

[메타데이터] 빅데이터 시대에서 메타데이터의 활용법은?

빅데이터 시대의 도래

지난해에 이어 올해의 큰 화두 역시 빅데이터이다. 빅데이터 기술은 기존의 하둡(Hadoop), 맵리듀스(MapReduce) 등의 거대한 양의 데이터를 얼마나 빠르게 처리할 수 있을 지에 대한 기술들을 지나 데이터들 간의 연결을 통해 상황을 파악하고, 미래의 현상을 예측하며, 개인화된 추천에까지 이를 수 있는 지를 분석하는 기술들이 큰 이슈가 되고 있다. 이에 ISO/IEC JTC1에서는 빅데이터에 대한 연구그룹(Study Group)을 설립하고 지난 3월과 5월, 6월에 각각 미국, 네덜란드, 중국에서 ISO Study Group on Bigdata(SGBD) 워크숍을 개최하는 등 전 세계적으로 빅데이터에 대한 표준화를 꾀하였다.

ISO/IEC JTC 1/SC 32/WG 2의 표준화 동향

ISO/IEC JTC1 산하의 SC32는 데이터의 관리 및 교환에 관한 표준을 담당하며 4개의 작업그룹(e-Business, Metadata, Database Language, SQL Multimedia & Application Packages)으로 구성되어 있다. 특히 제2 작업그룹(WG2)에서 개발하는 표준화 주제인 메타데이터는 상위 레벨에서의 데이터를 다루며, 개념적이고 의미적으로 데이터를 처리 및 교환하기 위한 표준화를 진행하고 있다. 메타데이터는 “다른 데이터를 정의하고 서술하는 데이터”로 정의되며 거의 모든 정보통신 영역에서 응용되고 있다.

지난 2014년 11월 캐나다 토론토의 토론토 대학에서 ISO/IEC JTC 1/SC 32/WG 2의 인터림 회의가 개최되었다. 이 회의에서는 대한민국을 포함한 총 6개국 16명의 표준 전문가들이 모여 현재 진행 중인 표준들과 추후 메타데이터 영역에서 필요한 표준화 이슈들을 놓고 회의가 진행되었다. WG2는 지금까지 ISO/IEC 11179 메타데이터 레지스트리(Metadata Registry, MDR)와 ISO/IEC 19763 상호운용성을 위한 메타모델 프레임워크(Metamodel Framework for Interoperability, MFI)에 대한 표준을 주력으로 표준화를 진행시켜왔다. ISO/IEC 11179 MDR은 메타데이터의 관리 및 등록을 위한 레지스트리로 지난 2013년 2월 MDR의 핵심 표준인 ISO/IEC 11179-3의 3판이 IS로 출판되었으며, 현재는 이를 기반으로 한 다른 파트들을 3판으로 업그레이드 하기 위한 개발에 힘쓰고 있다. 또한 ISO/IEC 19763 MFI는 온톨로지, 프로세스, 서비스와 더불어 서로 다른 MDR 간의 데이터 교환을 통한 상호운용성 증진을 꾀한다. MDR 3판 출판과 그 동안의 표준화의 노력과 따라 MFI의 세부 파트들은 DIS 및 PDTR로 진척되는 등 단계별로 표준화 프로세스를 밟아가고 있다.

메타데이터의 활용과 데이터 프로비던스

지난 WG2 토론토 인터림 회의에서는 기존 표준들의 편집회의뿐만 아니라 새로운 표준화 이슈들에 대한 열띤 논의도 이어졌다. 첫 번째 이슈는 2014년 2월에 NP를 통과하여 새로운 프로젝트로 출범한 ISO/IEC 19583 메타데이터의 개념과 활용(Concepts and Usage of

Metadata)이다. 그간 WG2에서는 주력으로 진행 중이던 MDR과 MFI의 표준화가 점차 마지막 단계로 진행되어감과 동시에 어떻게 하면 이 표준들을 활용하고 확장시킬 수 있을까에 대한 논의가 이어져 왔다. 이에 메타데이터에 대한 명확한 정의와 함께 MDR 및 MFI의 구현 모델 등에 대한 기술 보고서(Technical Report)를 개발하기로 결정하고 새로운 프로젝트인 ISO/IEC 19583을 착수하였다. 이를 위하여 한국, 미국, 캐나다, 일본, 영국 등 WG2에 참여하고 있는 각국의 대표(National Body)들이 이 표준에 참여하기로 결정하였다. 따라서 ISO/IEC 19583을 통한 메타데이터에 대한 정확한 정의와 함께 MDR, MFI등의 실질적 활용이 기대된다.

두 번째 이슈로는 데이터 프로비넌스(Data Provenance)이다. 데이터 프로비넌스란 데이터의 출처 및 근원정보를 의미하며, 데이터의 버전관리, 기관 인증, 출처 추적 등과 같은 데이터의 신뢰성(Trust) 문제를 해결하기 위한 메타데이터의 한 종류로 볼 수 있다. 특히 이번 회의에서는 데이터 프로비넌스의 정의와 표준화 범위가 논의되었으며, WG2에서 프로비넌스를 어떻게 적용할 수 있는지에 대한 내용들이 논의되었다.

메타데이터의 활용법은?

추후 MDR, MFI과 같은 ISO/IEC JTC 1/SC 32의 주력 표준들이 실질적으로 활용된다면 헬스케어, 의료, 환경, 보안, IT 등 정보통신 시스템이 기반이 되는 거의 모든 분야에서 무수한 활용 가능성을 지니게 될 것으로 보인다. 각 기업들은 메타데이터를 이용하여 데이터의 신뢰성 및 상호운용성을 보장받게 되며, 시스템들은 서로 연결되어 사용자들은 더욱 풍부하고 편리한 서비스를 제공받게 될 것이다.

이는 빅데이터에서도 마찬가지이다. 대용량의 데이터 저장 및 처리뿐 아니라 데이터들 간의 연결 및 의미처리를 통한 데이터 분석에도 메타데이터는 핵심적인 역할을 하고 있다. 메타데이터는 다양한 분야에서 넘쳐나는 데이터들이 각각의 영역을 넘어 서로 다른 데이터 영역과 연결됨으로 그 의미를 도출해 내며 분석될 수 있도록 도움을 줄 것이다. 이처럼 영역을 넘나들며 빅데이터의 상호운용 및 일관성을 보장받기 위하여 메타데이터의 활용은 필수가 될 것이다.

따라서 빅데이터를 통한 분석 및 상호운용은 MDR, MFI와 같은 표준들이 그 기반이 될 수 있다. ISO/IEC JTC 1/SC 32/WG 2에서는 메타데이터의 활용을 위하여 현재 개발되고 있는 표준들의 구현 모델을 개발하고, 이를 널리 퍼뜨리기 위하여 노력하고 있으며 향후 5~6년 내로 괄목할만한 성과가 있을 것으로 보인다.

백두권 (고려대학교 정보대학 컴퓨터학과 교수, baikdk@korea.ac.kr)