

시맨틱 디지털도서관 서비스를 위한 서지 온톨로지 구축*

Implementing Bibliographic Metadata Model for Social Semantic Digital Libraries

이유진(You-Jin Lee)**

양성권(Sungkwon Yang)***

송민아(Mina Song)****

김흥기(Hong-Gee Kim)*****

초 록

MARC, DC, MODS, MarcOnt 등의 서지 메타데이터 모델과 소셜 시맨틱 디지털도서관 사례인 JeromeDL의 메타데이터 모델 및 서지적 개념모델인 FRBR모델 등의 분석을 통해 국내 디지털도서관의 서지메타데이터에 적용 가능한 온톨로지 모델을 제안하고자 한다. 이는 디지털도서관의 다양한 자원형식과 특성을 고려하고, 기존의 서지메타데이터들을 내포시켜 발전시킴으로써 서지자원에 대한 범용성과 상호운용성이 높은 서지 온톨로지를 구축하여 의미적인 검색과 서비스를 확보하고자 하였다.

ABSTRACT

We propose semantic model that is possible to apply for the bibliographic metadata of domestic digital library by analysing bibliographic metadata models like MARC, DC, MODS, JeromeDL's metadata model MarcOnt as the representative case of semantic digital library and FRBR model as the conceptual model.

키워드: 서지메타데이터, 디지털도서관, 서지온톨로지, 온톨로지모델링
bibliographic metadata, digital library, bibliographic ontology, ontology modeling

-
- * 본 연구는 2008년도 산업기술평가원 "산업고도화 기술개발사업"의 연구개발과제로 수행되었음.
 - ** (주) 아이네크 차장(yjlee@inek.co.kr) (제1저자)
 - *** 서울대학교 치과대학 의생명 지식공학연구실(sungkwon.yang@snu.ac.kr) (공동저자)
 - **** 서울대학교 치과대학 의생명 지식공학연구실(songmina@snu.ac.kr) (공동저자)
 - ***** 서울대학교 치과대학 교수(hgkim@snu.ac.kr) (공동저자)
- 논문접수일자: 2009년 2월 16일 ■ 최초심사일자: 2009년 2월 19일 ■ 게재확정일자: 2009년 2월 27일
■ 정보관리학회지, 26(1): 215-230, 2009. [DOI:10.3743/KOSIM.2009.26.1.215]

1. 서론

정보기술의 발전으로 구조화된 문헌정보 뿐 아니라 다양한 형태의 비구조화된 멀티미디어 자료의 표현이 가능하게 되었다. 서지자원을 표현하기 위한 기반기술인 메타데이터나 온톨로지는 자원에 대한 물리적 표현 뿐 아니라 의미적 표현을 위한 방법의 일환으로 보편화 되고 있다.

개념정보모델링 기술인 온톨로지는 컴퓨터가 정보자원의 의미구조나 개념구조를 이해하고 처리하는 것을 목표로 하는 시맨틱 웹(Semantic Web)에서 지능화된 지식처리 시스템을 구축하기 위한 기반기술로써 요구되고 있다. 특히 도서관정보시스템 분야에서는 MARC와 같은 기계가독형 목록 형태에 온톨로지 기술을 적용 시킴으로써, 기계와 인간이 목록 지식의 개념을 공유하고 기계간에도 개념 의미를 이해할 수 있게 하여 도서관정보시스템을 지능적인 형태로 발전시키기 위한 연구가 이루어지고 있다.

시맨틱 웹에서는 표현방식 및 구문체계(syntax)의 표준화를 통해 데이터의 의미적 상호연동성을 확보함으로써 각 자료의 독자적인 기술과 다양한 정보시스템 간의 상호운용성이 가능하다. 이를 위한 표준들 중 RDF/OWL은 특정 스키마에 제한을 두지 않고 자원에 관한 메타데이터를 제공하기 때문에 자료의 보유기관에 관계없이 자원 기술간 상호연동성을 확보할 수 있다. 특히 RDF 스키마를 이용해 자원에 대한 메타데이터 스키마의 클래스와 속성을 정의하고, OWL을 이용해 시맨틱 웹에서의 RDF 스키마의 표현력을 확보함으로써 자원에 대한 메타데이터의 의미적 상호 연동성 강화와 기계가

독률을 높일 수 있다.

본 연구의 목적은 MARC(Machine Readable Cataloging, 이하 MARC)의 의미모델 한계와 FRBR(Functional Requirements for Bibliographic Records, 이하 FRBR)모델의 메타데이터로의 실 적용 한계를 보완하고, 서지온톨로지인 JeromeDL의 MarcOnt(MARC bibliographic Ontology, 이하 MarcOnt)의 한계로 지적되어 온 서지자원에 대한 범용성을 확보하는 서지 온톨로지 모델을 구성하고 실제 디지털도서관의 서지 메타데이터를 대상으로 변환 구축 및 적용하여 봄으로써 서지 온톨로지에 대한 객관성과 타당성을 증명하고 차세대 시맨틱 디지털 도서관 서비스로 발전을 전망해 본다.

2. 이론적 배경 및 한계

2.1 서지메타데이터의 의미적 표현

서지자원간의 의미적이고 동적인 상호연동에 대한 어려움을 해결하기 위해 방대한 분량의 문헌자료와 다양한 형식의 디지털 정보자원에 대한 의미적이고 구조적인 표현을 위한 많은 서지메타데이터 연구가 이루어지고 있다.

MARCXML은 기존의 MARC를 XML 스키마를 기반으로 정보자원의 메타데이터를 의미적으로 표현할 수 있도록 개발되었다. MARCXML의 기반인 MARC는 컴퓨터로 자원의 서지데이터를 식별하고 배열하기 위한 포맷의 집합으로 모든 유형의 자료를 수용할 수 있고, 목록규칙이나 전거제어기법등을 통해 자원에 대한 기술의 일관성과 표준화가 가능하다. 그러나 MARC는

기술 요소들 간의 관계가 단순하여 기술된 자원에 대한 개념적 확장이 어렵고, 의미적인 표현기술에 있어 한계를 가지고 있다.

이에 비해, MARCXML은 MARC 레코드데이터를 XML기술을 응용하여 변환하여 표현하거나 정보자원의 메타데이터를 의미적으로 표현하기 위해 개발되었다. 그러나 MARCXML은 MARC요소를 XML구조로 단순 변환하여 표현하고 있어 XML의 근본 취지인 정보자원의 의미적인 기술에 부합하지 못하고 있으며, 메타데이터의 개념이 결여되어 있어 실용화하는데 한계가 있다.

MARC21의 서지자원요소를 기반으로 한 메타데이터 형식의 표준안인 MODS(Metadata Object Description Schema, 이하 MODS)는 MARCXML과 달리 서지자원에 대한 메타데이터모델을 기반으로 설계되어 있어 서지자원 요소 간 계층구조와 그 의미를 기술하여 표현할 수 있고, 구조의 확장성과 유연성을 확보할 수 있어 XML을 기반으로 한 응용환경을 활용할 수 있다. 또한 디지털 자원에 대한 서지자원 표준 메타데이터인 MODS는 기존의 서지메타데이터에 비해 디지털자원 형태에 대한 서지자원 요소들을 많이 포함하고 있으며, MARC21에 비해 서지자원요소에 대한 의미적 기술방식을 제공하고 있다. 하지만 MODS는 MARC의 주요요소의 부분만을 수용하고 있어 표현의 완전성에 한계를 가지고 있어 근원적인 개념 메타데이터로는 완전하지 않다.

2.2 서지의 개념모델

MARC 뿐 아니라, MARCXML과 MODS

등과 같은 기존 서지메타데이터의 한계적인 개념모델의 부재를 해결하기 위해 개념 구조적 프레임워크인 FRBR모델이 제안되었다. FRBR는 방대한 양의 서지자원에서 상호 관련된 개체들의 속성과 관계를 명확히 기술하기 위한 방법이다. 기존의 MARC, MARCXML, MODS 등의 서지 메타데이터는 개념 모델의 부재로 인해 서지자원으로부터 관련 저작간의 의미적, 통합적 유추가 힘들었으나, FRBR 모델은 메타데이터로 서지 요소와 그 의미관계를 표현할 수 있어 서지자원에 대한 의미적, 통합적 검색이 가능하다.

특히 동일한 내용의 원저작(work)이 디지털 정보매체를 통해 다양한 형태로 제작되어 다양한 방식으로 보급되는 경우, FRBR모델의 유용성이 높다. 하나의 원저작은 번역판, 개정판, 편집 등의 다양한 형태의 저작물로 파생되어 저작물간 매우 복잡한 관계를 가지게 된다. 이러한 저작물들은 원저작과 밀접한 관련이 있으므로, 번역판, 개정판 등에 상관없이 원저작물과 서지적으로 관련된 모든 자료를 한 번에 검색하여 접근할 수 있도록 하는 것이 FRBR의 취지이다. FRBR는 원저작과 서지자원과의 관계성을 표현할 수 있어 이러한 요구에 효과적이고 유용하다. 특히 저작 간의 집합형성과 전거제어에 유용하며, 멀티미디어나 네트워크 기반 디지털자료와 같은 자원을 수용하고 표현할 수 있게 된다.

하지만 개념 모델인 FRBR는 추상적인 것을 구체적인 수준에 맞추어 표현하는 방법으로, 이 모델을 제대로 이해하고 해석하여 운용하는 것은 간단하지가 않다. 또한 기존 서지자원의 FRBR로의 불완전한 변환과 구체화할 목록규

칙의 부재 등은 FRBR의 한계로 남아있다.

2.3 시맨틱 디지털도서관의 사례 및 한계점

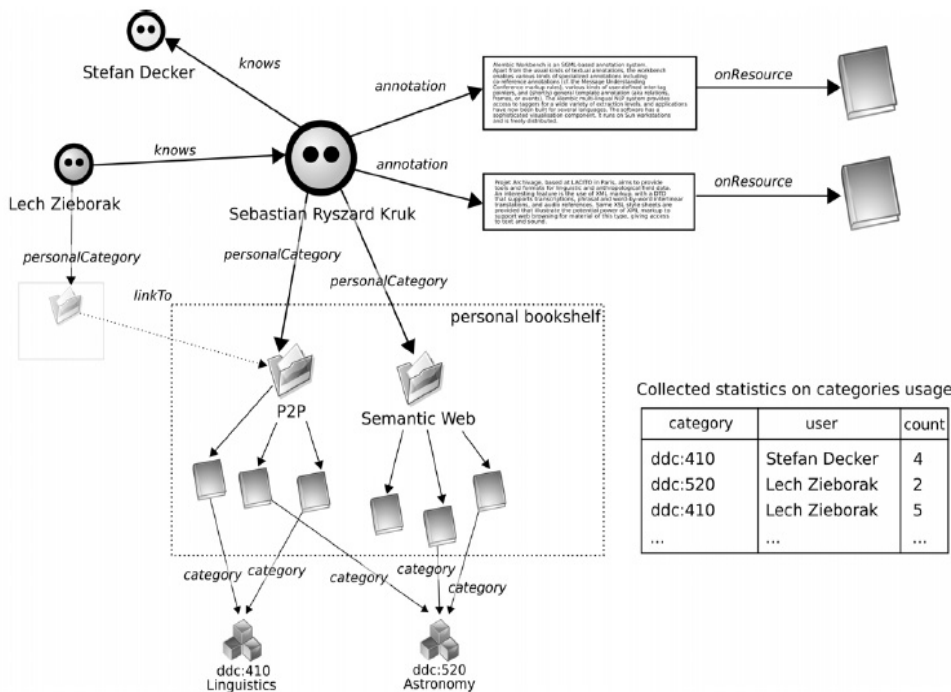
DERI 연구소와 Gdansk 공과대학 중앙도서관이 진행한 협동 프로젝트인 JeromeDL의 분석을 통해 시맨틱 디지털도서관 시스템을 살펴봄으로써 시맨틱 디지털도서관이 가지는 의미와 한계점을 파악하고자 한다.

JeromeDL은 소셜 시맨틱 디지털도서관의 요건으로 1) 전통도서관 자원의 활용 및 지원, 2) 다양한 형식의 자원 지원, 3) 사용자 정보를 활용한 의미적이고 효과적인 검색결과 제공, 4) 타 디지털도서관시스템과의 연계와 5) 사용자

정보의 보안과 계정 제한 등을 들고 있다.

특히 사용자에게 의미적이고 효과적인 검색결과를 제공하고 커뮤니티 연계 등의 서비스를 지원하기 위해 FOAF기반의 사용자 프로파일, 로그정보, 사용내역 정보 등을 활용하여 의미적으로 유용한 자원을 필터링하는 방식인 소셜 시맨틱 협업 필터링(Social Semantic Collaborative Filtering)을 통해 서비스를 제공하고 있다.

또한 시스템 관리와 서지정보, 커뮤니티 관리 등에 FOAF와 FOAFRealm을 적용하여, 사용자 프로파일의 관리와 의미적으로 관련한 사용자 간의 연계, 서지자원과 관계된 해당 인 물정보의 의미적인 관리 및 연계가 가능하도록 하였다(그림 1 참조).



<그림 1> JeromeDL의 사용자 프로파일

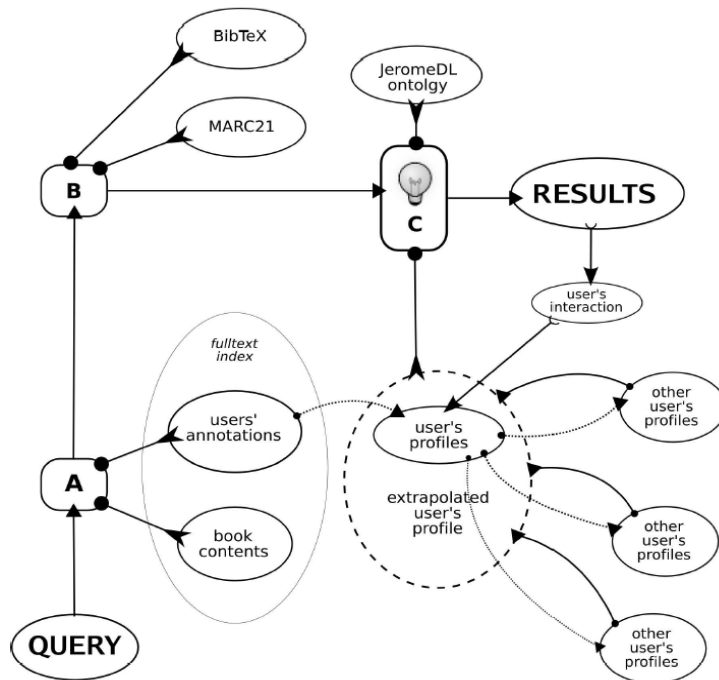
끊임없이 생성되고 확장되는 서지자원을 사용자 개인 책꽂이에 반영하고 자원을 분류하기 위해서 JeromeDL은 일반 메타데이터인 MARC21, BibTex, 더블린코어 등으로 기술된 기존의 서지정보와 온톨로지 기반의 서지정보를 함께 사용할 수 있도록 MarcOnt를 적용하였다. 이를 위한 검색 알고리즘은 3개 단계로 구성되어 있으며, 각 단계는 순차적으로 진행된다 (그림 2 참조). 각 단계에 대한 설명은 다음과 같다.

- 단계 A: 도서관 자원과 사용자가 자원에 단 주석에 대한 검색
- 단계 B: MARC21과 BibTex 형식으로 구성된 서지자원에 대한 검색
- 단계 C: 마지막 단계로서, 사용자 검색어

중 가장 높은 연관성이 있는 카테고리에 기반한 사용자 중심의 의미적인 검색

하지만 MARC21, BibTex, 더블린코어 등의 메타데이터를 사용하는 JeromeDL은 MarcOnt를 적용하였다 하더라도 서지자원에 대한 총체적이고 통합적인 의미기반 검색에 있어 한계가 있다.

따라서 본 연구에서는 개념모델인 FRBR에 기존 메타데이터 서지모델들을 적용시킴으로써 광범위하고 다양한 서지자원의 의미관계에 기반한 통합적인 검색이 가능하도록 접근할 것이다.



<그림 2> JeromeDL의 시맨틱 검색 알고리즘

3. 서지 온톨로지의 설계 원칙

3.1 목적

앞서 언급하였듯이 현재 사용하고 있는 MARC나 MODS는 의미적 상호운용성을 충분히 보장해주지 못하고 있으며, RDF로 표현하고자 한 최근의 연구도 XML로 표현된 MARC 데이터를 RDF로 표현한 것에 그치고 있다. 또한 이러한 메타데이터 모델이 가지고 있는 서지에 대한 개념적인 모델도 FRBR과 같이 검증된 개념 모델을 포괄하고 있지 못하는 실정이다. 이러한 한계점을 보완하기 위하여 본 연구에서는 기존에 검증된 개념적인 모델과 풍부한 메타데이터를 정의하는 모델을 적극 활용하고 의미적으로 표현하여 디지털도서관에서 활용할 수 있는 서지 온톨로지를 설계하였다.

3.2 설계 원칙

다음과 같은 설계원칙들을 기준으로 서지 온톨로지를 설계하였다.

3.2.1 상호운용성(interoperability)

서로 다른 서지 데이터베이스, 서비스 간에 서지 데이터 교환에 있어 의미적인 유실이나 충돌이 없도록 하기 위하여, 의미적 상호운용성을 충분히 제공하는 OWL을 사용하였다.

3.2.2 범용성(generality)

서지 모델에서 주요한 개체들을 추상적인 계층에서 구조화시켜, 특정 서지 데이터베이스나 서비스에 한정되지 않고 범용적으로 사용할 수

있도록 설계하였다.

3.2.3 호환성(compatibility)

서지 온톨로지의 범위 내에 있는 개념들 중 검증된 온톨로지에 이미 존재하는 개념들은 새롭게 정의 하지 않고, 참조하여 사용하거나 상속 받음으로서 다른 온톨로지의 인스턴스들과 쉽게 호환할 수 있도록 하였다. DC나 DCTERM의 일반적인 메타데이터 속성들을 상속 받고, FOAF에서는 인물과 기관과 관련한 개념과 속성들을 참조하였다.

3.2.4 현실성(reality)

서지를 단순히 하나의 개념으로 표현하는 MARC와 같은 서지 메타데이터 모델과 달리 보다 본질을 잘 반영하는 개념모델을 채택하여 서지를 현실적으로 표현할 수 있도록 설계하였다.

FRBR 모델을 적용하여 서지를 '저작', '표현', '구현', '개체'로 나누어 표현하였다. 현실을 반영한 모델을 통하여, 보다 의미적으로 정확하고, 다양한 표현이 가능하다.

3.2.5 사용성(usability)

네임스페이스를 사용하여 주제별로 개념들을 분류하여 온톨로지의 확장과 재사용을 용이하게 하도록 설계하였으며, 온톨로지를 사용하는 전문가들의 가독성을 위하여 일반적으로 RDF/OWL에서 사용하는 이름 규칙을 사용하였다. 또한, 서지 전문가들로 하여금 이전에 사용하던 메타데이터 모델과의 친숙함을 위하여 되도록 개념이나 속성, 관계의 이름을 동일하게 설계하였다.

4. 서지 온톨로지 구축

4.1 디지털도서관 자원의 종류

디지털도서관에서 다루는 자원은 그 범위가 매우 방대하다. 도서관이라는 특성상 이용자 대상에 따라 다루어지는 콘텐츠가 달라질 수 있으며, 특수도서관 등 성격에 따라 콘텐츠 유형도 다양해질 수 있다. 여기서는 대학도서관을 기본으로 하여 살펴보기로 한다.

대학 도서관들도 디지털화를 시작하면서 기존의 인쇄 자료형태에서 e-book, e-journal 등의 디지털자원의 비중이 크게 늘어나고 있다. 대학도서관의 장서구성을 보면, 전공과 관련해서 전 학문분야에 걸친 주제의 장서를 지니고 있으며, 형태상으로도 책자형, 비디오형, 오디오형, 지도, 고문서, 전자형태 등 다양하고, 인문, 사회, 자연과학 서고 외에도 멀티미디어실, 고서실의 형태를 두고 주제에 따라, 장서의 특성에 따라 분류하여 배치되고 있다.

디지털도서관에서는 모든 장서의 서지메타데

이터를 데이터베이스화하여 이용자에게 서비스하고 있으며, 메타데이터의 범위도 점차 확장되면서, 표지이미지, 목차, 초록, 원문까지 확대서비스를 하고 있다. 특히 웹 자원이나 전자저널과 같은 전자형태의 자료와 학술지, 논문 등의 대학도서관에서 학술적 목적으로 이용이 잦은 자료의 경우는 전자형태로 구축하여 원문서비스가 활성화되고 있으며, 이는 자관구축 자료 뿐 아니라, 타기관의 장서들까지도 URL방식으로 링크서비스로 원문열람까지를 제공하고 있다.

그 사례로 국립대학도서관과 사립대학도서관의 자원의 종류를 기술해보면 <표 1>과 같다.

4.2 디지털도서관 자원의 특성

앞에서 디지털도서관 자원의 종류를 살펴 보면서 표면적으로 나타나는 형태적 특성을 확인할 수 있다. 즉 책자형, 비도서, 전자자료의 형태적 특성과 구조적 측면에서 일회성 발간자료와, 연관관계를 가진 연속성 발간자료를 볼 수 있다. 연속간행물이란, 동일제호를 가지고 발행간

<표 1> 도서관 자원의 종류 비교

S 국립대학	구 분	S 사립대학
단행본, 학위논문	책자형 단행본유형	단행본, 학위논문
연속간행물, 대학신문, 학술지 기사색인 e-Book	책자형 연속간행물유형 전자형 단행본유형	연속간행물, 학술지 기사색인, 정부간행물, 보고서 e-Book, e-Book공급원
e-Journal, 학내간행물	전자형 연속간행물유형	e-Journal, e-Journal공급원, 전자저널, e-Learning
비도서전체, 디지털사진관, VOD	비도서유형	비디오, 오디오, 마이크로폼, 슬라이드, 지도, VOD
고문헌실 자료, 기록사료, 대학사료	고서유형	고서
e-Resource, CD-ROM, WebDB	Database	컴퓨터파일, 인터넷정보원, WebDB
학술행사	행사자료	행사자료
음악작품, 미술작품, 의학자료, 곤충자료, 농학자료, 지리학자료, 박물관, 한국학	특성화자료	교사자료, 기증자료, 교수연구업적물, 삼성경제연구소
타기관 지식정보	외부자원	SCI, 포털사이트 등 타기관 정보

기가 정기적 또는 부정기적으로 순차적인 발행 번호를 가지고 계속적으로 발행되는 간행물을 말한다. 신문, 잡지, 연감, 회보 등을 들 수 있다. 이 연속간행물의 특성을 보면, 동일제호에 순차적인 발행번호를 갖는 각 권호들이 존재한다. 또 각 권호에는 여러 개의 기사색인으로 구성되어 있다. 기사색인은 각각 저자와 서명, 주제 등을 갖는 독립적인 내용의 서지 메타데이터를 포함하고 있는 것으로 각각이 장서로서의 의미를 지니게 된다. 즉, 동일제호로 대표되는 collection의 집합이 있고, 발행간기를 가지는 권호단위의 물리적인 Material이 있고, 각 권내에 기사단위의 독립적인 메타데이터를 지니는 색인단위의 Part가 존재하고 있다. 도서관의 서지 메타데이터 중에는 이러한 전체-부분의 collection 특성

을 가지는 자원이 존재한다.

본 연구에서는 디지털도서관 자원의 다양한 특성을 고려하여 온톨로지를 구성하고자 분류기준을 만들어 자원의 의미적 계층분류를 시도하였다. 분류의 기준은 먼저 광범위한 형태적 분류로 Analog와 Digital로 분류하고, 각각에서 간행형식에 따라 Press, Tape, e-Press, MultiMedia, Web Resource로 분류하였다. 여기에 계층적 분류는 아니지만, 자원의 구조적 특성인 전체-부분의 분류를 collection, Material, Part를 각각에 부여하였다. 마지막으로 세부적으로 사용되는 디지털도서관 자원의 콘텐츠 유형들을 그룹핑하여 온톨로지 객체로 정의하였다.

디지털도서관 자원의 특성들을 정리하여 보면 <표 2>와 같다.

<표 2> 디지털도서관 자원별 분류

분류기준	형태적 분류	간행형식에 따른 분류	전체-부분에 따른 분류	세부적 콘텐츠 유형별 분류(예)
Resource	AnalogResource	Press	PressCollection	AcademyConference
				BookSeries
				JournalSeries
				NewspaperSeries
				ResearchReportSeries
			PressMaterial	Book
				JournalVolume
				Map
				Newspaper
		PressPart	PresentationNote	
ResearchReport				
Tape	TapeCollection	Thesis		
		BookChapter		
		JournalArticle		
		CassetteTapeCollection		
		MicroformCollection		
SlideCollection				
VideoTapeCollection				
CassetteTape				

(표 2 계속)

분류기준	형태적 분류	간행형식에 따른 분류	전체-부분에 따른 분류	세부적 컨텐츠 유형별 분류(예)
Resource	AnalogResource	Tape	TapeMaterial	MicroForm
				Slide
				VideoTape
Resource	DigitalResource	DigitalMedia	MediaCollection	AudioCDCollection
				CDCollection
				DVDCollection
			MediaMaterial	AudioCD
				CD
				DVD
		e-PressCollection	VOD	
			e-AcademyConference	
			e-BookSeries	
			e-JournalSeries	
		e-PressMaterial	e-NewspaperSeries	
			e-ResearchReportSeries	
			e-Book	
			e-JournalVolume	
			e-Learning	
			e-Newspaper	
e-PressPart	e-ResearchReport			
	e-Thesis			
WebResource	WebResourceCollection	e-BookChapter		
		e-JournalArticle		
	WebResourceMaterial	WebDatabase		
		Course		
		PresentationSlide		
WebResourcePart	UniversityMaterial			
	Lecture			
			Item	

4.3 서지 온톨로지 구성

서지 온톨로지의 설계는 온톨로지 개발에 있어서 기본적으로 가장 빈번히 참조되는 “Ontology Development 101”에서 제안하는 다음과 같은 절차를 따른다.

- 1) Determine domain and scope
- 2) Consider reusing existing ontology

- 3) Enumerate terms in the ontology
- 4) Define classes and hierarchy
- 5) Define properties
- 6) Refine properties
- 7) Create instance

4.3.1 영역과 범위 결정

제목에서 알 수 있듯이 구축하고자 하는 온톨로지의 도메인은 디지털도서관 서비스에서

이용되는 콘텐츠를 대상으로 한다. 대상 콘텐츠의 범위는 디지털도서관에서 서비스되고 있는 다양한 콘텐츠 유형의 자료가 가진 서지 메타데이터 들을 주요 온톨로지 구축범위로 한다. 현재 디지털도서관에서 이용되는 대상 콘텐츠에 대해서는 앞에서 자원의 종류로 분류하여 기술하였고, 서지메타데이터의 요소들은 온톨로지를 구성하면서 기술한다. 서지 온톨로지의 구축에서 고려한 또 한가지의 대상은 FRBR 개념모델의 요소들도 포함하고 있다.

4.3.2 기존 온톨로지의 재사용

기존에 존재하는 온톨로지의 활용을 고려한다는 원칙은 구축된 온톨로지의 메타데이터의 상호운용성을 높여주고 다른 온톨로지와의 호환성을 고려할 때 꼭 필요한 원칙이 된다. 따라서 기존에 서지 메타데이터를 운영하고 있는 형식은 MARC와 DC(Dublin Core)가 있다. 최근의 디지털 또는 웹 자원에 대한 콘텐츠를 도서관에서는 XML방식으로 기술하고, 그 내용과 의미는 주로 DC로 표현하고 있으며, 그 외의 도서, 비도서 등 전통적인 콘텐츠는 MARC로 표현되고 있다. MARC 형식은 서지 메타데이터의 의미가 기호화되어 있기 때문에 이를 의미화하여 기술한 MODS를 개념적으로 활용하였다. 하지만 MARC, MODS 모두 기존 온톨로지는 아니므로, 온톨로지 모델링 틀에서 DC를 import해서 추출한 주요 메타데이터 요소와 매핑되어 재사용한다. 또한 JeromeDL에서 사용하고 있는 MarcOnt는 MARC와 BibTex 등의 다양한 서지 메타데이터 요소를 기반으로 구축된 RDF 온톨로지이고 디 팩토 표준(de facto standard)을 준수하고 있으므로 이에 대한 스

펙을 내포시켜 매핑하였다. MarcOnt는 DC와 마찬가지로 다양한 자원의 유형을 표현하고 있으며, 서지 메타데이터 요소에 대한 정의와 용어들 간의 관계도 정의되어 있는 온톨로지이지만, 국내의 Kormarc이 갖는 요소들과 자료유형들을 포함하기에는 연구자료적인 요소로 한정되어 있어서 문제가 있었다. 또한 기존의 MarcOnt 요소들의 대부분은 DC의 요소나 MODS요소와 매핑이 되고 있으므로 한두 가지 소수의 용어를 위하여 하나의 온톨로지를 재사용하기에는 부적합하여 기존의 DC와 MARC에 흡수 통합하여 사용하였다.

4.3.3 용어 열거

서지 메타데이터 요소들을 모두 열거하여 브레인스토밍하면서 개념과 용어를 조직화하는 과정을 거쳤다. 먼저, MARC에서 주로 사용되는 용어와 MODS의 요소들을 바탕으로 열거하고, 이를 DC용어들과도 매핑하였다. 다음으로 소셜 시맨틱 디지털 라이브러리를 대표하는 JeromeDL의 요소와 MarcOnt의 온톨로지 스키마도 함께 내포하여 공통요소들을 통일하고 필요한 요소를 추가하고, 연구소 성격이 강하거나 로컬적인 요소들은 배제하는 용어구축의 단계를 거쳤다. 마지막으로 실제 대학의 디지털도서관 서비스에서 사용되고 있는 콘텐츠의 유형과 메타데이터 요소들을 파악하여 기존 구축한 용어와 매핑하고 필요한 사항들을 추가하는 과정을 거쳐서 온톨로지에 사용될 용어를 열거하고 용어들 간의 상관관계를 규정짓는 단계를 거쳤다.

예를 들어, MarcOnt의 resource 클래스에 연관된 여러 가지 속성요소들을 MARC, MODS, DC에 <표 3>과 같이 매핑된다. MarcOnt 온톨

〈표 3〉 용어 간의 의미적 매핑

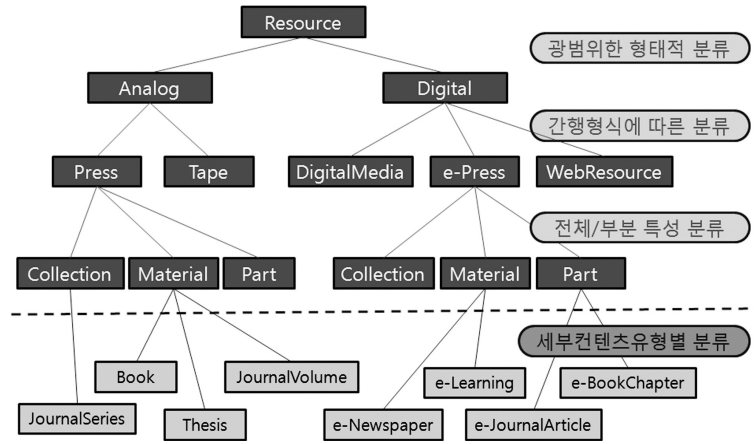
MarcOnt	MODS	MARC	DC
hasPublisher	publisher	260 \$b	publisher
hasContents	targetOfContents	500 \$a	Contents
hasHumanCreator	name	100 \$a	Creator
hasCopyright	-	-	copyright
hasEdition	originInfo > edition	250 \$a	
hasISSN	-	022 \$a	-
hasKeyword	subject	650 \$a	keyword
hasJournal	originInfo > name	440 \$a	-
hasSchool	-	502 \$b	-
hasSponsor	-	-	-
hasTagging	-	-	-
⋮			

로지를 기준으로 살펴보았을 때, resource에 연결된 속성 리스트 중 일부인데, 이들이 모두 MARC나 다른 메타데이터 표현과 매핑되지 않는다. 특히 hasSchool, hasSponsor 등은 특정 콘텐츠 유형에 국한될 수 있는 요소이며, 이에 대한 매핑요소도 의견이 다를 수 있다. hasTagging과 같이 현재의 소셜 네트워크 서비스에서 새롭게 나타나는 특징적인 요소들도 존재한다. 최근의 디지털 라이브러리 서비스에서는 web2.0 기반의 서비스를 구성하면서 콘텐츠 각각마다 이용자가 입력하고 공유하는 태깅 서비스 채택이 늘어나면서 필요성이 대두되고 있다. 따라서 이러한 요소들은 SemanticDL이라는 서지 온톨로지를 구성하는 네임스페이스에 재정의하여 사용하도록 한다.

4.3.4 클래스 및 계층구조 정의

다음단계로 온톨로지의 클래스와 클래스간의 계층관계를 정의한다. 위에서 열거된 여러 용어들 가운데에서 클래스를 추출하고 클래스 간의 관계를 정의하는 단계이다. 우선 하나의 서지를

resource라고 기준으로 하여 정의할 클래스의 계층구조를 구성해 보면 〈그림 3〉과 같다. 앞서 살펴본 디지털도서관 자원의 여러 종류와 특성을 고려하고 구성한 클래스 계층구조를 정의한 것이다. 다음으로는 서지의 개념모델을 반영하여 work-expression-manifestation-item의 제1개체 클래스와 책임개체 클래스, 그리고 subject, event와 관련된 제3개체 클래스에 대한 정의와 관계를 적용해 본다. 이번 서지 온톨로지에서는 최상위의 원저작에 대한 work, 추상적 개념인 work를 서지로 구체화시킨 manifestation 간의 관계를 클래스로 정의하고 두 클래스간의 계층관계를 구성하였다. work의 또다른 expression의 개념은 work의 속성과 work들간의 속성으로 정의하여 포함하기로 한다. FRBR모델의 책임개체인 person과 corporateBody는 각각 클래스에 해당하며, 이들은 기존 서지요소에서도 주요한 클래스로 추출된 공통적인 개체들이다. 개념, 대상, 사건, 장소에 해당하는 3개체 또한 서지요소에서 클래스로 또는 속성으로 정의되어 흡수하고 있다. 구체적인 구현물인 Manifestation이 서



〈그림 3〉 서지레코드의 계층구조 정의

지 레코드와 직접 매핑되고 있으므로 기존의 resource라 정의된 서지 콘텐츠 유형에 의한 클래스 계층관계가 manifestation이 된다. 이를 도식화 시켜보면 〈그림 4〉와 같다. 명명규칙의 기술(記述)적인 측면에 있어서 클래스의 첫 이니셜은 일반적으로 대문자로 기술한다.

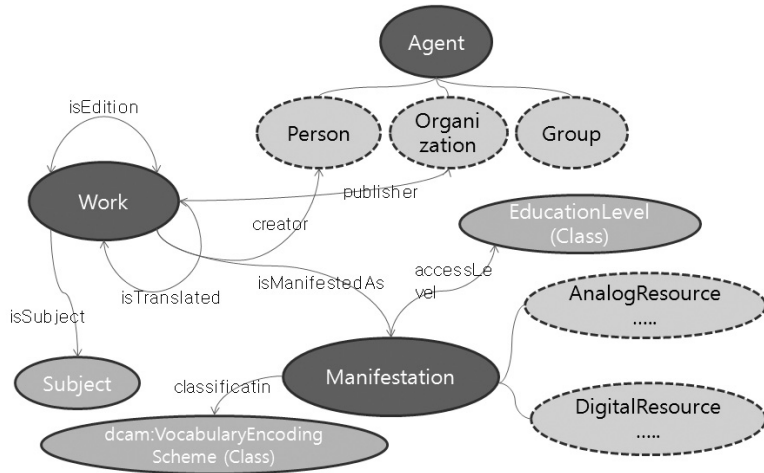
4.3.5 속성 정의

클래스와 클래스간의 계층구조의 정의와 함께 클래스의 속성을 생성, 수정, 삭제하는 단계이다. 속성을 정의함에 있어서 속성의 종류가 여러 가지 있다. DatatypeProperty, SubProperty, ObjectProperty, inverseProperty 등을 들 수 있다. 온톨로지를 정의함에 있어서 하나의 속성에는 domain과 range의 트리플 관계를 구성하게 된다. domain은 속성이 사용될 대상이 되는 클래스 범주를 가리키며, range는 속성의 값을 가질 수 있는 범위를 가리킨다. 앞의 다양한 속성의 종류도 이 range에 올 수 있는 값에 따라 나누어지는 것이다. 서지 온톨로지 중에서 속성을 가장 많이 가진 Manifestation domain

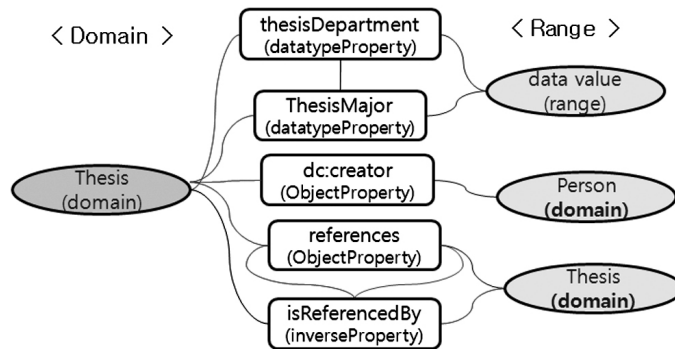
에 대한 속성 중 dcterms: abstract과 같은 text type의 초록을 가지는 경우, dcterms: created와 같이 날짜형식이 오는 생성일 등이 모두 DatatypeProperty이다.

논문과 같은 콘텐츠에서 사용되는 thesisDepartment(대학 및 학과) 속성의 SubProperty로 thesisMajor(전공) 속성을 구성하는 경우가 있으며, 이 두 속성은 모두 Manifestation의 도메인 중에서도 subclass인 Thesis를 도메인으로 가지고, DatatypeProperty이면서 두 속성 간에는 subProperty 관계를 구성하고 있는 사례이다.

ObjectProperty는 속성의 range 값으로 또 다른 domain이나 동일 domain내의 속성 값을 가지는 경우를 가리키며, Manifestation 도메인 내에 hasPart 속성이 가지는 range는 또 다른 Manifestation의 연계 값을 가지고, isPartOf 속성은 hasPart와 서로 역관계로 domain과 range를 지니게 된다. 이러한 속성이 InverseProperty이다. 이해를 돕기 위해 그림으로 표현해보면 다음과 같다.



〈그림 4〉 클래스와 클래스간의 관계 정의



〈그림 5〉 속성의 종류와 특성

〈그림 5〉와 같은 속성의 특성을 고려하면서 서지 온톨로지의 속성들을 정의한다. Manifestation이 가장 큰 도메인으로 많은 속성들을 가지게 되지만, Manifestation의 계층 클래스에 대해서도 상위에서 상속받지 않는 고유한 메타데이터 요소에 대해서는 개별적인 콘텐츠 유형 클래스에 정의하도록 한다. 예를 들면 위의 thesisMajor와 같은 thesis 클래스에서만 사용되는 요소나 series에서만 나오는 ISSN 속성 등은 포괄적으로 상위에서 보다는 해당되는 특정 콘텐츠 클래스로

한정하여 정의하였다.

4.3.6 속성 및 모델링 정제

온톨로지를 구성하고 모델링을 완성하기까지 많은 정제과정을 거치게 된다. 일차적으로 모델링 툴 내에서 트리플 관계생성에서나 논리적 오류에 의한 속성값들의 정제를 거친다. 다음으로 인스턴스를 생성하고 디지털도서관서비스 대입해 나가면서 다시 한 번 정제과정을 거치고, 구축된 온톨로지에서의 질의, 추론을 해 나

가는 과정에서도 모델링 전반에 걸친 크고 작은 정제작업이 계속될 것이다. 따라서 이 단계의 모델링 정제과정은 반복적 루틴을 돌면서 지속되는 단계라 할 수 있다. 서지 온톨로지 구축에서도 MARC, DC 등의 다양한 메타데이터 요소를 반영하여 구성된 온톨로지 모델에서 FRBR 모델을 거치면서 여러번의 정제과정을 거쳐왔고, 이후 인스턴스의 생성과 데이터 변환과정에서, 그리고 서비스를 위한 SPARQL 질의를 시행하는 과정에서 지속적으로 정제단계를 거치고 있으며, 본 과제가 완료되는 시점까지 계속될 것으로 예상하고 있다.

4.3.7 인스턴스 생성

1차적으로 온톨로지의 구성이 어느정도 완성되었다고 판단하고, 디지털도서관 서비스를 실행하면서 해당되는 서지 메타데이터 요소들을 직접 인스턴스로 입력하였다. 서지 온톨로지의 특성상 Manifestation 도메인 내에 연결된 많은 속성값들을 입력하기에는 결코 쉽지 않은 인터페이스를 가지고 있으며, 서지데이터 내에는 특수문자, 기호등의 값이 제한이 없이 자유롭게 기술될 수 있으므로, 모델링 툴에서 이를 지원하는 한계가 나타났다. 따라서 모델링 툴을 통한 인스턴스 생성은 온톨로지를 검증하는 단계적 의의를 수행하는데 그치고, 이후 실제 메타데이터를 바탕으로 매핑에 의한 변환으로 인스턴스를 일괄 생성해 나가기로 한다.

4.4 서지 온톨로지 확장

서지 온톨로지의 확장은 두 가지 측면을 고려하고 있다. 첫 번째는 구축한 서지 온톨로지

를 실제 시맨틱 디지털도서관 서비스에 적용, 활용될 수 있도록 데이터 레벨까지 직접 매핑, 변환하여 서지 온톨로지도 더 효율적으로 정제할 뿐 아니라, 실제로 온톨로지를 구축함으로써 서비스가 시맨틱화 되고, 이용자의 만족을 가져다 줄 수 있음을 검증하고자 한다. 서지 온톨로지의 확장으로 메타데이터를 변환하는 방법은 상호 매핑에 의한 변환 알고리즘을 구성하여 변환에 적용하려고 한다. 입력대상이 되는 기존 디지털도서관의 콘텐츠 서지 메타데이터 들은 주로 MARC와 DC형식을 많이 취하고 있으며, 이는 모두 XML 형식으로 다운로드가 가능하다. 기호화된 MARC를 그대로 XML 형식으로 옮긴 MARCXML에 의하여 가능한 것이다. 산출결과 대상인 서지 온톨로지도 RDF/OWL로 구성함으로 RDF XML 형식으로 나타나게 된다. 따라서 의미적 정의는 다르지만 형식이 동일하므로 이 사이에 의미적 매핑만 전달할 수 있도록 xslt(eXtensible Stylesheet Language Transformation)화일을 구성하여 변환하고자 한다. 이러한 구현을 위해서는 서지메타데이터에 대한 온톨로지 변환 알고리즘 개발이 요구된다.

두 번째 서지 온톨로지의 확장은 FRBR 온톨로지를 내포, 반영하여 개념모델을 기존 온톨로지와 결합하는 부분과 이후 소셜 시맨틱 디지털도서관 서비스를 위하여 FOAF, SIOC 등과 같은 이용자의 참여와 공유를 위한 서비스 요소들의 반영을 위한 온톨로지 모델의 병합, 또는 e-Book, e-Learning 등의 콘텐츠가 확대되면서 이에 대한 새로운 온톨로지가 구축되어 이를 추가로 반영하면서 점차로 온톨로지의 범위가 점차 커져가고 도메인도 다양해진

지식정보 서비스 기반의 온톨로지로 확장되는 부분을 고려하고자 한다. 이를 위하여 디팩토 표준 스키마들을 수용하고, 상호운용성, 범용성, 호환성, 현실성, 사용성의 원칙 하에 온톨로지 구축이 요구된다.

5. 결론

지금까지 소셜 시맨틱 지식서비스 개발을 최종 목표로 하여 시맨틱 디지털도서관 서비스를 위한 서지 온톨로지 구축에 대하여 기술하였다. 시맨틱 웹이 등장하고 이를 통한 서비스가 한 차원 성장할 수 있다는 기대와 기존의 서지메타데이터가 가지는 의미적 한계와 MARC를 대상으로 한 기존의 연구한계를 넘어서 진정

서지 온톨로지가 구축되고 실용화 될 수 있는 방안을 고려한 모델을 제시하고자 하였다. 뿐만 아니라, 독창적이기 보다는 상호운용성과 범용성, 호환성을 최대한 살려 현실성과 이용자의 사용성을 확보한 서지 온톨로지 모델을 구축함으로써 실제 디지털도서관 자원을 대상으로 적용하여 시맨틱 서비스가 가능하도록 하는데 목적을 두고 온톨로지 모델을 제시하였다. 또한 도서관 자원의 실제 메타데이터와 서지 온톨로지 간의 매핑에 의한 변환 알고리즘 개발과 FOAF, 이용자 서비스 기반의 온톨로지 로 확대한 서지 온톨로지의 구축과 확장으로 시맨틱 디지털도서관 서비스를 구현하여 이용자에게 의미검색 및 효율향상과 서비스 만족과 함께 서지 온톨로지가 적용 확대되어 국제 표준스펙으로 채택되기를 기대해 본다.

참 고 문 헌

- 김성희. 2007. 디지털도서관 상호운용성 기술 요소에 기반한 기술참조모델 설계에 관한 연구. 『한국정보관리학회지』, 24(4): 239-254.
- 노지현. 2008. KORMARC 레코드에 대한 FRBR 모델의 적용 실험. 『한국도서관·정보학회지』, 39(2): 291-312.
- 오삼균. 2007. 이용자 참여형 시맨틱 디지털도서관 아키텍처 설계. 『한국비블리아학회지』, 18(2): 229-251.
- 오삼균. 2002. 디지털도서관에서의 메타데이터 역할. 『한국정보과학회지』, 20(8): 45-57.
- 유사라. 2007. 특정주제 정보관리를 위한 온톨로지 모형 연구. 『한국정보관리학회지』, 24(1): 31-53.
- 이성숙. 2006. FRBR 모형의 적용 저략에 관한 연구. 『한국문헌정보학회지』, 40(3): 305-331.
- 이현실. 2004. OWL을 이용한 온톨로지 기반의 목록시스템 설계 연구. 『한국정보관리학회지』, 2(2): 249-267.
- 이현실. 2005. MARC의 개념 모델링 연구. 『한국

- 도서관·정보학회지』, 36(3): 275-289.
- 이혜원. 2008. 학교도서관과 MLA 연계를 위한 교육 온톨로지 모형 구축에 관한 연구. 『한국비블리아학회지』, 19(1): 19-36.
- 정은경. 2008. 디지털도서관의 의미적 상호운용성 향상을 위한 메타데이터 요소와 활용에 관한 연구. 『한국문헌정보학회지』, 42(1): 193-211.
- 조현양. 2004. 시소러스와 온톨로지의 상호 호환성에 관한 연구. 『한국정보관리학회지』, 21(4): 27-47.
- 한성국. 2005. MARC의 개념 모델링 연구. 『한국도서관·정보학회지』, 36(3): 275-289.
- 한성국. 2007. 시맨틱 라이브러리를 위한 아키텍처 참조 모델. 『한국정보관리학회지』, 24(1): 75-101.
- Marcos Andre Goncalves. 2008. "Towards a digital library theory: a formal digital library ontology." *Int J Digital Libr*, 8: 91-114.
- Sebastian Ryszard Kruk. 2005. "JeromeDL- Adding Semantic Web Technologies to Digital Libraries." *DEXA*, LNCS 3588, 716-715.
- Simon Buckingham Shum. 2000. "ScholOnto: an ontology-based digital library server for research documents and discourse." *Int J Digital Libr*, 3: 237-248.
- Rob Styles. 2008. "Semantic MARC, MARC21 and the semantic web." *Linked Data on the Web Workshop*, ISSN 1613-0073, online CEUR-WS.org/Vol-369/
- Richard Newman. 2005. *Expression of Core FRBR Concepts in RDF*.
<<http://vocab.org/frbr/core>>.
- Natalya F. Noy. 2001. *Ontology Development 101: A Guide to Creating Your First Ontology*. Technical Report SMI-2001-0880, Stanford Medical Informatics.