

공공도서관 빅데이터 정보서비스에서 나타나는 윤리적 문제점에 대한 연구

A Study on Ethical problems of Big Data Information Services in Public Libraries

박지은, 성균관대학교 일반대학원 문헌정보학과, bagjieun5@gmail.com

배소영, 성균관대학교 일반대학원 문헌정보학과, ssoy1201@gmail.com

Park, Ji-Eun. Dept. of Library and Information Science, Sungkyunkwan University

Bae, So-Young. Dept. of Library and Information Science, Sungkyunkwan University

사회 전반에 걸쳐 많은 영역에서 빅데이터 기반의 기술이 개발되고 활용됨에 따라 이러한 기술이 가져올 다양한 문제들에 대하여 인식하기 시작하였다. 이에 본 연구는 현재 공공도서관이 제공하고 있는 빅데이터 기반 정보서비스 현황에 대해 정리하고, 이러한 정보서비스를 위한 데이터 수집 및 활용 과정에서 일어날 수 있는 전반적인 윤리 문제에 대해 영국의 데이터 프레임워크 등을 적용하여 5가지 측면으로 살펴 보았다.

1. 서론

빅데이터에 관한 관심이 높아지게 되면서 국내외 다양한 영역에서 빅데이터를 수집하고 분석하여 실제 업무 환경에서 운영되는 데이터 기반 활용 사례들이 급증하고 있으며, 이는 도서관 분야에서도 예외는 아니다. 싱가포르 국립도서관(National Library Board)에서는 빅데이터 기술을 활용한 도서 및 정보자원 추천 서비스를 제공하고 데이터 기반의 전략적 도서관 운영에 활용하고 있다. 또한 국내에서는 KAIST 도서관이 웹 로그 분석 시스템을 구축하여 웹서비스를 개선하는 데 활용하고 있으며 문화체육관광부에서는 2014년부터 ‘도서관 빅데이터 분석활용 구축’ 사업을 통해 분석 플랫폼 구축 및 서비스 방향을 제시하는 등의 움직임을 보이고 있다. 이처럼 많은 사례들이 도서관 빅데이터 활용의 유용성을 보여주고 있지만

이와 관련된 문제점도 무시할 수 없다.

그동안 진행되었던 도서관 빅데이터 관련 연구는 공공도서관 빅데이터 기반 서비스의 활용과 개발에 관한 연구가 주를 이루었으나 윤리에 대한 인식은 부족하였다. 따라서 본 연구에서는 현재 공공도서관이 제공하고 있는 빅데이터 기반 정보서비스 현황에 대해 정리하고, 이러한 정보서비스를 위한 데이터 수집 및 활용 과정에서 일어날 수 있는 전반적인 윤리 문제를 고찰하는 데 의의를 두었다.

2. 공공도서관 빅데이터 수집 및 활용 유형

2.1 공공도서관 빅데이터 수집 유형

빅데이터란 기존 데이터보다 방대하여 기존의

방법이나 도구로 수집·저장·분석 등이 어려운 정형 및 비정형 데이터들을 의미하며 데이터의 크기(Volume), 다양한 형태(Variety), 빠른 생성과 유통, 그리고 속도(Velocity)라는 뜻에서 ‘3Vs’라고도 한다. 도서관 분야의 빅데이터는 도서관 내·외부에서 수집 및 생성되는 데이터 중 빅데이터의 특징인 3Vs를 만족하고 도서관 분야의 가치를 창출할 수 있는 데이터로 정의할 수 있다(표순희, 2015).

이수상(2014)은 공공도서관에서 수집되는 빅데이터를 데이터 종류 및 속성에 따라 정형데이터, 반 정형데이터, 비정형데이터로 구분하였다. 따라서 본 연구에서도 도서관 빅 데이터의 유형을 <표 1>과 같이 정의한다.

2.2 공공도서관 빅데이터 활용 유형

현재 빅데이터 정보서비스를 실시하는 공공도서관은 대부분 한국과학기술정보연구원이 사업에 참여하는 도서관의 데이터, 공공데이터, 기타 데이터를 수집하여 분석한 데이터를 바탕으로 정보서비스를 개발하고 운영하고 있다. 2017년도 공공도서관 빅데이터 활용사례집을 통해 공공도서관이 빅데이터를 활용하여 제공한 정보서비

스(한국과학기술정보연구원, 2017)에 대해 살펴보면 다음과 같이 크게 5가지로 범위를 요약할 수 있다.

- 1) 맞춤 서비스 및 프로그램 개발: 요일·연령·시간대·주제별 빈도를 파악하여 프로그램을 기획 및 개발하고 진행하였다.
- 2) 맞춤서비스 운영: 대출 상위 순위 주제와 관련된 정보지원 서비스를 집중적으로 실시하거나 연령별, 성별 대출 순위를 제공하였다.
- 3) 장서운영: 장서회전율에 따라 서가를 재구성하여 미대출장서의 이용률을 제고하고자 하였으며 대출 분석을 통해 도서 구입비용 등을 수서정책에 반영하기도 하였다.
- 4) 홍보수단: 빅데이터를 이용한 도서관 실태 분석에 대한 내용이 실린 도서관 소식지를 발간하고 지역분포를 분석을 통해 지역 내 홍보의 범위를 설정하였다.
- 5) 도서관 환경개선: 이용률과 장서회전율이 높은 주제에 대하여 접근 편리성과 독서환경을 개선하고자 하였다.

<표 1> 공공도서관 빅데이터 수집 유형(이수상 2014)

구분	내용	종류
도서관 정형데이터	- 도서관 홈페이지 및 정보시스템을 통해 수집되는 데이터 중 의미 해석이 직접적으로 가능한 데이터 - 주로 고정된 필드에 저장된 데이터, 도서관에서 관계형 데이터 베이스나 스프레드시트 등으로 관리 중인 데이터	-도서관 장서데이터 -도서관 이용데이터
도서관 반정형데이터	- 도서관 홈페이지 및 정보시스템을 통해 수집되는 정보 중에서 축적은 되지만 바로 처리가 어려운 데이터 - 도서관에서 수집되는 정보를 바탕으로 도서관에서 자체 적으로 생성되는 데이터	-로그데이터 -통계데이터 -공공데이터
도서관 비정형데이터	- 도서관 내/외부에서 생성되는 도서 및 도서관 프로그램 관련 텍스트 데이터 - 분석기법에 의해 의미있는 데이터가 될 수도, 버려지는 데이터가 될 수 있는 데이터	-도서관 관련 텍스트 정보

3. 공공도서관 빅데이터 서비스에서 나타나는 윤리적 문제점

3.1 선행연구

Wang 외(2016)는 도서관에서 빅데이터를 활용할 시 나타날 수 있는 문제점으로 ‘데이터 과학자의 부족’, ‘빅데이터 적용 능력’, ‘예산 문제’, ‘기술적 문제’, ‘프라이버시’, ‘조직별 빅데이터 적용 여부 적합성’을 지적하였다. 이는 빅데이터를 활용하는 대부분의 조직에서 발생할 수 있는 이슈들로 공공도서관 또한 예외는 아니다.

또한 영국에서는 데이터 윤리를 데이터의 생성, 분석 그리고 배포할 때 취할 수 있는 가치에 대한 판단과 접근방식으로 정의하고 데이터 윤리 프레임워크를 제시하였으며 이는 크게 7가지 원칙이 있으며 다음과 같다.

- 1) 사용자의 명확한 니즈와 공공의 이익으로부터 시작하라
- 2) 실무자는 관련 법규와 실무에 대한 규칙을 숙지하라
- 3) 사용자의 요구와 비례하는 데이터를 사용하라
- 4) 데이터의 한계를 이해하라
- 5) 일정한 스칼렛 내에서 연습하고 작업하라 (그렇지 못한다면 전문가의 조언을 구하라)
- 6) 데이터 수집과 활용 작업에 있어 투명성을 보장하고 책임감 있게 수행하라
- 7) (내장)데이터를 책임감 있게 사용하라

3.2 윤리적 문제점 도출

Wang 외(2016)가 제시한 빅데이터 관련 이슈들과 영국의 데이터 윤리 프레임워크를 통해 국내 공공도서관에서 도출될 수 있는 윤리적 이슈들은 다음과 같다.

1) 개인정보보호

반정형화, 비정형화된 데이터의 비중이 커지면서 개인정보의 식별 가능성 여부를 따지는 데에 어려움이 있어 해당 정보가 개인정보에 해당하는지에 대하여 식별하고 규제 기준을 명확하게 잡기 어려워졌다. 또한 빅데이터의 실시간성은 정보의 주체로부터 개인정보를 수집하면서 동의를 구하고 그 수집 목적과 범위에 대하여 알릴 기회와 시간적 여유가 부족할 수 있다.

2) 투명성의 보장

공공도서관의 빅데이터 관련 사업은 데이터의 수집부터 분석, 활용단계에 있어서 사용하는 기술과 과정, 그리고 내용 등을 공개함으로써 투명성을 확보해야 한다. 단 목적이 위태로워서는 안되며, 개인정보와 같은 사항은 법에 준수해야 한다. 더 나아가 우수한 사례와 알고리즘 등을 공유하고 더 나은 방향을 추구할 수 있다.

3) 저작권과 소유권

도서관에서 빅데이터를 수집하는 과정에는 수집 프로그램이나 시스템에 일시적으로 데이터가 저장된다. 현행 저작권법에서 다루는 복제의 개념은 영구적인 것뿐만 아니라 일시적으로 유형물에 고정하는 것을 포함하고 있다(이진태 2013). 도서관에서 취급하는 빅데이터는 처리과정에서 개인 저작물의 복제를 수반하기 때문에 이로 인한 저작권 문제가 발생할 수 있다.

4) 감독과 규제의 부재

빅데이터 활용에 관한 윤리적 교육을 받았다 하더라도 교육을 받은 이가 교육을 받지 않은 이보다 윤리적 규범에 입각하여 업무를 수행한다고 보기 어렵다(임진택과 김양우, 2015). 그러므로 담당자들의 업무 처리와 관련된 윤리 서약과 같은 강제적 성격을 가진 장치를 추가적으로 설치해야 할 것이다.

5) 사용자 윤리

이용자와 사회에 어떤 영향을 미칠지에 대해 고려해야 한다. 공공도서관은 그 특성상 이용자의 범위가 폭넓고 다양하며 목적이 공중의 정보 이용에 있다. 빅데이터를 활용한 혁신과 가치 창출에만 집중하여 그것이 초래하는 잠재적 부작용이나 역기능에 대해 관심이 없다면 차별적인 판단을 내릴 수도 있다.

4. 결론

최근 빅데이터 기술을 활용하는 데에 있어 윤리 문제가 활발히 논의되고 있다. 현재까지는 공공도서관이 빅데이터를 활용하는 데 있어 데이터의 수집과 활용의 과정과 결과를 공개하고 있을 뿐 윤리적인 접근은 미미한 상태다. 때문에 본 연구는 도서관 정보서비스를 위한 데이터 수집 및 활용 과정에서의 윤리 문제를 상기시켰다는 데에 의의가 있다.

마지막으로 데이터 윤리는 제도만으로는 해결할 수 없는 조직적인 과제를 안고 있다. 그렇기 때문에 데이터 분석에 참여하는 공공도서관은 윤리적인 관리를 위해 내외부의 유기적인 논의를 통해 조치를 마련할 필요가 있을 것이다.

참고문헌

이수상(2014). 도서관과 빅데이터-도서관과 빅데이터 분석. 圖書館文化, 55(8), 14-25.

이진태(2013). 빅데이터 활성화와 저작권 문제:

하둡(Hadoop)을 중심으로. 계간 저작권, 26(2), 136-173.

이혜진 외(2015). 도서관 빅데이터 활용을 위한 분석 플랫폼 구축 및 서비스 방안: 도서관 빅데이터 분석 활용 체계 구축 사례를 중심으로. 디지털도서관, 79(1), 3-14.

임진택, 김양우(2015). 공공도서관의 개인정보 보호 현황분석 및 개선방안 연구: 서울·경기지역을 중심으로. 정보관리학회, 32(1), 85-108.

표순희 외(2015). 도서관 빅데이터 서비스 모형 개발에 관한 연구: 공공도서관을 중심으로. 정보관리학회지, 32(2), 63-86.

한국과학기술정보연구원(2017). 2016 도서관 빅데이터 분석활용 체계 구축 최종보고서.

한국과학기술정보연구원(2017). 2017 도서관 빅데이터 활용 사례집.

한국정보화진흥원(2015). 2015년 빅데이터 글로벌.

Data Ethics Framework. Retrieved from <https://www.gov.uk/>

Lim, Chee Kiam, Chinnasamy, Balakumar. (2013). Connecting library content using data mining and text analytics on structured and unstrured data. Paper presented at: IFLA WLIC 2013 - Singapore - Future Libraries: Infinite Possibilities in Session 152 - Reference and Information Services.

Wang et al., (2016). Exposing Library Data with Big Data Technology: A Review. In Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference.