RDF/RDFS를 이용한 연구성과물정보 메타데이터 모델링에 관한 연구

A Study on the Metadata Modeling for Research Result Information Using RDF/RDFS

박동진*

목 차

Ⅰ. 서론
Ⅱ. 연구의 배경
1. 연구성과물 범위
2. 메타데이터의 표현
3. 메타데이터 상호운용성을 위한 기반 기술
Ⅲ. 연구성과물 메타데이터 스키마
1. 연구성과물 데이터모델
2. 연구성과물 메타데이터구성
Ⅳ. RDF/RDFS 기반의 메타데이터 모델링
1. RDFS를 이용한 모델링
2. RDF를 이용한 메타데이터 기술
Ⅴ. 결론

Key Words: 연구성과물, 메타데이터, 시맨틱 웹, RDF/RDFS

Abstract

The purpose of this paper is to develop the metadata on the information of research result in Science and technology and to design the domain knowledge structure using semantic web technology for further implementation. In this paper, we first analyze the existing theories and techniques related to the metadata in such fields as R&D research result, international standard, and semantic web. Then, we extract and group the relevant factors from Dublin Core, CERIF, and the research results for building the integrated metadata framework. Based on our proposed metadata, we design a domain knowledge structure which employs RDF/RDFS as knowledge representation tool. Therefore, we can implement the ontology which produce the 'intelligent' information service and improve the interoperability between the research institutions. Also, the metadata can be used as the basis for developing National R&D Performance Information, and in terms of research institutions, can be used as tools for managing the their own research results information systematically and consistently.

* 공주대학교 산업시스템공학과 부교수

Ⅰ. 서 론

연구개발 성과정보들의 수집, 교환 및 평가를 가능하게 하는 성과관리 정보시스템의 구축에 앞서 필수적으로 선행되어야 할 것이 연구성과 정보에 대한 저장 체계의 표준화이다. 특히 연구성과물 공유 차원에서 각 연구기관별로 생산된 연구성과물을 상호연계하고 유통을 하기 위해서는 상호운용성이 확보되어야 한다.

본 연구의 목표는 과학기술 분야에 있어서 연구개발 성과물의 상호운용성 제고를 위하여, 첫째, 연구성과로 인정되는 각종 산출물에 대하여 표준 메타데이터를 개발하는 것이며, 둘째는 시맨틱 웹 환경을 고려하여 연구성과물 메타데이터 도메인을 모델링하는 것이다.

제2장 에서는 연구성과물 범위 및 메타데이터에 대하여 알아보고, 제3장에서는 연구성과물 데이터 모델을 기반으로 한 메타데이터를 구조를 제시하고, 실제 데이터 요소들을 소개한다. 제4장에서는 RDF/RDFS를 이용한 메타데이터 도메인 모델링을 제시하고 마지막으로 결론을 맺는다.

Ⅱ. 연구의 배경

1. 연구성과물의 범위

연구성과물은 R&D 전주기에 걸쳐 다양한 문서 및 매체로 발생되고 연구성과물의 범위나 대상에 있어서도 통일된 의견은 없다. 과학기술분야에 있어서 결과물은 일반적으로 저널이나 기타 문헌들을 통한 연구성과물의 발표(publication)나 실험과 관찰을 통하여 획득한 데이터세트이다(Mattews et al., 2002).

문헌을 통한 연구성과물은 학술적 검토 과정을 거치며 공식적인 채널을 통하여 입수가능 한 저널 논문(journal articles)과 생산, 유통, 자료의 이용자 및 서지 기술적 특성이 독특하여 공식적인 출판 경로를 통해서는 입수하기 어려운 회색문헌(grey literature)으로 구분된다(김미영, 2000). R&D 관점에서의 회색문헌은 출판전 배포문, 연구계획 제안서, 연구보고서, 일반적으로 컨퍼런스 발표논문, 기술보고서, 학위논문, 회의/세미너/워크샵 자료, 기술계약에 관한 보고서, 특허 등이다(Almeida, 1999:Auger, 1996). Jeffery(1999)은 R&D에서 생성된 대부분의 연구결과물은 회색문헌이다 라고 주장하였다. 국내 연구소에 대한 조사의 결과에 따르면(남영준, 2002), 회색문헌은 최신의 연구결과이므로 과학기술의 발전을 위해서 회색문헌의 학술적 가치 및 활용상 가치가 매우 크며, 각 기관은 물론 국가적 차원에서 이를 체계적으로 관리해야할 필요가 있음을 주장하였다.

2. 메타데이터의 표현

메타데이터는 데이터에 관한 구조화된 데이터로서 자원과는 독립적으로 존재하면서 자원에 대한 다양한 접근점과 네트워크 주소를 포함하는 데이터이다(Miller, 1998). 메타데이터는 다음과 같이 활용된다. 첫째, 자원의 의미를 요약해 준다. 둘째, 자원의 검색을 가능하게 한다. 셋째, 자원의 필요성에 대한 판단을 가능하게 해 준다. 넷째, 다른 자원들과의 연관성을 말해준다(Steinacker et al., 2001).

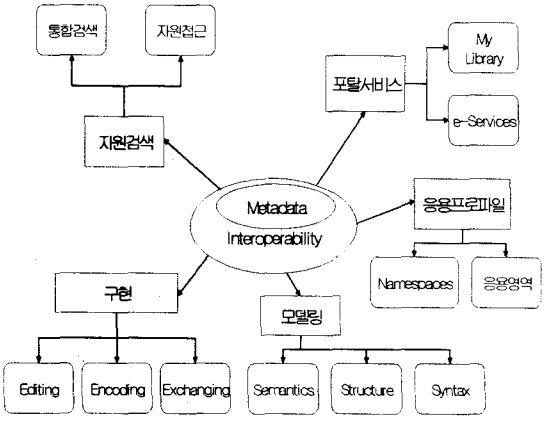
대상이 되는 자원을 메타데이터로 기술하기 위해서는 먼저 적절한 메타데이터 요소의 집합(element set)을 선택하고, 각 데이터 요소에 대하

여 어휘를 정의하고 그리고 표현 방법인 스킴(scheme)을 결정한다. 많은 경우 메타데이터 스킴은 기존의 표준이나 온톨로지(ontology)들로부터 도입한다. 대표적인 웹자원을 기술하는 표준 메타데이터가 Dublin Core 이다(Dublin Core Metadata Initiative, 2005). 특히 과학기술분야의 연구개발 성과물을 위해서는 최근에 CRIS에서 CERIF-2004 메타데이터를 제시하였다(CRIS, 2004).

3. 메타데이터 상호운용성을 위한 기반기술

이수상(2003)에 따르면 메타데이터의 상호운용성 보장을 위한 요소기술은 아래 (그림 1)과 같다.

응용프로파일(application profiles)은 응용영역의 목적에 부합하도록 메타데이터 스키마를 결정하는 것을 말한다. 메타데이터 스키마는 기존의 표준 메타데이터들로 부터 추출한 메이터 요소들과 그들의 스킴을 도입하고 추가적으로 응용영역에서 적합한 데이터 요소들을 포함시키는 것이다.



(그림 1) 메타데이터 상호운용성을 위한 기반기술

다음으로 모델링(modeling)은 응용프로파일에서 정의된 메타데이터 요소를 실제의 시스템의 데이터형식에 맞춰 상세화하는 과정이다. 즉 다양한 메타데이터 요소간의 연결을 위하여 메타데이터에 대한 의미론(semantics), 물리적구조(structure), 그리고 구문론(syntax)에 대하여 결정하고, 실제의 개체의 메타데이터가 시스템에서 어떻게 표현되어야 하는가를 제시하는 것이다. 메타데이터의 표현방식으로는 RDF 모델, RDF Schema 모델, XML DTD 모델, XML Schema 모델 등 다양한 방법이 있다

구현(implementation)은 프로그램 코딩영역으로 메타데이터 편집기를 제작하는 것이다. 마지막으로 검색시스템(resource discovery)의 구현 및 포탈서비스 개발도 메타데이터 상호운용성을 위해서 필요한 기술들이다. 본 논문의 3장은 응용프로파일의 개발과 4장은 모델링과 관계된다.

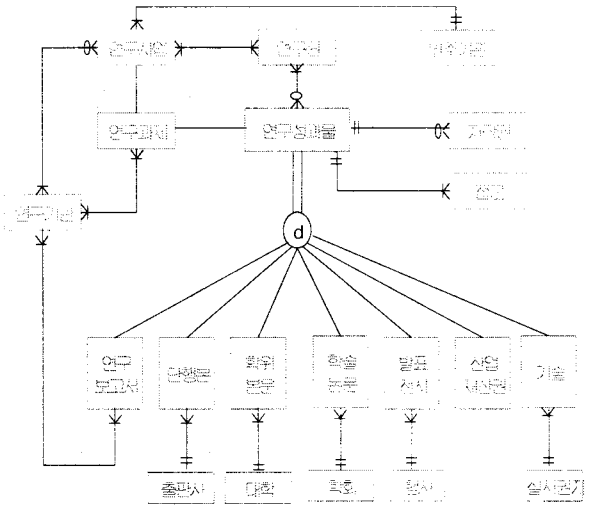
Ⅲ. 연구성과물 메타데이터 스키마

1. 연구성과물 데이터모델

데이터 모델은 메타데이터의 정적인(static) 면을 보는 것으로 메타데이터에 포함될 개체 파악하고 데이터 모델링을 통하여 개체 유형의 구분, 연관관계, 세부 데이터 요소들을 파악하는 것이다.

연구성과물에 대한 메타데이터는 크게 ① 연구성과물 자체, ② 연구성과물 생산주체, 그리고 ③ 연구성과물 유통과 관련된 요인들로 그룹화 할 수 있다.

이상의 내용을 ERD 로 표현하면 아래 (그림2)와 같다.



(그림 2) 연구성과물 데이터 모델

각각의 연구성과물을 세분하면 아래와 같다.

① 연구보고서(중간보고서/최종보고서/n차보고서)

② 학술논문(국내대학 및 기관논문지, 학술진흥재단등록, 학술진흥재단후보, 전국규모 학회지, 지역학회지, 해외 학회.)

③ 각종발표회 및 전시회(국내학술대회발표/해외학술대회발표/심포지움/포럼/세미나/강연회/회의/전시회)

④ 단행본 및 간행물

⑤ 학위논문

⑥ 산업재산권(국내특허/국제특허/국내실용신안/국제실용신안/디자인/소프트웨어등록)

⑦ 기술개발(기술등록/기술이전)

2. 연구성과물 메타데이터 구성

연구성과물 메타데이터 요소는 Dublin Core의 기본요소 및 하위요소를 근간으로 하고, 국제적으로 연구성과물 관리의 대표적인 메타모델인 CERIF에서 사용되는 요소를 수용하며, 국내 연구기관에서 연구성과물 관리에 적합한 기술요소를 선택적으로 적용하여 구성하였다.

연구성과물 메타데이터는 다음과 같은 범주로 구분되고, 아래의 〈표 1〉과 같이 총 82개의 주요소, 47개의 하위요소로 구성되어 있다.

〈표 1〉 메타데이터 구성

	주요소	하위요소
공통 메타데이터 요소		
식별기호범주	1	2
연구성과물 일반범주	14	21
개별 연구성과물 메타데이터 요소		
연구보고서범주	2	0
학술논문범주	8	0
각종발표회 및 전시회범주	7	5
간행물범주	3	0
학위논문범주	5	0
산업재산권등록범주	5	5
기술개발범주	5	4
연구주체 메타데이터 요소		
연구과제범주	11	8
연구기관범주	5	0
발주기관범주	5	0
연구원범주	11	2
총	82	47

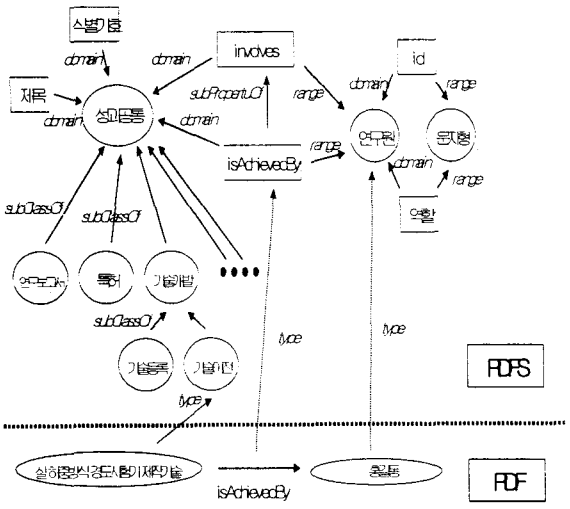
Ⅳ. RDF/RDFS 기반의 메타데이터 모델링

1. RDFS를 이용한 메타데이터 모델링

본 절에서는 시맨틱 웹 환경의 메타데이타 언어인 RDF/RDFS로 연구성과정보를 표현하고자 한다. 이는 구조화된 메타데이터를 기계가 이해할 수 있도록 기호화하고, 교환하고, 재사용하는 것을 가능하게 하는 기반구조이다. RDFS와 짝을 이루는 RDF는 필요에 따라 확장할 수 있는 모델링 근원어(primitives)를 제공하며, 문법적으로 RDF 모델은 XML로 저장하고 교환할 수 있다. 본 연구에서는 장기적 측면에서 연구성과물의 상호운용성

제고를 위하여 검색엔진, 지능형 에이전트 등의 개발과 시맨틱 웹환경을 고려하여 RDF/RDFS를 이용하기로 한다. 특히 RDFS는 용어들 간의 관계를 표현하는 기본 어휘들과 제약방식을 활용하여 도메인의 기본적인 온톨로지를 구성할 수 있게 한다.

(그림 3)은 RDF/RDFS를 이용하여 연구성과물 영역의 도메인을 모델화 한 것의 일부이다. RDFS로는 3장에서 파악된 연구성과물 클래스의 계층적 구조, 클래스와 클래스, 클래스와 요소(properties), 요소와 요소간의 관계를 표현한 것의 일부이다.



(그림 3) 연구성과물 메타데이터 도메인 모델링

예를 들면 기술개발 클래스는 성과공통 클래스의 하위클래스(subClassOf)이고 기술등록은 기술개발의 하위클래스이다. 그리고 연구성과물은 연구원에 의해서 완성된다는 관계를 보여주는 "관계형요소"인 isAchievedBy는 involves 의 하위요소(subPropertyOf)이다. 또한 각 관계형요소는 관계상의 주체(subject)가 되는 클래스와 값(value)을 갖는 클래스에 대하여 domain과 range을 이용하여 대상에 있어서 제약을 가할 수 있다. 예를 들면 연구성과물은 연구원에 의해서만 작성된 것 만을 인정한다라는 제약이 있다면 IsAchievedBy 의 주체인 성과공통이 domain이 되고, 연구원이 range가 된다. 또한 type을 이용하여 RDFS에서 정의된 객체들과 RDF의 실제 데이터간의 관계를 맺어준다.

아래 (그림 4)는 (그림 3)에서 방향성 그래프로 모델링한 도메인의 일부를 XML 타입으로 변환한 것이다. 즉 연구성과물공통요인("researchResultCommon") 클래스와 하위클래스인 기술개발("technologyDevelopment") 클래스, 그리고 연구원("researcher") 요소와 연구성과물공통요인 클래스간의 관계를 설명하는 "isAchievedBy" 요소를 정의한다. 특히 "isAchievedBy" 에는 domain과 range를 이용하여 적용되는 대상 클래스를 구체적으로 표시한다.

```
<rdf: RDF
    xmlns:rdf:
http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

    <rdfs:Class rdf:ID="researchResultCommon">
        <rdfs:comment>
            연구성과물 공통요인 class
        </rdfs:comment>
    </rdfs:Class>
    <rdfs:Class rdf:ID="technologyDevelopment">
        <rdfs:comment>
            기술개발 class
        </rdfs:comment>
        <rdfs:subClassOf
rdf:resource="#ResearchResultCommon"/>
    </rdfs:Class>
            ... 중략 ...
    <rdfs:Property rdf:ID ="researcher">
        <rdfs:comment>
            연구원 property
        </rdfs:comment>
    </rdfs:Property>
    <rdfs:Property rdf:ID ="isAchievedBy">
        <rdfs:subPropertyOf rdf:resource="#involves"/>
        <rdfs:domain
rdf:resource=#researchResultCommon"/>
        <rdfs:range rdf:resource=#researcher"/>
    </rdf:Property>
</rdf:RDF>
```

(그림 4 RDFS로 기술한 도메인 모델)

도메인 전체에 대한 명세(specification)는 본 연구의 범위를 벗어나지만, 이상과 같은 방법으로 메타데이터의 모든 어휘들과 그들 간의 관계 및 구조(scheme)를 정의한다.

2. RDF를 이용한 메타데이터 기술

다음으로 RDF 편집기를 이용하여 아래 (그림 5)와 같은 RDF 메타데이터를 생성한다. 먼저 선언부를 기술하고 모델링 스키마를 준수하며 메타데이터를 생성한다. 선언부는 RDF 문서가 XML 기반이고 이름공간(namespece)를 이용하므로 반드시 선언부에 명시하여야 한다. Dublin Core 표준인 dc 와 (그림 5)에 포함된 result 가 이름공간이다. 선언부를 표시한 후 〈rfd:Description〉과 about 속성을 이용하여 자원을 표시한다. (그림 5)에는 기술이전 성과인 식별기호가 200410-4112인 "실 하중방식 경도시험기 제작기술" 에 대한 메타데이터 기술이다.

```
<xml version="1.0" encoding="UTF-16"?>
<rdf:RDF
    xmlns:rdf:                                                    "
http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:result="http://www.mydomain.or.kr/result-ns
#>

  <rdf:Description                                    rdf:about="
http://www.mydomain.org/rd_sub/#200410-4112">
      <dc:title> 실하중식 경도시험기 제작기술 </dc:title>
      <result:              isAchievedBy>              홍길동
</result:isAchievedBy>
        .....
  </rdf:Description>
```

(그림 5) RDF 기반의 메타데이터

비록 본 논문에서는 연구성과물 메타데이터 스키마 전체와 RDFS로 표현한 도메인 구조 및 RDF로 기술한 연구성과물 메타데이터를 언급하지는 못하였지만, 연구성과물의 상호운용성을 제고를 위한 기반구조구축에 대한 방향을 구체적으로 언급하였다. 특히 RDFS/RDF에 의한 메타데이터 표현은 해당분야의 온토로지의 개발을 가능케 하고 궁극적으로는 "지능적인(intelligent)" 정보 서비스를 가능케 할 것이다.

V. 결 론

현재 모든 국내연구기관 및 연구정보서비스기관에서는 매우 제한적으로 연구성과물을 관리 및 정보서비스를 하고 있다. 즉 관리범위도 공식적인 검토를 거친 성과물에 국한될 뿐 아니라, 서로 상이한 체계로 관리하고 있어서 연구성과물의 상호연계와 정보유통을 위한 상호 운용성이 매우 떨어지고 있다. 이러한 상황에서 본 연구는 관리대상이 되는 연구성과물의 범위를 확장하고 이들을 통합한 표준 메타데이터를 처음으로 제시함으로써 국내 R&D 연구성과정보 관리의 기틀을 마련했다는 데 의의가 있다. 또한 Dublin Core와 CERIF 등과 같은 사실상의 국제 표준(de facto international standard)과의 호환성을 가짐으로 해서 국제적 교류도 가능하게 될 것이다. 또한 본 연구에서 RDF/RDFS를 이용한 도메인 모델의 개발은 연구성과물 검색과 유통에 있어서 '지능적인(intelligent)' 정보서비스를 가능하게 하는 시맨틱 웹환경의 구축에 기반이 될 것이다.

각 연구기관 관점에서는 연구성과물에 관한 정보를 일관성 있고 체계적으로 관리할 수 있는 시스템 개발의 도구로 활용되어 질 수 있다. 구체적으로 연구기관별 자체 연구성과물 정보관리를 위한 표준구조로서의 역할을 하며, 연구성과물 측정 및 평가시스템용 DB 설계 시 참조될 수 있다. 국가적으로는 현재 추진 중에 있는 R&D 종합관리체제 개발에 있어서의 기반 인프라로서 활용되어 질 수 있다.

참 고 문 헌

1. 김미영, 회색문헌의 생산현황과 관리-부산대학교내 37 연구소를 중심으로, 석사학위논문, 부산대학교 대학원, 2002
2. 남영준, 디지털 시대의 회색문헌 이용 활성화에 관한 연구, 한국정보관리학회지, 제19권, 4호. pp.234-255.
3. 이수상, 메타데이터의 상호운용성 보장을 위한 요소기술, 한국도서관 · 정보학회지, 제34권, 제1호, 2004, pp.91-109.
4. Almeida, Mario do G.G., "Control Access for Grey Literature in Brazil: A Proposal", In fourth International Conference on Grey Literature, 1999.
5. Auger, C. P., Information Sources in Grey Literature, 4th ed. London, 1996 CRIS, CERIF(Common European Research Information Format) 2004, <http://www.eurocris.org/en/taskgroups/cerif>
6. Dublin Core Metadata Intiative, <http://dublincore.org>, 2005
7. Jeffery. K., "An Architecture for Grey Literature in a R&D," In fourth International Conference on Grey Literature, 1999
8. Matthews, B., Wilson, M., and Kerstin Kleese-van Dam, Access the Outputs of Science Projects, Proceedings, 6th International Conference on CRIS, 2002
9. Miller, E., "an Introduction to the Resource Description Framework", <http://www.dlib.org/dlib/may98/miller/05miller.html>
10. Steinacker A., A. Ghavam and R. Steinmetzm Metadata Standards for Web-Based Resource, IEEE Multimedia, Vol. 8, Issue 1, pp. 70-76