

Ontology를 이용한 이종 메타데이터 검색 시스템의 설계 및 구현¹⁾

최현종, 김태영
한국교원대학교 컴퓨터교육과

요 약

웹은 이제 더 이상 무의미한 정보의 바다가 아니라, 의미있는 정보를 사용자에게 정확하게 전해줄 수 있는 새로운 Semantic Web 시대를 맞이하고 있다. 그 출발점이 XML과 메타데이터였다면 웹의 자원을 연결시켜 주는 RDF를 거쳐, 이제는 의미와 논리가 Ontology를 통해 자리 잡고 있다. 현재 국내의 많은 웹 교육 자료들이 한국 교육학술정보원의 메타데이터형식(KEM)에 메타데이터가 구축되었다. 따라서 우리나라도 이제는 의미와 논리의 형식을 연구해야 될 시기가 되었다. 하지만, 국내의 많은 연구자들은 KEM보다 DublinCore의 DC와 SCORM의 LOM의 연구에 더 적극적이다. 따라서 현재는 의미와 추론의 연구보다 세 개의 서로 다른 메타데이터를 엮어 서로 공유하고 통합할 수 있는 방법의 연구가 선행되어야 한다. 이에 본 연구는 세 메타데이터 형식을 통합해 주는 Ontology를 설계해 보고, 프로토타입 검색 시스템을 구현해 보았다. 연구 결과 세 메타데이터의 많은 요소들이 서로 같은 요소 이름과 의미를 가지고 있었고, 몇 개의 요소들은 서로 같은 의미이지만 다른 요소 이름을 가지고 있었다. Ontology를 설계할 때 이렇게 동일한 의미를 가진 다른 이름의 요소들은 일대일 매핑 기법을 적용하여 설계하였다. 본 연구에서 설계한 Ontology가 세 가지 다른 형식의 메타데이터를 통합하는 통합 스키마로 활용될 수 있음을 확인하였다. 더불어 검색 결과를 좀 더 다양하게 동의, 반의, 계층, 연관 관계까지 제공하기 위해서는 좀 더 정교한 속성 설계가 필요하였다.

주제어 : 메타데이터, 온톨로지, 통합 스키마

A Design and Implementation of Heterogeneous Metadata Searching System using Ontology

Hyunjong Choe, Taeyoung Kim
Dept. of Computer Education, KNUE

ABSTRACT

World Wide Web is not more meaningless sea of information but is becoming the Semantic Web that provides many users with meaningful information. The starting point is the XML and metadata, RDF is a stopover which gives technique to relate arbitrary web resources. And now, the semantic and logic of web resources can be settled in the Ontology. A lot of educational multimedia web resources in Korea have produced their metadata with KERIS's KEM(Korea Educational Metadata). Therefore our country have to start the study of the semantic and logic in web resources. But, many researchers in Korea are more eager to study DublinCore's DC and SCORM's LOM metadata specification than KEM. Thus the study of method about sharing and integrating these three metadata specifications should be performed before the study of semantic and logic in web resources in Korea. We design the Ontology to integrate these three metadata specifications and implement the prototype system using this Ontology. These three metadata have some elements that have same labels and meanings, and other elements have different labels and same meanings. To match these different labels which have same meanings, we adapted the one-to-one mapping technique in designing our Ontology. This designed Ontology was imported as "integrated schema" in our prototype searching system to integrate three different metadata in databases. Moreover we know that the more specific property design of class in Ontology was needed in order to provide users with more informed searching results such as synonym, antonym, hierarchy and associations.

Keywords : Metadata, Ontology, integrated schema

1. 서론

웹(Web)은 세계의 수많은 사용자와 정보들로 인해 단기간에 괄목할만한 성장을 이루었다. 그러나 십여 년 전에 개발되어서 지금까지 사용되고 있는 웹 기술은 현재의 이런 복잡한 웹 정보를 충분히 반영하지 못하고 있는 것 또한 사실이다. 즉 아직까지 웹의 수많은 자료에서 유용한 정보에 접근하고, 추출하고, 번역하고, 유지 보수하는 일이 사용자들의 손에 의지하고 있다는 것이다. 때문에 W3C의 의장인 Tim Berners-Lee는 사람의 손에서 기계(machine)의 자동화된 처리 능력에 정보를 맡기기 위해 웹 자원들이 스스로 의미를 가지고 있어 이를 처리할 수 있는 새로운 Semantic Web이라는 개념을 등장시켰다[22]. 웹 자원이 적절한 구문(syntax)과 의미(semantics)를 가지게 된다면 웹 자원의 정보 관리와 통합, 지식 관리 등과 같이 가치 있는 정보처리영역에서 인간은 훨씬 더 쉽고 효율적으로 지식을 관리할 수 있게 될 것이다.

웹 자원의 구문은 이미 XML(eXtensible Markup Language)를 통해 그 기반이 마련되어가고 있고, 의미를 통한 접근은 현재 Ontology를 통해 접근 중이다. Ontology를 통해 웹을 비롯한 많은 자원들이 서로 공유할 수 있는 용어를 가지게 되고, 이 용어를 통해 객체와 객체의 관계를 표현할 수 있게 되었으며 또한 객체들 간의 관계를 통해 의미있는 유추(infering)까지 가능하게 된 것이다.

우리나라의 경우 다른 어느 나라보다 발전된 웹 인프라(infrastructure)로 인해 전국의 16개 시,도교육청과 많은 교육자료 개발회사 및 개인들이 수많은 웹 교육 자료를 개발하고 있다. 이렇게 개발된 교육 자료들을 자체적으로 배포하는 것을 뛰어넘어, 이제는 웹에서 통합적으로 서비스하고자 하는 노력을 한국교육학술정보원(KERIS)과 한국교육과정평가원(KICE)에서 하고 있다. 그 일차 작업으로, 현재 개발된 교육 자료에 대한 메타데이터를 작성하는 작업이 마무리되고 가고 있다. 하지만, 이 작업에 사용되고 있는 메타데이터는 2001년 KERIS에서 발표한 국가표준 교육정보 메타데이터(Korea Educational Metadata, KEM)를 이용하고 있는데, 이 모델은 DC, DCED, GEM, LOM 등의 모델에서 몇 가지 사용될 수 있는 요소(element)들을 추출하여 국내의 실정에 맞게 다시 재구성하여 만든 모델이다. 따라서 현재 국내외에서 가장 많이 연구 및 사용되고 있는 ADL(Advanced Distributed Learning)의 SCORM(Sharable Course Object Reference Model) 모델이나 IEEE의 학습 객체 메타데이터(LOM)와는 많은 부분에서 서로 다르게 정의되고 있다. 그러므로 KEM으로 정의된 자료와 LOM, DC로 정의된 자료를 서로 공유하여 통합하기 위해서는 서로 다른 메타데이

터 스키마를 통합해 주는 Ontology가 필요하게 된다.

따라서 본 논문은 Ontology와 Semantic Web에 대해 알아보고, 국내에서 사용되는 메타데이터 모델인 KEM과 LOM을 서로 비교해 본 다음, 이를 공유하기 위한 Ontology를 설계하고 이를 구현해 보았다. 본 논문의 2장에서는 Ontology의 의미와 웹에서의 위치를 살펴보고 국내외의 관련 연구를 고찰해 보았다. 3장에서는 이중 메타데이터를 통합하기 위한 Ontology를 설계하고, 4장에서는 이를 이용한 검색 시스템을 설계 및 구현해 보았다.

2. 관련연구

2.1. Ontology

Ontology는 그리스어 ontos(being)와 logos(word)에서 유래한 단어로 철학 분야에서 실세계에 존재하는 것에 대한 사물들간의 유기적 관계를 설명하는 개념으로 사용되었다. 이 개념이 정보처리 분야에서 지식의 재사용, 서로 다른 시스템들간의 상호 운영, 개발자들간의 커뮤니케이션 도구, 특정 도메인(domain) 영역 지식에 대한 공통적 이해를 표현하는 목적으로 사용되어지고 있다[23]. 특히 1990년대에 들어서면서 인공지능에서 지식 표현, 자연어 처리 등의 분야에서 연구되어지다가 최근에는 Semantic Web 분야에서 지식을 표현하고 처리하기 위한 수단으로 연구되고 있다. 인공지능이나 Semantic Web 분야에서 Ontology는 특정 영역의 지식을 표현하기 위한 객체(object), 릴레이션(relation), 공리(axiom) 등으로 표현된다. 실세계의 사물이나 사건은 객체로 표현하고, 객체들의 관계는 릴레이션, 특정 영역의 지식을 표현하기 구성된 객체와 릴레이션으로 이루어진 계층적 구조를 분류(taxonomy)라고 표현하기도 한다. 이 계층화된 구조속에서 일정한 패턴의 로직(logic)이나 공리가 사용되어 새로운 지식을 생성하거나 추론할 수 있다.

Ontology는 어떻게 표현하느냐와 어느 정도의 일반성과 구체성을 갖느냐에 따라 분류될 수 있는데, 일상어와 로직 체계를 갖춘 표현 정도에 따라 highly informal, semi-informal, semi-formal, rigorously formal Ontology로 나눌 수 있고[16], 표현의 일반성과 구체성에 따라 top-level, domain, task, application Ontology로 나눌 수 있다[15]. Ontology가 이전의 정보 시스템에서 사용되어 왔던 로직, 의미망(semantic network), 프레임(frame), 퍼지(fuzzy) 등과 같은 지식 표현 방법이 있음에도 불구하고 활발하게 연구되고 있는 이유로 Natalya는 다음과 같은 이유를 들고 있다 [14].

첫째, 사람들이나 소프트웨어들간에 서로 다르게 이해할 수 있는 정보의 구조에 대해 공통적 이해를 갖게

1) 이 논문은 2004년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2004-002-B00249)

된다.

둘째, 지식의 재사용이 가능하다.

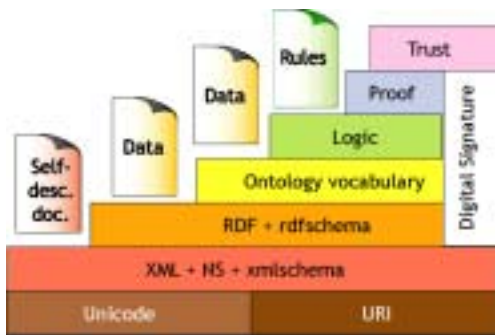
셋째, 기존의 지식을 이용한 가정(assumption)이 가능하다.

넷째, 특정 도메인의 지식 그 자체와 지식의 조작을 분리하고 있다.

다섯째, 도메인 지식을 분석할 수 있다.

2.2. 웹에서의 Ontology

웹에서의 Ontology의 적용은 현재 Semantic Web에서 가장 활발하게 연구되고 있는 분야이다. Semantic Web은 새로운 웹 기술의 패러다임으로 기존의 웹이 가지고 있는 내용과 형식으로 표현된 구조에 형식적인 의미를 부여해 주었다. 내용에 부여된 의미를 이용하여 특히 웹 에이전트들(agents)이 웹의 내용을 이전보다 더 정확하게 추출(extract) 및 추론할 수 있기 때문에 사용자 입장에서 보면 더 좋은 서비스를 받을 수 있을 것이라고 예상하고 있다. 여기에는 두 개의 문제가 해결되어야 한다. 그 첫 번째는 의미를 어떻게 표현할 것이냐의 문제이고, 두 번째는 표현된 의미를 이용하여 어떻게 추론할 것이냐이다. 이에 대해 Tim Berners-Lee는 Semantic Web의 구조를 (그림 1)과 같이 몇 개의 계층(layer) 구조를 통해 설명하고 있다 [22].



(그림 1) Semantic Web의 구조

XML+NS(NameSpace)+xmlschema는 구문(syntax) 계층이고, RDF+rdfschemata 계층은 데이터 계층을 형성하고, Ontology vocabulary 계층은 데이터의 의미를 규정하는 공통적인 형식을 표현해주는 계층이고, Logic 계층은 추론을 할 수 있도록 해 주는 규칙(rules)을 담당하는 계층, Proof 계층은 서로 다른 웹 에이전트들간에 통신과 추론된 식의 검증에 사용되는 계층이다. Trust 계층은 서로 다른 네트워크의 자료를 검증된 계층과 전자 서명을 통해 Trust Semantic Web을 구축할 수 있는 기반을 제공해 주는 계층이다.

자료와 표현 양식을 함께 제공했던 HTML의 한계를 극복하고자 자료 처리가 가능한 XML이 등장하게 되고, XML의 자료에 통일된 구문을 제공하고자 RDF

가 등장하게 되었다. 하지만 RDF 역시 웹 자료의 메타데이터로 이용할 수 있는 어휘, 구조, 제약조건 몇 가지에 대한 구문만을 제공하기 때문에 웹 에이전트나 응용 프로그램에서 지식 처리가 어렵다. 예를 들어 동의어(synonym), 반의어(antonym), 분류(classification), 집합(aggregation) 등의 기본적 추론을 위한 공리가 부족하다. 따라서 XML/RDF를 기반으로 한 웹 Ontology인 OIL(Ontology Inference Layer), DAML+OIL, OWL(Ontology Web Language) 등이 제안되었고, 이외에도 XOL(Ontology Exchange Layer), OML(Ontology Markup language), SHOE(Simple HTML Ontology Extensions) 등이 등장하게 되었다.

D. Fensel은 Ontology가 갖추어야 될 세 가지 조건으로 높은 직관성, 잘 정의된 형식적 의미, 웹 표준언어와의 호환성을 제시하고 있는데, 이는 현재 웹 인프라의 성장속도와 그 활용 범위를 살펴 볼 때 반드시 충족되어야 할 요건이라고 할 수 있다[12]. 현재 가장 활발하게 연구되고 있는 웹 Ontology는 W3C에서 표준화 작업 중인 RDF, OIL, DAML+OIL, OWL로 분류해 볼 수 있다. <표 1>은 네 가지 Ontology에 대해 각각에서 사용되는 구문 측면에서 서로 간략하게 비교한 것이다.

<표 1> 네가지 Ontology의 구문 비교

RDF	OIL	DAML+OIL	OWL
class	class-of	Class(OWL Lite
subclassof	subclass of	intersectionOf,	(Class
	class-expression	unionOf	Individual
	and	complementOf	sameAs
	or	hasValue	AllDifferent
	not	...)	...)
	slot-constraint	axiom(OWL DL,Full
	has-value	subClassOf	(disjointWith
	value-type	sameClassOf	unionOf
	max-cardinality	subPropertyOf	complementOf
	min-cardinality	inverseOf	intersectoinOf
	cardinality	...)	...)

OIL은 On-To-Knowledge 프로젝트를 수행한 IST(Information Society Technologies)에서 개발된 웹 Ontology로 웹기반의 지식 표현과 추론을 위해 설계되었다. DARPA 프로젝트에 의해 추진되어진 DAML(DARPA Agent Markup Language)에 OIL 개발이 합쳐 DAML+OIL로 명명되어 개발되고 있는데, DAML+OIL에서는 객체를 DAML 클래스(class), 속성, 데이터 유형 등으로 XML 스키마를 이용하여 클래스를 정의한다. 하지만 DAML+OIL이 언어 표현력에서 우수하지만 처음 사용하는 사용자, 개발자에게는 용이하지 않은 측면이 있어, 이를 개선하여 보편적인 이용자를 확보하기 위해 간결하고 사용하기 쉬운 언어로 OWL이 개발되었고 현재 W3C에서 표준화 진행중이다 [10][17][19][21][25][26].

2.3. 국내외의 관련 연구

국내의 Ontology 관련 연구는 주로 정보처리 분야에서, 추론에 관한 연구는 인공지능 분야에서 연구되고 있다. 주로 웹 환경에서 Ontology를 이용한 지식관리 시스템 설계 및 구현이 있다[2][5][6]. 국외에서는 지식관리 시스템과 더불어 Semantic Web의 지식 표현과 추론을 위한 중요한 구성 요소로서 여러 방면에서 다양하게 연구되고 있다[9][13][18][20].

특히, 국내의 연구들을 살펴보면 문헌 정보학 분야나 지리정보 시스템 등 일부 특정 분야에서 사용되는 도메인(domain) Ontology로서의 연구가 대부분이기 때문에 연구되고 있는 분야가 특정적이며 특수한 지식에 국한된다는 특징을 가지고 있다[1][3][4]. 따라서 본 연구처럼 특정 부분에 국한되는 도메인 지식을 통합, 확장하는 연구가 부족한 실정이다.

이에 반해 국외에서는 이미 오래전부터 여러 연구단체들이 Semantic Web에 대한 연구를 활발하게 진행하고 있는데 주로 Ontology 설계과 이를 적용한 지식 기반 시스템, 각각 다른 도메인에서 설계된 Ontology의 통합 등에 연구가 활발하다[8][9][13][24].

3. Ontology 설계

3.1. 학습객체 메타데이터

학습객체에 대해 D. Wiley는 다양한 디지털 기반의 교육 자료로, 객체지향 프로그래밍 언어의 모듈처럼 한 번 만들어지면 여러 상황에서 다시 재사용하여 사용할 수 있게 하는 학습을 위한 작은 학습 컴퍼넌트(small instructional component)로 정의하였다[11]. 학습객체는 객체기반 프로그래밍의 발전과 더불어 국내외의 웹기반 학습 시스템에서 많이 연구되고 있는데, 우리나라에서도 2001년 KERIS에서 국가표준 교육정보 메타데이터의 형식인 KEM을 발표하였다. 국외의 다양한 모델을 검토하여 국내 실정에 맞게 수정을 한 결과 23개의 기술 요소와 51개의 하위 요소를 발표하였는데, 현재 전국의 16개 시,도 교육청에서 이를 기반으로 이미 생산된 많은 멀티미디어 교육 자료에 메타데이터를 기입하고 있다. 국외에서는 ADL의 SCORM 모델과 IEEE의 학습객체 메타데이터(LOM)가 가장 활발하게 연구되고 있다. <표 2>는 이 두 모델과 두 모델의 기반이 된 DC(Dublin Core) 모델을 서로 비교한 것이다.

DC를 기반으로 한 두 모델을 비교해 보면 표제, 저자, 주제, 개요 등의 일반적인 메타데이터 요소들은 서로 비슷하지만 KEM은 국내의 교육 현실을 감안하여 학교급, 교과목, 학기 등의 정보를 담고 있는 learningArea 요소를 추가하였으며, LOM에는 존재하지 않지만 GEM(Gateway to Educational Materials) 메타데이터에서 사용하는 있는 교수 방법에 관련된

pedagogy 요소를 추가하였다. 또한 저자라는 의미를 표현해 주는 메타데이터 요소로 KEM에서는 creator를 사용하고 LOM에서는 lifecycle.contribute를 사용한다. 따라서 두 개의 학습객체 메타데이터가 서로 다른 스키마를 가지고 있기 때문에 이를 공유하여 사용하기 위해서는 두 개의 서로 다른 메타데이터 스키마를 통합해 주는 새로운 Ontology 가 필요하다.

<표 2> DC, KEM, LOM 메타데이터 형식비교

요소명	DC	KEM	LOM
표제	title	title	general.title
저자	creator	creator	lifecycle.contribute
주제	subject	subject	general.keywords classification.keywords
개요	description	description	general.description
발행처	publisher	publisher	lifecycle.contribute
기여자	contributor	contributor	lifecycle.contribute
날자	date	date	lifecycle.contribute.date
자료유형	type	type	educational.learningresource type
자료형태	format	format	technical.format
접근정보	identifier	identifier	general.identifier
정보원	source	source	relation.resource
언어	language	language	general.language
내용범위	coverage	coverage	general.coverage
관련자료	relation	relation	relation.kind
교과목	-	learningArea	-
저작권	rights	right	rights.description
이용대상자	DCED.audience	audience	educational.typicalage range
주석	-	annotation	annotation.*
교수방법	GEM.pedagogy	pedagogy	-
메타데이터	-	metaMetadata	metametadata.catalogen try
학습시간	-	typicalLearning Time	educational.typicallearni ngtime
상호작용 유형	-	interActivityType	educational.interactivity type
상호작용 수준	-	interActivityLevel	educational.interactivit ylevel

3.2. 메타데이터 공유를 위한 Ontology 설계

Ontology 설계는 특정 도메인에 사용되는 지식 기반 시스템의 지식의 표현이자 방법이기 때문에 다른 어떤 소프트웨어 개발 방법론보다 신중한 접근이 필요하다[24]. 일반적인 Ontology 개발의 절차는 (그림 2)과 같다. 일반적인 개발 방법인 폭포수 모델이나 진화(evolutionary) 모델, 정형 모델보다 초기 설계 과정을 거친 정제(refine), 평가, 유지 보수 과정이 무척 중요하다. 설계된 Ontology를 평가, 유지 보수를 통해 계속 정제해 가는 과정이 필요한데, 본 연구는 초기 설계된 Ontology를 교육 전문가들과 프로토타입을 통해 평가해 보고, 이 결과를 정제 과정에 피드백하는 절차를

적용하였다.



(그림 2) Ontology 설계 방법론

현재 사용되고 있는 KEM은 <표 2>에서와 같이 DC, DCED, GEM, LOM 등의 몇 가지 메타데이터 형식을 이용하여 설계된 것이기 때문에 원활한 공유를 위해서는 관련된 모든 메타데이터 형식을 같이 표현해주는 것이 좋다. SCORM에서 사용하는 메타데이터는 IEEE의 LOM의 요소들과 비교해 보았을 때 같은 용어를 사용하고 있는 경우도 있고 전혀 다른 용어를 사용하고 있는 경우도 있다.

(그림 3)는 KEM 모델과 DC, LOM 모델 메타데이터를 공유하기 위해 필요한 Ontology를 Protégé 2000 Ontology Editor를 이용하여 설계한 클래스(class)와 하위클래스(sub-class)간의 관계를 계층적으로 나타낸 그림이다.



(그림 3) 설계된 Ontology의 클래스 계층구조

세 모델별로 각 하위 요소가 계층적으로 표현되고 있으며 동일한 클래스를 공유하는 경우에는 클래스에 M(multiple)로 표시되어 나타나있다. 예를 들어 DC 모델의 "title" 요소와 KEM 모델의 "title"은 하나의 클래스로, 이 클래스의 상위 클래스는 "DC"와 "KEM" 두 개의 클래스이다. DC에서 선언된 요소들이 KEM 모델에서 그대로 사용되는 경우가 많기 때문에 이런 경우가 비교적 많이 발생하고 있다. 하지만 "저자"를 표현

하는 메타데이터 요소는 DC와 KEM의 경우에는 같은 요소(클래스)인 "creator"를 사용하지만, LOM에서는 전혀 다른 용어인 "contribute"를 사용하고 있다.

```

<daml_owl:Class rdf:ID="typicalagerange">
  <daml_owl:sameClassAs rdf:resource="#audience"/>
  <rdfs:subClassOf rdf:resource="#educational"/>
</daml_owl:Class>
<daml_owl:Class rdf:ID="date">
  <rdfs:subClassOf rdf:resource="#DC"/>
  <rdfs:subClassOf rdf:resource="#KEM"/>
  <rdfs:subClassOf rdf:resource="#contribute"/>
</daml_owl:Class>
<daml_owl:Class rdf:ID="creator">
  <rdfs:subClassOf rdf:resource="#DC"/>
  <rdfs:subClassOf rdf:resource="#KEM"/>
</daml_owl:Class>
<daml_owl:Class rdf:ID="contribute">
  <daml_owl:sameClassAs rdf:resource="#creator"/>
  <rdfs:subClassOf rdf:resource="#lifecycle"/>
</daml_owl:Class>
<daml_owl:Class rdf:ID="title">
  <rdfs:subClassOf rdf:resource="#DC"/>
  <rdfs:subClassOf rdf:resource="#KEM"/>
  <rdfs:subClassOf rdf:resource="#general"/>
</daml_owl:Class>
  
```

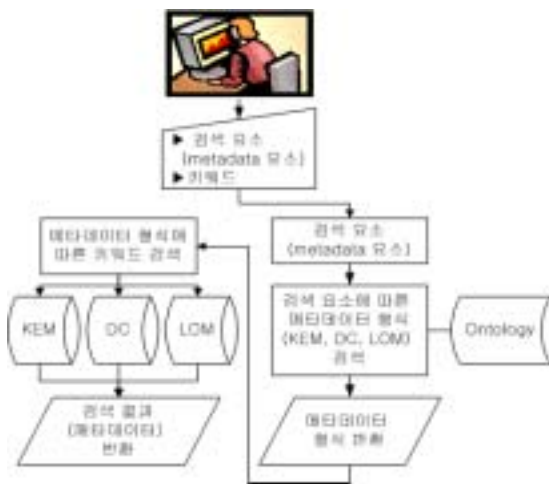
(그림 4) 설계된 Ontology의 DAML+OIL

최현중은 이런 경우에 서로 다른 용어를 사용하고 있지만 같은 의미를 가지고 있기 때문에 서로 일대일 매핑(one to one mapping)하여 같은 클래스인 것을 Ontology에 표현해 줄 수 있을 것이라고 하였다[7]. 미국과 오스트리아간의 서로 다른 컴퓨터과학 학부 부서의 Ontology 용어를 통합해주는 Ontology를 설계한 AnHai Doan의 GLUE 시스템과 비슷한 경우라고 할 수 있다[8]. GLUE 시스템의 경우 여러 다양한 용어들을 서로 유사한 정도에 따라 통합하기 때문에 용어 매핑에 있어 용어들간의 유사성(similarity)을 "Jaccard coefficient"에 의해 측정하여 매핑하였지만, 본 연구의 대상인 KEM, DC, LOM 메타데이터인 경우 각 용어들이 직접적으로 일대일 매핑되기 때문에 유사성을 측정하는 알고리즘은 필요하지 않다는 결론을 얻었다. 일대일 매핑을 위한 Ontology 설계로 본 연구에서는 DAML+OIL Ontology 언어를 적용하였는데, (그림 4)는 그 일부분이다. DAML+OIL은 클래스 표현법 중의 하나인 "sameClassAs" 요소를 적용하면 일대일 매핑을 표현할 수 있고 다양한 속성, 데이터 형을 정의할 수 있기 때문에 ontology 표현과 처리에 다양하게 사용된다. "creator" 메타데이터 요소인 경우 "DC", "KEM"의 하위 클래스인 동시에 "contribute" 클래스와 서로 동일한 개념임을 "sameClassAs" 요소로 나타내 주었다. 설계된 초기 Ontology는 지식 시스템 전문가와 교육 전문가의 도움을 받아 설계된 논리적 모델을 정제하였다.

4. 시스템의 설계 및 구현

4.1. 시스템의 개요 및 설계

본 연구에서 검색 대상이 되는 메타데이터 형식은 DC, KEM, LOM 세 가지로 이 세 메타데이터 형식은 모두 국내외로 많이 사용되는 메타데이터 형식이다. KEM은 국내의 실정에 맞게 수정된 메타데이터 형식으로 국내에서만 사용되지만, 에듀넷과 국내 16개 시도 교육청에서 개발한 자료는 모두 이 메타데이터 형식을 따르고 있기 때문에 실질적인 검색을 위해서는 이 메타데이터 형식이 반드시 포함되어야 하며 DC 및 LOM은 국내외로 학습 객체를 연구하는 많은 E-Learning 시스템에서 채택하고 있는 메타데이터 형식이다.



(그림 5) Ontology를 이용한 검색 과정

세 가지 메타데이터 형식을 이용하여 서로 다르게 저장된 메타데이터를 검색하기 위해서는 각각의 메타데이터 형식과 요소를 통합해 주는 통합 스키마(schema)인 Ontology가 필요하다. 본 연구에서 설계된 Ontology는 세 메타데이터 형식의 각 요소(element, 클래스)들에서 동의어에 속하는 요소들을 추출하여 이를 동의어 클래스(sameClassAs)로 선언하고, 동일한 상위 클래스를 함께 가지고 있는 클래스의 경우에는 동일 클래스를 상속받고 있는 형태(subClassOf)로 선언해 주었다.

이렇게 설계된 Ontology를 이용하여 이종(heterogeneous) 메타데이터를 검색하는 과정은 (그림 5)와 같고 전체 시스템의 구조는 (그림 6)와 같다. 본 시스템의 검색 과정은 아래와 같다.

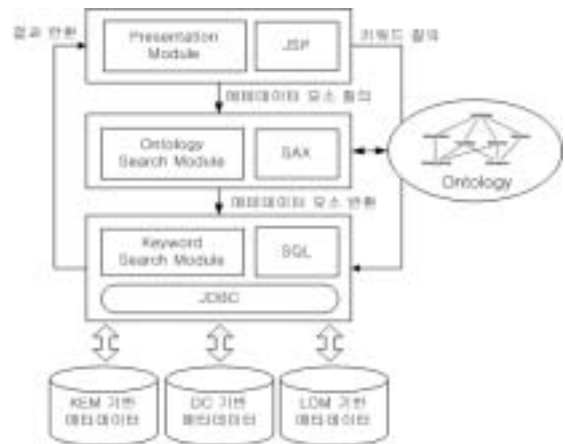
- ① 사용자는 검색요소(주제, 제작자, 파일형태 등)와 검색요소에서 검색하고 싶은 키워드를 입력한다.
- ② 사용자가 입력한 검색요소는 Ontology 검색 모듈에서 세 가지 메타데이터 형식에 맞는 메타데이터요소들을 추출한다.
- ③ 추출된 세 가지 메타데이터 형식에 맞는 메타데

이터 요소가 검색 모듈로 반환된다.

④ 반환된 메타데이터 요소를 이용하여 각각의 메타데이터가 저장된 데이터베이스에서 메타데이터를 추출한다.

⑤ 추출된 메타데이터 결과를 사용자에게 반환한다.

본 시스템에서 제일 중요한 모듈은 Ontology 검색 모듈로 사용자가 검색하고 싶은 요소를 세 메타데이터 형식에 맞게 적절하게 매핑하는 역할을 한다. 예를 들어 사용자가 검색 요소로 “저자”를 선택했다면 Ontology 검색 모듈은 DC와 KEM의 경우 검색 메타데이터 요소로 “creator”를 반환해 주어야 하지만, LOM의 경우에는 “lifecycle.contribute”를 반환해 주어야 한다. 또한 “교과목”을 검색요소로 선택하였다면 교과목에 대한 메타데이터 형식은 KEM의 “learningArea” 요소만 선택되어 반환되어야 한다.



(그림 6) Ontology를 이용한 검색 시스템 구조

4.2. 시스템의 개발 환경 및 구현

시스템의 개발 환경은 <표 3>과 같다.

<표 3> 시스템의 개발 환경

Hardware	Client	IBM Compatible PC
	Web Server	IBM Netfinity 3500
	DB Server	IBM xSeries 235 SUN UltraBlade 3000
Software	Client Application	MS Explorer 6(sp1)
	Web Server	Tomcat 4.0
	Programing Lang.	Java 1.4.2/JSP JDOM
	Data Middleware	JDBC(Type 4)
	DBMS	MS SQL Server 2000(sp2) Oracle 9i(release 2)

사용자 인터페이스는 웹 브라우저를 통해 JSP와 JavaBean 로 구성된 Ontology 검색 모듈과 키워드 검색 모듈로 검색된 결과를 확인해 볼 수 있다. 웹 서버는 Java와 JSP 연동이 가능한 Tomcat 4.0을 사용하였

고, 데이터베이스에 접속하여 키워드 검색을 실시하는 키워드 검색 모듈은 JDBC(Type 4)를 사용하였다. 데이터베이스에 저장된 세 메타데이터 형식은 모두 테이블 형태로 저장하여 SQL을 이용하여 추출하였으며 대상 데이터베이스는 Oracle 9i와 MS SQL Server 2000을 사용하였다. DAML+OIL Ontology에 저장된 메타데이터 요소를 질의하는 언어로는 SAX를 사용하여 처리하였다.



(그림 7) 구현된 검색 시스템

DAML+OIL의 경우 표현된 Ontology가 루트요소의 자식요소(child element)에 모든 메타데이터가 클래스로 정의되어 있고, 정의된 클래스의 속성들은 모두 그 클래스의 자식요소로 정의되어 있기 때문에 사용자가 원하는 요소와 속성을 추출하는데, SAX/DOM을 사용하여 쉽게 처리할 수 있다는 장점이 있다. 따라서 본 시스템의 경우 JDOM의 SAX를 이용하여 Ontology에 정의된 요소와 속성을 처리하여 그 결과를 검색 모듈에게 반환해 주었다. (그림 7)는 사용자가 자료유형에서 "애니메이션"을 검색했을 때 이에 해당하는 DC, KEM 메타데이터를 검색한 결과이다. 프로토타입을 적용해 평가한 결과 서로 다른 이종 메타데이터를 통합하여 잘 검색할 수 있었다.

4. 결론

본 연구는 국내에서 가장 많이 사용되는 교육자료 메타데이터 모델인 KEM과 DC, LOM을 서로 통합하여 검색할 수 있는 시스템을 설계 및 구현해 보았다. 서로 다른 학습객체 메타데이터를 통합하여 검색하기 위해서는 각각의 메타데이터에서 서로 다르게 정의된 메타데이터 요소를 올바르게 매핑해 줄 수 있는 통합 스키마로서의 Ontology 설계가 필요하다. 따라서 본 연구는 세 메타데이터 형식을 비교하여 서로 다른 요소 이름으로 정의된 것들을 일대일 매핑하여 세 메타데이터 형식을 통합해 주는 Ontology를 설계해 보았고, 이를 이용하여 서로 다르게 저장된 메타데이터를 검색해 보는 프로토타입을 구현해 보았다.

세 메타데이터를 통합하는 Ontology의 경우 KEM와 LOM 두 모델의 경우 DC를 기반으로 제작되었기 때문에 표제, 저자, 주제, 개요 등의 일반적인 메타데이터 항목을 서로 비슷한 경우가 많았지만, KEM은 국내의 교육 현실을 감안하여 학교급, 교과목, 학기 등의 정보를 담고 있는 learningArea 요소를 추가하였으며, LOM에는 존재하지 않지만 GEM 메타데이터에서 사용하는 pedagogy 요소가 추가되었다는 것을 확인할 수 있었다. 이 몇 가지 요소만을 제외한다면 KEM, DC, LOM의 요소들과 무리없이 일대일 매핑이 가능하였다. 하지만 본 연구의 Ontology는 세 메타데이터간의 단순한 공유를 위해 설계되었기 때문에 아직 추론을 위한 설계는 고려되지 않았다. 따라서 사용자들에게 검색 결과를 다양하게 제공해 주려면-예를 들어, 동의, 반의, 계층, 연관 등-본 Ontology에 좀 더 다양한 속성들(properties)이 추가되어야 할 것이다.

또한 본 연구에서 사용된 DAML+OIL 언어 이외에 최근 W3C에서 표준화된 OWL 웹 Ontology언어를 이용한 설계 및 구현도 추가로 이루어져 그 장단점을 비교해 보아야 할 것이다. 더불어 본 연구에서 사용된 DAML+OIL의 경우 Ontology 표현이 루트요소의 자식요소와 속성에서 대부분 이루어졌기 때문에 SAX를 통해 Ontology 처리를 했지만, 대부분의 XML 형태의 자료는 요소와 속성을 비교적 자유롭게 표현할 수 있기 때문에 특별한 경우 이를 고려한 처리도 고려해야 한다.

참고 문헌

- [1] 고세영, 토픽맵을 이용한 이기종 상품분류체계 Ontology 통합에 관한 연구, 숙명여자대학교 석사학위논문, 2003
- [2] 김종석, 온톨로지를 사용하는 전자책 어노테이션 시스템의 설계 및 구현, 연세대학교 석사학위논문, 2002
- [3] 송민선, 문헌정보학 분야의 강의자료 및 연구업적에 관한 Ontology 의미 정보에 대한 연구, 성균관대학교 석사학위논문, 2003
- [4] 이현실, Ontology기반 한의학 처방 지식관리시스템 설계에 관한 연구, 중앙대학교 석사학위논문, 2003
- [5] 정도현, 시소러스를 기반으로 한 온톨로지 시스템 구현에 관한 연구, 연세대학교 석사학위논문, 2003
- [6] 조진형, Ontology Inference Layer 및 XML 기반의 설계지식 관리, 명지대학교 석사학위논문, 2001
- [7] 최현중 외, 이종메타데이터 공유를 위한 Ontology 설계, 2004 한국컴퓨터교육학회 동계학술발표논문집 8(1), pp. 253-257, 2004
- [8] AnHai Doan et al., Learning to Map between Ontologies On the Semantic Web, The VLDB

Journal 12(4), pp.303-319, 2003

[9] A. Naedche, B. Motik, L. Stojanovic, Managing multiple and distributed ontologies on the Semantic Web, The VLDB Journal 12, pp. 286-302, 2003

[10] DAML+OIL (March 2001) Reference Description [online] available: <http://www.w3.org/TR/daml+oil-reference>

[11] David A. Wiley, II. Learning object design and sequencing theory. Brigham Young Univ., 2000.

[12] Deiter Fensel et al. OIL: An ontology infrastructure for the semantic web, IEEE Intelligent Systems 16(2), pp. 38-44, 2001

[13] Ljiljana Stojanovic, Nenad Stojanovic, Raphael Volz, Migrating data-intensive Web Sites into the Semantic Web, Proceedings of the ACM Symposium on Applied Computing, pp. 1100-1107, 2002

[14] Natalya F. Noy and Deborah L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, pp. 1-2, 2001

[15] Nicola Guarino, Formal Ontology and Information Systems, Proceedings of the 1st International Conference on Formal Ontologies in Information Systems, pp. 3-15, 1998

[16] Mike Uschold and Michael Gruninger. Ontologies: Principles, Methods and Applications, Knowledge Engineering Review 11(2), pp. 93-155, 1996

[17] Ontology Inference Layer [online] available: <http://www.ontoknowledge.org/oil>

[18] R. Guha, Rob McCool, Eric Miller, Semantic Search, Proceedings of the twelfth international conference on World Wide Web, pp. 700-709, 2003

[19] Resource Description Framework [online] available: <http://www.w3.org/RDF>

[20] Riichiro Mizoguchi, Ontological Engineering: Foundation of the next generation knowledge processing, N.Zhong et al.(Eds.) WI2001, LNAI2198, Springer-Verlag, pp.44-57, 2001

[21] The DARPA Agent Markup Language [online] available: <http://www.daml.org>

[22] Tim Berners-Lee, Semantic Web-XML 2000 [online] available: <http://www.w3.org/2000/Talks/1206-xml2k-tbl>

[23] Tom Gruber, What is an Ontology? [online] available: <http://www.w3.org/2001/sw/WebOnt>

<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

[24] York Sure and Rudi Studer, A Methodology for Ontology-based Knowledge Management, In John Davies et al., Towards The Semantic Web. Wiley, pp. 33-46, 2003.

[25] W3C Semantic Web [online] available: <http://www.w3.org/2001/sw/>

[26] Web Ontology Working Group [online] available: <http://www.w3.org/2001/sw/WebOnt>

저자 소개

최 현 종



1993년 : 공주교육대학교 졸업

2001년 : 한국교원대학교

컴퓨터교육과 교육학석사

2002년 ~ : 한국교원대학교

컴퓨터교육과 박사과정

관심분야: 컴퓨터교육, Semantic

Web, 분산 객체

E-Mail: blueland@blue.knue.ac.kr

김 태 영



1985년 : 한양대학교 졸업

1990년 : Texas A&M

Univ. 공학석사

1994년 : 1994 Texas A&M

Univ. 공학박사

1994년 ~ : 한국교원대학교

컴퓨터교육과 부교수

관심분야: 데이터베이스, 컴퓨터교육, 데이터통신

E-Mail: tykim@knue.ac.kr