

과학데이터 관련 Z39.88 KEVFormat:Sch-Svc 확장 연구 A Study on the expansion of the Z39.88 KEVFormat:Sch-Svc for Scientific Data

김 선 태, 이 태 영*

한국과학기술정보연구원, 전북대학교

Kim sun-tae, Lee tae-young*

Korea Institute of Science and Technology
Information, Chonbuk National Univ.*

요약

DataCite 메타데이터 요소를 분석하여 OpenURL 학술 서비스 유형을 기술하기 위한 메타태그를 Key/Encoded-Value (KEV) 형식으로 확장 제안하였다. 학술 서비스 유형 분석을 위해 Scopus와 Web of Science, NDSL 서비스를 비교 검토하여 8개의 학술서비스 유형을 도출하였다. 또한 과학데이터 기술을 위한 DataCite 컨소시엄의 <RelatedIdentifier> 메타데이터 요소를 집중적으로 분석하여 9개의 대표속성을 도출 하였다.

I. 서론

세상의 모든 것이 도처에서 데이터를 생산하고 유통시키고 있는 것이다. 이러한 시대적 환경에 발맞추어 과학연구자가 수행하는 연구 패러다임이 지속적으로 변하여 최근에는 데이터가 연구의 중심이 되고 있다. 이것은 여러 가지 측면에서 큰 변화를 가져왔다. 특히 연구의 과학적 모델 측면과 학술정보 유통 측면에서 커다란 변화를 요구하고 있다. 학술 커뮤니케이션에 있어서도 많은 변화를 요구하고 있다. 문헌중심의 유통에서 과학데이터 등 다양한 자원으로서의 연계를 통해 연구자 검색이후 자연스러운 관련 자료의 발견을 이끌어 낼 수 있는 서비스 개발을 요구하고 있는 것이다. 예를 들어, 연구자가 관심을 갖는 특정 논문과 관련되어 Web 2.0 도구들을 이용하여 연구자들이 생성해 놓은 콘텐츠들 간의 연결과 이중 콘텐츠 간의 연계·융합 서비스가 요구되고 있는 것이다. 이것은 도서관이나 정보센터가 연구자들의 새로운 연구 패러다임을 지원하기 위하여 기존의 학술논문이나 특허, 연구보고서 중심의 서비스에 데이터¹⁾ 서비스를 추가 확대할 필요가 있음을 의미하기도 한다. 즉, 문헌위주의 과학기술정보 서비스 체계를 구성하는 각각의 구성요소가 과학데이터도 처리할 수 있게 확장되어야 함을 의미한다. 정확한 학술 커뮤니케이션의 보장은 연구자의 요구가 서비스 제공 시스템에 정확하게 전달될 때 가능하다. 그러기 위해서는 여러 가지 커뮤니케이션 상황을 분석하여 서비스의 핵심역할을 수행하는 링크서버의 역할이 매우 중요하다. 또한 링크서버와 타깃서비스를 제공하는 시스템 간 상호운용성 확보를 위해 시스템 간 교

환하는 메타태그의 확장도 수반되어야 한다.

1. 과학데이터 유통과 링크서버 역할

과학데이터의 유통에서 활용측면의 핵심요소는 정보의 출판, 식별, 인용이라 할 수 있다. 데이터의 출판은 연구결과를 공개하는 필수 절차로서 연구의 마지막 단계라 할 수 있다. 데이터 식별의 경우 출판된 데이터의 영속적 접근점을 제공하는 수단으로서 그 중요성이 날로 더해지고 있다. 데이터 인용은 데이터 활용의 확산을 이끌어 낸다. 이러한 과학데이터의 유통에 링크서버의 역할 강화가 요구되고 있다. 문헌콘텐츠 중심의 학술정보 유통에서 링크서버는 세 가지 측면에서 역할을 수행하고 있다. 첫째, 가장 기본적인 역할로서 정보소스와 타깃서비스를 연결해주는 매개자 역할을 수행한다. 둘째, 연구자가 활용할 수 있는 관심 콘텐츠의 복사본 문제를 해결하는데 중심적인 역할을 수행하고 있다. 문헌콘텐츠의 확산으로 이해관계를 가진 문헌콘텐츠 유통 주체들이 연구자가 관심을 둔 콘텐츠를 중복으로 서비스 할 경우에 발생 할 수 있는 문제를 해결하는 것이다. 셋째, 연구자가 소속된 기관의 라이선스 정보를 종합하여 연구자에게 발견의 기회를 제공하는 역할을 수행하고 있다. Ware와 Mabe (2009)는 STM 보고서에서 데이터가 독립적인 출판물이 될 것 이라 예측하였다. 따라서 과학데이터 유통에서 링크서버의 역할은 변함없이 요구되고 있다.[3]

2. 학술서비스 유형 및 관련 스킴 분석

2.1 학술서비스 유형 분석

범용적으로 사용될 수 있는 OpenURL 학술 커뮤니티

1) 연구자가 연구과정을 통해 생성한 모든 데이터는 연구데이터(Research Data)와 과학데이터(Scientific Data)로 불린다. 본 논문에서는 이를 혼용하여 사용한다.

서비스 유형의 핵심 메타태그를 개발하기 위하여 대표적인 학술정보 서비스인 해외 Elsevier 출판사의 Scopus와 ISI의 Web of Science (이하 WoS), 국내 과학기술정보 대표포털서비스인 NDSL (National Discovery for Science Leaders, 이하 NDSL)을 중심으로 조사하였다. 각각의 서비스는 다양한 유형의 서비스를 제공하고 있지만 본 연구에서는 OpenURL 메타 태그로 적합한 항목만 도출 하였다.

2.2 DataCite 스킴 분석

〈RelatedIdentifier〉요소와 관계유형 속성을 함께 사용하면 학술논문과 그 연구의 기초자료가 되었던 데이터의 관계를 기술할 수 있다. 또한 여러 버전의 자원들 관계와 다양한 유형의 자원 관계 또한 기술할 수 있다. 하나의 자원을 구성하는 요소자원을 기술할 수 있으며, 기술된 자료의 보충 자원도 기술 가능하다. 이 요소는 반복적으로 기술이 가능하며, 폭넓은 자원을 대상으로 하고 미묘한 관계의 유형을 자세하게 기술할 수 있게 한다.[1]

3. KEVFormat:Sch-Svc 현황

과학데이터는 생성된 데이터 자체가 연구의 결과물인 경우도 있지만 대부분 연구의 내용을 정리한 보고서, 논문 등에 더욱 자세하게 기술되어 있다. 따라서 논문과 관련된 데이터에 대한 이용자의 수요가 반드시 존재하며, 이러한 이용자의 요구가 서비스 서버에 전송될 수 있어야 한다. 하지만 현재의 6가지 메타태그로서는 이를 만족시킬 수 없다. 'any' 키를 활용 할 수도 있지만 ContextObject에 정확한 요구를 담기에는 한계가 있다.

표 1. 학술적 서비스 유형을 위한 key 구성
(NISO AX Committee 2007, 15)

Key	내용
초록 (abstract)	Referent의 초록정보
인용 (citation)	Referent의 서지 인용정보
원문 (fulltext)	Referent의 원문정보
소장정보 (holdings)	Referent 관련 소장정보
도서관상호대차 (iii)	Referent에 대한 도서관 상호대차
기타 (any)	다른 유형의 서비스

4. KEVFormat:Sch-Svc 확장

메타태그 설계원칙은 다음과 같다. 최근의 국내·외 학술 서비스 유형을 파악하고 과학데이터 기술을 위한 최근 메타데이터 포맷을 반영하였다. 또한 기존 Z39.88을 구현한 링킹서버와 연동되는 레거시 시스템의 호환성을 위해 기존 표준 메타태그를 모두 포함하였다. 과학데

이터 기술을 위해 2011년 2월에 발표된 DataCite 메타데이터 〈RelatedIdentifier〉요소의 속성 값을 모두 수용하였다. 핵심 메타태그 도출을 위해 학술서비스 분석결과 세계적인 초록 인용데이터베이스인 Scopus와 WoS, 국내의 대표 과학기술정보 포털 서비스인 NDSL의 학술 서비스 유형을 비교한 결과 SCOPUS 서비스의 경우 비교대상 서비스와 크게 차별화 할 수 있는 것은 보다 구체적인 인용정보 제공을 꼽을 수 있다. WoS 데이터베이스의 경우 타 서비스에서 제공하지 않는 인용정보 지도 (Citation Map) 서비스를 제공하고 있다. 이것은 Forward 인용(Citing)과 Backward 인용(Cited) 정보를 지도로 살펴볼 수 있는 서비스로서 인용의 깊이를 2단계 까지 지정하여 논문의 인용정보를 제공하는 것이다. NDSL은 네 가지의 특화 서비스를 제공하고 있다. 소장처 리스트 제공과 원문복사, NTIS 과제정보, 연관검색어 리스트 이다.

DataCite 스킴 중 〈RelatedIdentifier〉 요소의 type속성 값과 그에 대한 설명 그리고 사용가능 값을 그룹화 과정을 통해 대표속성을 도출한 결과를 보여준다. 자원 A와 B가 특정 관계를 맺고 있다고 가정할 경우 'A 속성 값 B' 구문의 의미를 설명하고 있다. 대표속성 중 citation은 기존 표준 메타태그 KEVFormat: Sch-Svc에 포함되어 있으며, Reference는 학술 서비스 유형 메타태그에서도 도출되어 그 중요성이 매우 크다. 총 18개의 사용가능 값은 두 개의 쌍으로 구성되어 있는데 각각을 대표하는 하나의 대표속성을 도출 하였다.[2]

OpenURL 학술 서비스 유형 메타태그 정의에 사용한 제한 언어(Constraint Language)로는 OpenURL 레지스트리에서 관리되고 있는 "The Z39.88-2004 Matrix Serialization (식별자: info:ofi/fmt:kev:mtx)" 형식을 따랐다. Key 이름은 표준에 맞게 변경하였다. 모두 소문자로 변경하였으며, '_' 문자와 공백을 사용하지 않았다. 〈표 5〉는 "Z39.88-2004 Matrix Serialization" 언어에 맞게 작성된 확장된 OpenURL 학술 서비스 유형 메타태그 정의 이다.[4] 실제 정의되어 있는 모습은 <http://bit.ly/eddqUB> 에서 확인 가능하다. Min, Max 열의 값은 학술 서비스 유형 메타태그의 특성 상 서비스 요청 여부만을 나타내기 때문에 모두 '0'과 '1'로 설정하였다.

■ 참고 문헌 ■

- [1] Joan Starr, Angela Gastl, "isCitedBy: A Metadata Scheme for DataCite," D-Lib Magazine, Vol. 17, No. 1/2, 2011
- [2] Joan Starr외 14인, "DataCite Metadata Kernel for the Publication and Citation of Research Data," DataCite Metadata Working Group, 2010
- [3] Mark Ware, Michael Mabe, "An overview of scientific and scholarly journal publishing," The stm report, 2009
- [4] OCLC, "Process for the evaluation and approval of new OpenURL community profile," Z39.88 OpenURL Maintenance Agency, 2006