

기초학문자료 메타데이터 설계 분석 및 온톨로지 적용 방안 연구

A Study on Design and Analysis of Metadata and Ontology based on Humanities and Social Sciences

이 정 연(Jung-Yeoun Lee)*

김 정 민(Jung-Min Kim)**

최 석 두(Suk-Doo Choi)***

김 이 겸(Lee-Kyum Kim)****

목 차

- | | |
|-----------------|------------------------|
| 1. 서 론 | 4. 시소러스의 확장과 메타데이터의 연계 |
| 2. 메타데이터 형식의 설계 | 5. 온톨로지 기반 정보검색 |
| 3. 메타데이터의 구축 | 6. 결 론 |

초 록

기초학문자료의 특성인 복잡한 관계의 개념구조, 자료유형 및 자료간의 의미적 상관관계 등을 표현할 수 있는 기초학문자료 메타데이터 모형을 설계하였다. 설계된 메타데이터 모형의 정당성 및 효율성을 평가하기 위해 실제 구축된 자료의 분석을 토대로 실제적인 메타데이터 구성요소를 제시하였다. 메타데이터의 키워드 기반검색의 한계를 극복할 수 있는 확장된 시소러스를 설계하였으며 도메인 온톨로지를 구축하는 방안을 모색해 보았다. 또한 철학 종교학 분야의 주제분류 중심의 확장시소러스를 설계하고 토픽맵 기반 시스템으로 구현하여 주제 중심의 메타데이터 검색이 가능함을 보였다.

ABSTRACT

The purpose of this study is to design metadata model for describing different kinds of concepts, properties, and semantic relationships of result materials of researches. We examine our metadata model to evaluate correctness and efficiency of the model through contents analysis of a constructed database. From the results of examination, we suggest more effective structure of metadata schema. Domain ontology could be constructed by the enlarged thesaurus in order to overcome the limitation of the keyword search, therefore we design a philosophy and religion ontology based on subject classification to improve information retrieval and implement it using XML/Topic Maps to improve retrieval functionality of our database.

키워드: 기초학문자료, 메타데이터 설계, 메타데이터 분석, 시소러스, 주제분류 온톨로지, XML/Topic Maps
Metadata Design, Thesaurus, Ontology, Information Retrieval, XML/Topic Maps

* 한국학술진흥재단 지식정보센터(shampoo@krf.or.kr)

** 서울대학교 언론정보학과 정보문화학 연구교수(jmkim@idb.snu.ac.kr)

*** 한성대학교 지식정보학부 교수(sdchoi@hansung.ac.kr)

**** 광주대학교 문헌정보학과 교수(leekyum@kwangju.ac.kr)

논문접수일자 2007년 5월 18일

게재확정일자 2007년 6월 20일

1. 서론

오늘날 국제 경쟁 시대에서 국가의 안정적 성장과 번영을 위해서는 인문사회과학 분야를 비롯한 기초학문 분야에 대한 지속적인 지원이 요구된다. 국가발전은 단순히 경제적 성장만을 의미하지 않으며, 사회 구성원의 정신적 풍요로움을 함께 고려하여 지속가능한 발전을 도모해야 하는 것이기 때문이다. 여기에서 기초학문은 경제성장의 목표와 방향을 제시하며, 사회 성원의 일체감 고양 및 삶의 질 향상에 기여하고, 사회 성원들이 세계화 시대에 요구되는 의식과 태도를 갖추는 데 이바지한다.

이러한 필요성에 따라 한국학술진흥재단(이하 학진이라 칭함)에서는 2002년부터 기초학문육성사업 등을 통해 매년 수천 억 원의 예산을 투입하여 인문사회과학 분야를 비롯한 기초학문 분야의 학술연구를 지원하고 있다. 그 결과 매년 다양한 기초학문 분야에서 서로 다른 내용과 형태의 연구결과물 및 중간산출물이 생산되었다. 여기에는 완성된 논문 및 저술뿐만 아니라, 사회과학의 각종 통계자료를 비롯하여 인문학 분야의 연구, 조사과정에서 수집된 고문서자료, 구술채록자료, 사진자료, 탁본자료, 유물자료, 민속자료 등이 포함되어 있다. 이와 같은 자료들을 기초학문 분야 및 관련 분야의 발전과 후속 연구를 위한 '기초학문자료'라고 할 수 있다.

이에 따라 2002년도 이후 시행된 대규모 연구지원의 결과로 생산된 자료들을 체계화하여 효율적으로 관리할 필요성이 증대하였다. 그리고 앞으로 더욱 증가할 자료의 양을 감안할 때, 자료의 효율적 관리를 위해서 종래의 간단한

설명과 파일만 등록하는 방식으로 수집 및 정리하는 것으로는 충분하지 않고, 온라인 매체의 특징을 최대한 활용하여 이미 획득한 기초학문의 자료를 지식자원으로 구축하고, 이를 필요로 하는 연구자들에게 보급함으로써 학문 연구의 지원효과를 극대화할 필요가 있다.

이와 같은 근거에 의해 학진의 기초학문육성 지원사업에 의해 지원된 기초학문 분야에서 산출된 "기초학문자료"들을 체계적으로 관리하고, 효과적으로 활용하기 위한 기구로서 "기초학문자료센터"의 설치를 요구하게 되었고(백종현 2005), 이에 학진의 지원으로 2005년 12월부터 2006년 6월까지 약 7개월간 인문사회과학분야 기초학문자료센터 설립을 위한 시범사업이 수행되었다. 사업의 구체적인 목적은 인문사회과학분야의 기초학문자료를 수집하여 통합데이터베이스를 구축하고, 동 사업에서 획득한 경험과 지식을 바탕으로 기초학문자료센터의 설립 및 운영안을 마련하는 데 있었다(한국학술진흥재단 2006).

기초학문자료의 특징은 인문, 예술, 사회 등의 학문 분야의 다양성에서 비롯되는 복잡한 관계의 개념구조와 보고서, 단행본, 고문서, 녹음자료, 사진자료 등의 복합적인 자료유형 및 과제와 과제, 자료와 자료 사이의 의미적 상관관계 등으로 구체화 된다. 따라서 시범사업 기간에는 복잡하고 다양한 자료의 체계적인 수집과 저장 및 메타데이터 기술에 초점을 맞추고 표준화 관점에서 연구를 수행하였다.

본 연구에서는 인문사회과학분야 기초학문자료센터 설립을 위한 시범사업에서 저자들이 수행한 기초학문자료 통합 메타데이터 설계의 원칙과 내용을 소개하고, 이에 따라 구축된 메

타데이터를 분석하였다. 나아가 메타데이터 간의 개념관계를 시소러스로 정의하고 시범적으로 철학·종교학 주제분류에 기반한 도메인 온톨로지를 설계함으로써 구축된 메타데이터를 효율적으로 이용하기 위한 방안을 제시하였다.

발음: 곱따

활용형 발음: 곱는 [곱는]

의미: 이익을 보려다가 도리어 손해를 입게 되다

예문: 그는 큰돈을 벌겠다고 주식에 손을 대었다가 도리어 곱고 말았다.

2. 메타데이터 형식의 설계

2.1 메타데이터의 정의

메타데이터에 관한 정의는 매우 다양하며 공동체나 맥락에 따라 이 용어의 의미가 다르게 이해되고 있다. 많이 인용되는 것으로 “데이터에 관한 데이터, 즉 다른 데이터를 설명(기술)하기 위하여 사용되는 데이터”이며, 그 조건은 “정보는 반드시 구조화 되어야 하며, 메타데이터는 반드시 어떤 정보자원을 기술해야 한다”(Caplan 2004, 19-24)는 정의이며 그 전형적인 예로 단행본의 목록데이터를 든다. 보다 구체적인 정의로는 “입력시스템, 응용프로그램, 데이터베이스 혹은 간단한 처리엔진에 입력된 데이터의 상세한 설명; 입력된 데이터가 가지고 있는 형식과 특성; 메타데이터 수용기의 역할에 맞추어 실제로 입력된 데이터”(Tannenbaum 2002, 91-93)라는 정의가 있다. 그러나 이들 정의는 모든 경우에 적합한 것은 아니다.

예컨대, 국어사전(국립국어연구원 2000)의 한 항목인 “곱다”의 설명의 일부에 대하여 메타데이터 요소명을 붙여 보면 다음과 같다.

곱다

품사: 동사

이 예에서 “곱다”에 대한 데이터는 이것이 전부이며 어디까지가 데이터이고 어디까지가 메타데이터인지 구별할 수 없다. 이를 데이터라고 할 수도 있으며 메타데이터라고 할 수도 있다. 따라서 상기 “곱다”와 같은 데이터를 메타데이터로 취급하기 위해서는 메타데이터의 정의를 보다 넓히고 일반화 시킬 필요가 있다.

가령, 어떤 액체가 든 플라스틱 병이 있다고 하자. 병 속에 든 액체가 무엇인지 모르는 상황에서 이것은 데이터이다. 그런데 그 병에 “식용수, ○○회사, 보존기간” 등의 정보가 포함된 레이블을 붙이는 순간 메타데이터가 된다. 즉, 레이블을 붙인 그 플라스틱 병 자체가 메타데이터가 된다. 특히 이 경우에는 레이블과 병을 분리할 수가 없으며 분리되어 있는 경우에는 그 존재이유가 없어진다. 초록이 대상인 경우에는 초록이 메타데이터에 포함될 수 있을 것이며, HTML 태그가 포함된 문헌이 바로 메타데이터가 될 것이다. 이 외에도 메타데이터는 수치데이터, 텍스트데이터, 문서, 보고서, 프로그램 코드, 데이터베이스 스키마, XML 태그, HTML, 시스템 매뉴얼, 데이터 사전, 데이터모형 등에서 출현하며 다양한 경우가 발생한다. 즉, 대상 데이터가 포함된 데이터도 메타데이터라고 보아야 한다. 그리고 실제의 환경에서는 이러한 경향이 점점 늘어나게 될 것이다.

본 연구에서는 대상 원문이 포함되는 메타데이터 운용시스템을 목표로 메타데이터서비스 시스템을 설계한다. 따라서 원칙적으로 어떤 형태의 자료라 할지라도 메타데이터 구축의 대상이 되는 것은 모두 디지털화한다는 것을 전제하며, 파일형식을 불문하고 대상 메타데이터에 디지털화된 원파일을 첨부하고 있다.

2.2 설계 원칙

데이터베이스화 할 연구지원 연구성과물은 연구 초기단계에서부터 종료단계까지 발생한 모든 원자료를 포함하는 것을 그 대상으로 하였다. 또한 데이터베이스 구축의 기반이 되는 메타데이터 요소 및 형식의 표준안 개발을 위하여 다양한 분야와 정보매체를 통합적으로 처리할 수 있는 메타데이터 스킴을 개발하였고 개발된 메타데이터를 적용한 입력시스템에 메타데이터를 입력하고 원문파일을 업로드하며 검증할 수 있도록 구축하였다.

기초학문자료센터 시범사업단 메타데이터 요소, 형식, 기술규칙을 포함하는 메타데이터 스킴의 개발원칙은 다음과 같다. 첫째, 기존 연구결과를 도출하게 된 모든 유형의 토대자료를 다른 연구자가 새로운 연구의 원자료로 사용할 수 있도록 한다는 본 메타데이터의 독특한 사용목적을 최대한 반영한다. 특히, 동분야의 연구자가 요구하는 정보는 가능한 한 반영하는 것을 원칙으로 한다. 둘째, 메타데이터 형식설계에는 공통요소의 경우 국제적으로 공통요소의 집합으로 인정되는 더블링크어를 기본적으로 포함시키되, DCMI에서 기타 요소로 확장한 표준요소와 표준 검토요소를 반영하였다. 다만,

본 형식의 목적상 하위요소인 경우에는 가능한 더블링크어의 하위요소를 그대로 수용하였으나 필요에 따라 이를 더욱 세분화하였다. 셋째, 자료 유형별 요소의 경우에는 국내에서 현재 사용하고 있는 형식들을 검토하여 이들을 반영하였고, 이들 중에 더블링크어 요소로 포함해도 그 호환성에 문제가 없는 것은 하위요소로 포함하였으며 나머지는 새로운 요소들로 반영하였다. 넷째, 본 메타데이터는 기초학문자료센터의 자료들을 처리하기 위한 것이기 때문에 그 자료들의 소유 특성(소유기관, 소유기관 자료번호)이나 다른 메타데이터와의 연계적인 특성(학진의 연구과제 번호) 및 관리적인 측면(입력일, 입력자, 수정일, 수정자)을 고려하여 특수한 요소들을 추가하였다.

2.3 기술요소

앞서 기술한 메타데이터의 원칙에 근거하여 26개의 기술요소를 기본으로 하고 하위요소를 개발하였다(표 1 참조). 표제, 저작자, 발행자, 기타 저작자, 내용유형, 내용범위, 날짜, 설명, 형태, 자료링크, 언어, 이용조건, 정보원, 관련자료, 주제, 이용대상, 표준번호, 이미지 정보, 녹음정보, 동영상 정보, 조사자료 정보, 주식, 과제번호, 프로그램 정보, 관리정보, 소유기관 정보 등이 기술요소의 기본 요소이다.

이들을 보고서, 단행본, 학술지, 보고서, 조사자료, 사진 및 이미지, 동영상, 녹음자료, 웹사이트, 낱장자료 등 9개의 자료유형별로 그룹화하고 기술요소를 필수, 해당 시 필수, 선택 항목으로 구별하여 개발하였다. 특히 표시상수를 미리 정하여 데이터의 출력 시 이용자의 편의

<표 1> 기초학문자료센터 시범사업단 메타데이터 기술요소

번호	기술요소명	표시상수	하위요소	표시상수	하위요소	표시상수		
1	Title	표제	properTitle	본표제	하위요소 없음			
			parallelTitle	대등표제				
			originalTitle	원표제				
			otherTitle	기타 표제				
2	Creator	저작자	personal	개인	role	역할어		
					name	인명		
					identifier	전거ID		
		corporate	단체	role	역할어			
				name	단체명			
identifier	전거ID							
3	Publisher	발행자	하위요소 없음					
4	Contributor	기타저작자	personal	개인	role	역할어		
					name	인명		
					identifier	전거ID		
		corporate	단체	role	역할어			
				name	단체명			
identifier	전거ID							
5	Type	내용유형	하위요소 없음					
6	Coverage	내용범위	spatial	지리/공간	하위요소 없음			
			temporal	시기			period	기간
							periodName	시대명
7	Date	날짜	created	저작일	하위요소 없음			
			modified	수정일				
			issued	발행일				
			available	이용기간				
			valid	유효기간				
8	Description	설명	generalNote	일반주기	하위요소 없음			
			tableOfContents	목차			caption	목차제목
			abstract	초록			contents	목차
9	Format	형태	medium	유형	하위요소 없음			
			extent	크기				
			color	색채유무				
10	Identifier	자료링크	bibliographicCitation	서지인용	하위요소 없음			
11	Language	언어	하위요소 없음					
12	Rights	이용조건	license	저작권				
			accessRights	자료공개구분				
			accessRightsNote	공개제한 사유				
13	Source	정보원	하위요소 없음					
14	Relation	관련자료	isPartOf	종합자료	하위요소 없음			
			hasPart	부분자료				

			isFormatOf	대본자료		
			hasFormat	다른형태 자료		
			isVersionOf	후속판		
			hasVersion	선행판		
			replaces	대체자료		
			isReplacedBy	대체된 자료		
			requires	필요한 자료		
			isRequiredBy	요구되는 자료		
			isReferencedBy	참고된 자료		
			references	참고자료		
			conformsTo	적용표준자료		
15	Subject	주제	classification	분류기호		
			keyword	주제 용어		
16	Audience(유보)	이용대상	mediator	중개이용자		
			educationLevel	교육수준		
17	NumberInfo	표준번호	하위요소 없음			
18	ImageInfo	이미지정보	location	촬영장소		
			object	촬영대상		
			background	촬영배경정보		
19	SoundRecordingInfo	녹음정보	location	녹음장소		
			soundType	녹음내용유형		
20	MovingImageInfo	동영상정보	location	촬영장소		
21	SurveyInfo	조사자료정보	object	조사대상	하위요소 없음	
			period	조사기간		
			surveyMethod	조사방법		
			sampleSize	표본크기		
			samplingMethod	표본추출방법		
			surveyArea	조사지역		
			unitOfAnalysis	분석단위		
			question	설문항목	firstVariable	사회인구학적변수
					secondVariable	주요변수
22	Annotation(유보)	주석	person	비평자		
			date	비평일		
			description	비평		
23	LinkingEntry	과제번호	하위요소 없음			
24	SoftwareInfo	프로그램정보	하위요소 없음			
25	MetaMetadata	관리정보	contributeEntity	입력/수정자		
			contributeRole	역할어		
			contributeDate	일시		
26	LocalInfo	소유기관정보	owner	소유기관명		
			localNumber	소유기관 자료 번호		

를 고려하였으며 필요시 언제든지 이를 변경할 수 있다.

3. 메타데이터의 구축

3.1 메타데이터 구축현황

2002-2004년에 걸쳐 학진에서 지원한 인문 사회과학 기초학문육성 지원 611건 과제에 관한 중간산출물 및 연구성과물에 대하여 구축한 메타데이터는 2006년 7월 25일 현재 34,022건이다. 철학·종교학, 역사학, 어문학, 사회과학 등 네 개의 주제분야 구축팀별로 메타데이터의 유형별 구축현황은 다음의 <표 2>와 같다.

자료유형별로 보면 낱장문서가 16,568건(49.25%)으로 가장 많고 사진 및 이미지자료가 6,146건(18.06%), 학술지논문 5,007건(14.72%), 녹음자료 2,219건(6.52%), 동영상 1,404건(4.13%), 보고서 1,246건(3.66)%, 단행본 1,043건(3.07%), 조사자료 150건(0.44%), 웹사이트 50건(0.15%)으로 나타났다. 이 중 텍스트 형태(낱장자료, 학술지, 단행본, 보고서)(70%)와 멀티미디어 형태(사진 및 이미지, 녹음자료, 동영상)(29%)가

대부분을 차지하고 있었으며, 조사자료와 웹사이트는 극 소수였다. 이것은 인문사회과학의 성과물이 보편적으로 텍스트형태일 것이라는 가정을 입증한 반면, 멀티미디어자료도 상당부분 차지하는 것을 볼 수 있다.

각 분야별 자료유형을 보면, 철학·종교학 분야는 학술지논문, 보고서, 사진 및 이미지, 단행본, 동영상, 낱장자료, 녹음자료, 조사자료, 웹사이트 순이었으며, 텍스트 형태가 주류를 이루고 있다. 역사학 분야는 고문서를 중심으로 하는 낱장자료가 월등히 많으며, 그 다음은 학술지논문, 사진 및 이미지, 단행본, 보고서, 동영상, 녹음자료, 웹사이트 순이었다. 어문학 분야는 사진 및 이미지자료가 월등히 많으며, 녹음자료, 낱장자료, 학술지논문, 동영상, 보고서, 단행본, 웹사이트, 조사자료로 이어지고 있다. 어문학 분야에 사진 및 이미지자료가 대부분인 것은 근대의 신문기사나 고문헌들을 이미지로 디지털화한 경우가 대부분이었기 때문이다. 또한 사회과학 분야는 학술지논문이 월등히 많으며, 낱장자료, 동영상, 사진 및 이미지, 녹음자료, 보고서, 단행본, 조사자료, 웹사이트로 이어지고 있다. 사회과학은 학술지 논문을 위주로 자료가 구성되어 있는데 보고서 역시 학술논문

<표 2> 구축팀별 자료유형별 메타데이터 구축 건수

구축 분과팀	대상 과제수	텍스트 자료				멀티미디어 자료			조사 자료	웹사 이트	합계
		보고서	학술지	단행본	낱장 문서	사진 이미지	동영상	녹음			
철학·종교학	101	524	550	388	252	487	273	168	8	4	2,654
역사학	86	139	1,359	157	14,412	898	40	15	0	11	17,031
어문학	178	145	955	106	986	3,925	206	1,315	4	23	7,665
사회과학	246	438	2,143	392	1,107	836	885	721	138	12	6,672
합계	611	1,246	5,007	1,043	16,757	6,146	1,404	2,219	150	50	34,022

의 모음집인 경우가 다수 있었으며, 낱장자료가 많은 것은 질적 연구를 위한 고문서류 등이 많았기 때문이다.

3.2 메타데이터 관계유형 분석

본 시범사업의 메타데이터 특징은 기술대상 메타데이터와 관련된 자료에 대하여 식별정보를 기술하고 이를 통하여 관련자료들을 검색할

수 있도록 설계하였다. 관련자료 요소의 12개 하위요소명과 표시상수는 <표 3>과 같다.

관련자료 맺기는 한 과제의 성과물에는 여러 종류의 자료들이 있고 이들 간에는 특별한 관계가 있다는 것을 보여주기 위한 기능이다. 관련자료를 기술하는 방법은 전체부분관계, 형태관계, 판관계, 대체관계, 필요관계, 참고자료 관계를 수용할 수 있도록 하였다. 총 34,022건의 메타데이터 중에서 서로 관계를 갖고 있는 메

<표 3> 관련자료 하위요소 및 기술용법과 예제

관련자료 하위요소		기술방법	예
관련자료	전체부분 관계 종합자료 (isPartof) 부분자료 (hasPart)	- 종합자료는 해당 메타데이터의 상위개념의 자료를 표시한다. - 부분자료는 해당 메타데이터의 하위개념의 자료를 표시한다.	예) 종합자료, 부분자료 예) 가계 안 모습1 사진자료를 기술한 후에 종합자료를 추가할 경우 종합자료: 하동군의 생활모습에 관한 연구
	형태관계 대본자료 (isFormOf) 다른 형태자료 (hasFormOf)	- 대본자료는 대본이 된 자를 말한다. - 다른 형태자료는 대본이 된 자료를 판단하기 어려운 경우에 기술한다.	예) 대본자료/다른 형태자료 『칸트: 실천이성비판』의 학술지 논문 hwp 자료를 pdf로 제작한 경우 대본자료: hwp 다른 형태자료: mp3
	판 관계 선행판 자료 (isVersionOf) 후속판 자료 (hasVersionOf)	- 선행판자료 isVersionOf는 기술대상 자료가 관련 자료의 후속판일때 선행판에 적용 - 후속판자료 hasVersionOf는 기술대상 자료가 관련 자료의 선행판일때 후속판에 적용	예) 선행판 자료/후속판 자료 『1920년대 중국 여성소설의 섹슈얼리티 II』의 메타데이터를 작성하고 『1920년대 중국 여성소설의 섹슈얼리티 I』을 관련 자료에서 선행판에 기술한다.
	대체관계 대체한 자료 (replaces) 대체 자료 (isReplacedBy)	- 대체관계는 기술대상 자료가 관련 자료에 의해 대체되었거나 관련 자료를 대체하였을때 적용하는 요소이다. - 대체된 자료는 이와 상반된 경우에 적용한다.	예) 대체한 자료/대체된 자료 기술대상 자료: 최종보고서 대체한 자료: 중간보고서
	필요관계 필요한 자료 (requires) 요구되는 자료 (isRequiredBy)	- 필요관계는 물리적으로나 논리적으로 관련 자료를 필요로 하거나 관련 자료에 의해 요구될 때 적용하는 요소이다	예) 필요한 자료/요구되는 자료 『칸트: 실천이성비판』의 단행본 중에서 딸림자료로 녹음자료가 있을 경우에 녹음자료를 필요한 자료로 선택하여 적용한다.
	참고자료 관계 참고된 자료 (isReferencedBy) 참고자료 (references)	- 참고자료 참고된 자료는 자료가 인용관계에 있거나 내용상 관련이 있어서 참고자료를 제시할 필요가 있을 때 적용한다.	예) 참고자료 하위요소들의 참고문헌을 제시한다.

타데이터 수는 12,438건(36.5%)이었으며 관계를 갖는 메타데이터의 평균관계 수는 3.4개였다. 유형으로 보면 참고자료/참고된 자료 간의 관계 수와 부분자료/종합자료 간의 관계가 가장 많았다(그림 1 참조). 이것은 메타데이터 간의 복합적인 관계를 나타내고 있으므로 다른 연구의 토대자료로서 매우 유용하게 제공될 수 있다는 것을 상정할 수 있다.

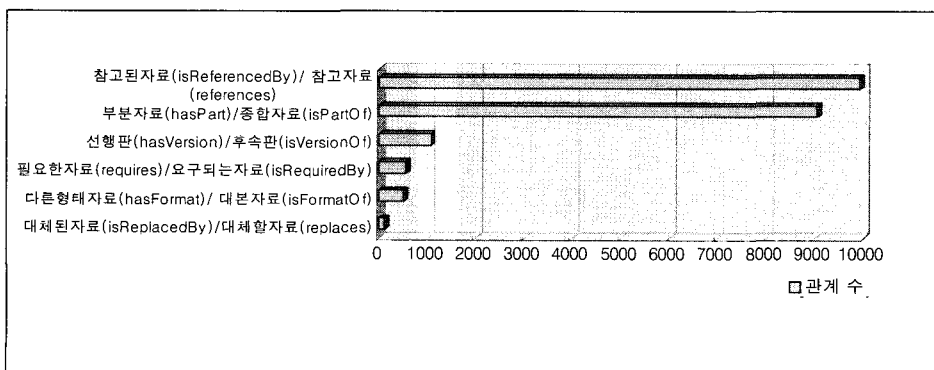
3.3 메타데이터 내용분석

구축된 메타데이터의 자료유형별로 내용분석을 할 필요가 있다. 이를 통해서 다음과 같은 측면에서 피드백을 얻을 수 있을 것이기 때문이다. 첫째, 기술요소를 유형별로 필수, 해당 시 필수, 선택의 세 가지로 설계하였는데 구축된 메타데이터를 분석함으로써 이들의 적합성 여부와 해당 범주를 조정할 수 있다. 둘째, 본 메타데이터에서 기술하기 어려운 부분을 설명(Description) 하위요소인 일반주기(generalNote)에 기록하도록 하였다. 이를 분석함으로써 자료유형별 기술요소의 추가, 확장 및 삭제 여부를 가늠할 수 있다.

3.3.1 학술지

학술지의 일반주기를 분석한 결과 다음과 같은 특징을 발견하였다. 첫째, 학술지 논문의 종류에는 학회지 논문 뿐 아니라 프로시딩 발표문, 미출판 논문, 강연문, 세미나 발표문, 학술발표회 논문, 심포지움 발표회 논문 등 다양한 유형을 보이고 있다. 특히 각종 발표회 논문인 경우에는 발표일시, 장소, 후원 및 협찬 기관 등의 기술요소 또한 필요하였다. 둘째, 논문의 일부가 단행본의 한 부분으로 간행되기도 하는데 이것은 같은 내용에 관해 표현형식을 달리 한 것이므로 관련 자료의 대체논문으로 연결될 수 있게 하였다. 셋째, 논문의 미확보, 저작권이 출판사 혹은 학회에 있는 경우에는 비공개 설정을 하는 등 디지털 원문에 대하여 해결되지 않은 부분이 산적해 있었다. 넷째, 논문 형식에 기준이 없었기 때문에 저자가 주제어를 기술하지 않은 경우도 다수 있었다. 이런 경우에는 메타데이터 구축자가 본문의 내용이나 서문 등을 분석하여 주제어를 선정하였다.

학술지의 필수항목 중에서 과제번호, 논문명, 발행일, 주제어, 자료공개구분은 100%, 그리고



<그림 1> 메타데이터 간의 관계유형

개인저자의 역할 및 개인저자명 또한 90% 이상이 입력되었다. 그러나 연구책임기관과 그 역할명은 각각 19%, 22%를 보였는데 이것은 학술지의 경우에는 단체명으로 학술지를 게재하는 경우보다는 개인명으로 게재하는 경우가 많다는 것을 보여주고 있다.

해당 시 필수 항목 및 선택 항목 중에서 50% 이상 입력된 기술요소로는 정보원(학회지명, 권호, 권차, 페이지 수)84%, 목차 75%, 국문초록 68%, 외국어 논문명 67%, 발행처 66%, 외

국어 초록 56%, 물리적 크기 51%로 나타났다 (표 4 참조). 50% 이상 기술한 요소는 향후 이들 요소를 필수항목으로 변경하는 것이 바람직할 것이다.

3.3.2 단행본

단행본의 필수 요소인 과제번호, 표제명, 발행일, 초록(국문), 주제용어는 100% 입력되었고 개인저자와 단체저자의 비율은 각각 59%, 31%로 나타났다. 단체저자는 개인저자 없이 단

〈표 4〉 학술지 하위기술 요소별 평균 메타데이터 입력비율

하위 기술요소명	표시상수	필수여부	평균 메타데이터 입력 비율
/LinkingEntry	과제번호	필수	100%
/Title/properTitle	논문명	필수	100%
/Date/created	발행일	필수	100%
/Format/color	색채유무	필수	100%
/Creator/personal/role	제1저자 역할	필수	99%
/Creator/personal/name	제1저자명	필수	91%
/Creator/corporate/role	연구기관 역할	필수	22%
/Creator/corporate/name	연구책임기관	필수	19%
/Subject/keyword	주제어	필수	100%
/Rights/accessRights	자료공개구분	필수	100%
/Contributor/personal/role	공동저자역할	해당시필수	17%
/Contributor/personal/name	공동저자명	해당시필수	16%
/Relation	관련자료	해당시필수	0%
/Contributor/corporate/role	기타연구기관	해당시필수	11%
/Contributor/corporate/name	기타연구기관명	해당시필수	8%
/Date/issued	디지털제작일	해당시필수	6%
/Description/tableOfContents/contents	목차	해당시필수	75%
/Source	정보원	해당시필수	84%
/Description/generalNote	주기(비고)	해당시필수	19%
/Publisher	발행처	선택	66%
/Title/parallelTitle	외국어논문명	선택	67%
/Rights/accessRightsNote	자료공개조건	선택	3%
/Subject/classification	주제분류	선택	21%
/Description/abstract/korean	초록(국문)	선택	68%
/Description/abstract/foreign	초록(외국어)	선택	56%
/Format/extent	크기(물리적)	선택	51%
/NumberInfo	ISSN	선택	30%

체가 저자로 되어 있는 경우에 기술하도록 하였는데 다른 유형의 자료에 비해 단행본은 감수, 번역, 저술, 편집 등의 역할어를 가진 단체저자의 비율이 높음을 보여주고 있다. 선택요소 중에서 50% 이상의 입력률을 보인 요소는 목차 73%, 발행처 77%, 그리고 물리적 크기 58%이다. 따라서 목차, 발행처, 물리적 크기는 단행본의 필수요소로 고려해 볼 수 있을 것이다.

특히 출판된 단행본인 경우 서문과 목차 정도만 공개 가능하였고, 저작권 문제로 인하여 원문파일을 첨부하지 못한 경우가 많았다. 그러므로 연구자와 과제 계약 전에 연구성과물의 결과인 단행본의 디지털 저작권에 대한 협의를 충분히 가져야 할 것이다. 초록부분과 주제어 부분 역시 저자가 직접 제공하지 않은 경우는 메타데이터 구축자가 직접 발췌하여 작성하였다. 이 역시 의무적으로 저자가 단행본 내용에 대한 주제어를 제시할 수 있는 시스템을 설계해야 함을 보여주고 있다.

3.3.3 보고서

보고서인 경우, 과제번호, 보고서명, 자료공개구분은 100% 입력되었으며, 저작 책임여부에 관한 항목인 연구책임자(개인)가 96%가 입력되었고, 단체가 책임을 맡을 경우인 연구수행기관에 32%가 기술되었다. 그 외에 제출일, 주제어, 국문초록 등은 98% 이상 입력되었다. 따라서 이들의 필수요소는 그대로 유지하면서 선택 항목이었던 목차(66%), 물리적 크기(51%)는 필수요소로 재조정할 필요가 있을 것이다. 공동연구원이 있는 경우가 47%나 있었는데 이는 선택 항목에서 해당 시 필수 항목으로 조정하여 공동연구원의 역할과 함께 이를 기술하도

록 해야 한다.

보고서 형태이지만 내용상으로 학술지에 발표되지 않은 발표논문의 모음집인 경우도 많았으며, 워크샵자료의 모음집인 경우도 있었다. 이 경우에 구축자는 학술지로 취급하여 개별적인 메타데이터를 구축하고 별도로 전체 보고서를 대상으로 메타데이터를 작성하여 이들 간의 관계를 종합자료와 부분자료 형식으로 관련을 맺고 있었다. 따라서 보고서는 내용 및 형태 등 보고서의 성격을 규명할 수 있는 특징을 명확하게 명시해 주어야 하며, 연구종료 시 저작책임자가 보고서 제출과 함께 보고서 내용에 대한 기본적인 메타데이터를 구축할 수 있는 시스템을 마련해야 할 것이다. 또한 본 기초학문자료는 학진의 과제와 관련된 정보구축이었으므로 과제 수행자의 식별이나, 연구책임자의 소속 부분은 전거처리를 통해 일관성 있는 메타데이터 구축을 기해야 할 것이다.

3.3.4 동영상

동영상의 필수요소인 과제번호, 자료명, 주제어, 촬영 및 제작일, 촬영장소는 100% 기술되었고 자료공개구분 역시 99%가 기술되었다. 그리고 언어는 75%, 국문초록은 74%의 입력률을 보이고 있었다. 저작 책임의 개인명, 기타 개인명, 제작기관명은 선택항목이었음에도 불구하고 60%이상의 입력률을 보이고 있는 것은 동영상의 특성을 나타내 주는 표시항목과 더불어 제작, 촬영, 편집, 구술, 지휘, 연주 등의 역할구분을 명확히 하기 위한 것이라 사료된다. 따라서 개인명, 기타개인명, 제작기관명, 국문초록 등을 필수요소 항목으로 조정할 필요가 있을 것이다.

동영상의 일반주기분석의 결과 다음과 같은 추가 및 조정이 필요하다고 생각된다. 첫째, 기술소로 추가되어야 할 부분은 동영상의 음성부분이다. 유성인 경우와 무성인 경우에 이를 명시할 수 있는 항목이 요구된다. 둘째, 코미디, 드라마, 다큐멘터리 등 동영상 장르를 구분하여 기술할 수 있는 항목이 요구된다. 셋째, 촬영 목적 등을 명시하여 동영상의 원자료가 과제와 관련된 촬영목적을 제시해 주는 것이 필요하다.

넷째, 원작을 각색하거나 작품의 형태로 만들었을 경우에는 원작을 기술할 수 있는 요소도 선택항목으로 제시해 주는 것이 바람직할 것이다(표 5 참조).

3.3.5 사진 및 이미지

사진 및 이미지 자료의 필수요소인 과제번호, 색채유무, 자료명, 자료공개구분, 촬영대상, 촬영장소는 100% 기술되었고 주제용어와 촬

〈표 5〉 동영상 하위기술 요소별 평균 메타데이터 입력비율

하위 기술요소명	표시상수	필수여부	평균 메타데이터 입력 비율
/LinkingEntry	과제번호	필수	100%
/Title/properTitle	자료명	필수	100%
/Language	언어	필수	75%
/Rights/accessRights	자료공개구분	필수	99%
/Format/color	색채유무	필수	100%
/Subject/keyword	주제용어	필수	100%
/Date/created	촬영/제작일	필수	100%
/MovingImageInfo	촬영장소	필수	100%
/Date/issued	디지털제작일	해당시필수	18%
/Description/tableOfContents/contents	목차	해당시필수	30%
/Rights/license	저작권정보	해당시필수	18%
/Source	정보원	해당시필수	16%
/Description/generalNote	주기(비고)	해당시필수	32%
/Creator/personal/role	개인	선택	63%
/Creator/personal/name	개인명	선택	60%
/Contributor/personal/role	기타개인 역할	선택	66%
/Contributor/personal/name	기타개인명	선택	66%
/Contributor/corporate/role	기타기관 역할	선택	27%
/Contributor/corporate/name	기타기관명	선택	37%
/Publisher	발행처	선택	27%
/Title/parallelTitle	외국어자료명	선택	39%
/Rights/accessRightsNote	자료공개조건	선택	18%
/Creator/corporate/role	제작기관 역할	선택	50%
/Creator/corporate/name	제작기관명	선택	77%
/Subject/classification	주제분류	선택	16%
/Description/abstract/korean	초록(국문)	선택	74%
/Description/abstract/foreign	초록(외국어)	선택	21%
/Format/extent	크기(물리적)	선택	21%

영 제작일도 각각 99%, 98%가 입력되었다. 그러나 언어부분은 75% 밖에 기술되지 않았다.

초기 메타데이터 설계 시 이미지자료 뿐 아니라 텍스트형 낱장자료 형태의 자료를 사진 및 이미지에 포함했기 때문에 언어항목이 필수 항목이었으나 텍스트형태의 낱장자료가 다양하게 나타남에 따라 추가되어 후에 낱장자료를 새로운 자료유형으로 설계하였다. 따라서 사진 및 이미지자료 유형은 텍스트형태가 아닌 이미지 자체로서의 유형만을 포함하는 것이 바람직할 것이다. 선택항목에서는 촬영자 혹은 제작자 등의 역할어를 포함하여 개인명을 선택사항으로 기술하기로 하였는데 60% 가까이 기술한 것으로 보아 저작자 기술항목을 필수항목으로 재조정하는 것이 요구된다.

내용 및 일반주기분석의 결과 다음과 같은 조정이 필요하다고 생각된다. 첫째, 사진 및 이미지자료에도 촬영목적이 필수항목으로 요구된다. 둘째, 이미지자료는 개별적으로 메타데이터를 작성하기 보다는 그룹화 할 수 있는 주제를 모아서 메타데이터를 작성하는 경우가 많다. 이러한 그룹을 대표할 수 있는 주제어가 자료명이 될 수 있다. 이 때 대표 자료명을 부여하는 일반적인 가이드라인이 필요하다. 셋째, 이미지자료 내의 등장인물이나 장면을 기술할 수 있는 항목이 선택항목으로 필요하다. 넷째, 인물의 초상권에 대한 면밀한 검토가 요구된다 (표 6 참조).

3.3.6 녹음자료

녹음자료의 경우에는 다음과 같은 특징을 가지고 있었다. 첫째, 학문영역별로 주류를 이루는 자료유형이 상당한 차이를 보이고 있다. 철

학·종교학 분야는 인터뷰 내용이 주로 활용되었고, 역사학 분야는 자료설명에 대한 녹음자료들이 있으며, 어문학 분야는 말뭉치가 주가 되고 있었다. 또한 사회과학 분야는 질적 연구를 위한 인터뷰자료들이 있다. 둘째, 녹음자료의 음성형 자료 외에 텍스트형태의 녹취록의 경우 녹음자료로 처리한 경우도 있고 문서형태이기 때문에 낱장자료로 취급한 경우가 있다. 이에 관해 보다 명확한 기준이 요구되는데 녹취록은 녹음자료의 일부로 확장하여 재설계함이 필요하다. 즉, 녹취록에서 다루어야 할 문서의 형태적 측면과 내용적 요소를 녹음자료의 기술요소에 추가하고 녹음자료인지 녹취록인지를 선택할 수 있는 항목 제시가 필요하다. 또한 특정한 녹음자료와 연관된 녹취록은 관련자료로서 관계설정이 필요하다. 셋째, 자료에 대한 구체적인 내용과 관점을 용이하게 파악할 수 있도록 녹음 및 녹취 목적도 기술할 수 있는 항목이 요구된다. 넷째, 녹음일자 등은 일정한 기간을 두고 반복적으로 녹음할 수 있으므로 녹음일자를 반복요소로 처리해야 한다. 다섯째, 녹음자료 역시 구술자의 동의하에 자료를 공개해야 하므로 이러한 부분 역시 디지털 저작권 문제에서 다루어야 할 것이다.

녹음자료의 필수항목인 과제번호, 자료명, 녹음내용유형, 자료공개구분, 녹음일, 녹음장소는 100%기술되었다. 그러나 언어 부분은 37% 입력률을 보이고 있는데, 선택항목인 개인명(96%), 기타 개인(91%), 제작기관(76%)은 높은 입력률을 보이고 있다. 이것은 녹음자료에 있어서 필요한 제작자, 구술자, 연주자, 녹음자, 면담자에 관한 역할어와 이름을 기술한 것이다. 따라서 저작부문과 관련된 기술요소명은 모두 필수

〈표 6〉 사진 및 이미지 하위기술 요소별 평균 메타데이터 입력비율

하위 기술요소명	표시상수	필수여부	평균 메타데이터 입력 비율
/LinkingEntry	과제번호	필수	100%
/Format/color	색채유무	필수	100%
/Language	언어	필수	75%
/Rights/accessRights	자료공개구분	필수	100%
/Title/properTitle	자료명	필수	100%
/Subject/keyword	주제용어	필수	99%
/Date/created	촬영/제작일	필수	98%
/ImageInfo/object	촬영대상	필수	100%
/ImageInfo/location	촬영장소	필수	100%
/Relation	관련자료	해당시필수	3%
/Coverage/temporal/period	기간	해당시필수	25%
/Contributor/personal/role	기타개인역할	해당시필수	15%
/Contributor/personal/name	기타개인명	해당시필수	16%
/Contributor/corporate/name	기타단체명	해당시필수	2%
/Date/issued	디지털제작일	해당시필수	9%
/Coverage/temporal/periodName	시대명	해당시필수	25%
/Source	정보원	해당시필수	14%
/Description/generalNote	주기(비고)	해당시필수	42%
/Creator/personal/role	개인역할	선택	65%
/Creator/personal/name	개인명	선택	59%
/Description/tableOfContents/contents	목차	선택	3%
/Title/parallelTitle	외국어자료명	선택	30%
/Date/available	이용가능기간	선택	2%
/Rights/accessRightsNote	자료공개조건	선택	6%
/Creator/corporate/role	제작기관역할	선택	46%
/Creator/corporate/name	제작기관명	선택	42%
/Subject/classificationTable	주제분류표	선택	6%
/Description/abstract/korean	초록(국문)	선택	62%
/Description/abstract/foreign	초록(외국어)	선택	5%
/Format/extent	크기(물리적)	선택	27%
/ImageInfo/background	피사체배경정보	선택	13%

항목으로 조정해야 할 것이다. 초록(국문)인 경우에도 85%의 입력률을 보였는데 이것은 녹음의 내용을 기술할 수 있는 항목으로 활용한 것이다. 따라서 녹음자료에서 이 항목의 표시상수를 조정하고 이를 필수항목으로 변경해야 할 것이다.

3.3.7 조사자료

조사자료는 양적 연구의 통계분석자료를 의미하는데 초기 조사자료의 기술요소 설계 시 필요한 요소들을 거의 반영하였기 때문에 추가할 요소는 거의 없다(표 7 참조). 그러나 조사자료에 첨부될 설문지와 코드북, 데이터파일의 유무에 대한 기술요소를 추가하는 것이 바람직

〈표 7〉 조사자료 하위기술 요소별 평균 메타데이터 입력비율

하위 기술요소명	표시상수	필수여부	평균 메타데이터 입력 비율
/LinkingEntry	과제번호	필수	100%
/Rights/accessRights	자료공개구분	필수	100%
/Date/created	작성일	필수	99%
/Title/properTitle	조사자료명	필수	100%
/Subject/keyword	주제용어	필수	98%
/Contributor/personal/role	공동연구자	해당시필수	72%
/Contributor/personal/name	공동연구자명	해당시필수	75%
/Coverage/temporal/period	기간	해당시필수	24%
/Contributor/corporate/role	기타기관	해당시필수	52%
/Contributor/corporate/name	기타기관명	해당시필수	53%
/Date/issued	디지털제작일	해당시필수	8%
/Description/tableOfContents/contents	목차	해당시필수	50%
/Coverage/temporal/periodName	시대명	해당시필수	5%
/Source	정보원	해당시필수	18%
/Description/generalNote	주기(비고)	해당시필수	56%
/Creator/personal/name	연구책임자(개인)	선택	98%
/Creator/personal/role	개인연구자	선택	97%
/Title/parallelTitle	외국어 조사자료명	선택	74%
/Description/abstract/korean	연구목적	선택	69%
/SurveyInfo/surveyArea	조사지역	선택	65%
/SurveyInfo/object	조사대상	선택	65%
/SurveyInfo/surveyMethod	조사방법	선택	64%
/SurveyInfo/period	조사기간	선택	61%
/SurveyInfo/sampleSize	표본크기	선택	58%
/SurveyInfo/unitOfAnalysis	분석단위	선택	55%
/SurveyInfo/question/firstVariable	주요변수	선택	51%
/Description/abstract/foreign	초록(외국어)	선택	51%
/SurveyInfo/samplingMethod	표본추출방법	선택	47%
/SurveyInfo/question/secondVariable	사회인구학적변수	선택	40%
/NumberInfo	보고서번호	선택	29%
/Creator/corporate/name	연구책임자(단체)	선택	25%
/Format/extent	크기(물리적)	선택	25%
/Creator/corporate/role	연구기관	선택	20%
/Rights/accessRightsNote	자료공개조건	선택	9%
/Subject/classification	주제분류	선택	4%

하다. 해당 시 필수 및 선택 항목에서 50%이상 입력되어 필수항목으로 변경되어야 할 요소들은 다음과 같다. 즉, 연구책임자(개인), 공동연구자, 외국어 조사자료명, 연구목적, 조사지역,

조사대상, 조사방법, 조사기간, 표본크기, 분석단위, 주요변수, 외국어 초록, 목차, 기타기관명(조사수행, 연구비지원) 등이 그것이다.

3.3.8 웹 사이트

지원과제의 성과물로서의 웹사이트는 일정한 관리기준 없이 구축되어 있으므로 구축된 웹사이트 내에 검색기능을 두는 경우 보다는 단순히 자료를 모아두는 용도로 설계된 경우가 대부분이었다. 시범통합시스템에서 기존에 구축된 웹 사이트와의 연결방법은 메타서치 형태로 검색부분을 링크시키는 방법과 웹 사이트 자체의 자료들을 메타데이터로 구축하는 방법이 논의되었으며, 메타서치가 가능한 웹사이트를 골라서 시범적으로 통합검색을 시도하였다. 그러나 많은 경우 웹 사이트의 검색이 불가능하였고 표준화된 웹 사이트 구축의 가이드라인이 제시되지 않았기 때문에 기존에 웹 사

이트에 구축된 자료들을 대상으로 메타데이터를 제작성하였다.

메타데이터 표준설계에서 필수항목으로 제시된 사이트명, 사이트주소, 자료공개구분, 최초개설일은 모두 100% 입력되었고 주제어는 98%가 입력되었다. 그밖에 선택항목이었던 주제분류는 75%, 운영책임자명은 63%, 운영기관명은 60%, 사이트맵은 50%로 입력되었다(표 8 참조). 이들 항목들은 추후에 필수항목으로 조정할 필요가 있다고 사료되며 추후에는 이러한 부분들을 고려하여 웹사이트 구축을 위한 기본 가이드라인을 제시해야 할 것이다.

〈표 8〉 웹사이트 하위기술 요소별 평균 메타데이터 입력비율

하위 기술요소명	표시상수	필수여부	평균 메타데이터 입력 비율
/LinkingEntry	과제번호	필수	100%
/Title/properTitle	사이트명	필수	100%
/Identifier	사이트주소	필수	100%
/Format/color	색채유무	필수	100%
/Rights/accessRights	자료공개구분	필수	100%
/Subject/keyword	주제용어	필수	98%
/Date/created	최초개설일	필수	100%
/Description/tableOfContents/contents	사이트맵	해당시필수	50%
/Source	정보원	해당시필수	4%
/Description/generalNote	주기(비고)	해당시필수	42%
/Contributor/personal/name	공동운영자명	선택	33%
/Contributor/corporate/role	기타기관	선택	22%
/Contributor/corporate/name	기타기관명	선택	17%
/Title/parallelTitle	외국어 사이트명	선택	46%
/Creator/corporate/name	운영기관명	선택	60%
/Creator/personal/name	운영책임자명	선택	63%
/Rights/accessRightsNote	자료공개조건	선택	25%
/Subject/classification	주제분류	선택	75%
/Description/abstract/korean	초록(국문)	선택	39%
/Description/abstract/foreign	초록(외국어)	선택	26%
/Format/extent	최적화면/페이지수	선택	31%

3.3.9 낱장자료

낱장자료는 메타데이터 표준설계 시 초기에 설계된 자료유형이 아니라 고문서 등 낱장문서 형태의 성과물 자료들이 수집됨으로써 추후 포함되었다. 필수요소 구분 역시 다른 자료유형에 비해서 입력률이 낮는데 그 이유는 고문서 위주로 낱장자료의 기술요소를 설계하였기 때문이다. 필수요소인 문서명, 문서작성일, 자료공개구분, 주제용어는 100% 입력되었으나 문서작성

자(개인) 및 문서작성자(단체), 문서작성자, 언어구분은 필수요소임에도 불구하고 모두 60% 미만이 입력되었다. 선택 및 해당 시 필수 항목 역시 50% 이상 입력된 항목은 초록(국문) 뿐이었다. 따라서 낱장자료의 메타데이터 표준양식은 고문서뿐만 아니라 낱장형태로 된 자료들이 낱장자료 유형에 준하여 구축되었음을 알 수 있는데 다양한 형태의 낱장자료를 재구분하여 설계해야 함을 보여주고 있다(표 9 참조).

〈표 9〉 낱장자료 하위기술 요소별 평균 메타데이터 입력비율

하위 기술요소명	표시상수	필수여부	평균 메타데이터 입력 비율
/LinkingEntry	과제번호	필수	100%
/Title/properTitle	문서명	필수	100%
/Creator/personal/name	문서작성(개인)	필수	56%
/Creator/corporate/name	문서작성(단체)	필수	51%
/Creator/corporate/role	문서작성기관	필수	60%
/Date/created	문서작성일	필수	100%
/Creator/personal/role	문서작성자	필수	57%
/Language	언어	필수	51%
/Rights/accessRights	자료공개구분	필수	100%
/Subject/keyword	주제용어	필수	100%
/Relation	관련자료	해당시필수	1%
/Coverage/temporal/period	기간	해당시필수	29%
/Date/issued	디지털제작일	해당시필수	11%
/Description/tableOfContents/contents	목차	해당시필수	12%
/Coverage/temporal/periodName	시대명	해당시필수	54%
/Source	정보원	해당시필수	32%
/Description/generalNote	주기(비고)	해당시필수	42%
/Contributor/corporate/name	기타문서작성기관	선택	28%
/Contributor/corporate/role	기타문서작성기관	선택	31%
/Contributor/personal/role	기타작성자	선택	33%
/Contributor/personal/name	기타작성자명	선택	34%
/Title/parallelTitle	문서외국어명	선택	22%
/LocalInfo/owner	소장기관명	선택	19%
/LocalInfo/localNumber	소장기관분류번호	선택	5%
/Date/available	이용가능기간	선택	5%
/Rights/accessRightsNote	자료공개조건	선택	5%
/Subject/classification	주제분류	선택	23%
/Subject/classificationTable	주제분류표	선택	19%
/Description/abstract/korean	초록(국문)	선택	52%
/Description/abstract/foreign	초록(외국어)	선택	9%
/Format/extent	크기(물리적)	선택	23%

날장자료의 주기사항을 분석해 보면 녹취록, 구술자료(면담자료), 신문기사, 말뭉치자료, 목록자료 그리고 고문서 형태의 자료군으로 범주화 할 수 있다. 따라서 이들 자료유형에 관한 상세한 메타데이터 항목을 개발하여 제시하는 것이 바람직할 것이다.

4. 시소러스의 확장과 메타데이터의 연계

모든 이론과 그 이론을 바탕으로 하는 응용은 상황에 따라 진화한다. 전통적인 시소러스의 용어관계가 단순하면 이를 확장하면 되는 것이며, 정보처리시스템의 목적에 따라 적절히 수정하여 최적의 데이터를 구축할 필요가 있다. 용어나 개념의 관계에 대한 연구는 무수히 많다(Green et al. 2002; Bean and Green 2001). 이들 개념관계를 시소러스의 진화측면에서 보았을 때, 1차 개념관계, 2차 개념관계 및 속성관계로 크게 그룹화할 수 있을 것이다. 1차 개념관계는 일반적으로 표준에서 정하고 있는 관계가 중심이 되는 기본관계이며, 2차 개념관계와 속성관계는 확장관계이다. 2차 개념관계는 1차 개념관계가 정하고 있는 것 외에 계층관계, 동등관계, 관련관계를 각각 더 확장한 것이며 계층관계의 확장은 다시 의미역관계와 패식관계로 나눌 수 있다. 속성관계는 지금까지 고유명사류의 전거에서 일부 적용하고 있었으나 주제명 전거나 시소러스에서는 볼 수 없었던 새로운 관계로 무거운 온톨로지를 위하여 중요한 역할을 하게 될 것이다(최석두 2006).

메타데이터의 기본목적은 정보의 소재를 파

악하고, 정보를 해석하며, 데이터를 통합하기 위한 것이다. 그러나 메타데이터를 구축하는 것만으로는 전통적인 서지데이터베이스와 별로 다를 것이 없으며 그 서비스는 한계를 가질 수밖에 없다. 기초학문자료센터의 메타데이터 구축 대상은 학진이 지원한 연구결과물이며 모든 연구물은 서로 의미관계를 가지고 연구자에게 제공되어야 한다. 이를 위해서는 확장된 시소러스, 즉 개념(용어)의 메타데이터를 추가함으로써 비로소 지식지도의 역할을 할 기본적인 환경이 갖추어 지게 된다.

4.1 클래스와 속성

확장 시소러스는 클래스와 속성을 가지며 클래스는 확장관계 중 속성관계를 정의하기 위한 것이다. 예컨대, 공통속성을 갖는 것을 묶어 클래스로 만들 수 있다. 기초학문자료센터의 메타데이터를 시범적으로 처리하기 위해서 클래스와 각 클래스에 대한 속성을 정의하였다. 클래스는 포괄적으로 '주제, 인명, 기관/단체, 문헌, 지명, 사건, 시대, 회의/집회, 국보/보물/기념물, 법률, 동물, 식물, 상품, 활동'의 14개로 정의하였으며 각 클래스마다 전체 클래스의 공통속성 및 그 클래스만 갖는 개별속성리스트를 갖고 있다. 기초학문자료센터의 메타데이터 중이 14개의 범주에 포함되는 모든 개념이 시소러스의 대상이 된다.

다만, 각 클래스의 속성의 세분정도는 목적에 따라 달라질 수 있다. 예를 들어, 같은 지명의 속성으로 '정의, 이칭/이형, 위치, 인구, 면적' 등이 설정되어 있을 때, '국가, 행정구역' 등은 문제가 없으나 '산, 강, 사찰, 해협' 등인 경우

에는 ‘인구, 면적’이 아니라 ‘높이, 길이, 창건연도’ 등이 필요하게 되므로 이들의 서브클래스를 만들어 속성을 세분할 수도 있을 것이다. 또한, 인명에 대하여 다음 절의 예에서 보이는 카뮈[Camus, Albert] 정도의 속성을 가질 수도 있으나 목적에 따라서는 개인의 기본사항(이름, 생년월일, 출생지, 본관, 신체관련정보, 취미 등), 현직사항, 학력사항, 경력사항, 작품사항, 가족사항, 활동사항, 취미사항, 사망사항 등 얼마든지 세분할 수 있을 것이다. 기초학문자료센터를 위하여 시범적으로 정의한 속성은 차후 시스템을 운영하면서 그 깊이를 조정할 수 있을 것이다.

4.2 확장 시소러스의 예

시범적으로 정의한 각 클래스 중, 전술한 1

차 및 2차 개념관계, 그리고 속성관계를 갖는 시소러스의 예로서 인명클래스에 속하는 ‘카뮈 [Camus, Albert]’를 중심으로 속성과 관계를 정의하면 <표 10, 11>과 같다. 카뮈는 시소러스의 용어이면서 인스턴스가 될 수 있다. <표 10>에서 항목명인 ‘인명-정의’에서 ‘인명’은 클래스를 ‘정의’는 속성을 나타낸다. 속성값은 단순한 값이거나 오브젝트가 될 수 있다. <표 11>에서 가장 왼쪽 열은 1차 개념관계를 나타내는 관계명이며 그 중 TIT는 ‘저작’을 의미한다. 가장 오른쪽 열은 2차 개념관계를 나타내는 관계명이며 두 용어의 대응관계를 나타낸다. 즉, ‘결혼(산문:카뮈)[結婚] - 산문/저자’는 카뮈의 산문이며, 그 산문의 저자는 카뮈라는 것을 나타낸다. 또한 <표 11>의 가운데 열에 있는 용어는 모두 별도의 클래스와 속성을 가질 수 있다.

<표 10> 클래스와 속성

카뮈[Camus, Albert]	
인명 - 정의	프랑스의 소설가극작가(1913~1960). 평론 <시지프의 신화>에서 부조리의 철학을 논하여 실존주의를 더욱 심화시켰고, 전후(戰後)의 사상과 문학에 크게 영향을 끼쳤다. 1957년에 노벨 문학상을 받았다. 작품에 <이방인>, <페스트> 따위가 있다<표준국어대사전>.
인명 - 생몰년	1913-1960
인명 - 이칭/이형	까뮈[Camus, Albert]: 까뮈, 알베르[Camus, Albert]: 알베르 까뮈[Camus, Albert]: 알베르 카뮈[Camus, Albert]: 카뮈, 알베르[Camus, Albert]: Camus, Albert: Albert Camus
인명 - 출생지	알제리 몽드비
인명 - 직업	소설가, 극작가
인명 - 지위	
인명 - 소속기관	
인명 - 저작	결혼(산문:카뮈)[結婚]: 계엄령(희곡:카뮈)[戒嚴令]: 독일인 친구에게 보내는 편지[獨逸人親舊--便紙]: 반항적 인간[反抗的人間]: 시지프의 신화[--神話]: 오해(희곡:카뮈)[誤解]: 이방인(소설:카뮈)[異邦人]: 전락(소설:카뮈)[轉落]: 정의의 사람들[正義--]: 칼리굴라[Caligula]: 페스트(소설:카뮈): 표리(소설:카뮈)[表裏]
인명 - 국적	프랑스
인명 - 시대	

〈표 11〉 1차 및 2차 개념관계

카뮈[Camus, Albert]		
UF	Albert Camus	프랑스어명/이름
	Camus, Albert	프랑스어명/이름
	까뮈, 알베르[Camus, Albert]	음차명/이름
	까뮈[Camus, Albert]	음차명/이름
	알베르 까뮈[Camus, Albert]	음차명/이름
	알베르 카뮈[Camus, Albert]	음차명/이름
	카뮈, 알베르[Camus, Albert]	음차명/이름
RT	노벨 문학상[--文學賞]	상/수상자
TIT	결혼(산문: 카뮈)[結婚]	산문/저자
	계엄령(희곡: 카뮈)[戒嚴令]	희곡/저자
	독일인 친구에게 보내는 편지[獨逸人親舊--便紙]	편지/저자
	반항적 인간[反抗的人間]	평론/저자
	시지프의 신화[--神話]	평론/저자
	오해(희곡: 카뮈)[誤解]	희곡/저자
	이방인(소설: 카뮈)[異邦人]	소설/저자
	전락(소설: 카뮈)[轉落]	소설/저자
	정의의 사람들[正義--]	희곡/저자
	칼리굴라[Caligula]	희곡/저자
	페스트(소설: 카뮈)	소설/저자
	표리(소설: 카뮈)[表裏]	소설/저자

4.3 시소러스와 메타데이터의 관계

〈표 10, 11〉에서와 같은 데이터를 기술함으로써 단순하지만 ‘카뮈[Camus, Albert]’라는 개념(용어)에 대한 메타데이터가 만들어지게 된다(문헌의 메타데이터와 개념의 메타데이터가 혼동되므로 이하 편의상 문헌의 메타데이터는 메타데이터로, 개념의 메타데이터는 확장 시소러스라 칭하기로 한다). 확장 시소러스와 메타데이터의 가장 전형적인 관계는 메타데이터요소 중 주제에 해당하는 용어가 시소러스에 포함된다는 것이다.

더 바람직한 것은 메타데이터의 어느 요소에 포함되어 있는 것이라도 그것이 용어로서 의미가 있는 것은 모두 시소러스에 포함시키는 일

이다(Choi 2006). 예컨대, 상기 카뮈[Camus, Albert]의 정의, 출생지, 직업 등에 포함되어 있는 ‘프랑스, 소설가, 극작가, 평론, 부조리, 철학, 실존주의, 전후(戰後)의 사상, 문학, 노벨문학상, 작품, 알제리 몽드레’ 등이 이에 해당한다. 다만, 이들 중 ‘카뮈[Camus, Albert]’와 관계가 있는 것도 있겠지만 모두 반드시 관계를 가져야 한다는 말은 아니다.

이와 같은 데이터를 구축하는 일은 노동집약적인 방대한 작업이다. 컴퓨터시스템이 인간세계의 지식을 마술처럼 이해하고 처리하는 인공지능시스템이란 존재하지 않는다. 인간이 데이터를 잘 만들고 처리방법을 잘 정의해야 기계가 문제를 풀 수 있는 것이다. 시소러스의 1, 2차 관계뿐만 아니라 속성관계를 얼마든지 확장

할 수 있으므로 시소러스의 범위는 그만큼 넓어지며 목적에 따라 조정할 수 있는 융통성을 갖게 된다. 한정된 도메인의 메타데이터와 확장시소러스의 결합은 여러 가지 서비스에 따른 추론규칙 등을 추가함으로써 도메인 온톨로지로서의 역할을 충분히 할 수 있을 것이며, 도메인들을 통합함으로써 범용온톨로지가 가능하게 될 것이다.

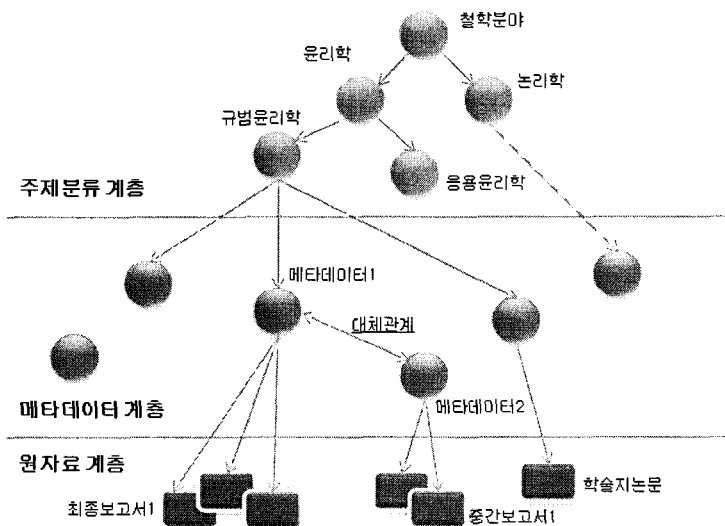
모형을 주제분야 분류표를 기반으로 설계하고 토픽맵 기반으로 구현함으로써 주제 중심의 자료검색이 가능함을 제시하고자 하였다. 이는 키워드 및 주제어 기반 정보검색은 동일한 키워드를 포함하고 있으나 내용상으로 서로 상이한 자료를 검색한다는 점과 상이한 키워드를 포함하고 있으나 내용상으로 관련성이 깊은 자료들을 검색하지 못하는 한계를 극복하기 위해서이다.

5. 온톨로지 기반 정보검색

이와 같이 메타데이터 상위에 도메인 온톨로지를 분야별로 구축함으로써 주제어들 사이의 상하 계층 관계 및 의미적 연관관계를 정의하고 주제어와 메타데이터를 연결하는 것이 가능하다. 본 연구에서는 기초학문자료센터의 대상 분야인 철학·종교학 분야의 도메인 온톨로지

5.1 철학·종교학 도메인 온톨로지 설계

도메인 온톨로지 구축 모형은 <그림 2>와 같이 3계층 구조로 설계할 수 있다. 최하위 계층은 자원 계층(resource layer)으로서 보고서, 학술지 논문, 단행본, 동영상 등의 원자료 파일들이 저장된 계층이다. 메타데이터 계층(metadata layer)은 자원 계층의 상위에 존재하는 계층으



<그림 2> 3계층 구조의 도메인 온톨로지 및 메타데이터

로서 원자료의 메타데이터가 등록되어 있으며 메타데이터와 원자료 사이에 연결되어 있고 메타데이터 상호 간에도 부분-집합관계, 대체관계, 요구관계 등의 관계로 연결되어 있다. 메타데이터 계층 상위에는 도메인 온톨로지가 정의되어 있는 계층으로서 학문분야별 온톨로지와 이들을 통합할 수 있는 상위수준의 온톨로지가 정의된다. 이 부분은 전문한 확장시소러스와 톱클래스에 대응된다. 본 연구에서는 철학·종교학 도메인 온톨로지를 주제분류 수준에서 시범적으로 설계하였으며 향후 추가적인 속성 및 관계 정의를 통하여 보다 정밀한 온톨로지 확장을 할 수 있도록 하였다.

철학·종교학 도메인 온톨로지는 주제 중심(subject-oriented)의 온톨로지로서 크게 철학 카테고리과 종교학 카테고리로 구성된다(그림 3 참조). 철학 카테고리의 주제 분류는 동양 및 서양 철학의 공통적인 철학 사상, 동양의 시대적 및 지역적 철학 사상, 서양의 시대적 및 지역적 철학 사상을 기준으로 체계화하였다. 예를 들어, 윤리학의 경우 동양과 서양의 공통적인 철학 사상으로서 상위 주제는 철학이고 하위 주제는 규범 윤리학, 메타 윤리학, 응용 윤리학이다. 응용 윤리학은 그 하위에 생명의료윤리, 환경윤리 등 6가지 세부 주제를 가진다. 종교학 카테고리는 종교학 및 전반적인 이론 분야와 한국종교, 유교, 불교, 도교, 기독교, 카톨릭 등의 세부 종교 분야로 계층화하였다.

5.2 토픽맵 기반의 도메인 온톨로지 구현

ISO 표준인 토픽맵은 영역 지식(domain knowledge)을 추출하여 개념화한 다음 이들 사이

를 의미적 관계로 연결함으로써 지식 네트워크를 구성할 수 있다. 토픽맵은 RDF, OWL과 함께 온톨로지를 표현할 수 있는 데이터 모델로서 주로 지식 관리, 콘텐츠 관리, 문헌 관리 등에서 의미 기반으로 자료를 연결하고 검색하기 위한 인간 중심의 온톨로지 구축을 위해 사용된다(Graham 2000). 따라서 본 연구에서는 기초학문자료의 도메인 온톨로지를 표현하고 구축하기 위해 토픽맵을 사용하였다.

설계된 도메인 온톨로지를 토픽맵 기반으로 구현하는 과정은 크게 XTM(XML Topic Maps)(Steve and Graham 2001) 구문으로 온톨로지를 기술하는 단계와 XTM 문서들을 해석하여 토픽맵 모델을 구축하는 단계, 그리고 구축된 토픽맵을 데이터베이스에 저장하는 단계로 나누어진다. 먼저, XTM 구문으로 온톨로지를 기술하는 단계에서는 메타데이터를 토픽맵 구조로 매핑하는 과정과 도메인 온톨로지를 XTM 구문으로 작성하는 과정을 수행한다. 기초학문자료 메타데이터와 토픽맵 구조 사이의 매핑 관계는 <표 12>와 같다.

<표 12>의 매핑 관계에 따라 기초학문자료 데이터베이스의 메타데이터 인스턴스들이 대응되는 XTM 구문으로 자동으로 변환시킬 수 있는데 이를 위해 주제분류-XTM 템플릿과 DC-XTM 템플릿이 필요하다(그림 4 참조). 설계된 주제 분류는 메타데이터와 함께 데이터베이스에 저장되므로 주제분류-XTM 템플릿은 주제 분류 목록을 이와 대응하는 XTM 구문으로 생성하기 위한 변환 규칙 및 예외 규칙, 계층적인 문서 구조를 정의하고 있다. DC-XTM 템플릿은 메타데이터 테이블의 proper Title, Parallel Title, Creator 등의 DC 요소들을 어커런스 타



〈그림 3〉 철학·종교학 도메인 온톨로지 주제분류

〈표 12〉 기초학문자료 메타데이터와 토픽맵 구조 사이의 매핑 관계

메타데이터	토픽맵
메타데이터 인스턴스	토픽(Topic)
메타데이터 요소	어커런스 타입(OccurrenceType)
메타데이터 요소 값	어커런스 값(resourceRef 또는 resourceData)
메타데이터 관계	연관관계(Association)

입 역할의 토픽으로 변환하기 위한 변환 규칙을 정의하고 있다. 이 템플릿은 데이터베이스 시스템 카탈로그의 메타데이터 테이블 정의를 입력으로 하여 메타데이터요소들을 추출한 다음 이들을 대응되는 토픽으로 변환한다.

XTM 변환 규칙은, 1) 토픽 생성 규칙, 2) 어커런스 생성 규칙 3), 주제 분류와의 연결 규칙, 4) 메타데이터 관계를 연관관계로 변환하기

위한 규칙 등으로 나누어진다. 이를 세분화하면 토픽 생성 규칙에는 토픽ID 정의 규칙, 토픽명 정의 규칙 등이 있으며 어커런스 생성 규칙에는 어커런스 타입 정의 규칙, 어커런스 값 정의 규칙 등이 있다. 주제 분류와의 연결 규칙은 메타데이터와 연결된 다중 상위 분류들을 XTM의 상위주제-하위주제(superclass-subclass) 관계로 변환하기 위한 규칙으로 세분화되며 메타데

이터 관계를 연관관계로 변환하기 위한 규칙은 isPartOf, hasPart, isVersionOf, hasVersion, isFormatOf, hasFormat, isReplacedBy, replace, isRequiredBy, requires, isReferencedBy, reference, conformsTo 등의 메타데이터 관계를 토픽들 사이의 연관관계로 변환하기 위한 규칙

으로 세분화된다.

변환된 XTM 문서들은 토픽맵 네비게이션 프로그램인 Ontopia Omnigator에서 오류 여부 확인과 함께 주제 중심의 자료검색의 결과를 확인하였다. <그림 5>는 조선불교통사 연구 과제의 최종보고서 메타데이터를 검색한 것으

분류수준	상위분류명	순번	하위분류명	한글분류명(영문)	토픽ID	상위토픽ID
1	인문학	1	철학	Philosophy	Philosophy	Humanities
2	철학	1	논리학	Logic	Logic	Philosophy
3	논리학	1	논리학사	History of Logic	History_of_Logic	Logic
4	논리학사	1	고대논리학	Ancient Logic	Ancient_Logic	History_of_Logic
5	고대논리학	1	그리스논리학	Greek Logic	Greek_Logic	Ancient_Logic
5	고대논리학	2	인도논리학	Indian Logic	Indian_Logic	Ancient_Logic
5	고대논리학	3	중국논리학	Chinese Logic	Chinese_Logic	Ancient_Logic
4	논리학사	2	중세논리학	Medieval Logic	Medieval_Logic	History_of_Logic
4	논리학사	3	근대논리학	Modern Logic	Modern_Logic	History_of_Logic
4	논리학사	4	현대논리학	Contemporary Logic	Contemporary_Logic	History_of_Logic
3	논리학	2	윤리철학	Philosophy of Logic	Philosophy_of_Logic	Logic
3	논리학	3	철학적논리학	Philosophical Logic	Philosophical_Logic	Logic
3	논리학	4	수리논리학	Mathematical Logic	Mathematical_Logic	Logic


```

<?xml version="1.0" encoding="UTF-8" standalone="yes" ?
<topicMap id="Philosophy_and_Religion" xmlns="http://www.topicmaps.org/xt" xmlns:xlink="http://www.w3c.org/1999/X"
  <baseNameString
  <topic
  <baseName
  <scope
  <baseNameString
  <instanceOf
  <topicRef xlink:href="#1#*+topicId"
  
```

<그림 4> 주제분류-XTM 템플릿

근기념인 실학파 예학관 연구;성호학의 실학성과 그 예학 Type(s): KRF-2003-074-AS0014; 보고서; 유교의류; 유교이론; 조선시대유가철학; 한국정리사

Untyped Names (4)

- 근기념인 실학파 예학관 연구;성호학의 실학성과 그 예학

Associations (3)

- 선행판
- 근기념인 실학파 예학관 연구;성호학의 실학성과 그 예학
- 종합자료
- 星峯 聖學의 展開와 變遷

Identifiable Resources (2)

- 철학종교학 기초학문자료 지식지도
 - 철학
 - 동양철학
 - 한국철학
 - 한국유가철학
 - 조선시대유가철학
 - ...

Identifiable Resources (16)

- 개인저자
- 백도근
- 목차
- 1. 서론 2. 성호학의 두길 1) '解以敎用'의 길 2) 博探佛證의 길 3. 성호학의 조곡 1) 古談學術의 문지점 2) '星峯文庫'의 왜곡 4. 星峯聖學의 실학성 1) 朱子家禮 비판 2) 厚人家禮 제작 3) 성호예학의 실학성 5. 결론
- 주제명어
- 성호학; 수사학; 가례; 서인가례; 실무
- 초록
- 본 논문의 목적은 성호예학이 오직 근신부자의 의리구현이 목적이었던 전통예학과는 달리, "참하에 굶주린 백성이 없게 한다(는 養天下 養萬民)"는 이념의 지배 아래에 있는 실학의 한 영역임을 밝히는 데 있다. 성호학은 성호학당의 우파제자들이 위해 왜곡되어 근기념적학으로 분류되어 왔으나 근본적으로 성리학과는 이념을 달리하는 실학임을 밝히고, 성호의 예학도 이념상 전통예학과 다른, 실학의 영역에 있는 학문임을 밝혔다.
- 크기(용량적)
- 16쪽(A4)

<그림 5> 토픽맵 기반의 정보검색

로 상위타입으로 연구과제번호, 자료유형(보고서), 주제 분류(한국근현대철학, 한국불교철학, 한국철학사)를 가지고 있으며 개인저자, 개인 역할 등의 상세한 속성들이 검색될 수 있음을 보이고 있다.

6. 결 론

학문 분야별로 중점적으로 관리하고자 하는 기초학문자료의 자료 유형과 자료 표현 및 구성 방식에 있어서 차이를 가진다. 예를 들어, 사회과학에서는 설문에 의한 조사 및 통계 자료, 역사학에서는 과거의 행정 문서, 도서 등을 디지털화한 이미지 자료를 중요하게 여긴다. 반면에 철학, 어문학 등 인문학에서는 모든 연구의 토대가 되는 과거 문헌들의 원본, 번역본, 해설본 등의 텍스트 자료가 중요한 정보원이다. 따라서 동일한 자료 유형일지라도 학문 분야별 관점에 따라 메타데이터 기술 양식에 대한 요구가 다양하다.

본 연구에서는 학문 분야별 각기 다른 관점에서의 다양한 자료 유형을 통합하기 위한 메타데이터 형식을 설계하고 메타데이터 구축을 위한 자료 분석 및 메타데이터 기술 가이드라인을 작성하였다. 설계된 메타데이터의 특징은 표준화, 계층화, 확장성의 세 가지로 요약할 수 있다. 표준화는 다양한 유형과 분야의 원자료들을 통합하고 다른 메타데이터 모델과 매핑이 가능하도록 한 것이고 계층화는 연구과제, 연구자, 주제어, 자료 유형 사이에 계층적 구조를 둬으로써 메타데이터들 사이의 상호 연결이 가능하도록 한 것이다. 그리고 확장성은 학문 분

야의 확장과 자료 유형의 확장 그리고 메타데이터 요소의 확장이 가능하도록 한 것이다.

또한 설계된 메타데이터 모델의 정당성 및 효율성을 평가하기 위해 이를 토대로 구축된 기초학문자료 데이터베이스와 메타데이터를 분석하였으며 분석 결과를 바탕으로 보다 실제적인 인문사회과학 원자료를 위한 메타데이터 구성요소를 제시하였다. 기초학문자료와 같이 전문적인 분야의 자료에서는 메타데이터 설계자와 학문 분야 전문가인 메타데이터 구축자 사이에 자료 해석의 차이가 발생할 가능성이 높으며 이에 따라 실제 구축된 자료의 분석에 의해 제시된 메타데이터 모델은 이미 구축된 자료의 보정과 새롭게 구축될 자료의 정확한 메타데이터 기술에 활용될 수 있다.

메타데이터는 자원 중심의 서술이므로 키워드 검색 기반의 정보검색에 있어서 한계를 가진다. 이러한 문제를 해결하는 방법은 확장시소러스를 설계하고 이를 메타데이터와 결합하여 도메인 온톨로지를 구축하는 것이다. 본 연구에서는 철학·종교학 분야의 도메인 온톨로지 구축을 위해 시범적으로 주제 분류 중심으로 확장시소러스를 설계하였으며 ISO 표준인 토픽맵을 기반으로 시스템 상에 구축하고 이를 기반으로 주제 중심의 메타데이터 검색이 가능함을 보였다. 토픽맵 구축을 위해 XTM 변환 템플릿들을 개발하였으며 변환 규칙 및 예외 규칙 등을 정의하여 자동화된 XTM 문서 생성이 가능하도록 하였다.

향후 연구 과제는 향상된 메타데이터 설계에 따라 지속적으로 기초학문자료 데이터베이스를 구축하고 용어 온톨로지와 학문 분야별 주제 온톨로지를 설계 및 구축하여 하위로는 자

료들 사이의 의미적 연관관계를 구축하고 상위로는 동일 학문 내 주제들 사이에 또는 서로 다른 학문 주제들 사이에 의미적 연관관계를 구축하는 것이다. 또한 이러한 의미망을 기반으

로 지식검색이 가능하도록 함으로써 기초학문 자료 데이터베이스가 후속 연구를 위한 토대 자료로 활용될 수 있도록 하는 것이다.

참 고 문 헌

- 국립국어연구원. 1999. 『표준국어대사전』. 서울: 두산동아.
- 백종현. 2005. 『기초학문자료센터 설치를 위한 기획연구』. 서울: 한국학술진흥재단.
- 최석두 등. 2006. 대단위 우리말 온톨로지 구축을 위한 시소러스 개발. 『情報管理學會誌』, 23(4): 147-164(2006.12).
- 한국학술진흥재단. 2006. 『기초학문자료센터 설립 및 시범사업』. 서울: 동 재단.
- Bean, Carol A. and Rebecca Green. ed. *Relationships in the Organization of Knowledge*. Boston: Kluwer Academic Publishers.
- Caplan, Priscilla 저. 오동근 역. 2004. 『메타데이터의 이해』. 서울: 태일사.
- Choi, Suk-Doo. 2006. "Ontology development using thesaurus." *The Second International Symposium on Knowledge Processing and Service for China, Japan and Korea: Metadata and Ontology, October 15-16, 2006, Beijing, China*, 43-54.
- Dublin Core Metadata Initiative.
<<http://www.dublincore.org>>
- Graham, M. 2000. "Topic Map technology - the state of the art." *XML 2000 Conference & Exposition*. Washington, USA, December 2000.
- Green, Rebecca, Carol A. Bean, and Sung Hyon Myaeng. ed. 2002. *The Semantics of Relationships: An Interdisciplinary Perspective*. Dordrecht: Kluwer Academic Publishers.
- Steve, P. and M. Graham. 2001. "XML Topic Maps(XTM) 1.0", TopicMaps.Org, 2001.
- Tannenbaum, Adrienne. 2002. *Metadata solutions: Using Metamodels, Repositories, XML, and Enterprise Portals to Generate Information on Demand*. Boston: Addison-Wesey.