

빅데이터의 이해와 도서관 정보서비스에의 활용*

Understanding Big Data and Utilizing its Analysis into Library and Information Services

이 정 미 (Jeong-Mee Lee)**

초 록

본 연구는 빅데이터 시대의 도래와 함께 다양한 분야에서 연구의 초점이 되고 있는 빅데이터를 재조명한 것이다. 빅데이터의 개념, 제기되는 중요한 이슈, 도서관 정보서비스에의 활용이라는 세 가지 연구문제를 각종 문헌과 사례를 검토하여 밝히고자 하였다. 연구결과 빅데이터의 개념은 현실사회의 맥락에서 분석, 가치가 부가되어야 그 의미가 있음을 밝혀냈으며, 데이터의 정확성이나 신뢰성에 대한 문제, 개인정보보호, 보안, 윤리적 이슈, 지적재산권 등 다양한 이슈가 빅데이터 연구와 연관되어 제시됨을 알 수 있었다. 또한 이러한 이슈에 대한 문제제기를 이해한 가운데 빅데이터는 도서관 정보서비스에 활용될 수 있다고 판단했는데 도서관의 비전과 전략 수립에 도입하거나 지역사회를 지원하고 맞춤형 정보서비스를 제공하기 위한 근거를 제시하는 등 효과적으로 활용될 수 있다고 결론지었다.

ABSTRACT

This study revisits issues for Big data. Three research questions, understanding the concept of Big data, important issues of Big data research and utilization methods for library information services, are explored by the literature and practice reviews. Study results revealed several important issues of Big data including the concept in the context of real world situation, the problems with the accuracy and reliability of the data, privacy and ethical issues, and issues of intellectual property rights. With understanding these issues, a few utilization methods were introduced for Library and Information services. It was included using its analysis for developing vision, adopting Library management, supporting community services, and providing customized information services for various users. The study concluded Big data analysis would effectively provide valid evidences for all those services.

키워드: 빅데이터, 빅데이터 분석학, 정보서비스, 데이터마이닝, 문헌정보학
Big Data, Big Data Analysis, Information Services, Data Mining,
Library and Information Science

* 이 논문은 2013학년도 서울여자대학교 교내학술특별연구비의 지원을 받았음.

** 서울여자대학교 문헌정보학과 부교수(jmlee@swu.ac.kr)

논문접수일자 : 2013년 11월 23일 논문심사일자 : 2013년 11월 26일 게재확정일자 : 2013년 12월 18일
한국비블리아학회지, 24(4): 53-73, 2013. [http://dx.doi.org/10.14699/kbiblia.2013.24.4.053]

1. 서론

1.1 연구의 필요성

인터넷의 대중화와 정보기술의 혁신적 발달 및 모바일 기기의 급속한 보급은 언제 어디서나 네트워크에 접속할 수 있고 이로 인해 사람들이 좀 더 쉽게 정보처리가 가능한 환경을 제공하고 있다. 정보 기술의 활용이 당연해지면서 각종 SNS(Social Network Service 이하 SNS)의 활용과 동시에 사람들은 단순히 SNS를 거쳐가는 것이 아니라 SNS를 통해 수집된 정보를 이용해 자신만의 정보를 재생산하고 타인에게 전달하는 과정을 자연스럽게 행하고 있다.

이와 같은 정보기술의 발전과 이용확산은 언제부터가 데이터의 양적 팽창을 가져왔으며 더불어 거시적 입장에서 데이터의 흐름을 주시할 필요성이 여러 분야에서 제기되었다.

이런 사회적 흐름을 바탕으로 가트너 그룹은 전략적 빅데이터를 2013년 기업의 전략적 대응 10대 기술 및 트렌드로 선정했으며(Gartner 2012) 우리나라도 2013년 국가정보화 트렌드 설문조사 결과를 바탕으로 2013년 10대 국가정보화 키워드를 선정하면서 클라우드, 소셜네트워크, 태블릿PC 등과 함께 중요한 하나로 빅데이터를 선정하고 더불어 빅데이터 분석이 중요한 이슈로 대두될 전망이다(김진숙 2013).

빅데이터가 활발히 형성되고 알려진 가장 큰 이유는 정보통신기술의 발전과 괄목할 만한 성장에 있다고 할 것이다. 특히 SNS의 성장은 빅데이터 연구의 발전에 큰 요인이 되고 있다(강학주 2012). SNS 사용자들은 자신의 일상적 사회 활동을 하며 스마트폰과 SNS를 통해 다른

사람들과 소통하고 정보를 교환하는 등의 다양한 상호작용을 함으로써 오프라인의 생활과 온라인에서의 소통이 결합된 삶의 패러다임을 영위하고 있다.

이러한 패러다임의 변화는 초기에는 단순한 소통 미디어였던 소셜미디어를 정보기술의 중심 주제이며 대상으로 올려놓았으며 당연히 소셜미디어는 정보기술과 사회를 이야기할 때 화제의 중심에 있다. 소셜미디어 기반 데이터는 기존에 우리가 접할 수 있었던 데이터보다 훨씬 방대하고 다양한 구조를 가지고 있다. 또한 사용자들 사이에서 직접적인 상호 소통을 통해 생성되고 공유되며 교환되는 형태를 띠고 있기 때문에 상호작용이 일어나는 바로 그 시점에서의 사회 모습이나 사람들의 행동 양식을 나타내고 반영하고 있다고 볼 수 있다. 따라서 소셜미디어에 대한 이해는 해당 시점에 속하는 사회 전반의 모습을 이해할 수 있다는 것과 같으며 바로 이런 이유 때문에 비구조적이고 복잡한, 방대한 데이터를 이해하고 필요한 지식을 뽑아내서 현재 주요하게 떠오르는 이슈들을 파악하고 미래를 예측하기 위한 연구, 즉 빅데이터 활용연구가 활발히 진행되고 있는 것이다(이충희 외 2013).

정보기술의 발전은 인간에게 좋다 나쁘다의 단편적 판단이 아니라 이를 어떻게 활용하느냐가 중요한 관건으로 이해되어야 한다. 빅데이터 시대도 같은 맥락에서 이해되어야 할 것이다. 빅데이터 분석이 단지 데이터 양이 많으니까 좋을 것이다 라는 막연한 기대가 아니라 그 많은 양의 데이터를 기반으로 새로운 방법을 모색하고 활용함으로써 현재 사회의 모습을 또는 현상을 이해하는데 유용한 맥락적 데이터를 생산하

는 것이 더욱 중요한 것이다.

많은 연구자들이 빅데이터에 관심을 가지고 활발히 연구를 진행 중이다. 이는 또 다른 한편으로는 빅데이터가 가져오는 문제점에 대한 우려를 내비치기도 한다. 우리가 빅데이터를 어떻게 활용할 것인가 하는 문제는 매우 중요한 문제로서, 이는 빅데이터 현상이 불확실하기도 하고 급격한 변화의 가능성도 있으며 지금 우리가 어떻게 활용할 것인가에 대해 내린 결정이 미래를 좌우하므로 중요하게 인식되어야 한다.

빅데이터 연구의 궁극적인 목적은 빅데이터 분석을 통해 해당 현상, 시점을 이해하고 그 흐름을 파악해 미래를 예측할 수 있는 근거를 찾음으로써 그 가치를 추출하는 것이다. 간단히 말해 폭증하는 데이터를 효과적으로 수집, 관리하여 그 속에서 유용한 정보를 추출해내는 것이 관건이다. 빅데이터 연구는 빅데이터 자체에 주목하는 것이 아니라 빅데이터를 이용한 분석이 가져올 수 있는 데이터의 가치에 주목해야 하며 이는 빅데이터를 활용하면 더 좋은 예측, 더 현명한 의사결정이 가능해질 수 있음을 의미하는 것이라 하겠다.

2012년 3월 오바마 행정부는 빅데이터의 효과적인 이용을 목적으로 하는 Big Data Research and Development Initiative를 발족했다(Kalil 2012). 우리나라의 경우 같은 해 11월 “스마트 국가 구현을 위한 빅데이터 마스터플랜”을 발표했다. 이는 빅데이터 분석에 대한 연구가 사회에 미치는 영향이 국가적 차원에서 활용될 필요가 있음을 보여주는 한 단면이기도 하고 국가 경쟁력을 높이기 위해서도 반드시 필요하다고 할 것이다.

현재 세계 각국은 마케팅, 비즈니스 차원에

서도 빅데이터 연구를 근거로 한 예측, 전략수립을 활발히 진행 중이며 다른 많은 분야로 확산되고 있는 실정이다. 문헌정보학 입장에서 이 흐름은 간과할 수 없다. 이는 문헌정보학이 밟고 있는 현실의 많은 이용자와 그들의 정보추구행태에 대한 것이기 때문이다. 바로 이런 현실이 빅데이터를 이해하고 발생가능한 문제점에 대해 살펴봄과 빅데이터 연구를 문헌정보학과 도서관 정보서비스에 활용할 방안을 고민해야만 하는 이유인 것이다.

1.2 연구목적 및 연구질문

본 연구는 빅데이터의 개념을 이해하고 문제점을 살펴본 가운데 도서관 정보서비스에 적절히 활용할 방안을 고민하는 것이 중요한 목적이다. 이에 다음과 같은 세 가지 연구질문을 제시하여 살펴보고자 한다.

연구질문 1. 빅데이터의 개념은 무엇이며 왜 문헌정보학에서 주시해야 하는가?

연구질문 2. 빅데이터 연구에서 주목해야 할 문제점 또는 제기되는 이슈는 무엇인가?

연구질문 3. 빅데이터 연구를 도서관 정보서비스에서 적용한다면 그 활용방안은 무엇인가?

2. 빅데이터

2.1 개념 및 기술적 전제

빅데이터는 그 용어의 생소함 때문에 이 세상

에 존재하지 않다가 갑자기 나타난 신개념이라고 생각하게 될 수도 있으나 이는 사실 그렇지 않다. 실제로 빅데이터는 그 이전부터 개념적, 실체적으로 존재했으나 당시에는 데이터의 방대함 때문에 이 데이터를 어떻게 다루어야 할지 몰랐던 개념이라고 생각하는 것이 더 적절할 것이다.

Rousseau(2012)는 1990년대 초반부터 빅데이터 연구의 기반이 다져졌다고 본다. 알고 있고 보고는 있으나 제대로 처리하지 못했던 빅데이터를 하드웨어와 소프트웨어, 네트워크를 비롯한 정보기술의 현저한 발전으로 인해 제대로 제어할 수 있는 능력이 갖춰져서 실행에 옮길 수 있게 된 것이라는 해석이다.

동일한 맥락에서 Miller(2013) 또한 자신의 최근 연구에서 빅데이터를 “수년 동안 존재해 왔으나 지금에 와서야 더 빨리, 더 큰 규모로 적용되어 더 많은 이용자가 접근 가능한 분석적 기술”이라고 정의 내렸다.

Loukides(2012)는 오늘날의 big이 내일은 medium, 다음주는 small이 될 수 있기 때문에 “빅데이터”의 가장 효과적인 개념은 데이터의 사이즈 자체가 연구문제의 일부일 경우일 것이라고 주장하기도 했다.

이와 같이 빅데이터의 개념은 데이터의 크기를 이야기하는 것에서 시작하지만 단지 그 크기의 중요성에 비롯되지는 않으며 어디에서 갑자기 나타난 새로운 개념이 아니라 오래전부터 존재는 해왔으나 적절한 처리, 통제 방안을 몰라 방치해두었던 방대한 양의 데이터를 처리·통제하는 실제적인 분석활동이 가능해졌기 때문에 최근 몇 년간 재조명되고 활발한 논의가 시작됐으며 관심의 집중을 받게 된 개념으로

이해하는 것이 타당할 것이다.

이런 의미에서 빅데이터는 그냥 커다란 데이터라는 개념으로 단순하게 이해될 것이 아니라 많은 정보기술 연구의 경계선을 넘나드는 하나의 동력이며 활동 같은 움직임과 함께 통합적인 이해가 필요할 것이다.

우리나라에서 받아들여지는 빅데이터의 개념은 연구자들마다 조금씩 다르게 정의내리고 있으나 가장 간단한 개념정의는 다음과 같이 내려지고 있다.

“일반적인 데이터베이스 소프트웨어가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터”(김연진 2013)

“기존의 방법이나 도구를 사용하여 데이터의 수집, 저장, 분석 및 관리가 어려운 방대한 정형 또는 비정형 데이터”(강윤희 2013)

즉 빅데이터의 가장 단순한 개념은 직접적이며 단순한 형태로 다루고 처리하기 어려운 방대한 양의 데이터를 의미한다. 기존 데이터에 비해 너무 방대해 일반적으로 사용하는 방법이나 도구로 수집, 저장, 검색, 분석, 시각화 등을 하기 어려운 정형 또는 비정형 데이터 집합을 의미하는 것이다.

최성곤과 그의 동료들은 현재 빅데이터의 개념은 외부데이터, 비정형, 실시간 데이터 및 서로 상이한 정보의 결합으로 인한 새로운 지식의 창출을 포함할 수 있는 영역으로까지 그 의미가 확대 해석되고 있는 실정(최성곤 외 2012)이라 주장하며 가장 간단한 빅데이터 개념에서 한발자국 나아가 있는 입장이다.

2011년 McKinsey 보고서에서는 빅데이터

를 “보통의 데이터베이스 소프트웨어 도구로 수집, 저장, 관리, 분석하기엔 그 능력의 한계를 넘어서는 방대한 사이즈의 데이터셋”을 말하는 것이라 규정(Manyika 2011)했는데 이 개념정의는 “비용효과적이며 혁신적인 형태의 정보처리를 요구하는 방대한 양, 속도와 다양한 정보로서 통찰력을 증진시켜주고 자동처리됨으로써 의사결정을 증진시킨다”는 가트너 연구소의 정의(LeHong and Laney 2013)와 더불어 연구자들에게 가장 많이 사용되고 있다.

빅데이터는 그냥 방대한 양(big)인 것이 아니라 다양한 데이터 타입과 스트리밍 데이터들이 존재한다. 그래서 빅데이터를 연구하는 학자들은 빅데이터 연구는 단순히 빅데이터 연구다 라고 정의하는 것보다는 빅데이터 분석학이라고 부르는 것이 훨씬 타당할 것이라고 주장하고 있으며 빅데이터 분석학은 이 빅데이터 세트를 적용해 “의미를 해석”하고 “가치를 부여”하는 고급분석기술이다.

빅데이터 분석연구들은 연구 네트워크와 클라우드 컴퓨팅 기술을 바탕으로 한 mapping-oriented 시각화 연구들과 e-research 사례들이 급격하게 발달함으로써 2000년대 후반부터 활발해지기 시작했다(Jankowski 2009).

이러한 기술의 발달은 감당하기 어려운 데이터의 처리를 가능하게 했기 때문에 실제적으로 빅데이터 처리의 기본 전제가 되는, 동반자가 되는 기술들이다.

빅데이터는 다양한 방법으로 처리될 수 있지만 빅데이터 분석의 가장 기본적인 목적은 기술적, 사회적, 경제적 환경에서 존재하는 방대한 데이터를 뽑아내서 흐름을 파악하고 그 안에 숨겨진 패턴을 찾아내는 것에 있다(Park and

Leydesdorff 2013).

빅데이터의 처리를 가능하게 한 전제는 많은 기술적인 발전이다. 이를 통해 감각, 수집, 저장과 분석이 가능한 통합적 기술 처리가 가능해졌기 때문에 빅데이터 분석연구가 제대로 이루어질 수 있었다. Bryant와 그의 동료들(Bryant, Katz, and Lazowska 2008)은 빅데이터 처리를 가능하게 하는 기술적 전제들을 6가지 대표 기술로 표현한다. Bryant와 그의 동료들은 이외에 이 기술들을 둘러싼 다양한 기술적 이슈들도 제시하고 있다는 점에서 흥미롭다. Bryant와 그의 동료들이 제시한 6가지 대표 기술은 <표 1>에 기술되어 있다.

빅데이터란 오늘날의 조직들이 방대할 뿐 아니라 너무 빨리 성장하는 데이터 정보원이나 정보원을 취급하는데 직면하는, 또한 직면하는 문제들을 분석하고 이용하는 복잡한 범위를 표현하는 데 있어서도 더욱 커져가는 도전(Villars, Olofson, and Eastwood 2011)이라 이해되기도 한다.

어느 분야에서 빅데이터 연구가 진행되던 분야와는 상관없이 빅데이터 연구의 실행에 있어서 중요한 부분인 가치에 대한 평가는 다음 중 하나 이상의 평가로 이루어질 수 있을 것이다(Villars, Olofson, and Eastwood 2011과 McAfee and Brynjolfsson 2012에서 발췌, 재구성).

- ① 빅데이터 분석 연구의 결과가 우리가 기존에 진행했던 다른 연구결과들과 비교해볼 때 좀 더 유용한 정보를 제공해줄 수 있는가?
- ② 빅데이터 분석 연구의 결과로서 제시된 정보가 해당 주제에 대한 정보의 적합성을 높여주는가?

〈표 1〉 빅데이터 분석 연구를 가능케 한 전제 기술

전제기술	내용
Sensors	다양하고 방대한 데이터를 찾아낼 수 있는 다양한 형태의 센서 기술의 발전
Computer networks	인터넷 뿐 아니라 지역 센서 네트워크를 통해 수집되는 방대한 양의 다양한 데이터를 수집할 수 있도록 하는 컴퓨터 네트워크 기술의 발전
Data storage	많은 양의 데이터를 저장할 수 있는 저렴한 하드웨어 저장소의 생산 기술의 발전
Cluster computer systems	방대한 양의 데이터를 저장할 뿐 아니라 저장된 데이터를 그룹핑, 즉 분류하고 분석하여 조직할 수 있는 컴퓨팅 기술의 발전
Cloud computing facilities	새로운 방식의 저장방식으로 일종의 사이버 공간 렌트, 이를 통한 유비쿼터스식 데이터 접근을 가능하게 만든 기술의 발전
Data analysis algorithms	패턴을 알아내고, 지식을 추출하기 위한 데이터 분석 알고리즘으로 통계 분석과 최적화, 인공지능 등을 통해 방대한 데이터 집합을 처리할 수 있는 통계적 모델을 개발할 수 있도록 한 기술의 발전

(Bryant, Katz, and Lazowska 2008에서 발췌, 재구성)

- ③ 빅데이터 분석 연구 결과로 제시된 정보가 의사결정에 도움을 주는가?
- ④ 의사결정과 더불어 해당 문제에 대처하는 신속성을 개선시킬 수 있는가?

맥킨지 보고서에서는 빅데이터를 데이터베이스의 규모와 업무 수행에 초점을 맞춰 “일반적인 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터”로 정의하고 있다(Manyika 2011).

Gantz와 Reinsel(2011)은 빅데이터 기술을 아주 방대한 양의 다양한 데이터를 매우 빨리 캡처함으로써 그냥 지나칠 수 있던 것을 발견해내고 또는 분석함으로써 그 안에 존재하는 가치를 경제적으로 추출할 수 있게끔 디자인되어있는 차세대 기술이며 구조라고 묘사하고 있다. 또한 빅데이터를 이해하는데 있어서 저장되거나 소비되는 원래의 콘텐츠뿐 아니라 해당 콘텐츠가 사용되는 실제 흐름 안에서 살펴볼 수 있는 주변 정보까지도 이해해야만 한다고 주장한다. 이는 빅데이터 분석이라는 것이 사

회에 바탕을 두고 있으며 콘텐츠의 흐름이 중요하기 때문에 해당 콘텐츠의 이해를 위해서 주변 정보까지도 함께 분석의 테두리 안에 두어야 한다는 의미일 것이다.

앞서 말한 것처럼 빅데이터라는 개념은 사실상 새로운 것이 아니다. 그러나 최근 빅데이터 이슈가 더욱 주목받고 있는 것은 정보기술의 발전에 따른 데이터 정보원의 다양화와 이런 데이터를 통해 새로운 가치를 창출하려는 기업이나 정부의 요구가 증가되었기 때문이다. 빅데이터를 가장 단순한 용어를 사용해 개념 정의한다고 해도 간과할 수 없는 부분은 빅데이터 연구라는 것이 빅데이터에 대한 분석과 함께 만들어지는 가치 창출이라는 것이고 이는 또한 많은 사례들을 통해 그 유용성이 증명되고 있기도 하다.

2.2 특징 및 분석기법

현재까지 정의되고 있는 ‘빅데이터’의 특징은 크게 3가지, Volume(대용량), Variety(다

양성), Velocity(실시간성)의 3V로 일컬어진다 (McAfee and Brynjolfsson, 2012). 여기서 가트너 연구소는 한 가지 특성 Complexity(복잡성)를 덧붙여 설명하고 있다(이미영, 최완 2012).

대부분의 연구자들이 빅데이터의 특징을 앞서 말한 3V 또는 3V+1C로 설명하고 있으며 대용량, 다양성, 실시간성이 복합적으로 이루어지는 복잡성을 가졌다는 측면에서 4가지 측면 관점으로 설명하고 있다. 즉 규모, 다양성, 복잡성, 속도의 증가 특성을 중심으로 각각의 범주에서 원하는 가치를 얻을 수 있는 정도의 상대적 해석이 이루어지고 있다(최성곤 외 2012).

<표 2>는 빅데이터의 특징을 해당 요소에 대한 설명과 함께 기술한 것이다.

빅데이터의 가장 큰 특징은 취급하는 텍스트와 이미지가 정형적이지 않다는 것이다. 또한

데이터 자체의 양이 방대할 뿐 아니라 빠르게 전파되기 때문에 중요한 패턴을 찾기가 쉽지 않다. 다른 한편으로는 유용한 정보의 증가만큼이나 불필요한 정보도 급증하고 있어 방대한 데이터 안에서 의미가 있는 것을 찾아내고 분석하는 것이 아주 중요하다.

빅데이터의 분석기법은 각 분석기법의 내용에 따라 몇 가지로 구분할 수 있는데 간단히는 4가지(강만모 외 2012)에서부터 6가지(윤홍근 2013) 분석기법까지 거론되고 있다.

<표 3>은 보편적으로 알려져 있는 빅데이터 분석기법과 내용을 정리한 것이다.

빅데이터 분석 단계에 대해서 윤용익(윤용익, 김스베틀라나 2013)은 빅데이터를 보유하는 단계를 시작으로, 보유한 데이터에서 각종 데이터의 차이를 알아보고 뽑아내는 최적화단

<표 2> 빅데이터 특징의 요소별 설명

특징(요소)	내용
Volume (대용량)	데이터의 규모, 막대한 양의 데이터 사이즈를 가지고 있으며 수십 테라 바이트는 기본으로 사용한다. 데이터 분석을 위해 MapReduce 기반의 분산컴퓨팅 프레임워크를 활용한 클라우드 기술을 필요로 한다. 인터넷, 모바일폰 및 센서 등의 활용이 늘어나면서 데이터 생성 속도가 가속화되고 있기 때문에 데이터의 규모는 방대할 수 밖에 없다.
Variety (다양성)	데이터의 다양성. 놀랄만큼 다양한 정보원을 포함하며 또한 너무 다양한 각종 데이터타입, 개체들을 포함한다. SNS상의 다양한 텍스트, 이미지 및 동영상 자료 외에도 각종 센서에서 수집된 데이터로 인해 종류가 더욱 다양해졌다. 로그기록, 소셜미디어, 위치정보, 소비데이터 등 각종 데이터와 CC카메라에서 생성되는 동영상, 개인의 디지털 파일, 각종 센서로부터 발생되는 센서값 등 정형 데이터뿐 아니라 다양한 비정형 데이터도 생성되며 이를 다 포함한다.
Velocity (실시간성)	데이터 생성/처리 속도, 실시간과 같은 느낌의 빠른 속도의 처리 능력을 보유한다. 급속한 데이터의 증가에 따라 복잡한 이벤트 데이터에 대한 실시간에 가까운 처리를 위해 대규모 데이터센터를 활용한다. 자동으로 데이터를 생성하는 센서 장치, 스마트폰 등과 같이 데이터 생성 및 유통 채널의 변화로 지속적으로 생성되는 스트림 데이터가 증가하고 있으며 데이터 생성속도도 빨라지고 있다. 데이터 생성 속도의 가속화는 이를 활용하기 위한 처리 및 분석 속도의 가속화도 요하고 있다.
Complexity (복잡성)	데이터 활용 환경의 복잡성을 내포한다. 데이터 종류에 따라 적용되는 표준, 도메인별 규칙, 저장 형식 등이 서로 다르며 데이터 안에 존재하는 가치있는 정보를 얻기 위해서는 이와 같은 이질의 데이터를 통합하여 활용할 필요가 있다. 데이터구조화의 차이, 저장방식의 차이, 중복성 문제 등까지 데이터 종류 및 관리대상의 증가로 보다 복잡한 새로운 처리 기법이 요구된다.

(McAfee and Brynjolfsson 2012와 이미영, 최완 2012에서 발췌, 재구성)

〈표 3〉 빅데이터 분석기법과 내용

분석기법	내용
데이터마이닝	<ul style="list-style-type: none"> • 통계와 같은 수학적 기술과 함께 패턴인식 기술도 이용해 대용량 데이터를 조사 분석하는 기법 • 데이터 안에서 숨겨진 패턴 발견이 목적 • 가치 있는 정보는 의사결정에 적용
텍스트 마이닝	<ul style="list-style-type: none"> • 비정형 또는 반정형 텍스트 데이터에서 자연어 처리 기술에 기반해 유용한 정보를 추출, 가공하는 기법 • 자연어처리기술에 기반 • 대용량 언어자원과 통계적·규칙적 알고리즘 사용 • 문서분류, 문서 군집, 정보추출, 문서요약 등에 응용
평판분석	<ul style="list-style-type: none"> • 웹사이트와 소셜 미디어에 나타난 여론과 의견(긍정 및 부정, 중립 등 선호도)을 분석하여 실질적으로 유용한 정보로 재가공하는 기술 • 특정 서비스 및 상품에 대한 반응 및 입소문 분석 등에 활용 • 전문가에 의한 선호도를 나타내는 표현 및 단어 자원의 축적이 선행되어야 분석이 용이
소셜분석	<ul style="list-style-type: none"> • 일명 소셜네트워크분석 • 소셜네트워크의 연결구조 및 강도 등을 기반으로 소셜 네트워크상의 위치를 규명 • 소셜 미디어의 글과 사용자를 분석해 소비자 흐름이나 패턴 등을 분석하고 판매나 홍보에 적용
클러스터분석	<ul style="list-style-type: none"> • 일명 군집분석 • 비슷한 특성을 가진 개체를 개별 클러스터로 만들어 묶어내는 것이 시작 • 클러스터로 나누는 방법은 통계기법을 활용해 유사성 판단
현실마이닝	<ul style="list-style-type: none"> • 사람들의 일상 또는 특정한 시기 행동패턴 예측을 위한 분석 • 사회적 행동과 관련된 정보를 휴대폰이나 GPS등의 기기를 통해 수집 • 현실에서 발생하는 정보를 기반으로 인간관계와 정보추구 또는 이용행태 등을 추론

(강만모 외 2012와 윤홍근 2013에서 발췌, 재구성)

계, 데이터의 상관관계를 발견하고 간단한 모델링과 분석이 가능한 분석단계, 기존의 데이터분석과 함께 예측 분석 모델을 개발하는 빅데이터 최종분석단계라는 네 단계로 구분했다.

위와 같이 빅데이터의 분석은 학자별로 각 단계에 해당하는 내용에 따라 작은 차이는 존재하지만 일반적으로 데이터입수, 정보추출, 데이터분석과 결론추출이라는 4단계로 이루어진다고 받아들여지고 있다.

최근 문헌정보학 연구에서 빅데이터와 많이 겹쳐서 거론되는 개념 중의 하나로 클라우드 컴퓨팅이라는 것이 있다. 앞서 〈표 1〉에서 이미 빅데이터 연구를 위한 전제 기술로 선보인 바 있으나 타 연구에서 명확한 구분이 이루어지지 않

아 다소 설명이 필요하다 생각되어진다. 흔히 빅데이터는 인터넷 기반 컴퓨터네트워크를 통해 사용자가 필요하면 언제든지 인터넷 접속을 통해 데이터를 공유할 수 있는 데이터의 커다란 덩어리라고 이해해야 하며 클라우드 컴퓨팅은 이런 빅데이터 활용을 가능하게 만드는 전제기술로 다양한 클라이언트 기기에서 필요한 시점에 인터넷을 이용해 공유플에 있는 서버, 저장공간, 애플리케이션, 서비스 등과 같은 정보기술 자원에 쉽게 접근할 수 있도록 사이버 저장공간의 렌트를 가능하게 하는 기술로 빅데이터라는 개념과는 구분해 이해되어야 한다(최성, 우성구 2012).

클라우드 컴퓨팅과 도서관서비스에 관한 연

구로는 김용(2012)의 연구가 있는데 도서관서비스에 클라우드 컴퓨팅 도입을 제안하면서 관련 기술에 대한 광범위한 소개를 하고 있다.

2.3 활용사례 및 연구동향

빅데이터 분석의 활용 사례는 빅데이터분석을 이용한 공공부문 또는 비즈니스 마케팅의 성공사례와 더불어 빅데이터를 이용한 소셜네트워크분석, 빅데이터의 연구 근원을 찾고자 하는 연구까지 다양한 방법으로 다양한 분야에서 활발하게 이루어지고 있음을 볼 수 있다. 현재까지 발표된 사례는 상당히 많으나 본 연구에서는 문헌검토를 통해 볼 수 있는 빅데이터 활용 사례 연구 중 유명한 몇 사례와 빅데이터 연구의 주제적 다양성을 볼 수 있는 사례들 위주로 간추려 정리하였다.

가장 대표적인 사례로 미국 대선에서 오바마 캠프의 대선전략 수립에 빅데이터를 활용한 것인데 이는 대상 선거운동 집단을 작은 단위의 타겟으로 나누어 특징에 맞게 맞춤형 선거공약을 진행해 큰 성과를 보았다고 평가되고 있으며 의류회사인 ZARA는 빅데이터 분석을 통해 최적의 분배 시스템 개발함으로써 상품의 무재고 운영 정책이 실현될 수 있도록 하는 정책을 실행한 바 있다(김연진 2013).

의료부문과 관련해서는 미국에서 진행된 분석으로 크리스마스 연휴 이후 발생율이 늘곤 하는 율혈심부전증에 관한 빅데이터 분석을 볼 수 있는데, 분석결과 크리스마스때 상대적으로 음식 섭취량이 많고 더불어 엽분 섭취가 늘면서 생기는 크리스마스 후유증일 수 있다는 재미있는 연구결과를 볼 수 있었다. 구글의 독감예상,

아시아나의 기상데이터에 바탕을 둔 항공운항 정보수립 등이 빅데이터를 이용한 비즈니스 마케팅 적용 사례들이다(김원호 2013).

Park과 Leydesdorff의 2011년 연구는 소셜네트워크 분석을 볼 수 있는 연구인데 저자들은 SCI저널 데이터를 광범위하게 수집해 소셜네트워크분석과 공저자네트워크 분석 진행했다. 분석결과 세계 빅데이터의 연구 중 미국은 데이터마이닝이라는 개념을 중심으로 독자적 행보를 걷고 있음을 볼 수 있었으며, 주제상으로는 시맨틱웹, 링크드데이터 등이 빅데이터와 강한 연관성을 보여주고 있음을 알 수 있었다(Park and Leydesdorff 2011).

이외에도 광범위한 병리학 자료의 활용을 통해 각종 의문을 해석하고자 시도한 병리학에서의 빅데이터 활용연구(Jacobs 2009), 자연환경에서 존재하는 미생물을 배양하지 않고 직접적으로 분석하기 위한 연구분야로 생명과학적 접근이 돋보이는 메타지노믹스연구(오정수, 조완섭 2013) 등이 생물학적 접근점을 보여주고 있으며 빅데이터를 기반한 다양한 콘텐츠개발로 맞춤형 서비스, 학생별 맞춤형학습, 학교업무 지원서비스, 교수법 지원 등 맞춤형 교육서비스에 대한 연구(홍언주 2013; 권영옥 2013) 또는 정보부처와 공공기관의 인포그래픽을 사례로 연구한 빅데이터 이용 시각커뮤니케이션을 위한 인포그래픽 자료 조사 연구(최진원, 김이연 2012) 등 다양한 학문분야에 적용되어 연구되고 있다. 또한 강윤희(2013)는 그의 논문에 각종 빅데이터 관련 과학프로젝트를 정리해놓았는데 이 중에는 극지방 협력 연구과제로 극지 빙하의 손실이 해수면 상승 및 지구 온난화에 미치는 영향을 연구한 PolarGrid프로젝트는 빅데이터 분

석 연구의 다양성을 볼 수 있는 연구이다.

이러한 구체적인 적용사례들 이외에 빅데이터 연구사례 중 흥미로운 것은 빅데이터연구와 “정보” 또는 “facts”라는 개념을 사용해 현실과 빅데이터와의 관계를 생각하게끔 하는 Losee의 연구(2013)이다. 이 연구는 빅데이터의 연구가 얼마나 다채로울 수 있는지 보여주는 흥미로운 연구라 할 수 있다. 그의 연구는 Facts에 대한 재조명을 중심으로 전개하여 Informational Facts (IFacts), Regular Facts(RFacts)라는 두 가지 개념을 소개하고 있다. 여기서 IFacts는 세상 또는 세상의 파생물에 대한 실험(사례적) 관찰 사이의 관계를 표현하는 사실로 인간이 세상을 관찰한 결과와 세상자체(RFacts)와의 관계가 IFacts라고 할 수 있다. 빅데이터는 facts, IFacts와 RFacts 사이의 차이를 인지하고 추적할 수 있게끔 하는 것이다. 이 둘의 차이가 있어 야만 해석이 가능하고, 그런 해석이 빅데이터를 필요로 하는 이유라고 판단하는 것이다.

이를 재해석하면 빅데이터는 실시간으로 내용물이 축적되는 동시에 사용되는 시공간압축의 저장고가 될 수 있다. 그것은 이용자들이 여러 가지 형태로 올린 인터넷 상의 말과 생각, 의견과 감정이 쌓이는 실시간 저장고이다. 이용자가 올려놓은 디지털 콘텐츠는 그것이 저장되는 물리적 장소와 상관없이 네트워크 안에서 사이버스페이스라는 동일한 가상의 공간 안에 축적된다. 빅데이터는 앞서 본 IFacts와 Rfacts와 같은 가상공간과 실제공간을 연결하여 차이를 해석하고 이해할 수 있게 한다는 것이다.

수많은 사례 연구들이 존재하지만 사실상 빅데이터 분석을 위한 전문인력은 상당히 부족한 현실이다. 몇몇 연구들이 빅데이터 분석과 분석

가의 필요성을 강조(Chen, Chiang, and Storey 2012)하고 있으며, 이러한 현실은 실제로 문헌정보학계의 정보전문가나 학생에게 또 다른 중요한 기회로 여겨질 수 있음을 연구결과로 피력하고 있다.

빅데이터는 그 자체가 중요한 게 아니라 분석의 결과로 주어지는 가치부여가 더욱 중요한 문제이다.

지난 추석 때 교통정보를 보면 여느 해와는 달리 지난 몇 년간의 데이터를 총체적으로 분석해 언제가 가장 많은 교통정체를 보일 것인지, 서울에서 부산까지 몇 시간이나 걸릴 것인지 등에 대해 데이터를 근거로 구체적으로 알려주는 교통정보서비스를 본 적이 있다. 또한 공공부문에서 세계 각국의 정부들이 주도적으로 공공부문 중장기 전략수립, 항공, 국토, 지질, 기후, 대선, 교육, 교통, 마케팅, 스포츠 등 다양한 분야의 연구를 주도하고도 있다. 이러한 현실은 빅데이터 분석연구가 일상생활에서부터 우리가 직접적으로 볼 수는 없으나 많은 영향력을 받는 일상 곳곳에서 우리와 아주 가까이 친밀한 변화를 보여주고 있음을 보여주는 것이라 하겠다. 또한 빅데이터 분석 연구의 무한한 가능성을 암시해주는 것이라고도 할 수 있겠다.

3. 빅데이터의 이해와 도서관 정보서비스에의 활용

3.1 빅데이터 활용의 의의

빅데이터 분석 연구란 방대한 데이터의 흐름을 분석해 내재해있는 패턴과 그 연결성을 인

식하고 분석할 수 있는 연구이다. 문헌정보학에 있어서 빅데이터 연구는 빅데이터라는 말을 사용하지 않았을 뿐이지 이전부터 유사 연구는 진행되어 왔다고 볼 수 있다. 이는 webometrics나 다른 정보 기반 연구들을 통해 확인할 수 있는데(Thelwall, Wouters, and Fry 2008) 이 연구들의 대다수는 주어진 정보를 가지고 어떻게 현상을 파악할 것인가, 흐름을 볼 것인가와 맥락을 같이 하기 때문에 이와 유사한 흐름의 연구라는 측면에서 볼 때 빅데이터 연구는 우리에게 새로운 연구라고 할 수는 없다.

데이터의 폭발적 증가와 다양화된 데이터 유형, SNS를 필두로 하는 이용자 생성 데이터의 급증 등은 도서관과 같이 정보원을 수집, 보존, 저장 및 서비스 하는 조직에게 있어서 빅데이터 활용을 다소 두려운 현실로 받아들이는 것도 사실이지만 이는 또한 반드시 헤쳐 나가야만 할 도전이기도 하다. 또한 이용자의 정보활동 자체가 문헌정보학의 중요한 연구주제라는 학문의 주제적 측면에서 보았을 때도 빅데이터 연구는 절대적으로 피할 수 없는 주제이기도 하다.

문헌정보학은 전통적으로 정보를 조직, 관리, 검색, 수집, 기술하고 보존하는 학문이다. 여기서 정보라는 것이 고전적으로는 책이나 논문 등의 형태, 즉 주로 텍스트로 만들어진 학술 출판물이던 것이, 이후 출판의 형태와는 상관없이 연구자에 의해 출판되는 데이터의 다양한 형태를 포함하는 것까지로 확산되어 이해되고 있다. 또한 학문적 연구영역을 비추어볼 때도 문헌정보학이라는 학문은 역사적으로 다학제적

이고 협업적인 학문분야이며 다양한 기술적, 주제적 결합을 피해왔다(Lariviere, Sugimoto, and Cronin 2012). 문헌정보학 연구자들이 시맨틱 웹이나 링크드데이터와 같은 연구분야들에 많은 관심을 가지고 연구를 진행해 왔던 것도 이러한 기술적, 주제적 결합을 시도한 연구들의 사례라고도 할 수 있다.

지금 도서관은 Web 2.0 또는 Library 2.0¹⁾이라는 흐름의 한가운데에 존재해있다. 이는 이용자가 언제 어디서든 스스로 콘텐츠를 생산 또는 배포할 수 있는 정보기술 환경에 살고 있다는 것이고 이렇게 생산된 소셜 미디어 기반 데이터가 지금 현사회의 중요한 정보원중 하나로 존재한다는 것이다. 소셜미디어 기반 데이터는 기존의 데이터보다 훨씬 방대하고 다양한 구조를 가지고 있지만, 현재 시점에서 이용자들끼리의 직접적인 소통을 통해 만들어지는 것으로 현재 시점의 사회 동향이나 행동 패턴을 표현한다고 인식되고 있으며 이를 이해하는 것은 미래를 예측하기 위한 중요한 근거가 될 수 있는 것이다. 바로 이런 이유 때문에 이용자 생산 콘텐츠의 흐름, 이용자의 주제를 이해할 수 있는 또는 이해하도록 도움을 주는 빅데이터 연구는 문헌정보학계의 중요한 연구주제가 될 것임은 당연하다.

3.2 빅데이터 활용에 있어서 제기되는 이슈

앞서 말한 바와 같이 빅데이터 분석 연구는 그 유용성 하나만 보아도 연구의 의의가 높다.

1) 이미 Web 3.0, Library 3.0이라는 새로운 개념에 대한 많은 논의가 있는 것도 사실이나 이 논문에서는 이용자 생산 콘텐츠라는 내용의 부각을 위해 현재 도서관이 직면하고 있는 Web2.0, Library 2.0 관련 현상을 부각하고자 3.0의 논의는 차후로 미룬다.

그러나 많은 빅데이터 관련 연구들의 생산과 더불어 이로 인해 비롯되는 비판적 이슈도 적지 않다. 이러한 이슈들은 현재 다양한 분야에서 연구되고 있으며 그 내용에 있어서는 상당히 유사한 부분도 많다. 본 연구에서는 현재 제시된 다양한 관련 이슈를 다룬 boyd와 동료들의 논문들을 중심으로 이외 관련 이슈를 제시한 다른 논문을 광범위하게 검토하여 수집된 이슈 중 도서관 정보서비스에 있어 중점적으로 논의되어야 할 이슈들을 종합해 정리하고자 하였다.

빅데이터의 활발한 적용과 관련해 흥미로운 이슈 중 하나는 boyd와 Crawford(2012)에 의해 제기된 '지식'의 개념 변화라는 문제제기이다. boyd와 Crawford는 빅데이터가 현재 우리 사회가 이해하는 지식이라는 개념의 변화를 야기했다고 주장한다. 이 연구는 전통적으로 지식이라는 개념이 단순 사실 정보와 그에 따라 정해진 확실한 경우에 대해, 어떤 문화적 차이도 상관없이 받아들여질 수 있는 것이 지식이었다고 이야기 했던 반면에, 현재 빅데이터 이후의 지식은 데이터 자체보다 집합으로서 데이터세트의 흐름을 보는 것으로 이해하고 있다. 이 연구에 따르면 현재의 지식은 하나의 덩어리로 데이터세트의 흐름을 파악하고 전반적인 맥락을 파악한 후에 만들어진 예측이다. 이러한 방법으로 만들어진 예측은 해당 사회의 맥락 속에서 이루어지게 되기 때문에 당연히 해당 사회가 아니면 의미가 없다고 이해하는 것이다(boyd and Crawford 2012). 이는 전통적인 데이터, 정보, 지식의 개념에 대한 재구성을 필요로 하는 것으로 판단된다. 물론 여기서 지식은 우리가 기존에 이해하는 진리와 차별성을

가지고 이해되어야 하며 이 연구에서 말하는 지식이 다소 진리와 유사하게 설명되고 있기는 하지만 그럼에도 불구하고 전통적으로 지식이라는 개념이 정확한 사실정보에 기반해 확실한 경우에 대해 문화적 차이에 상관없이 받아들여졌다는 부분에 있어서는 논란의 여지가 없을 것이다. 바로 이런 이유 때문에 전통적인 지식의 개념과 다르게 빅데이터 분석학 이후의 지식은 반드시 기반한 사회의 맥락적 이해를 바탕으로 구성되고 해당 사회를 벗어나서는 그 의미를 잃게 된다는 문제제기는 아주 흥미로운 발상이다. 그러나 이들의 연구에서 제기하는 문제가 진정한 의미에서 지식의 개념이 변화했음을 이야기하는 것인지에 대해서는 다시 생각해볼 여지가 있다고 판단된다. 해석해보자면 중요한 문제제기는 지식의 개념이 진짜로 변했다는 주장이라기 보다는 지식을 이해하는 것에 있어서도 이제는 빅데이터가 불러올 맥락적 분석과 그에 따른 이해가 중요하다는 것으로 이해되어야 한다는 것이 더 중요한 논지가 아닐까 여겨진다.

제기되는 두 번째 이슈는 빅데이터 분석과 함께 만들어진 분석결과의 객관성·정확성에 대한 판단에 대한 문제제기이다(boyd and Crawford 2012). 빅데이터 분석은 해당 시점, 해당 사회에 기반해 수집된 각종 데이터를 바탕으로 이루어진다. 이로 인해 당연히 빅데이터를 통해 내려진 분석은 해당 사회의 맥락속에서 이해되어야 하고 이는 빅데이터의 분석과 함께 내려진 판단은 당연히 주관성을 내재할 수 밖에 없음을 이야기하는 것과 동일하다. 아주 단순하게는 데이터 오류의 가능성이 충분히 있을 수 있다는 것이고 또는 해당 시점 수집된 데이터

를 분석하는 방법이나 분석자의 관점에 따른 오해석의 가능성도 그만큼 충분히 존재한다는 것이다. 또한 분석하는 데이터의 양이 많다고 해서 이 데이터 집합이 해당시점 해당사회의 전부를 보여줄 수 있는 것은 아님을 이해할 필요가 있다. 빅데이터 분석이라는 것도 결국 해당 주체의 데이터 전부를 수집한 결과로 이루어지는 것이 아니며 해당 주제에 대한 데이터 중 일부(단지 그 데이터의 양이 방대하다는 것일 뿐이지)이기 때문에 이 또한 샘플데이터임을 간과해서는 안된다. 커다란 코끼리의 코만을 보고 이 동물이 어떤 동물이다 이야기하는 것과 다리를 보고 이야기하는 것과 같은 차이를 만들어낼 수도 있다는 것이다. 이런 점에서 빅데이터 분석의 결과는 어찌보면 그 중요성만큼 객관성·정확성에 대한 문제도 충분한 고려사항이 되며 그만큼의 논의도 필요하다.

다음은 분석에 사용되는 데이터 자체의 가치나 신뢰성 판단에 대한 문제제기이다. 빅데이터 분석에서 중요한 고려대상 중 하나의 데이터군이 소셜미디어 데이터이다. 소셜미디어 데이터는 그 자체로 현재 시점 이용자의 생각을 가장 신속하고 손쉽게 수집가능하고 또한 이해할 수 있는 데이터라고 볼 수도 있으나 또한 가장 왜곡이 많고 손쉬운 데이터이기도 하다. 간단히 긍정과 부정에 대한 분석의 경우에도 짧은 하나의 문장에서 보여지는 언어법이나 물음표 없는 질문등은 분석방법이나 분석자에 따라 잘못 받아들여질 소지가 다분하다. 또한 소셜미디어 이용자의 의도적 오정보의 생산과 그에 대한 재생산·확산은 분석의 결과만으로 해당 빅데이터 분석의 결과를 판단할 때 그 가치나 신뢰성을 부여하기가 부적절하게 되기 때문이다.

다음으로 제기되는 문제는 빅데이터가 해당시점 해당사회의 부산물임을 전제로 할 때 필수적으로 제기될 수밖에 없는 문화적 상대성에서 비롯되는 문제이다. 빅데이터 분석의 전제는 맥락에 대한 이해가 필수적이며 해당시점 해당사회가 어떠한 사회인가를 이해해야 적절한 분석이 가능하다. 이런 점에서 비추어볼 때 문화적 상대성에 대한 이해가 제대로 이루어지지 않으면 주어진 빅데이터 분석의 결과가 잘못 받아들여질 수 있는 소지가 충분하며 이는 같은 데이터를 가지고 상반된 분석결과를 낳게 될 가능성도 있으며 이후 이를 바탕으로 내려진 의사결정이 커다란 사회문제로 비화될 소지도 없지 않다.

다섯 번째 문제제기는 빅데이터가 가지는 윤리적 문제이다. 빅데이터 분석 기관들이 수집하는 데이터들은 사방에 이미 존재하기 때문에 이용이 가능하게 된 것이다. 그러나 접근 가능한 데이터라고 해서 이를 수집, 분석하는 것이 윤리적인 것은 아니다(boyd and Marwick 2011). 이용자가 자신의 소셜미디어에 올린 글들 하나 하나가 빅데이터 분석에 이루어질 가능성에 대해 인지한 가운데 분석에 사용가능, 사용불가능에 대한 전제를 표시하고 자신의 글을 게시하는 것은 아니다. 현재로서는 빅데이터 분석을 위해 쓰여지는 데이터에 대해, 특히나 소셜 미디어 데이터에 대한 법적, 윤리적 테두리는 거의 없다고 할 수 있으며 이러한 점 때문에 빅데이터 분석에 있어서의 윤리적 문제는 더욱 많은 논의가 필요하다. 또한 개인정보보호, 보위의 문제도 윤리적 문제와 연결선상으로 제기되는 중요한 문제 중 하나라고 할 수 있다.

여섯 번째 중요한 문제제기는 빅데이터로 인

한 빅데이터 감시의 문제이다(이광석 2013). 이광석(2013)은 자신의 연구에서 과거 사람이 직접 특정 대상이나 집단을 감시하며 이루어지는 감시에서 사람의 눈과 함께 정보기술의 혜택을 통해 이루어지는 감시체제가 이루어질 가능성에 대한 문제제기인 것이다. 그는 이를 빅데이터의 빅브라더화 가능성으로 이야기 하고 있는데 정보기술에서 나온 데이터가 인간을 통제하기 위한 감시 데이터로 사용할 수 있게 됨을 이야기하는 것이라 하겠다. 실제로 빅데이터 분석에서 소셜미디어뿐 아니라 CCTV 데이터 등이 함께 이용된다는 사실에 견주어 볼 때 결코 불가능하거나 무시할 수 있는 문제제기는 아닌 것으로 판단되며 이는 사회적 차원에서의 고민과 논의를 통한 접근이 필요하다 하겠다.

다음은 빅데이터에의 접근 가능성에서 파생되는 문제이다. 빅데이터는 현재 수집할 수 있는 사람도 얼마 되지 않을뿐더러 수집된 빅데이터를 다룰 수 있는 개인도 그리 많지 않다. 이로 인한 빅데이터의 제한적 접근 양상은 새로운 형태의 디지털 정보격차를 생산할 가능성이 적지 않다(boyd and Crawford 2012). 빅데이터에 접근 가능한 개인이나 집단은 이로 인한 의사결정에 있어 그렇지 않은 개인과 집단에 비해 엄청난 혜택을 보게 되는 것이 사실이다. 시험을 볼 때 문제를 미리 알고 가는 것과 그렇지 않은 것과의 차이는 점수에 있어서 엄청난 차이를 만들게 되는 것은 당연한 이치인 것이다. 인터넷을 비롯한 정보기술의 확산과 더불어 문헌정보학계에서 많은 고민과 노력을 요했던 부분이 디지털 정보격차의 해소와 같은 공익적인 공헌에 대한 부분이었다. 빅데이터로 인한 또 다른 양상의 디지털 정보격차에 대한 문제는 문헌정

보학계에서 또 한번 고민하고 논의해야 할 부분이 아닌가 판단된다.

여덟 번째 중요한 문제제기는 빅데이터 분석에서 사용되는 각종 텍스트, 이미지 파일등에 대한 소유권이나 지적 재산권에 대한 문제제기이다(Small, Kasianovitz, Blanford, and Celaya 2012). 이 문제는 앞서 다섯 번째로 제기된 윤리적 이슈와 조금은 맞물려 있는 문제제기인데 수집가능한 데이터라고 해서 이 데이터를 이용하는 것이 윤리적인 것은 아님을 이해한다고 했을 때 과연 이 데이터들을 의사결정에 이용하고 이익을 창출하는 개인이나 집단의 경우 이 데이터들에 대한 소유권이나 지적재산권을 주장할 수 있을까? 빅데이터 분석에 사용되는 데이터 전체가 특정 개인의 데이터가 될 수는 없지만 특정 개인 수십만 명, 수백만 명이 모여 이루어진 빅데이터에 대한 소유권이나 지적재산권은 누구에게 부여되어야 하는지 진지하고 엄격한 고민과 논의가 필요하다. 내 데이터가 들어가고 있다고 해서 빙산의 일각인 해당데이터만은 내 데이터이니 그만큼의 소유권을 달라고 주장할 수 있을까 또는 해당되는 모두의 집단적 소유권이 발동되어야 하는가, 아니면 그냥 떠돌아다니는 데이터를 모아 활용한 빅데이터 분석가 또는 분석 집단에게 부여되어야 하는가 등 복잡미묘한 문제들에 대한 논의가 반드시 이루어져야 한다는 것이다.

마지막으로 빅데이터 분석의 결과가 사용되는 방법에서 비롯되는 문제로서 빅데이터 이용의 전횡이 가져오는 범사회적 문제발생의 가능성에 대한 문제제기이다. 앞서 일곱 번째로 제시된 빅데이터의 제한적 접근양상에서 비롯된 또 다른 양상의 디지털 정보격차에 대한 문제

와는 다르게 이는 빅데이터가 의도적으로 잘못 이용되거나 조작되어 사용될 때 나타나는 문제로서 빅데이터 분석이 주로 특정 주제 이용자의 해당시점, 해당 사회적 맥락을 이해하고 이를 바탕으로 한 의사결정에 이용된다고 했을 때 자칫하면 커다란 범사회적 문제를 발생할 가능성도 없지 않다고 할 것이다.

이와 같이 빅데이터는 그 활용의 중요성만큼이나 고려해야 할 사회적 이슈들이 산적해있다. 이는 빅데이터 분석이 주로 의사결정에 활용되고 의사결정이라는 이 한마디는 이후 해당 집단, 사회의 장단기적인 변화를 내포하고 있다는 측면에서 빅데이터를 가지고 의사결정에 활용하면 그걸로 만족한다 또는 그걸로 충분하다고 안위할 수는 없다. 비록 거론된 이슈들이 단기간에 해결되거나 해결을 위한 제도가 쉽게 만들어질 수 있는 것은 아니다. 이는 범사회적인 논의와 동의를 바탕으로 이루어져야 하는 것이고 그렇기 때문에 더욱 빅데이터 분석으로 인해 파생될 문제들에 대해 고민해야 하고 논의해야 할 필요성이 우리에게 있는 것이다. 문헌정보학계의 경우 데이터·정보사용에 대한 이익창출이 우선인 일반 기업과는 다르게 도서관과 각종 정보센터를 필두로 한 공익적 차원의 데이터·정보 활용의 핵심에 있는 만큼 빅데이터 분석이 가져오는 문제에 대한 사전 인지와 논의는 필수적이라 하겠다.

3.3 정보서비스에의 빅데이터 활용

현재 문헌정보학계는 정보기술의 급격한 발전과 확산, 이에 따른 이용자의 정보기술력 향상평준화에 발맞춰 도서관·정보센터의 서비스

능력 향상에 많은 고민을 하고 있다. 현장 사서들은 이용자들이 일상생활에서 적극적으로 사용하는 각종 소셜미디어를 도서관 정보서비스에 적극 활용하며 이용자와 발맞추고자 하는 노력을 계속하고 있으며 고전적인 도서관 정보서비스와 새로운 정보기술의 도입 사이에서 많은 현실적인 한계에 봉착해 있는 것도 사실이다. 뿐만 아니라 앞서 기술한 바와 같은 빅데이터 관련 이슈들의 해결책이 범사회적으로 자리잡히지 않은 상태에서 도서관 정보서비스에 활용한다는 것은 분명히 한계가 존재한다.

빅데이터는 그 해석·분석과 활용의 중요성과 함께 가치를 중요시하는 부가가치 콘텐츠를 생산할 수 있다는데 초점을 두고 이해해야 한다. 또한 빅데이터의 본질이 이용자의 일상생활과 함께 하는 각종 데이터의 총체적 분석이라는 점, 이용자의 정보추구행동에 기반된 각종 데이터를 신속하고 손쉽게 수집해 작금의 현실 정보세계를 해당 사회의 관점에서 볼 수 있다는 점에 있다고 본다면 빅데이터 분석 결과는 도서관 정보서비스 정책에 반영, 양질의 정보서비스를 적시적소에 신속하고 적극적으로 제공하는데 큰 기여를 할 수 있을 것이라 여겨진다.

다음은 이를 전제로 빅데이터 분석결과를 도입, 도서관 정보서비스에 활용할 수 있는 몇 가지 방안과 그 의의에 대한 제안이다. 본 활용방안은 빅데이터 관련 광범위한 문헌검토를 바탕으로 한 내용분석의 결과로 제시되었으며 분석의 방법에 있어서 질적 대 양적 분석으로, 활용 목적에 있어서 의사결정 대 마케팅 방안으로 구분하였다. 활용방안의 첫 번째와 두 번째는 질적분석을 토대로 한 사회적 관계나 방향성, 현실 인식에 기반해 제시한 의사결정 관련 활용방

안이며 세 번째와 네 번째는 양적분석의 결과를 토대로 한 데이터 분포·선호도를 기반해 제시한 마케팅 관련 활용방안이다.

먼저 가장 중요한 부분으로 도서관 정보센터 자체에 대한 존재의 필요성과 인식에 따른 역할 규정 및 전략적 정책 수립에 도입하는 방안이다. 최근 몇 년 사이 더더욱 두드러진 부분이 도서관 이용률 저하²⁾에 대한 문제이다. 도서관 이용률 저하의 원인은 뉴스와 같은 미디어에서 많이 거론되는 독서인구 급감과 함께 PC와 모바일로 손쉽게 접근할 수 있는 각종 정보검색·독서 대체 매체의 대중화 및 활성화와 함께 이해되어야 한다. 오늘날의 정보이용자는 간단한 데이터나 정보가 필요할 때 스마트폰을 가장 먼저 꺼내 찾아본다고 한다. 불과 몇 년전까지만 해도 PC가 옆에 있어야만 가능했던 데이터·정보검색이 순간의 컴퓨터라 불리는 스마트폰 덕분에 언제 어디서나 원하는 데이터·정보를 검색하고 자신의 정보요구를 해소하는 상황에 이른 것이다. 또한 각종 모바일 콘텐츠의 확산과 더불어 스마트폰으로 가능한 각종 문화서비스들이 늘어나 책을 가지고 다니지 않아도 스마트폰을 이용한 전자책 또는 사운드, 이미지 파일등 각종 다양한 형태의 콘텐츠를 받아들임으로써 자신의 정보요구를 해소할 수 있는 것이다. 이러한 사회 현실속에서 도서관·정보센터는 어떠한 역할을 해야하며 활성화를 위한 전략적 방안은 무엇인가에 대한 진지한 고민이 필요하다. 이 고민들은 현실 이용자가 이해하고 바라는 정보요구와 도서관 자체에 대한 이해, 요구에서부터 비롯되어야 하며 빅데이터 분석은 이를 위한

중요한 판단 근거가 될 것으로 여겨진다. 한 가지 강조할 점은 빅데이터 분석을 통한 이용자의 현시점 이해와 요구가 무엇이든 무조건적으로 따라가려고만 해서는 안된다는 것이다. 이용자의 이해와 요구를 이해함과 더불어 도서관·정보센터의 존재가 가지는 의의를 정확히 인지한 가운데 전략적 대응이 모색되어야 한다.

다음 두 번째는 Birney(2012)가 자신의 연구에서 제시한 것처럼 지역사회 지원을 위해 활용하는 방안이다. 이는 각 지역사회의 요구를 파악하고 그에 부합되는 정보서비스를 기획하는데 빅데이터 분석결과가 중요한 근거가 될 수 있음을 의미하는 것이라 하겠다. 현재 지역사회의 요구를 파악하고 지원하기 위한 도서관계의 노력은 주로 이용자나 잠재적 이용자를 대상으로 한 설문조사와 같은 방법을 활용하는 것이다. 그러나 이러한 방법은 많은 재정적, 인력적 자원을 필요로 하며 일상의 업무를 계속해야 하는 도서관의 경우 또 하나의 업무부담으로 자리하는 것도 사실이다. 이런 현실에서 빅데이터를 활용한다면 신속하고 손쉽게 지역사회의 정보요구를 수집할 수 있을 것이며 이를 기반해 내려진 지역사회 지원 방안은 지역사회 이용자의 요구에 기반했다는 정당성을 확보하는 한편 지역사회 지원 정책의 일관성 확보에도 상당한 도움을 줄 수 있으리라 판단된다.

다음은 도서관·정보센터의 정보자원 개발 계획에 활용하는 방안이다. 현재의 이용자 정보요구는 이용자의 정보검색능력 상향평준화와 더불어 상당히 빠르고 전문화되어 변화하고 있다. 도서관에 있어서는 이용자의 정보요구는 알

2) 네이버뉴스에 따르면 독서인구급감과 더불어 출판생태계 위기, 서점의 붕괴, 도서관 이용률 저하라는 심각한 사회문제가 나타나고 있다고 한다. <<http://news.naver.com/main/tool/print.nhn?oid=021&aid=0002171691>>.

지만 구체적으로 어떤 주제로 어느 만큼의 수준에 맞춰 정보자원을 구비해야 하는가가 가장 커다란 숙제이다. 빅데이터 분석이 이용자가 필요로 할 만한 주제별, 이용자별, 사안별 정보서비스 자료의 비치 및 제공에 대한 방향을 제시하고 신속한 대응을 준비하는데 중요한 지침이 될 수 있을 것이다.

마지막으로 서비스 대상 이용자에 따른 맞춤형 정보서비스 계획에 활용하는 방안이다. 이는 앞서 세 번째로 말한 정보자원개발과 더불어 이해되어야 하는 방안으로 이용자가 원하는 정보와 서비스 방향에 대한 판단에 근거로 빅데이터 분석을 활용하는 것이다. 빅데이터 분석을 기반으로 신속한 이용자 정보요구 분석과 더불어 신속한 정보서비스 의사결정이 가능해질 것이다. 이는 특히 대학도서관이나 연구도서관과 같은 전문 도서관에게 유용하게 활용될 수 있으리라 여겨지는데 이는 다양한 수준의 정보요구를 가지고 있는 전문도서관 이용자를 위한 맞춤형 정보서비스에 대한 예측 및 정보서비스 제공이 가능해질 수 있을 것이라 생각되기 때문이다.

본 연구에서 제시한 활용방안 중 두 번째부터 세 가지는 실제로 도서관의 정보서비스에 활용하는 방안이지만 실제로 첫 번째 도서관의 전략적 정책수립에 대한 방안은 전체 도서관·문헌정보학계의 상당한 노력과 방대한 논의가 있어야 가능한 부분이다. 그럼에도 불구하고 이 연구에서 도서관의 전략적 정책수립에 빅데이터 분석을 활용하는 방안을 주장하는 이유는 현재 우리가 안고있는 위기에 대한 문제의식 때문이다. 어찌보면 빅데이터 분석이 도서관·문헌정보학계에게 이전에 어떤 분석들이 가져다주지

못했던 문제해결 방안을 제시해주지 않을까 하는 기대때문이기도 하다. 또한 본 연구에서 제시한 활용방안들 또한 앞서 제시한 빅데이터가 파생하는 각종 사회적·윤리적 이슈들을 반드시 고려해야 하기 때문에 신중하고 세심한 실행방안이 모색되어야 한다. 또한 문헌정보학과와 현장 모두 이에 대한 해결이나 사회적 제도의 정비와 같은 범사회적 움직임에 적극 개입하고 논의를 주도함으로써 빅데이터 활용이 당연해지는 사회적 동향에 발맞추어야 할 것이다.

4. 결론 및 제언

본 연구는 빅데이터 연구의 활성화와 함께 문헌정보학 관점에서 빅데이터를 이해하고 활용할 수 있는 방안을 모색하고자 하는 연구이다. 이를 위해 본 연구는 개념정의와 더불어 주목해야 할 문제점과 이슈, 도서관 정보서비스에의 활용방안이라는 세 가지 연구질문을 가지고 연구를 진행했다.

빅데이터란 방대한 데이터라는 간단한 개념 정의에서부터 사회의 맥락에서 이해되어야 하는 분석학적인 개념정의를 내포하고 있으며, 지식의 개념에 대한 재해석에서부터 데이터 처리에 있어서의 윤리적 문제, 객관성이나 정확성에 대한 판단의 문제, 또 다른 디지털 정보격차의 생산 등과 같이 빅데이터 분석 활용에 있어서 제기되는 이슈들을 정리해 보았다. 마지막으로 빅데이터 분석을 도서관의 정책수립과 지역사회 지원, 정보자원개발과 맞춤형 정보서비스 방향 수립을 위한 중요한 근거로 삼고자 하는 등의 활용방안을 제시했다.

빅데이터를 적절하게 활용하기 위해서는 앞서 제기되는 이슈를 고려하지 않을 수 없으며 도서관 입장에서는 데이터 취급행동윤리강령 등과 같은 실질적인 취급지침이 전제되어야 할 것이다(Birney 2012). 소셜 미디어 데이터를 다룰 경우, 기관의 사명이나 목적에 비추어 볼 때 해당 자료의 수집이 필요한지 명확히 해야 할 필요가 있으며 개인정보들이므로 전체적으로 어떤 위험 부담이 생길수 있는지에 대한 명확한 확인이 이루어져야 한다. 이러한 지침의 올바른 수립을 위해서는 초반부터 다양한 연구자와의 협업을 통해 문제발생의 가능성 낮추는 것이 현명한 방법이라 할 것이다.

빅데이터는 시맨틱웹이나 링크드데이터와 같은 문헌정보학의 또 다른 연구주제와도 연관성이 큰 연구주제이다. 이 주제들은 모두 인간의 상호작용 또는 커뮤니케이션 흐름을 보여주는 주제들이라 할 수 있다.

최근 가트너가 발표한 2014년 10대 전략기술에 전략적 빅데이터는 속해있지 않다.³⁾ 그러

나 여전히 모바일, 클라우드, 스마트 머신과 같은 기술 트렌드는 10대 전략기술에 속해있으며 퍼스널 클라우드의 시대와 같은 기술트렌드가 속해있는 것으로 보아 빅데이터 분석은 중요하지 않다는게 아니라 이미 기업이나 정부기관등이 사회를 이해하고 대응하는데 필수적인 인프라로 받아들여지고 있음을 의미하는 것이라고 받아들이는 것이 더욱 적절하다 하겠다.

구슬이 서 말이어도 꿰어야 보배다 라는 속담과 같이 지금까지 처리하지 못했던 방대한 양의 데이터들을 처리할 능력이 되는 지금 더욱 중요한 것은 어떻게 해석하고 어떻게 사용할 것인가에 대한 문제일 것이다. 빅데이터가 구슬이라면 이 구슬을 어떻게 엮어서 우리 목에 거느냐 이것이 바로 가치창출에 대한 문제이다. 세계 각국, 사회 전 분야에서 빅데이터를 활용하기 위한 경쟁체제에 이미 돌입했다. 도서관·문헌정보학계에서 어떻게 이를 도입하고 활용해 자체 경쟁력을 높일 것인가는 우리에게 맡겨진 숙제이다.

참 고 문 헌

- 강만모, 김상락, 박상무. 2012. 빅데이터의 분석과 활용. 『정보과학회지』, 30(6): 25-32.
- 강윤희. 2013. 빅데이터 처리를 위한 과학클라우드. 『한국콘텐츠학회지』, 11(1): 40-45.
- 강학주. 2012. Social Big DATA의 의미와 트렌드 분석. 『Local Information Magazine』, 74(May, June): 52-57.
- 권영욱. 2013. 빅데이터를 활용한 맞춤형 교육 서비스 활성화 방안연구. 『지능정보연구』, 19(2): 87-100.
- 김연진. 2013. 빅데이터 기반의 고도분석체계 도입을 통한 기업혁신: 사례와 방법론. 『Practice & 중소기업 포럼』, 20(1): 43-49.

3) <<http://www.bloter.net/archives/166671>>.

- 김용. 2012. 클라우드 컴퓨팅 기반의 도서관 서비스 도입방안에 관한 연구. 『한국비블리아학회지』, 23(3): 57-82.
- 김원호. 2013. 빅(Big) 데이터 vs 빈(空)데이터. 『마케팅』, 47(2): 9-15.
- 김진숙. 2013. 트렌드를 알면 해법이 보인다: 2013년 국가정보화 트렌드 설문조사 결과를 중심으로. 『Local Information Magazine』, 79(March, April): 56-59.
- 오정수, 조완섭. 2013. 메타지노믹스 - 빅데이터 시대의 새로운 도전. 『정보과학회지』, 31(8): 19-26.
- 윤용익, 김스베틀라나. 2013. 빅데이터와 클라우드 시대. 『정보와통신』, April: 3-6.
- 윤홍근. 2013. 문화산업에서 빅데이터의 활용 방안에 관한 연구. 『글로벌문화콘텐츠』, 10: 157-180.
- 이광석. 2013. 지배양식의 국면 변화와 빅데이터 감시의 형성. 『사이버커뮤니케이션 학보』, 30(2): 191-230.
- 이미영, 최완. 2012. 빅데이터 분석을 위한 빅데이터 처리 기술 동향. 『정보처리학회지』, 19(2): 20-28.
- 이충희, 허정, 오효정, 김현진, 류범모, 김현기. 2013. 소셜 빅데이터 이슈 탐지 및 예측분석 기술 동향. 『전자통신동향분석』, 28(1): 62-71.
- 최성. 우성구. 2012. 빅데이터 정의, 활용 및 동향. 『한국정보기술학회지』, 10(3): 10-19.
- 최성곤, 오진태, 장중수. 2012. 미래 지식 서비스를 위한 빅데이터 처리의 기술적 요구 사항. 『정보처리학회지』, 19(2): 5-10.
- 최진원, 김이연. 2012. 빅데이터 시대에 효과적인 시각커뮤니케이션을 위한 인포그래픽 연구. 『한국과학예술포럼』, 11(1): 165-175.
- 홍언주. 2013. 빅데이터 기반 맞춤형 교육 서비스. 『한국인터넷정보학회 학술대회 발표집』, 14(1): 77-78.
- Bimey, E. 2012. "Lessons for Big-data Projects." *Nature*, 489(September): 49-51.
- boyd, d. and K. Crawford. 2012. "Critical Questions for Big Data: Provocations for Cultural, Technical, and Scholarly Phenomenon." *Information, Communication & Society*, 15(5): 662-679.
- boyd, d. and A. Marwick. 2011. Social Privacy in Networked Publics: Teens' Attitudes, Practices, and Strategies. paper given at *Oxford Internet Institute*, [online]. [cited 2013.10.1]. <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1925128>.
- Bryant, R. E., R. H. Katz, and E. D. Lazowska. 2008. Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science, and Society. [online]. [cited 2013.10.1]. <<http://www.just.edu.jo/~amerb/teaching/2-12-13/cs728/20123173012.pdf>>.
- Chen, H., R. H. Chiang, and V. Storey. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS Quarterly*, 36(4): 1165-1188.
- Gantz, J. and D. Reinsel, 2011. "Extracting Value from Chaos." *IDC Iview*. [online].

- [cited 2013.10.1].
<<http://www.itu.dk/people/rkva/2011-Fall-SMA/readings/ExtractingValuefromChaos.pdf>>.
- Gartner. 2012. Gartner Identifies the Top 10 Strategic Technology Trends for 2013. [online]. [cited 2013.9.1]. <<http://www.gartner.com/newsroom/id/2209615>>.
- Jacobs, A. 2009. "The Pathologies of Big Data." *Communications of the ACM*, 52(8): 36-44.
- Jankowski, N. 2009. The Contours and Challenges of e-Research. In N. Jankowski (Ed.), *E-research: Transformations in Scholarly Practice*. London: Routledge, 3-34.
- Kalil, T. 2012. Big Data is a Big Deal. [online]. [cited 2013.9.16]. <<http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>>.
- Lariviere, V., C.R. Sugimoto, and B. Cronin. 2012. "A Bibliometric Chronicling of Library and Information Science's First Hundred Years." *Journal of the American Society for Information Science & Technology*, 63(5): 997-1016.
- Lehong, H. and D. Laney. 2013. *Toolkit: Board-Ready Slides on Big Data Trends and Opportunities*. Stamford, CT: Gartner.
- Loukides, M. 2012. *What is Data Science?* Sebastopol, CA: O'Reilly Media. [online]. [cited 2013.10.3]. <<http://radar.oreilly.com/2010/06/what-is-data-science.html#data-scientists>>.
- Losee, R. M. 2013. "Informational Facts and the Metainformation Inherent in Ifacts: The Soul of Data Sciences." *Journal of Library Metadata*, 13(1): 59-74.
- Manyika, J. et al. 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute. [online]. [cited 2013.10.9]. <http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation>.
- McAfee, A. and E. Brynjolfsson. 2012. "Big Data: The Management Revolution." *Harvard Business Review*, 90(10): 60-66.
- Miller, H. E. 2013. "Big-data in Cloud Computing: A Taxonomy of Risks." *Information Research*, 18(1) paper 571. [online]. [cited 2013.10.11]. <<http://InformationR.net/ir/18-1/paper571.html>>.
- Park, H. W. and L. Leydesdorff. 2013. "Decomposing Social and Semantic Networks in Emerging "Big Data" Research." *Journal of Informetrics*, 7: 756-765.
- Rousseau, R. A. 2012. "View on Big Data and Its Relation to Informetrics." *Chinese Journal of Library and Information Science*, 5(3): 12-26.
- Small, H., K. Kasianovitz, R. Blanford, and

- I. Celaya. 2012. "What Your Tweets Tell Us About You: Identity, Ownership and Privacy of Twitter Data." *The International Journal of Digital Curation*, 7(1): 174-197.
- Sugimoto, C.R., Y. Ding, and M. Thelwall. 2012. "Library and Information Science in the Big Data Era: Funding, Projects, and Future." *Proceedings of the American Society for Information Science and Technology*, 49(1): 1-3.
- Thelwall, M., P. Wouters, and J. Fry. 2008. "Information-Centered Research for a Large-scale Analysis of New Information Sources." *Journal of the American Society for Information Science and Technology*, 59(9): 1523-1527.
- Villars, R. L., C. W. Olofson, and M. Eastwood. 2011. Big Data: What It is and Why You Should Care. *White Paper, IDC*. [online]. [cited 2013.10.1]. <http://sites.amd.com/us/Documents/IDC_AMD_Big_Data_Whitepaper.pdf>.

