉽게 할 수 있구나 하는 생각을 하게 된다. 이것이 학습이다.

그런데 기계는 그렇지 못하다. 같은 일을 하면 매년 정말 똑같이 반복한다. 따라서 기계가 계속 변화할 수 있도록 방식을 바꾸고 내부적으로 동작하는 것이 과거의 경험을 반영하는 것이다. 이것이 기계 학습(machine learning)이다. 기계가 사람처럼 생각하고 행동하도록 하는 것이다. 따라서 빅데이터가 만드는 정보를 좀 더 정확하게 하기 위해서는 기계 학습이 병행되어야 하며, 생각하는 기계에 대한 이해도 여기서 출발해야 한다.

기계가 가지고 있는 근본적인 특징은 단순하고 반복적인 일을 잘 한다는 것이다. 나쁜 의미로 아무리 반복해도 좋아지지 않는다는 것이다. 그런데 이 근본적인 특징이 항상 나쁘지만은 않다. 사람은 학습에 의해서 발전하고 지능을 넓혀 가지지만, 반대로 자신의 생각 때문에 잘못된 판단을 하는 경우도 종종 있다.

예를 들면, 주변 친구들 중에 항상 이 친구가 만지면 컴퓨터가 고장이 나는 친구가 있다. 이 친구가 하는 말은 나는 한 것도 없는데 컴퓨터가 고장 났어? 라지만 기계(컴퓨터)는 거짓말을 하지 못한다. 어딘가 잘못된 것이 있다. 사람이 자신의 생각 속에서 스스로의 실수를 의식하지 못할 뿐 기계는 정확하게 반응한다. 기계는 입력에 대한 반응을 보이는 것으로 데이터를 가지고 이야기하지 감정이나 생각을 이입하지 않는다. 데이터가 잘못되어 있다면 결과도 틀리고, 데이터가 정확하다면 결과도 정확한 것이다.

빅데이터는 작은 데이터를 가지고 사람이 자의적으로 해석하는 것이 아니라 많은 데이터를 가지고 정확한 기계적 통계 바탕의 결론을 제공하는 것이다.

그러나 왜 데이터가 모이면 더 정확하고 심지어 미래를 예측하는 것도 가능한 것인가? 여기에 대한 답은 집단지성(collective intelligence)과 크라우드 소싱(crowd sourcing) 개념에서 출발한다.



(그림 2) 전문가 vs. 집단지성[10]

(그림 2)에서 보듯 집단지성은 다수의 사람들이 서로 협력하거나 경쟁해서 지식을 더욱 고도화해서 얻은 집단적 지능을 말한다. 비슷한 말로는 중지, 집단지능, 협업지성, 공생적 지능을 의미한다. 이러한 개념은 제임스 서로위키(James Surowiecki)가 실험을 통해 제시한 것으로 한 명의 우수한 사람의 대답보다는 평범한 여러 명의 대답이 더 우수한 결과를 만든다는 것[5]이다. 인기 TV 프로그램 중에 ‘1 대 100[6]’이라는 프로그램에서 100인의 찬스가 같은 개념이다.

온라인에서 검증된 집단지성의 대표적인 사례가 위키피디아(Wikipedia) 무료 온라인 사전이다[7]. 기존의 백과사전은 편집 위원들이 방대한 분량의 데이터를 분류, 가공, 추가한 다음 잘 정리된 책으로 출판한다. 반면 위키피디아[8]는 사용자가 검색하는 키워드를 페이지로 제공하고 누구나 정보를 올릴 수 있도록 하였다. 사용자들은 가장 타당성이 있고 잘 정리된 페이지를 투표하여 대표 페이지로 만든다. 처음에 위키피디아에 올려진 데이터는 작은 수준이었지만, 점차 사용자가 늘고 사람들의 참여가 많아지면서 위키피디아는 세계에서 가장 인정받는 사전으로 등극했다. 집단지성의 사례로 펩시가 Pepsi Refresh Project[9]에서 일반인들의 아이디어를 받고 투표를 통해 결정한 사례나 Dell이 자사의 컴퓨터에 대한 민원을 처리하고, 더불어 아이디어를 추천받는 IdeaStorm[10] 사이트를 운영하면서 큰 성공을 거둔 사례를 들 수 있다.

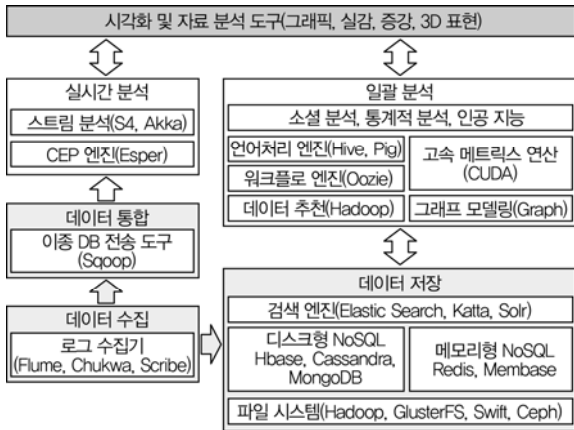
크라우드 소싱은 일을 대중(crowd)에게 일부 나눈다(sourcing)는 개념으로 기업이 상품 기획, 판매, 홍보 등에 소비자가 참여할 수 있도록 일부를 개방하고 소비자가 자발적으로 참여하여 기업의 수익을 올리면 기업이 수익을 공유하는 방식이다. 2006년 와이어드 매거진(Wired Magazine)의 제프 하우(Jeff Howe) 객원 편집자가 처음 제시한 개념으로 기존의 기업 활동이 내부의 인력으로 추진되다 보니 새로운 아이디어 도출이 어렵고 소비자의 다양한 욕구를 충족시키지 못했던 문제점을 개선한 것으로 우수 인력 중심의 경영이 아니라 다양한 인력의 의견과 생각을 모아 경영에 반영하는 것이다.

집단지성과 크라우드 소싱은 빅데이터와 밀접한 관련이 있는 것으로 빅데이터의 3V 특징 중에서 규모와 다양성이 중요하게 된 것은 집단지성과 크라우드 소싱의 특징을 반영한 것이다. 사람이 생성하는 데이터나 기계가 생성하는 데이터에서 일정 규모 이상 그리고 다양한 종류의 데이터를 확보하는 것은 결국 데이터에서 새로운 가치를 만들어 내는 데 중요한 역할을 하는 것이다.

1. 오픈 소스 중심의 빅데이터 처리 기술

빅데이터 분석의 기술적 배경이 되는 분산 처리 및 다양한 연산 기술은 오픈 소스의 역할이 크다. 이는 Facebook, Google[11]과 같은 글로벌 기업들이 자사의 초기 프로그램을 오픈 소스로 공개한 것에 큰 영향을 받았다. 오픈 소스는 프로그램을 공개하는 기구이지만, 그 분야의 실질적인 표준으로 자리를 잡는 경우가 많다. 하둡(Hadoop)의 경우도 이미 개발자들에게는 바이블로 통할 정도로 그 활용도가 높은 상황이다.

우선 빅데이터를 구성하고 있는 기술들을 보면 (그림 3)와 같이 데이터 수집, 저장, 분석, 시각화의 4가지 과정[12]을 거치게 되며, 실시간 분석의 경우 데이터 수집, 통합, 실시간 분석, 시각화의 단계를 갖는 반면 일괄 분석 방법은 데이터 수집, 저장, 일괄 분석, 시각화의 특징을 가진다. 일괄 분석은 전체를 다시 분석하는 것으로



(그림 3) 빅데이터 관련 오픈 소스 프로젝트

아직은 하루 정도의 시간이 걸린다. 반면 실시간 분석은 이전의 분석 결과에 새롭게 생성된 데이터를 추가 학습하는 방법으로 (그림 3)과 같이 데이터를 수집하고 스트림으로 제공하는 것이 중요한 기술이다.

2. 클라우드 기반의 빅데이터 처리 기술

빅데이터에서 또 다른 동향은 클라우드 기반의 빅데이터 처리 기술이다. 기업에서 빅데이터를 가장 어려워하는 부분은 하부의 저장 시스템을 구축해야 한다는 것이고, 지금은 데이터가 적지만 기하급수적으로 늘어나기에 시스템을 도입하려고 하는 기업에서는 부담이 될 수 밖에 없다.

최근의 인터 클라우드의 개념이 등장하면서 서로 다른 클라우드를 사용하는 경우 혹은 물리적으로 떨어진 기관이 데이터를 연동하기 위한 방법으로 (그림 4)와 같이 클라우드 기술이 많이 활용되고 있다. 다만 많은 기업들이 공공 클라우드보다는 사설 클라우드를 활용해서 자신들의 빅데이터를 분석하는 방법을 취하고 있다.

구글은 자사의 빅데이터 분석 서비스를 클라우드 기반에서 제공하는 빅쿼리[13] 제품을 출시하면서 국내외적으로 많은 기업들이 온라인 분석, 즉 클라우드 기반의 빅데이터 분석에 관심을 갖게 되었다. 클라우드 기반의



(그림 4) 클라우드 기반 빅데이터[13]

빅데이터 분석은 새로운 서비스를 제공하지만 내부의 데이터를 공개하지 않아도 되는 특징이 있다. KT-SKT가 사용자들의 이동 경로를 통해 상권을 분석하는 것도 사용자들의 이동 경로 정보를 내부에 가지고 있지만, 상권 분석을 제공하는 서비스에서 이 데이터를 공개하지 않는 특징이 있다. 만약 KT-SKT가 외주를 주고 기술 개발을 하는 경우 기존의 가입자의 동의를 받아야 하기 때문에 어려움이 예상되는 부분이다.

III. 표준화 기구별 동향

빅데이터 기술에 대한 표준화는 많은 기구에서 관심을 갖고 있지만 접근 방법은 표준화 기구별로 약간씩 상이한 특징을 가지고 있다. 특히 공적 표준화 기구와 사실 표준화 기구의 움직임이 상이한 구조를 가지고 있다. 다음에서는 각각의 표준화 기구별로 빅데이터 표준화를 추진하는 방향에 대해서 정리한다.

1. 공적 표준화 기구 동향

가. ISO/IEC JTC 1

ISO/IEC JTC 1에서는 빅데이터에 대한 본격적인 표준화 작업에 앞서 관련된 SC(Sub-Committee)의 활동 상황을 점검하고 전체 JTC 1 차원에서 대응하는

전략을 마련 중에 있다. 특히 SWG-P(Special Working Group-Planning)에서는 항시 이슈화되고 있는 기술에 대한 표준 항목을 도출하고 개별 SC와의 협력 방안을 마련하고 있다. 빅데이터 기술 역시 2012년 11월 제주 총회에서 언급된 후 구체적인 표준화 대응을 준비하고 있다. 공식적으로 관심을 표명하는 그룹은 SC32, SC29, SC27, SC23, SC7이다.

ISO/IEC JTC 1/SC32에서는 데이터 관리 및 교환에 대한 표준을 추진하고 있다. 특히 국가 간 데이터 교환이라는 취지에서 빅데이터의 데이터를 표현하는 방식에 대한 표준 제정에 관심을 갖고 있다. 특히 지난 2012년 6월 4일 회의에서 클라우드 컴퓨팅에서 데이터 교환을 위한 연구 그룹이 만들어지기도 했다.

ISO/IEC JTC 1/SC23에서는 정보교환 및 저장을 위한 디지털 미디어 표준을 추진하고 있다. 데이터 압축 기술 및 빅데이터 관련 데이터 저장 기술에 대한 관심을 갖고 있고, 세부적인 추진 전략이나 표준화 항목을 제시하고 있지는 않다. 하지만 클라우드 기반의 서비스에서 데이터의 압축 저장 방식에 대한 표준화가 추진될 가능성이 높은 상황이다.

ISO/IEC JTC 1/SC27에서는 IT 보안에 관련된 표준을 개발하고 있다. 빅데이터 분석이 활성화되면서 빅데이터 분석을 통한 개인의 사생활 침해 가능성에 대해서도 많은 논의가 진행되고 있다. 특히 정형 데이터에서 개인의 사생활 정보를 취급하는 방법이나 분석에서 개인 정보를 제거하는 방법에 대한 연구가 진행되고 있으나 아직 표준화 이슈가 되지 못했다. 다만 SC27에서는 개인의 사생활 문제 등 표준화가 필요한 항목에 대한 기본 조사를 추진하고 있다.

ISO/IEC JTC 1/SC29 WG11(MPEG)에서는 빅데이터의 비정형 데이터의 표현을 정규화하는 것을 추진하고 있으며, 의미 기반 데이터 처리를 위해 자연어 처리 기술을 바탕으로 데이터의 중의적 의미를 표현하는 표준을 추진하고 있다. 현재 MPEG Preservation 그룹에서

CFP가 발의된 상황이다.

ISO/IEC JTC 1/SC7(ISO/IEC 25012)에서는 소프트웨어와 시스템공학 기술 표준을 개발하고 있다. 빅데이터 시스템에 대한 컴포넌트 참조 시스템에 대해 관심이 있으며, 클라우드 기술과 빅데이터 기술에 관심을 두고 있다.

나. ITU-T

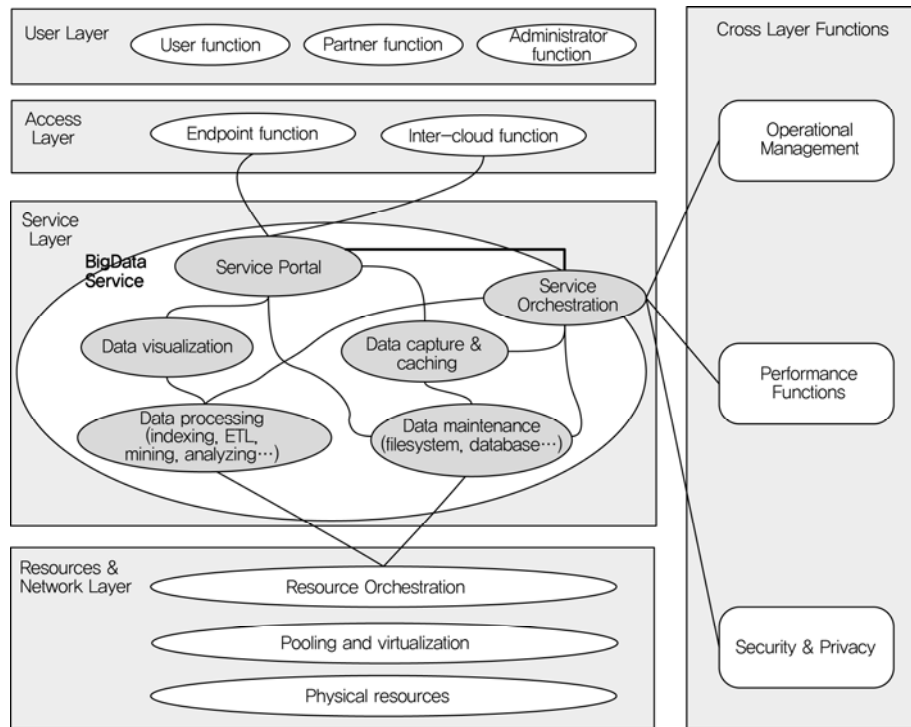
ITU-T에서는 SG13, SG16, SG17에서 관련된 표준화 활동에 대한 논의가 시작되고 있다. SG13은 미래 인터넷 측면에서 접근하기 때문에 클라우드 기반의 빅데이터 분석 시스템에 대해, SG16에서는 멀티미디어 빅데이터 분석에 관심을 갖고 있다. SG17에서는 빅데이터 분석에 따른 개인 사생활 보호에 대한 관심이 생기고 있다.

SG13은 미래인터넷에 대한 표준을 개발하는 곳으로 클라우드 기술에 대한 표준을 주도하고 있다. 빅데이터 기술은 클라우드 기반의 빅데이터 분석 기술을 추진하고 있으며, (그림 5)와 같이 2012년 여름에 중국에서 빅데이터 분석 플랫폼을 처음 제안하고 클라우드 플랫폼에서 빅데이터를 포함하기 위한 표준 개발에 착수했다.

향후 표준화 추진이 유력시되는 기술은 클라우드 기반의 빅데이터 시스템 참조, 클라우드 내부에서 빅데이터 지원을 위한 시스템 구조 표준, 클라우드 상위의 응용들 사이의 호환성 확보 표준이 가능하다. 현재 중국이 제안한 표준은 2013년 2월 회의까지 보류되어 있는 상황이며, 한국에서 전략적 대응을 준비 중에 있다.

3. 사실 표준화 기구 동향

사실 표준화 기구에서는 W3C에서 빅데이터 CG(Community Group)가 2012년 4월에 생성되었으며, 표준화 활동에 관심 있는 멤버를 모집 중에 있다. 한국은 2명이 참여하고 있으며 빅데이터 처리를 위한 표준 구조, 프로그램 API 정의를 목적으로 하며 상호 호환성, 보안, 저



(그림 5) 중국에서 제안한 BDaaS 구조

비용이라는 특징을 반영할 계획으로 있다.

ODCA(Open Data Center Alliance)는 클라우드 표준화 및 클라우드 기반의 오픈 서비스를 표준화하는 단체로 빅데이터의 중요성이 강조되면서, 2012년 4월 30일 데이터 서비스 작업반(Data Service Workgroup)을 만들고 안전하게 데이터를 수집, 관리, 분석할 수 있는 표준과 기존 BI(Business Intelligence) 솔루션 사이의 상호 운영성을 보장할 수 있는 표준 개발을 추진할 예정이다. 이 작업반에는 Cloudera, Hortonworks, MapR Tech., Teradata, SAS 등 빅데이터 기업들이 참여하고 있다.

ODI(Open Data Institute)는 영국 내각사무처에서 설립한 연구기관으로 공공데이터의 활용 기회를 높여 국가 전반의 혁신을 촉진하고 공공 데이터의 잠재력을 활용한 새로운 비즈니스 발굴을 위해 2012년 9월부터 향후 5년 동안 1,000만 파운드(약 180억 원)를 지원한다. ODI는 사설 표준화 기구는 아니지만, 공공 데이터를 공

유하고 개방하는 방법에 대해서 선도적 역할을 수행할 것으로 예상되기 때문에 산업계의 사실상의 표준으로 자리 잡을 가능성이 높다.

IV. 결론

빅데이터 기술은 ICT 전반에 걸쳐 다양한 이슈들이 관련되어 있는 기술이다. 특히 인공지능, 저장 시스템에 관련된 많은 연구가 진행 중에 있으며, 표준화 역시 관련 영역에서 병렬 추진되고 있다. 하지만 많은 표준화 기구에서 공격적으로 추진하기보다는 조심스럽게 접근하고 있는 것이 지금의 상황이다.

ITU-T SG13과 ISO/IEC JTC 1/SC32에서 지금까지는 가장 적극적인 대응을 하고 있다. 이는 빅데이터의 주요 이슈가 클라우드 기술과 결합한 시스템 표준화를 목적으로 빅데이터에서 분석의 대상인 데이터의 재활용 및 상호 호환에 초점을 두고 있기 때문이다.

하지만, JTC 1 산하의 SC32, SC29, SC27, SC23, SC7에서 빅데이터에 관심을 갖는 것과 같이 다양한 분야 다양한 영역에서 표준화가 확장될 것으로 판단된다. 하지만, 국내 기술이 아직 성숙되지 않은 상황이다.

따라서 국내 기술을 지원하는 표준화보다는 향후 기술의 발전 방향을 미리 점치고 선점하는 전략적 접근이 필요하다.

용어해설

원시 데이터 사람, 기계, 사회적 관계에서 생성된 데이터로 생성된 상태 그대로 혹은 낮은 수준으로 가공된 데이터

테라바이트 TB, 10^{12} bytes 혹은 1,000Gbyte 크기의 단위이며 약 20만 개의 MP3 파일을 저장하는 수준의 크기

약어 정리

BI	Business Intelligence
CG	Community Group
IEC	International Electrotechnical Commission
ISO	International Organization Standardization
ITU	International Telecommunication Union
JTC1	Joint Technical Committee 1
ODCA	Open Data Center Alliance
ODI	Open Data Institute
SC	Sub-Committee
SG	Study Group
SWG-P	Special Working Group-Planning

참고문헌

- [1] TDWI Research, "Big Data Analytics Report," 2011.
- [2] Narrative Science, "Narrative Science," We Transform Data into Stories and Insight. <http://www.narrativescience.com/>
- [3] IBM, "The DeepQA Project,". <http://www.research.ibm.com/deepqa/deepqa.shtml>
- [4] KBS, "생활의 달인,". <http://tv.sbs.co.kr/lifemaster/>
- [5] J. Surowiecki, *The Wisdom of Crowds Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business*, Doubleday, 2004.
- [6] KBS, "KBS 1 vs 100,". <http://www.kbs.co.kr/2tv/entert/1vs100/>
- [7] 유재훈, "대중의 지혜를 내 것으로, 클라우드 소싱에 성공하려면," LG Business Insight, Weekly 포커스, vol. 1089, LG경제연구원, 2010. 4.
- [8] Wikipedia, The Free Encyclopedia. <http://en.wikipedia.org/>
- [9] Pepsi, Pepsi Refresh Project. <http://www.refresheverything.com/>
- [10] Dell, IdeaStorm. <http://www.ideastorm.com/>
- [11] Google Translate. <http://translate.google.com/>
- [12] 이성춘, 임양수, 안민지, "Big Data, 미래를 여는 비밀 열쇠," KT경제경영연구소, 2011.
- [13] Answer Tough Big Data and Big Analytics Questions with Google BigQuery, 2012. <http://www.enor.com/blog/web-analytics/answer-tough-big-data-and-big-analytics-questions-with-google-bigquery>