

메타데이터의 연계성 확보를 위한 RDF활용에 관한 연구

A study on RDF for Metadata interoperability

김이겸, 광주대학교 문헌정보학과

김태수, 연세대학교 문헌정보학과

Kim Lee-Kyum, Dept. of Library and Information Science, Kwangju Univ.

Kim Tae-Soo, Dept. of Library and Information Science, Yonsei Univ.

본 연구는 웹상에 존재하는 메타데이터의 통합이나 그 활용성을 증진시키기 위해서 RDF를 활용하여 상이한 메타데이터 형식에 정의된 기술요소간의 어의적인 연계성을 확보하기 위한 방안에 대한 연구이다. 이를 위해 RDF스키마를 기반으로 연계성 확보의 기반이 되는 공통스키마를 설계하고, 이 스키마의 해당 클래스를 rdf:domain속성값으로 메타데이터 스키마를 설계함으로써 결과적으로 공통스키마를 기준으로 메타데이터 형식상호간 연계성을 갖도록 하였다.

1. 서론

정보통신기술과 웹(Web)의 비약적인 발전은 기존의 도메인(Domain) 중심의 정보환경에서 웹 중심의 통합정보환경으로의 급속한 변화를 가져왔다. 이와 함께 웹상에 존재하는 무제한의 정보자료로부터 비롯된 정보기술(情報記述) 환경의 변화는 전통적인 MARC형식의 비판과 더불어 웹에 기반한 새로운 메타데이터형식(Metadata Format)의 대량 출현을 가져오게 되었다. 이러한 상이한 형식의 난립은 웹상에 존재하는 메타데이터의 활용이나 통합에 저해요인이 되기 때문에 형식 상호간의 연계성을 확보하는 일은 대단히 중요한 과제가 아닐 수 없다.

메타데이터의 연계성은 상이한 형식의 기술요소들 간의 어의적인 연계성과 기술(記述)에 적

용되는 구문적인 연계성을 동시에 달성될 수 있어야 하며, 특정 형식에서 다른 형식으로서의 메타데이터 변환을 위한 수단으로서가 아니라, 모든 형식의 기술요소들을 그대로 수용하면서 동시에 어의적인 연계성을 확보할 수 있어야 한다.

이러한 측면에서 연결테이블(Crosswalk)은 구문적인 연계성을 제공하지 못할 뿐만 아니라, 어의적인 연계성 측면에서도 상이한 형식의 기술요소간 1 : 1 대응을 기준으로 하기 때문에 불완전하다고 할 수 있다. 특히, 동일 형식을 채택한 경우에도 그 변형들이 다양하게 나타나기 때문에 모든 형식을 적용한 연계테이블의 작성 또한 가능하지 않다고 할 수 있다.

따라서 본 연구는 연계테이블이 갖는 단점을 극복하고, 형식 상호간 연계성을 확보할 수 있는 방안으로 RDF(자원기술모형)스키마 명세를

기반으로 공통스키마를 설계하고, 이를 메타데이터 스키마 설계에 활용하는 방안을 제시하고자 한다.

2. RDF 개요

RDF는 모형, 구문, 스키마 3가지로 구성된다. RDF모형은 속성, 속성값의 집합을 갖는 복수의 자원으로 구성되며, RDF구문명세에서 Description으로 구체화된다. RDF구문을 적용한 메타데이터의 기술은 RDF스키마와 메타데이터 스키마에 기반하기 때문에 스키마는 RDF 기술에 핵심적인 것이라 할 수 있다.

RDF모형과 구문에서는 메타데이터의 기술에 사용되는 기술속성(Properties)을 정의하거나 속성과 자원과의 관계를 규정하지 않으며, RDF스키마명세에 의해 구체화된다. 그러나 RDF스키마명세는 서명이나 저자와 같은 기술요소를 직접 제공하는 것이 아니라 그러한 기술요소나 클래스를 정의하는 수단을 제공하는 스키마정의 언어라고 할 수 있다. 즉, 메타데이터 스키마와 같은 응용스키마 정의에 사용될 수 있도록 Class, Resource, Property 등의 클래스와 이들을 인스턴스화하여 새로운 자원을 정의할 수 있도록 유형시스템(Type System) 및 관계를 나타내는 속성들을 제공하고 있다.

그림1)은 RDF스키마명세에서 제공되는 자원들의 명세로서 rdfs:Class로부터 rdf:type속성을 사용하여 각각의 인스턴스(자원)가 생성되는 과정을 보여주고 있으며, rdf:Property는 rdfs:Class의 인스턴스로, rdfs:comment는 rdf:Property의 인스턴스로 생성된 것이라 할 수 있다. 따라서 rdf:type 속성은 새로운 인스턴스를 생성하기 위한 수단이라 할 수 있으며, 이는 객체지향시스템에서 객체의 선언을 통하여 새로운 객체를 생성하는 것과 개념적으로 동일하다 할 수 있다. 따라서 메타데이터 스키마에서 기술속성은 RDF스키마에 정의된 클래스의 인스턴스화 과정을 통하여 생성된 일종의 객체라고 할 수 있다.

RDF스키마는 속성을 정의할 때 그 속성과 관련한 제한을 정의하는 수단을 제공하고 있다. 이는 속성에 의해 클래스가 설계되는 다른 객체지향시스템과는 다르게, 클래스에 의해 속성이 정의되기 때문이라 할 수 있다. 즉, 이는 속성과 별도로 필요한 클래스를 정의하고, 속성 정의시 이들 속성이 속하거나 제한 받을 필요가 있는 클래스를 선택하는 방식이라 할 수 있다. 따라서 그 선택을 위해 rdfs:domain과 rdfs:range 속성이 제공되었다고 할 수 있다.

다음은 메타데이터 스키마 정의에 활용될 수 있는 RDF스키마의 자원들의 의미와 내용을 정리한 것이다.

1) rdfs:Resource

RDF에 의해 기술되는 모든 대상을 자원(Resource)이라 하며, 클래스 rdfs:Resource의 인스턴스이다. 이것은 자바에서 객체(Object)개념과 유사하다.

2) rdf:Property

이는 RDF 자원의 하위세트인 Property를 나타내며, 메타데이터 스키마 정의시 모든 기술속성은 이 클래스유형을 갖는다.

3) rdfs:Class

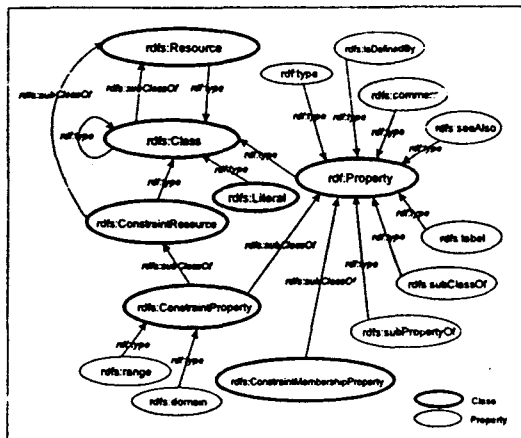


그림1) RDF스키마의 유형과 계층

이 클래스는 자바와 같은 객체지향 프로그래밍언어에서의 클래스(Class)와 유사하며, 새로운 자원을 클래스로 정의할 때, rdf:type의 속성값은 rdfs:Class가 된다.

4) rdf:type

자원이 특정 클래스의 구성원임을 지시하며, 이 때 자원은 클래스가 갖는 모든 특징을 상속받는다. rdf:type 속성값이 클래스인 자원은 그 클래스의 인스턴스(객체)가 된다. 따라서 rdf:type 속성은 클래스의 인스턴스를 만드는 수단(객체생성수단)이며, 동시에 이를 통하여 응용스키마에서 기술속성(인스턴스)을 정의할 수 있는 수단을 제공한다 할 수 있다.

5) rdfs:subClassOf

이 속성은 클래스간의 계층관계를 표현하기 위한 것으로서 클래스 확장 수단을 제공한다 할 수 있다. rdf:Class의 인스턴스만이 속성 rdfs:subClassOf를 갖을 수 있으며, 특정 클래스는 그 자신이 하위클래스로 정의될 수 없다.

6) rdfs:subPropertyOf

이 속성은 속성들간의 계층관계를 표현하기 위한 것으로 속성 확장 수단을 제공한다. 따라서 속성들은 하위속성이 없거나 subPropertyOf를 사용하여 복수의 하위속성을 정의할 수 있다.

7) rdfs:range

이는 클래스 ConstraintProperty의 인스턴스로서 속성값을 한정하기 위하여 사용된다. 이 속성값은 항상 클래스가 되며, 한 속성은 복수의 rdfs:range속성을 포함할 수 없다.

8) rdfs:domain

이 속성은 클래스 ConstraintProperty의 인스턴스로서 특정 속성이 사용되는 클래스를 한정하기 위해 사용된다. 특정 속성은 복수의 rdfs:domain속성을 포함할 수 있다.

4. RDF를 활용한 메타데이터 연계성 확보방안
RDF구문에서 사용되는 기술요소는 메타데이

터 스키마를 기초로 하기 때문에 형식 상호간의 어의적인 연계성도 그 스키마를 통하여 확보될 수 있어야 한다. 그러나 그림1)에 제시된 RDF스키마의 자원들에서 보는 바와 같이 어의적인 연계성을 위한 자원들을 제공하지 않기 때문에 RDF스키마명세에 기반한 메타데이터 스키마도 어의적인 연계성을 제공하지 못하는 문제가 있다.

따라서 모든 형식에 공통적으로 적용될 수 있는 요소들을 선별하여 집단화하여 이들을 연계성을 위한 자원으로 제공될 필요가 있다. 이러한 측면에서 더블린코어형식이 MARC에 대한 탐색요소집단으로, 다른 형식의 메타데이터에 대한 창으로 사용될 수 있으므로(Weiber, and Erway 1997) 더블린코어요소를 어의적인 연계성을 위한 기준요소로 활용하고 다른 형식과의 비교연구를 통하여 이를 보완할 필요가 있다. KORMARC과 더블린코어요소를 비교 분석한 결과 간행빈도, 지도수치정보 및 주기를 공통요소에 포함해야 할 것으로 연구되었다. 본 연구에서는 이들 공통요소들을 기준으로 어의적인 연계성을 위한 스키마를 설계하였다(공통스키마로 명명). 그림2)는 연구된 공통요소들 중 주제정보를 위한 공통스키마 계층을 나타낸 것이다.

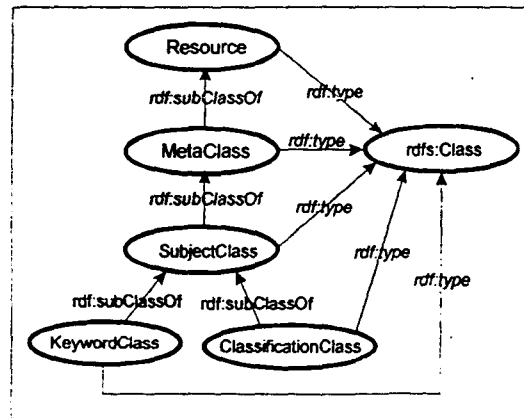


그림2) 공통스키마의 SubjectClass 계층

그림2)에서 Resource, rdfs:Class는 RDF스키마 명세에서 제공하고 있는 클래스이며, MetaClass, SubjectClass, KeywordClass, ClassificationClass 등은 공통스키마 설계를 위한 클래스이다. 이를 기반으로 축약구문방식을 적용하여 공통스키마를 설계하면 다음과 같다.

```
<?xml version="1.0" encoding="KSC5601">
  <rdf:RDF xml:lang="ko"
    xmlns="http://www.w3.org/1999/02-22-rdf-syntax-ns#"
    xmlns="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#">
    <rdfs:Class rdf:ID="MetaClass">
      <rdfs:label>공통클래스</rdfs:label>
      <rdfs:subClassOf="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#Resource"/>
    </rdfs:Class>
    <rdfs:Class rdf:ID="SubjectClass">
      <rdfs:label>주제클래스</rdfs:label>
      <rdfs:subClassOf="#MetaClass"/>
    </rdfs:Class>
    <rdfs:Class rdf:ID="ClassificationClass">
      <rdfs:label>주제-분류번호 클래스</rdfs:label>
      <rdfs:subClassOf="#SubjectClass"/>
    </rdfs:Class>
    <rdfs:Class rdf:ID="KeywordClass">
      <rdfs:label>주제-키워드 클래스</rdfs:label>
      <rdfs:subClassOf="#SubjectClass"/>
    </rdfs:Class>
```

한편, RDF스키마명세와 공통클래스를 활용하여 KORMARC의 056, 082, 650, 651필드를 대상으로 KORMARC 스키마를 설계하면 다음과 같다.

```
<rdf:Property ID="056">
  <rdfs:label>한국십진분류기호(KDC)</rdfs:label>
  <rdfs:domain rdf:resource="Books"/>
  <rdfs:domain rdf:resource="NonBooks"/>
  <rdfs:domain rdf:resource="Serials"/>
</rdf:Property>
<rdf:Property ID="056$a">
  <rdfs:label>한국십진분류기호(KDC)</rdfs:label>
  <rdfs:subPropertyOf rdf:resource="#056"/>
  <rdfs:domain rdf:resource="http://hosim.kwangju.ac.kr/~leekyum/meta_schema#ClassificationClass"/>
</rdf:Property>
<rdf:Property ID="082">
  <rdfs:label>듀이십진분류기호(DDC)</rdfs:label>
  <rdfs:domain rdf:resource="Books"/>
  <rdfs:domain rdf:resource="NonBooks"/>
```

```
<rdfs:domain rdf:resource="Serials"/>
  <rdfs:range rdf:resource="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#Literal"/>
</rdf:Property>
<rdf:Property ID="082$a">
  <rdfs:label>듀이십진분류기호(DDC)</rdfs:label>
  <rdfs:subPropertyOf rdf:resource="#082"/>
  <rdfs:domain rdf:resource="http://hosim.kwangju.ac.kr/~leekyum/meta_schema#ClassificationClass"/>
</rdf:Property>
<rdf:Property ID="650">
  <rdfs:label>주제명부출표목-일반주제명</rdfs:label>
  <rdfs:domain rdf:resource="Books"/>
  <rdfs:domain rdf:resource="NonBooks"/>
  <rdfs:domain rdf:resource="Serials"/>
  <rdfs:domain rdf:resource="http://hosim.kwangju.ac.kr/~leekyum/meta_schema#KeywordClass"/>
  <rdfs:range rdf:resource="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#Literal"/>
</rdf:Property>
<rdf:Property ID="651">
  <rdfs:label>주제명부출표목-지명</rdfs:label>
  <rdfs:domain rdf:resource="Books"/>
  <rdfs:domain rdf:resource="NonBooks"/>
  <rdfs:domain rdf:resource="Serials"/>
  <rdfs:domain rdf:resource="http://hosim.kwangju.ac.kr/~leekyum/meta_schema#KeywordClass"/>
  <rdfs:range rdf:resource="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#Literal"/>
</rdf:Property>
```

위의 KORMARC스키마에서 056\$a, 082\$a는 공통스키마의 ClassificationClass를, 650, 651은 KeywordClass를 각각 rdf:domain속성값으로 정의하였음을 보여주고 있다. 따라서 메타데이터의 기술에서 기술요소 056\$a와 082\$a가 사용되었다면 이들은 키워드로, 마찬가지로 650, 651은 분류기호로 통일되게 식별될 수 있다. 따라서 공통스키마를 활용하여 모든 메타데이터 스키마를 설계한다면 공통스키마의 각각의 클래스를 기준으로 형식 상호간의 어의적인 연계성을 갖게 된다.

5. 결론

RDF구문은 스키마 참조를 기초로한 구문이기 때문에 스키마는 기술에 있어 기술요소의 기반이 되지만, RDF스키마에 기초한 공통스키마설계를 통하여 상이한 형식간 어의적인 연계성을 확보할 수 있는 수단을 제공할 수 있다. 다만, 다수의 형식에 정의된 기술요소들을 대상으로

연구하여 공통스키마에 포함될 내용을 범주화하는 노력이 요구되며, 이를 RDF스키마명세서에 의의적인 연계성을 위한 기준으로 번역하여 그 활용성을 높여야 할 것이다.

참고문헌

- Alexandria Digital Library. 1997. Crosswalk from the Alexandria metadata schema to other schemas. [cited 1999.5.12]. <<http://www.alexandria.ucsb.edu/public-documents-metadata/crosswalks.html>>
- Berggren, M., and Anna Brumme. 1999. Design Considerations for the EULER project. [cited 1999.5.19]. <<http://www.emis.de/projects/EULER/Reports/pD31.html>>
- Brickley, D., and R.V.Guha. 1999. Resource description framework(RDF) Schema Specification. Proposed Recommendation (1999-03-03). [cited 1999.4.19]. <<http://www.w3.org/TR/1999/PR-rdf-schema-19990303>>
- Chang, W. W. 1998. A discussion of the relationship between RDF-Schema and UML. [cited 1999.4.19]. <<http://www.w3.org/TR/1998/NOTE-drf-uml-19980804>>
- Chapman, A. D. 1997. Dublin Core Metadata use in environment Australia. [cited 1999.3.11]. <http://www.environment.gov.au/epcg/erin/guidelines/www-standards/EA_DC.html>
- Clarke, Roger. 1997. Beyond the Dublin Core: Rich Meta-Data and convenience of use are compatible after all. [cited 1999.3.13]. <<http://www.anu.edu.au/people/Roger.Clarke/II/SublinCore.html>>
- Day, M. 1997. Interoperability between ROADS/IAFA templates and the Dublin Metadata Core Element set: mapping Dublin Core to ROADS templates. [cited 1998.7.29]. <http://www.ukoln.ac.uk/metadata/interoperability/dc_iafa.html>
- Dublin Core Metadata Initiative. 1997a. Dublin Core metadata element set: reference description. [cited 1999.2.11]. <http://purl.oclc.org/dc/about/element_set.htm>
- Dublin Core Metadata Initiative. 1997b. Coverage Element working draft 1997-09-30. [cited 1999.2.16]. <http://purl.org/DC/documents/working_drafts/wd-coverage-current.htm>
- Dublin Core Metadata Initiative. 1998a. A user guide for simple Dublin Core. draft version 5.1. [cited 1999.2.16]. <http://purl.org/DC/documents/working_drafts/wd-guide-current.htm>
- Hakala, J. 1998. The Nordic metadata project. Final report. [cited 1999.7.25]. <<http://linnea.helsinki.fi/meta/nmfinal.htm>>
- Iannella, Renato. 1998. A Idiot's Guide to the Resource Description Framework. [cited 1999.3.15]. <<http://www.dstc.edu.au/RDU/reports/RDF-Idiot/>>
- Lassila, Ora, and Ralph R. Swick. 1999. Resource description framework(RDF) model and syntax specification. W3C Recommendation. [cited 1999.4.19]. <<http://www.w3.org/TR/1999/PR-rdf-syntax-19990222/>>
- Mason, J. 1998. EdNA Metadata Elements. EdNA Metadata version 1.01. [cited 1999.5.25]. <<http://www.edna.edu.au/EdNA/genericpage.html?file=/edna/aboutedna/metadata/elements.html&sp=eec099e3eeeb>>
- Miller, Eric, and Renato Iannella. 1998. Dublin Core Examples in RDF. [cited 1999.5.26]. <<http://www.dstc.edu.au/RDU/RDF/dc-in-rdf-ex.html>>
- Miller, Eric. 1997. Monticello Electronic Library: Dublin Core Element Set Crosswalk. [cited 1998.7.29]. <<http://www.oclc.org:5046/~emiller/DC/crosswalk.html>>
- National Archives of Australia. 1998. The Australian Government Locator Service (AGLS) manual for users. [cited 1998.7.28]. <http://www.naa.gov.au/govserv/agls/user_manual.htm>
- Pierre M. S., and W. P. I.aPlant. 1998. Issues in crosswalking content metadata standards. [cited 1999.5.9]. <<http://www.nisc.org/crswalk.html>>
- Weibel, S., and Ricky Erway. 1997. "The news talks with metadata experts Weibel and Erway". *RLG News issue*, 44(1997. Fall). [cited 1999.2.24]. <<http://www.rlg.org/rlgnews/news44.html>>