

[ICT응용] 빅데이터 분석 서비스와 빅데이터 관점에서의 메타데이터 표준화

ISO/IEC JTC 1 / SGBD(Study Group on Big Data)는 빅데이터 표준화 현황에 대한 분석과, 공통 용어, 시장 요구사항 등에 따른 표준화 항목을 도출하고 향후 표준화 추진 방향을 제시하기 위하여 설립되었다. 본 고에서는 제1차 ISO/IEC JTC1/SGBD(Study Group on Big Data, 빅데이터 연구반) 회의(미국 샌디에고, 2014.3.18~21)를 통해 논의된 표준화 관련 이슈 중에서 빅데이터 분석 서비스(Data Analytics as a Service)와 빅데이터 식별 및 활용을 위한 메타데이터에 대하여 살펴본다.

빅데이터 분석 서비스

빅데이터를 정의하는 주요 특징인 3V(Volume, Variety, Velocity)는 로컬 영역에서의 데이터 분석 환경을 클라우드 기반의 데이터 서비스 형태로 변화시키고 있으며, 이에 따라 유통되는 방대한 실시간 데이터를 자신의 스토리지에 모두 저장하거나, 활용하기에는 하드웨어 및 네트워크 비용을 감당하기 어렵게 될 것이다. 빅데이터 분석 서비스는 빅데이터 서비스 제공자가 다양한 분석 기능들을 제공함으로써 스토리지 및 네트워크 비용을 절감할 수 있다는 장점을 갖는다. <그림 1>은 산업계의 빅데이터 아키텍처들을 기반으로 도출한 데이터 분석 기능을 도식화한 것으로 ETL(Extraction, Transformation, Load)/마이닝, 분석, 리포트/시각화, 거버넌스의 4개 카테고리에 따른 세부 기능 컴포넌트를 제시한다.

Big Data Analytics Components												
ETL & Mining			Analytics			Reports & Visualization			Governance			
Data Mining & Explorat.	Spatial	Text Mining	Predictive Analytics	Text Analytics	Content Analytics	BI Reporting	Scorecards & Metrics	Events & Alerts	Mobile BI	Data Profiling	Data Quality	Data Governance
Joins	Merges	Transformations	In-DB MapReduce	Semantic Analytics	Lifecycle Mgmt							
Clustering	Scaling		Sorts	Interactive Analytics & Interfaces	Real-time Analytics	Batch Analytics & Interfaces	Visual Exploration	Information Discovery	Data Conversion			
In Memory Data Grids												
Operational Database												
Analytics Database												

<그림 1> 빅데이터 분석 컴포넌트

(출처: Survey of Big Data Architecture and Framework from the Industry, NIST, 2014)

빅데이터 분석 서비스가 효과적으로 제공되기 위해서는 사용자들이 원하는 데이터가 어디에 존재

하며, 어떻게 접근할 것인지 또는 해당 데이터를 활용하여 어떠한 결과를 도출할 수 있는지 등을 제공하는지에 대한 지금까지의 개념과는 다른 개념의 메타데이터가 필요하다. 또한 다양한 데이터 포맷들을 그 활용 목적에 따라 변환할 수 있어야 하며, 상황에 따라 적합한 분석 방법과 질의 언어, 보안 정책 등을 제공할 수 있어야 한다. 그 뿐만 아니라, 데이터 분석 결과 역시 상당한 양의 데이터가 될 수 있으므로 분석 결과를 원격에서 시각화하여 전달 받는 방법 등의 기술적 고려가 필요하다.

빅데이터 관점에서의 메타데이터

기존의 구조화된 메타데이터들은 SQL 카탈로그 테이블, W3C의 XML 스키마나 XPath 등과 같이 데이터 인스턴스에 대한 가용성 식별을 위한 논리적 또는 물리적 구조를 기술하나 빅데이터 사용자 관점에서의 메타데이터는 다음과 같은 요구사항들을 반영해야 한다.

- 데이터 접근에 대한 비용(예. 접근 시간, 처리량, 비용)
- 데이터에 대한 접근성(예. 가용 시간, 신뢰성, 확실성)
- 소셜 메타데이터(예. FOAF: Friend of A Friend)

이와 같이 데이터와 메타데이터 간의 관계는 어떠한 관점에서 데이터를 바라보는가에 의존적이다. 데이터 분석 서비스 관점에서 빅데이터의 메타데이터를 기술한다면, 이는 더 이상 데이터 인스턴스 자체에 대한 개요 설명이 아닌, 데이터의 가용성, 용도, 동시에 제공 가능한 분석 모델 등을 아우르는 형태가 된다. 이러한 메타정보 접근 방법과 관련하여, 다음과 같은 기존의 표준화된 서비스 가용성을 기술하는 방법들이 있다.

- UDDI (Universal Description, Discovery and Integration), W3C
- WS-Discovery (Web Service Dynamic Discovery), W3C
- CMIS (Content Management Interoperability Services), OASIS
- GetCapabilities, OGC

빅데이터는 데이터 자체에 대한 제공과 분석 서비스를 제공하는 서비스의 두 가지 특징을 동시에 갖기 때문에, 보다 효과적인 빅데이터 공유 환경 조성을 위해서는 이러한 방법들에 대한 고려와 함께, 빅데이터 검색 및 활용과 관련된 유즈케이스, 시나리오 개발 및 분석의 병행이 필요하다.

결언

2014년 3월에 개최된 제1차 ISO/IEC JTC 1/SGBD 회의에서는 빅데이터와 기술적으로 연관성을 갖는 여러 분야들(보안 및 개인정보보호, 네트워크 가상화, 데이터 분석 서비스, SQL 등)에 대한 전반적 현황을 중심으로 진행되었다. 금번 회의가 향후 빅데이터 표준화 추진 방향에 대한 공식

적 논의 시작으로써 의의를 갖는다면, 차기 회의부터는 보다 구체적으로 표준화 방향에 대한 논의가 진행될 것으로 예상된다. 빅데이터 분석 서비스는 클라우드 기반의 신규 비즈니스 모델로 자리잡을 것으로 예상되므로, 조속한 국제 표준화 추진과 함께 관련 IPR을 확보하기 위한 노력이 병행되어야 한다.

차기 회의 일정은 다음과 같다.

- 2차 회의: 2014년 5월 13일~16일, 네델란드 암스테르담
- 3차 회의: 2014년 6월 16일~19일, 중국 북경

이강찬 (한국전자통신연구원 책임연구원, chan@etri.re.kr)

하수옥 (한국전자통신연구원 선임연구원, sw.ha@etri.re.kr)