

빅데이터 처리를 위한 메타데이터 표준화에 관한 연구

정동원*, 이석훈**, 정현준**, 온병원*

*군산대학교 통계컴퓨터학과

**고려대학교 컴퓨터·전파통신공학과

e-mail: {djeong,bwon}@kunsan.ac.kr, {leha82,darkspen}@korea.ac.kr

A Study on Standardization of Metadata for Bigdata

Dongwon Jeong*, Sukhoon Lee**, Hyunjun Jung**, Byung-Won On*

*Dept. of Statistics and Computer Science, Kunsan National University

**Dept. of Computer and Radio Communications Engineering, Korea University

요 약

빅데이터는 최근 몇 년 동안 많은 관심이 집중되었으며 다양한 분야에서 빅데이터 관련 연구가 활발하게 진행되어 왔다. 이와 함께 표준화 분야에서도 빅데이터 표준화에 대한 연구가 활발하게 진행되고 있다. 하지만 아직까지 빅데이터와 관련된 메타데이터의 수집 및 관리에 대한 연구 및 표준화가 제대로 이루어지지 않고 있다. 이는 추후 빅데이터 서비스의 질적 저하, 개발 비용 증가와 같은 다양한 문제를 야기할 수 있다. 따라서 이 논문에서는 국제 표준화 기구인 ISO/IEC JTC 1을 중심으로 현재까지의 표준화 동향을 분석하고 메타데이터 관점에서의 표준화 방향에 대하여 서술한다. 이 논문은 향후 빅데이터 분야에서의 표준화 추진을 위한 기초 자료로 활용될 수 있다.

1. 서론

컴퓨팅 패러다임의 변화와 함께 데이터의 생성 속도가 급속하게 빨라졌으며, 다양한 유형의 데이터가 생성됨과 동시에 생성 규모 또한 매우 방대해졌다. 이러한 데이터의 특성을 빅데이터로 정의하고 빅데이터 처리를 위한 다양한 연구가 진행되고 있다[1]. 최근 몇 년 동안 빅데이터에 대한 관심이 높아지면서 수많은 응용 시스템 및 서비스 개발이 이루어지고 있다. 아울러 표준 분야에서도 규범적인 빅데이터 관리 및 활용을 위한 표준화에 대한 논의가 지속적으로 진행되고 있다.

최근 빅데이터 표준화를 위한 활동이 활발하게 진행되고 있으며, 대표적으로 ISO/IEC JTC 1에서는 빅데이터 연구반(Study Group on Big Data)을 구성하여 빅데이터 표준화 방향에 대한 연구를 진행하였다[2]. 이러한 연구 결과를 토대로 JTC 1 산하에 WG9를 설립하여 빅데이터 표준화를 진행하고 있다[3]. 국내에서도 한국정보통신기술협회를 중심으로 기술위원회를 신규로 설립하여 표준화를 진행하고 있다.

이 논문에서는 현재까지의 빅데이터 분야에서의 표준화 동향을 소개하고 추가적으로 표준화가 요구되는 필수 항목에 대해 서술한다. 특히 빅데이터를 위한 메타데이터에 대한 표준화의 중요성과 함께 빅데이터 생성 객체 관

리를 위한 표준화 방향에 대하여 기술한다.

이 논문의 구성은 다음과 같다. 제2장에서는 국·내외 빅데이터 표준화 동향에 대하여 서술한다. 제3장에서는 표준화 동향 분석 결과를 토대로 빅데이터 분야에서의 메타데이터 표준화의 중요성에 의미에 대하여 조망한다. 제4장에서는 빅데이터를 위한 메타데이터 관리 방안과 표준화 방향을 제안한다. 제5장에서는 결론 및 향후 연구 방향에 대하여 서술한다.

2. 빅데이터 표준화 동향

ISO/IEC JTC 1은 2013년도 빅데이터 연구반을 신설하여 표준화 방향에 대하여 연구를 진행하였으며, 이 결과를 바탕으로 JTC 1 산하에 신규 작업반을 신설하였다. 신설된 ISO/IEC JTC 1/WG 9 빅데이터 작업반(Working Group on Big Data, WGBD)은 빅데이터 정의 및 용어, 빅데이터 참조 아키텍처 등의 기본 표준 개발, 빅데이터 관련 신규 표준화를 위한 요구사항 조사/분석, 빅데이터 관련 표준화 기구와의 협업 등을 수행하고 있다[4].

ISO/IEC JTC 1/WG 9의 주요 표준화 현황을 정리하면 다음과 같다. 먼저 2015년 4월 독일에서 열린 WGBD 회의에서 ISO/IEC 20546 빅데이터 - 개요 및 용어(Big Data - Overview and Vocabulary)와 ISO/IEC 20547 빅데이터 - 참조 아키텍처 (Big Data - Reference Architecture)가 정식 프로젝트로 승인받았다[5]. ISO/IEC 20546은 빅데이터의 개요와 용어와 정의들의 집합을 제공한다. 그리고 빅데이터 연관의 표준들을 위한 기본 용어를

* 이 연구는 산업통상자원부 국가기술표준원에서 지원하는 국가표준기술력향상 사업의 일환으로 수행되었음(과제번호: 10053645).

† 공동책임저자 : 정동원, 온병원

제공한다. ISO/IEC 20547은 빅데이터 참조 아키텍처를 명세하고 빅데이터의 역할, 활동, 기능 요소들과 상호 관계를 서술한다.

2015년 7월 스페인에서 열린 WGBD 회의에서는 ISO/IEC 20547의 세분화에 대한 논의가 이루어졌으며, 이를 정리하면 다음과 같다.

- ISO/IEC TR 20547-1 : Framework and Application Process
- ISO/IEC TR 20547-2 : Use Cases and Derived Requirements
- ISO/IEC IS 20547-3 : Reference Architecture
- ISO/IEC IS 20547-4 : Security and Privacy Fabric
- ISO/IEC TR 20547-5 : Standards Roadmap

WG9는 현재 지속적인 논의를 진행하고 있으며, 차기 회의에서 보다 구체적인 논의결과를 도출할 것으로 보인다.

국내에서도 빅데이터 표준화에 대한 연구가 활발하게 진행되고 있으며, 대표적으로 한국정보통신기술협회(TTA, Telecommunication Technology Association)를 중심으로 표준화가 진행되고 있다. TTA에서는 클라우드/빅데이터 특별기술위원회(STC2)를 2014년에 신설하고 클라우드 컴퓨팅(SPG21)과 빅데이터(SPG22)를 산하 프로젝트 그룹으로 생성하여 표준을 개발 중이다. 빅데이터 표준화 그룹에서는 빅데이터 기술 분야 표준화 항목, 추진전략 및 타 프로젝트그룹과 협력방안 도출을 담당한다. 빅데이터 기술 분야 표준화는 빅데이터 기술과 빅데이터 기반 응용 서비스, 공공데이터개방을 추진한다. 또한 한국 ITU연구위원회 관련 연구반(ITU-T SG12 등), JTC 1/SC 32, JTC 1/SC 38, W3C 등과 빅데이터 기술 분야 국제표준화 협력을 한다.

<표 1>은 현재까지 TTA에서 개발한 표준 목록을 보여준다[6].

<표 1> TTA의 빅데이터 관련 표준 개발 현황

표준번호	표준 제목	제/개정일
TTAK.KO-10.0802	빅데이터 기반 데이터 웨어하우스를 위한 요구사항	2015.6
TTAK.KO-10.0801	빅데이터 기반 데이터 웨어하우스 참조모델	2015.6
TTAK.KO-10.0800	모바일 클라우드 서비스를 위한 실시간 빅데이터 처리 요구사항	2015.6
TTAK.KO-10.0780	빅데이터를 이용한 개인 맞춤형 서비스: 로그 데이터 처리를 위한 시스템 요구사항	2014.12
TTAK.KO-10.0778	빅데이터 실시간 처리 - 제1부 : 기술 개요	2014.12
TTAK.KO-10.0779	빅데이터 실시간 처리 - 제2부 : 기능 요구사항	2014.12
TTAK.KO-10.0706	클라우드 기반 빅데이터 서비스 참조 모델	2013.12
TTAK.KO	클라우드 기반 빅데이터 서비스를	2013.12

-10.0705	위한 프레임워크 정의	
TTAK.KO-10.0700	클라우드 기반 빅데이터 서비스를 위한 기능적 요구 사항	2013.12

3. 빅데이터와 메타데이터

앞서 기술하였듯이, 최근 빅데이터 표준화가 활발하게 진행 중이다. 그러나 지금까지 진행된 빅데이터에 대한 표준화는 용어, 참조모델 및 기능 요구사항 등의 수준에 머물러 있다. 즉, 구체적인 빅데이터 관리 및 처리에 대한 표준화가 이루어지지 않고 있다.

빅데이터 특성을 3V(Variety, Velocity, Volume)으로 표현한다. 이러한 특성을 지닌 빅데이터 처리 및 활용을 위해서는 다양한 유형의 데이터의 효율적 관리, 빠른 처리 속도, 정확한 처리 등이 수반되어야 한다. 지금까지 비정형 데이터의 효율적 관리, 빠른 처리 성능 및 정확성 담보를 위한 많은 연구들이 진행되어 왔다. 따라서 표준화 또한 이러한 요구 사항을 충족시킬 수 있는 방향으로 진행되어야 한다. 무엇보다 실시간으로 생성되는 방대한 양의 데이터(빅데이터)를 정확하게 처리하여 서비스의 질을 향상시키기 위해서는 다양한 빅데이터에 대한 메타데이터를 수집, 관리하는 것이 필수적으로 요구된다. 그러나 현재 빅데이터를 위한 표준화 활동에서는 메타데이터 관리 및 활용 체계가 마련되어 있지 않으며, 이는 빅데이터 서비스의 질적 저하 문제, 높은 개발 비용 문제, 거버넌스 문제 등 다양한 한계에 직면할 것으로 보인다. 또한 현재 정부 3.0 등을 통하여 급속하게 확산되고 있는 빅데이터 환경을 고려할 때, 빅데이터를 위한 메타데이터 관리 체계 및 활용 프로세스의 연구개발 및 표준화가 요구된다.

4. 빅데이터를 위한 메타데이터 관리 프레임워크

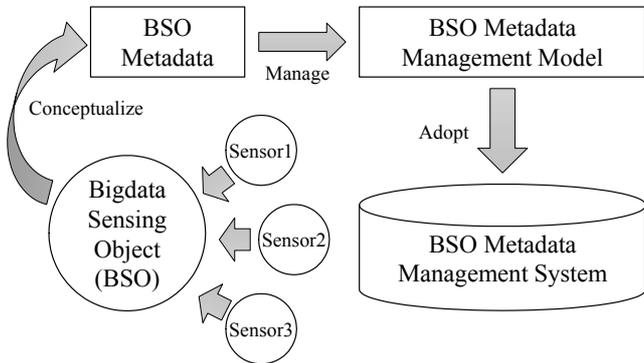
빅데이터의 메타데이터 관리 프레임워크를 위하여 이와 연관된 기존의 기술들은 다음과 같다.

- 빅데이터 정의 및 구축 기술: 빅데이터의 메타데이터를 정의하고 데이터를 구조화 하는 기술로 빅데이터 수집, 저장, 분석, 관리 및 가시화까지 메타데이터에 기반을 둔 작업들을 가능하게 함
- 메타데이터 레지스트리 기술: 메타데이터를 관리하고 공유하기 위한 기술로 메타데이터 분류 기준, 관리 구조, 명명 규칙, 권한 설정 및 생명 주기 등 메타데이터의 관리 프레임워크 제시함

이 논문은 이러한 기술들을 기반으로 빅데이터를 수집할 수 있는 영역을 특정하고 빅데이터를 위한 메타데이터 관리 프레임워크의 표준화를 제안한다. 빅데이터를 수집하는 영역을 특정 하는 이유는 빅데이터가 적용될 수 있는 영역이 광범위하기 때문이다. 모든 영역에 공통으로 적용할 수 있는 빅데이터의 메타데이터를 정의하기 위해서는 모든 영역이 지니는 공통 요소만 추출하거나, 각각의 특정된 영역의 요소들을 구체화해서 통합해야 한다. 하지만 공통 요소를 추출하는 방법은 생성되는 메타데이

터가 너무 단순하여 활용의 의미가 없고, 각 영역을 통합하는 방법은 메타데이터가 너무 복잡하여 이해하기 어렵기 때문이다.

이 논문은 메타데이터 정의 영역을 센서 네트워크와 같이 센서로부터 우리 주변의 환경 데이터나 행동 데이터 등을 수집하고 분석할 수 있는 영역으로 특정하고, 이를 빅데이터 센싱 객체(Bigdata Sensing Object, BSO)로 정의한다. 그림 1은 제안하고자 하는 빅데이터를 위한 메타데이터 관리 프레임워크를 보여준다. 먼저 빅데이터 센싱 객체는 센서들과 같이 주변 환경의 데이터를 수집할 수 있는 리소스를 의미한다. BSO들은 데이터 단계의 관점에서 공통적이고 일반적인 요소들이 추출되고 개념화된다. 이렇게 추출된 개념들은 BSO 메타데이터로서 BSO들을 표현하고 관리하는 대상이 된다. 이후 BSO 메타데이터로부터 이를 관리하기 위한 BSO 메타데이터 관리 모델을 정의한다. 정의된 관리 모델은 BSO 메타데이터를 관리 시스템에 적용되어 기본 기능 및 구조를 포함한다.



(그림 1) 빅데이터를 위한 메타데이터 관리 프레임워크

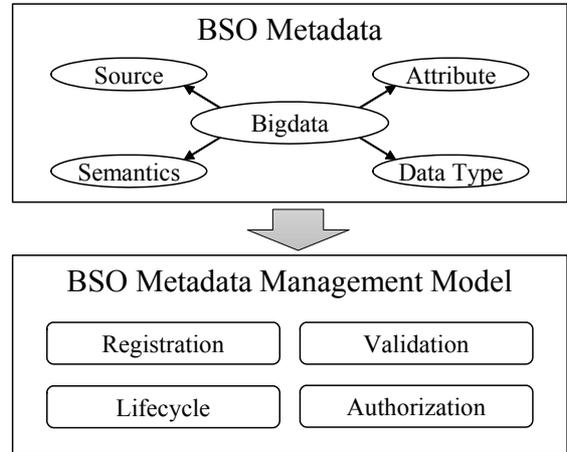
빅데이터를 위한 메타데이터 관리 프레임워크에서 핵심적으로 표준화 되어야 할 부분은 BSO 메타데이터와 BSO 메타데이터 관리 모델이다. 따라서 그림 2는 제안 프레임워크의 핵심 표준화 부분들을 보인다.

BSO 메타데이터는 어떠한 요소를 관리해야 하는지에 대해 추출해야 한다. 빅데이터가 수집되는 출처(Source), 속성(Attribute), 의미(Semantics), 데이터 유형(Data Type) 등과 같은 요소들이 구체화되어 BSO 메타데이터로 정의한다. 이때, 정의된 BSO 메타데이터는 다양한 도메인에서 적용가능 하도록 모델링해야 한다.

BSO 메타데이터 관리 모델은 BSO 관리 시스템이 지닐 수 있는 기능 및 구조를 설계한다. BSO 관리 시스템은 BSO 메타데이터를 등록(Registration)하고 검증(Validation)할 뿐만 아니라, 관리 시 필요한 BSO 메타데이터의 생명 주기(Life Cycle)와 관리 및 열람 권한(Authorization)과 같은 기능을 정의해야 한다.

5. 결론 및 향후 연구

이 논문에서는 빅데이터 분야 표준화 동향을 소개하고



(그림 2) 제안 프레임워크의 핵심 표준화 부분

현재까지의 표준화의 한계에 대하여 기술하였다. 이러한 문제점/한계를 극복하기 위한 표준화 방향을 서술하고 정확한 빅데이터 처리 및 활용을 메타데이터 관리 체계에 대한 표준화 방안을 제시하였다. 제안한 표준화 항목은 빅데이터 센싱 객체를 대상으로 하며, 향후 이에 대한 범위를 확장할 계획이다.

이 논문에서 제안한 표준화 방안은 급속히 확산되고 있는 IoT, SNS 등 다양한 빅데이터 생성 객체들을 효과적으로 통합할 수 있는 응용 서비스 개발에 활용될 수 있을 것으로 보이며, 다음과 같은 기대효과를 지닌다.

- 빅데이터 활용 및 응용 개발을 위한 규범적인 메타데이터 관리 인프라 제공
- 빅데이터 관리의 체계성 및 정확성 보장
- 보편화된 개념을 이용한 이해력 향상
- 표준 기반의 빅데이터 관리를 통한 분석 결과에 대한 신뢰성 향상
- 선진국과의 기술격차 해소 및 글로벌 경쟁력 확보

향후 연구로서, 제안한 표준화 항목에 대한 구체화 및 프로토타입 개발을 진행하고, 나아가 다양한 빅데이터 생성 객체로 범위를 확대할 계획이다.

참고문헌

[1] Wikipedia, "Big data", http://en.wikipedia.org/wiki/Big_data
 [2] ISO/IEC JTC 1/WG 9, WG9_N0001 Final SGBD Report (SGBD N0095) to JTC 1, ISO/IEC, April 2015
 [3] ISO/IEC JTC 1/WG 9, WG9_N0085 Meeting Report of 2nd JTC 1/WG 9 Meeting, ISO/IEC, July 2015
 [4] 하수욱, 이강찬, "빅데이터 상호운용성 프레임워크와 JTC 1 빅데이터 참조 아키텍처 표준화", ICT Standard Weekly, 2015-16호.
 [5] ISO/IEC JTC 1/WG 9, WG9_N0020 Report of 1st JTC 1/WG 9 Meeting, ISO/IEC, April 2015
 [6] 한국정보통신기술협회, www.tta.or.kr