

http://dx.doi.org/10.7236/IIBC.2014.14.1.189

IIBC 2014-1-25

빅데이터 분석/처리에 따른 생활밀착형 서비스의 프라이버시 보호 측면에서의 구조혈 연구

A Study on Structural Holes of Privacy Protection for Life Logging Service as analyzing/processing of Big-Data

강장묵*, 송유진**

Jang-Mook Kang*, You-Jin Song**

요 약 네트워크 서비스는 로컬서비스와 결합하면서 생활밀착형 서비스로 발전하고 있다. 생활밀착형 서비스는 기존의 모바일 서비스와는 달리, 위치정보와 로컬정보 그리고 소셜 네트워크서비스 정보 등을 모아 개인화된 서비스를 제공할 것으로 예상된다. 여러 정보를 모아 처리하는 과정에서 빅데이터 기술, 클라우드 기술 등이 필요하다. 이미 이에 대한 효율성 높은 알고리즘이 연구되고 있으나 반면, 생활 밀착형 서비스 모델 또는 빅데이터 환경에서의 프라이버시 보호 모델에 대한 연구는 상대적으로 미흡한 편이다. 이 글은 생활밀착형 서비스에 활용될 빅데이터 기술이 야기하는 프라이버시 문제에 대하여 구조혈 중심으로 다룬다.

Abstract SNS (Social Network Service) has evolved to life-friendly service with the combination of local services. Unlike existing mobile services, life-friendly service is expected to be personalized with gathering of local information, location information and social network service information. In the process of gathering various kinds of information, Big-data technology and Cloud technology is needed. The effective algorithm has researched for this already, however the privacy protection model hasn't researched enough in life-friendly service or big-data using circumstance. In this paper, the privacy issue is dealt with in terms of 'Structure hole', and the privacy issue comes from big-data technology of life-friendly service.

Key Words : Privacy, Personal Information, Structural Hole, Big-Data, Life Logging Service

1. 서 론

웹 2.0의 패러다임이 참여, 공유, 개방을 핵심 가치로 성장함에 따라, 미디어의 이용 기반이 집단에서 개인으로, 그리고 미디어의 주된 기능이 정보 전달에서 네트워크와의 상호작용으로 이동하게 되었다^[1].

2009년 아이폰이 국내에 소개된 후, 스마트폰은 위치 기반의 서비스와 소셜 네트워크 서비스(Social Network

Service, 이하 SNS)를 융합하면서 로컬 서비스 및 큐레이션 서비스로까지 진화하였다. 즉 SNS라는 관계망에 GPS값과 관심지역(Point of interest, 이하 POI) 콘텐츠가 융합됨에 따라 실시간 정보의 양은 빅데이터(big-data)로 발전하였다.

SNS에서 유입되는 빅데이터란 기존의 데이터베이스 소프트웨어가 저장, 관리, 분석할 수 있는 범위를 초과한 규모의 데이터를 뜻한다^[2]. 모바일 환경이 고도화됨에 따

*정회원, 고려대학교 정보창의교육연구소

**정회원, 동국대학교 정보경영학과(교신저자)

접수일자 : 2014년 1월 11일, 수정완료 : 2014년 2월 3일
게재확정일자 : 2014년 2월 7일

Received: 11 January, 2014 / Revised: 2 February, 2014

Accepted: 7 February, 2014

*Corresponding Author: song@dongguk.ac.kr

Dept. of Information Management, Dongguk University, Korea

라 빅데이터는 보다 개인화된 서비스를 큐레이션 하는 기술과 위치기반의 로컬 정보를 친구 공유로 확산 및 전파하는 특징을 갖게 된다. 빅데이터가 더욱 주목 받고 있는 것은 SNS와 같은 새로운 인터넷 서비스 등장에 따른 데이터 원천의 다양화와 이러한 데이터를 통해 새로운 가치를 창출하려는 기업이나 정부의 요구가 증가되었기 때문이다^[3].

이와 같은 전반적인 활동을 통칭하여 이용자의 생활 밀착형 서비스라고 부른다. SNS로 파악된 이용자의 관계망 정보로 위치와 관심 그리고 관련성 많은 정보를 큐레이션하여 제공하는 자동화 서비스의 백그라운드 작업으로 빅데이터가 활용되는 것이다. 따라서 빅데이터가 처리하고 분석하는 내용이 무엇이고 그 데이터가 이용자에게 언제, 어떻게, 어떤 의도로 전달되는지를 파악하는 것은 기업의 매출, 이용자의 관심뿐만 아니라 개인 정보 침해라는 염려 차원에서 중요한 프로세스이다.

이 글은 생활밀착형 서비스에서 메타 태그가 무엇이고 태그 간의 상관관계를 파악하여 빅데이터 처리 및 분석 환경에서의 프라이버시 보호를 고찰한다. 특히, 사회과학에서 다루고 있는 네트워크 구조형이 빅데이터 처리와 분석에 따른 프라이버시 침해에 미치는 관계를 분석한다.

제 2 장에서 빅데이터를 다루고 제 3 장에서 빅데이터 맥락에서의 구조형을 살펴본 후, 제 4 장에서 구조형과 프라이버시를 분석한다.

II. 빅데이터와 생활밀착형 서비스

1. 빅데이터 기술

빅데이터는 가치(value), 가변성(variability), 크기(volume), 속도(velocity), 다양성(variety)라는 5가지의 ‘V’를 특징으로 한다^[4]. 이 5가지 특징을 뒷받침하기 위한 기술은 다양하다. 예를 들어 대용량의 데이터를 빠르게 처리하기 위해서는 고성능 분산병렬 처리 기술이 보급되어야 한다. 뿐만 아니라, 빅데이터 처리를 위한 비용이 저렴하여야 한다.

이와 같은 시장과 표준 등 다양한 조건을 제외하고 비정형 대량의 데이터를 실시간으로 빠르게 처리할 수 있는 빅데이터 기술을 간략하게 소개하면 아래 표 1.과 같다.

표 1. 빅데이터 기술^[5]

Table 1. A Technology of Big-Data

범주	유형	사례
플랫폼 기술	Collect/Integrate	CEP, Crawling Robot
	Preprocess	Data Cleansing, ETL
	Store/Manage	MapReduce, NoSQL, In-memory DB
	Analyze	Text Mining, Machine Learning, Infographics,
	Visualize	Interactive Visualization
기본기술	Cloud computing, Clustering, Heterogeneous Computing, Visualization	

표 1.의 기술은 모바일 환경에서 애플리케이션을 통해서 구체화된다. 예를 들면, 전 세계적인 서비스로 성장한 네이버의 라인(line)은 메시징 프로그램으로서 다수 사람들이 실시간으로 메시지를 모바일/PC 환경 등에서 이용하게 되는데, 빅데이터 기술을 이용하면 실시간 시장 트렌드를 파악할 수 있다. 라인에서 실시간으로 급격하게 늘어나는 주요 어휘(식당 상호, 유명인 성명, 사건명 등)은 실제 이용자들의 관심사가 무엇인지를 알려준다.

그뿐만이 아니라, PC의 맥(MAC) 주소나 스마트폰의 GPS 값은 위치별로 주요 관심 키워드를 실시간으로 알려준다. 이를 토대로 SNS 서비스는 나와 유사하거나 내가 좋아하는 친구들 또는 주변인 등에 대한 정보를 공유하거나 이를 통해 생활에 아주 밀접한 정보 제공 서비스를 구현할 수 있다. 바로 빅데이터와 생활 밀착형 서비스가 연동되는 것이다.

2. 빅데이터와 생활밀착형 서비스

빅데이터 기술은 모바일 애플리케이션 개발에 적극 활용할 수 있다. 생활밀착형 서비스는 다양한 이벤트를 추출, 검색하는데 빅데이터 수집과 분석 기술을 이용하게 된다. 이벤트는 ‘주체(subject)’, ‘이벤트 속성(event-property)’, ‘객체(object)’ 등으로 구성된 트리플(triple)의 집합으로 구성된다.

흔히 개발의 편의를 위해 이벤트와 관련 속성 등에 대해 템플릿(template)로 정의하여 처리하게 되는데, 이때 템플릿 수준 또는 그 하위 수준에서 다양한 종류의 메타 태그를 삽입할 수 있다. 이 메타 태그는 생활 밀착형 서비스를 구성할 때, 서비스의 특징과 주요 내용을 결정하게 되는 값으로 메타 태그 서버를 따로 설치하여 태그만을 따로 분석하는 시스템도 가능하다.

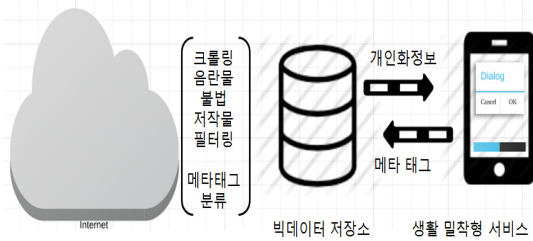


그림 1. 생활밀착형 애플리케이션의 아키텍처
Fig. 1. Architecture of Life Friendly App.

이와 같은 생활밀착형 애플리케이션의 아키텍처는 크게 SNS로부터 수집된 빅데이터를 데이터 크롤링(crawling)으로 수집한 후 음란 및 스팸 필터링(spam filtering) 등의 1차 여과 처리를 하도록 구성이 가능하다. 필터링된 빅데이터를 하둡(hadoop)의 빅데이터 저장소(repository)를 통해 저장하고 저장된 내용 중 생활 밀착형 서비스로부터 필요한 데이터를 빅데이터 저장소의 메타 태그로 매핑시켜 출력한다. Hadoop은 크게 분산 파일 시스템인 HDFS와 분산 프로그래밍 모델인 MapReduce의 두 가지 구성요소로 이루어진다.^[6]

생활밀착형 서비스는 빅데이터의 수집 및 분석을 전 처리로 하고 특정 생활밀착형 서비스의 응용 도메인에 필요한 메타값을 매핑하여 개발할 수 있다.

빅데이터가 생활밀착형 서비스에 실시간으로 상호작용하게 되는 것이다. 그림 1과 같은 빅데이터와 생활 밀착형 서비스 간의 상호작용은 메타 데이터로 처리된다.

예를 들어, 위치기반의 생활밀착형 서비스를 운영하는 회사의 경우, 고객이 현재 어느 위치에 있는지와 어느 시간대에 있는지를 중심으로 데이터가 처리되도록 설계할 수 있다. 이 경우에는 빅데이터에서 수집한 정보 중 특정 장소와 특정 시간대에 빈도수와 개인별 선호도 등을 계산하여 최적화된 정보를 이용자에게 제공할 수 있다. 이 과정에서 개인화된 서비스는 메타 데이터로 제공 가능하다. 즉 빅데이터에서 아무리 많은 정보가 존재한다고 할 지라도 결국 어떤 정보를 그 중에서 선택하느냐라는 메타값이 콘텐츠 수렴의 구조혈이 되는 것이다.

구조혈은 모든 정보를 수렴하는 역할을 한다. 반면 메타값은 특정 키워드를 미리 약속하여 정보를 수렴한다. 따라서 메타값은 구조혈을 형성하는 주요 값이 되거나 프라이버시 또는 개인정보를 수집하는 기준이 된다.

III. 구조혈(structure hole)과 프라이버시

1. 빅데이터 환경에서의 구조혈

메타 데이터란 구조화된 정보를 분석, 분류하고 부가적 정보를 추가하여 데이터를 보다 효율적으로 이용할 수 있도록 돕는 데이터를 위한 것이다. 메타 데이터는 데이터와 이의 구조에 대한 설명을 부가적으로 붙인 데이터이고 이용자가 직접 콘텐츠에 대한 정보를 붙이는 태그 등이 있다. 반면, 메타 태그란 HTML5 등에서 가장 일반적으로 사용되는 속성으로 name, content 등과 같은 속성값이다.

메타 데이터, 이용자가 입력한 태그, 브라우저에 표현될 정보의 레이아웃 등을 조정하는 메타 태그 등은 그 자체로서 새로운 가치나 제품의 성능을 향상시키지는 않는다. 그러나 빅데이터로 수집, 분석된 정보와 생활 밀착형 서비스가 융합된 환경에서 메타 데이터, 태그, 메타 태그 등은 빅데이터로부터 구조화된 질의에 응답을 신속하고 다양한 형태로 재사용되도록 돕는다.

표 2. 메타 DB와 빅데이터^[7]

Table 2. Meta DB and BigData

DB 종류	데이터	메타 태그로 관계 확장
관계 DB	관계형 데이터베이스	동일 DB 기반의 하이퍼링크(웹 1.0), 동일 DB를 기반으로 사람 및 사물 간 새로운 관계 제시(웹 2.0)
메타 DB	이용자 및 청취자에 의한 태그 분류(폭소노미), 전문가에 의한 디렉토리 분류(택사노미), 참여자에 의한 평점 및 추천으로 분류, 개인화된 포털 등 확장된 메타, 개별화되는 만물에 부여한 메타 DB	이용자에 의한 분류, 전문가 분류, 평가 및 리뷰, 댓글 및 트랙백 등 힘없고 나약한 사물 및 사람 중심의 메타 DB로의 확장
소셜 DB	지인(카페 정치원의 친구, 싸이월드 일촌의 일촌, 블로그 이웃의 이웃, 등)이 새로 업데이트한 데이터	이용자의 친구 또는 관계 맺은 지인, 참여자와 함께 구독하는 사람의 그룹, 시청자의 기사에 리뷰와 댓글을 쓴 다른 이용자
빅데이터	태그, 메타 태그, 메타 데이터 등을 통해 이용자가 입력, 디바이스가 자동으로 입력, 브라우저 등의 레이아웃에 개발자가 입력 등으로 데이터에 대한 부가적 정보가 급증	개인화된 식별 정보 추출, 개인의 맥락과 문맥을 인식한 데이터 제공

즉, 콘텐츠의 결합과 재구성 그리고 재조합을 가능하게 한다. 이 과정에서 여러 곳에 흩어진 개인정보가 수렴되거나 SNS에서 관계 데이터로 전환될 수 있다.

이와 같은 메타 데이터의 활용은 다양한 형태의 연결을 가능하게 한다. 즉, DB 확장을 통해 네트워크 효과와 서비스의 효용성을 높일 수 있다. 표 2. 네트워크 효과를 증진시키는 기술로는 RSS, ATOM, 신디케이션(syndication), OPEN API, 매시업(mashup)이 있다. 이들 기술은 정보를 물리·논리 계층에 전송하고, 다양한 이(異)기종 매체와 디바이스(device)로 연동시킨다^[8]. 이와 같은 메타 데이터, 태그, 메타 태그 등은 구조화된 데이터 표현을 기기, 사용자, 개발자가 한다는 점에서 다를 뿐, 빅데이터 서비스를 효율적으로 수행하기 위해 반드시 필요한 정보이다. 즉 빅데이터로 실시간 대규모로 유입되는 정보는 메타데이터, 태그, 메타 태그로 수렴되며, 이들 데이터에 대한 구조적 표현을 해석하고 가공하는 단계에서 구조홀이 발생하게 된다.

통상, SNS에서의 구조홀은 SSO(single sign on) 등의 기술로 여타 서비스의 프로필 기반 플랫폼이 되는 페이스북 등과 같은 서비스를 뜻한다. 사회과학에서 구조홀이란 경제, 문화, 사회, 정치 등의 현황 문제에서 인간의 관심사를 표현하는 주요 연결점이다.

빅데이터 환경에서는 이용자가 직접 입력한 태그, 개발자가 브라우저의 레이아웃 등을 잡은 메타 태그, 그리고 사진기 등과 같은 기기에서 영상정보를 저장할 때 자동으로 사진촬영날짜와 시간 등을 저장하는 메타 데이터에 의해 프라이버시가 위협받게 된다^[9].

즉, 빅데이터 환경에서 구조홀은 태그, 메타 태그, 메타 데이터에 의한 빅데이터 구조이다. 이 구조를 이용하여 다양한 서비스가 개발될 때 프라이버시는 새로운 위협에 빠지게 된다.

2. 구조홀에 의한 프라이버시 침해

구조홀의 대표적인 사례로 페이스북을 들 수 있다. 페이스북은 SSO기술로 주변 서비스와 연동된다. 예를 들면, 슬라이드웨어의 프레젠테이션 파일은 페이스북을 통해 로그인되고 공유된다. 페이스북은 친구 간의 소셜 관계를 형성하는 서비스이지만 프레젠테이션을 중심으로 한 친구맺기로 확장된다. 마찬가지로 유튜브, 인스타그램 등 다양한 분야의 도메인 서비스가 페이스북을 통해 정보를 확장한다. 즉 페이스북이 정보 가공 또는 확장

의 구조홀로 작용하는 것이다. 이 경우, 친구관계 등의 내용과 각 매체 생산 및 공유 등이 수렴됨에 따라 프라이버시는 개인의 사생활을 구체적으로 드러내게 된다. 즉 개인의 의도한 각 매체에서 흩어진 정보들이 한 장소에 모여짐에 따라 프라이버시의 위협이 높아지는 것이다.

IV. 결론

현재 우리나라는 사이버 공간에서의 자기정보 결정권이 심각하게 침해받고 있다. 구체적으로 주체로서의 이용자들에게는 개인정보의 자기결정권이, 도구로서의 컴퓨터에는 웹 접속에서의 자기 결정권이, 그리고 객체로서의 콘텐츠에는 정보 이주의 자기 결정권이 보장받지 못하고 있다^[10].

이와 같은 문제는 빅데이터 시대를 맞이하여 과거의 프라이버시 문제가 여전히 남겨진 가운데 새로운 문제로 중첩되거나 누적되는데 있다.

이 글은 빅데이터를 생활밀착형 서비스 등 여타 관계 기반 서비스들이 구현될 때 처리할 데이터에 대한 구조가 결국 프라이버시를 위협하는 구조홀이 될 수 있음을 분석하였다.

특히 이용자, 개발자, 기기가 각각 자동적으로 처리하는 개인정보 등이 빅데이터를 통해 하나의 서비스로 수렴될 때 프라이버시는 더욱 큰 위협에 빠질 수 있음을 고찰하였다.

References

- [1] Castells, M. "Communication Power", Oxford: Oxford University Press, 2009.
- [2] McKinsey Global Institute, "Big data : The Next frontier for innovation, competition, and productivity", 2011.
- [3] Hwansoo Lee, Dongwon Lim, Hangjung Zo, "Personal Information Overload and User Resistance in the Big Data Age", Journal Intell Inform Syst 2013 March: 19(1): 125-139.
- [4] JangMook Kang, "Bigdata and Privacy", 4slides, <http://www.slideshare.net/mooknc/130619ver1> or

<http://blog.naver.com/PostView.nhn?blogId=youngdisplay&logNo=60165775435>

- [5] Ahn, C. W. and S. G. Hwang, "Big Data technologies and main issues", Journal of Korean Institute of Information Scientist and Engineers, Vol.30, No.6(2012), 10-17.
- [6] Yun-Hee Kang, "Performance Analysis of MapReduce Application on Private Cloud by using OpenStack", Journal of KIIT. Vol. 11, No. 12, Dec. 31, 2013. p. 179.
- [7] Jang-Mook Kang, Woo-Jin Lee, You-Jin Song, "A Study for Vulnerability Analysis and Guideline about Social Personal Broadcasting Service based on Smart-Phone Environment (focus on SNS or U-Health)", Journal of the institute of internet, broadcasting and communication, vol. 10, (6), 2010. p. 164. table 1
- [8] Jang-Mook Kang, Woo-Jin Lee, You-Jin Song, "A Study for Vulnerability Analysis and Guideline about Social Personal Broadcasting Service based on Smart-Phone Environment (focus on SNS or U-Health)", Journal of the institute of internet, broadcasting and communication, vol. 10, (6), 2010. p. 164.
- [9]http://www.securingthehuman.org/newsletters/ouch/issues/OUCH-201204_kr.pdf
- [10] Kyung-Bae Min, Jang-Mook Kang, "Rights to Control Information and Related Security Technologies on the CyberSpace", Journal of the institute of internet, broadcasting and communication, vol. 10, 2ho, p. 136.

저자 소개

강 장 목(정회원)



- (현) 고려대학교 사범대 교수(정보창의교육연구소 소속)
- (현) 함께하는 시민행동 운영위원
- (현) 망중립성이용자 포럼 위원
- 2002년~2005년: 고려대학교 공학박사(정보보호)

<연구분야: 애플리케이션 설계, 빅데이터, 프라이버시 보호>

송 유 진(정회원)



- (현) 동국대학교 정보경영학과교수
- (현) 한국정보보호학회 부회장
- (현) 한국e-비즈니스학회 상임이사
- (현) 한국인터넷방송통신학회 이사
- 1992년~1995년: 일본 Tokyo Institute of Technology 박사

<연구분야: Privacy Protection, Secret Sharing, 클라우드 보안 및 응용, Context Aware Application Security>

※ 이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2013R1A1A2011581). 본 연구는 2013년 동국대학교 논문게재장려금 지원으로 이루어졌음