

온톨로지를 이용한 교육자료 관리 기능의 개선

장병철^{○*} 이재혁^{**}

* 한양대학교 대학원 정보통신학과, **한양대학교 정보통신학부

{bcchang[○], chajh}@hanyang.ac.kr

Enhancing the Education Resource management with ontology

Byoungchol Chang^{○*} Jaehyuk Cha^{**}

* Graduate School of Information & Communications, Hanyang UNIV.

**College of Information & Communications of Hanyang UNIV.

요 약

본 연구에서는 한국교육학술정보원의 교육 콘텐츠 관리 시스템인 에듀넷의 키워드 기반 콘텐츠 검색기능을 고도화하기 위해 온톨로지를 활용한 검색 프레임워크를 설계, 구현하였다. 에듀넷의 키워드 기반 검색 시스템에서 사용하는 KEM(Korea Education Metadata)을 owl을 이용하여 온톨로지로 바인딩하였으며, 의미 기반 교육 콘텐츠 검색이 가능하도록 중학교 수학의 일부 영역을 도메인 온톨로지로 구축하였다. 구축한 온톨로지에 실제 에듀넷에서 사용하는 콘텐츠 정보를 이용하여 인스턴스를 생성하였다. 사용자의 쿼리를 입력할 수 있는 인터페이스와 쿼리를 처리하고 추론할 수 있는 추론엔진을 사용하여 본 연구의 검색 시스템을 구축하였다. 실험을 통하여 본 연구에서 구축한 시스템이 키워드 매칭을 통한 검색 보다 사용자에게 의미 있고 유용한 결과를 도출함을 보였다.

1. 서 론

사이버 대학, 유비쿼터스 러닝 등 e-러닝 기술이 점점 더 발전하고 e-러닝이 광범위하게 활용되면서 e-러닝 콘텐츠에 대한 활용도 증가되고 있다. 따라서 e-러닝 콘텐츠를 체계적으로 관리, 검색, 활용하는 시스템의 필요성도 대두 되었다.

현재 e-러닝 콘텐츠는 각 기관, 단체별로 제공하고 있으며 대부분 사용자가 원하는 콘텐츠를 검색하기 위해서 키워드 기반의 검색 기능을 제공하고 있다.

초, 중, 고등학교용 e-러닝 콘텐츠를 가장 많이 보유하고 있는 국가 수준의 시스템은 한국교육학술정보원의 에듀넷(Edunet)[1] 이다. 이 시스템은 다른 정보 검색 시스템과 비슷하게 키워드를 기반으로 한 검색 기능을 사용자에게 제공하고 있으며, KEM(Korea Education Metadata)[2]이라는 메타데이터 구조를 이용하여 학습 콘텐츠에 대한 체계적인 메타정보를 기록하고 있다. 그리고 이 메타정보를 이용하여 통합 키워드 검색에 추가하여 학교급, 자료유형, 등록일, 자료의 관리기관에 따라 세부검색을 할 수 있는 기능을 제공하고 있다.

본 연구에서 실시한 사용자 요구 조사에 따르면 에듀넷 등을 이용하여 콘텐츠를 검색하는 사용자들은 검색 결

과 및 검색 방법에 대하여, '원하지 않는 결과가 도출된다', '너무 많은 결과가 나온다', '필요없는 데이터가 너무 많다' 정확한 검색을 하려면 비슷한 단어라고 판단되는 단어를 모두 입력하여야 한다', '원하는 데이터가 너무 없다', '검색 결과로 도출된 콘텐츠가 원하는 것과 다른 내용을 담고 있다' 등의 불편함을 표현하였다.

또한 본 연구에서 실시한 현재의 키워드 검색 기반에서 사용자들이 입력한 질의 유형 분석과 사용자에 대한 심층 면접의 결과를 보면 교사, 학생이 e-러닝 콘텐츠를 검색할 때는 지식의 구조를 이용하거나 콘텐츠 서로간의 상관관계가 있는 것을 검색하기를 원하는 경우가 많다. 그러나 현재 에듀넷을 비롯하여 여러 콘텐츠 검색 업체가 제공하는 키워드 기반 검색 시스템에서는 지식의 구조를 검색하거나 콘텐츠의 상관관계를 검색하는 것이 어렵다. 예를 들어, 중학교 2학년 수학의 도형 분야에서 평행사변형의 성질에 대하여 공부하려고 이와 관련된 콘텐츠를 검색하려고 키워드 검색 시스템에서 평행사변형이라고 입력하고 검색하면 기본적으로 '평행사변형'이라는 키워드에 매칭되는 콘텐츠만 검색이 되지만 지식의 구조 측면에서는 평행사변형의 관련 요소인 대각, 엇각 등의 개념을 설명하는 콘텐츠도 검색이 되어야 학습자가 체계적이고 효과적인 학습을 할 수 있다.

본 연구에서는 에듀넷을 이용하여 학습 콘텐츠를 검색하는 사용자에게 사용자가 원하는 정확한 콘텐츠 검색 결과를 제공하여 교육 콘텐츠 재활용과 콘텐츠 공유기능의 개선을 위해 온톨로지를 활용한 콘텐츠 검색시스템을 설계하고 구현하였다. 학습 콘텐츠를 위한 온톨로지는 먼저 현재의 키워드 기반에서 사용하는 KEM(Korea Education Metadata)을 OWL[4]을 이용한 온톨로지로 바인딩하고 이

* 본 연구는 한국교육학술정보원의 'e-러닝 표준화 관련 연구 사업' 및 서울시의 '서울형 미래 도시산업 육성 지원사업'의 지원을 받아 수행되었음.

온톨로지를 KEM 온톨로지로 정의 하였다. 또한 사용자에게 지식의 구조에 따른 체계적인 검색 결과를 제공하기 위하여 중학교 수학의 도형영역의 일부 개념을 온톨로지화하여 검색 시스템에서 활용하였다. 본 연구에서는 구축된 온톨로지의 추론을 위해서 별도의 추론 엔진은 제작하지 않았으며 기존에 개발된 OWL 추론 엔진인 KAON2[5], Bossam[6] 등을 이용하여 온톨로지를 이용한 검색 프레임워크를 제시, 구축하였다.

2. 관련연구

2.1 KEM(Korea Education Metadata)

메타데이터란 “데이터에 관한 데이터”로써 대상이 되는 자원의 속성과 특성 및 다른 자원과의 관계를 기술하여 이용자의 검색을 돕고 관련기관의 관점에서 자원의 제어와 관리를 돕는 역할을 하는 “데이터에 관한 구조화된 데이터”이다[2]. 이는 자원과 독립적으로 존재하면서 자원에 대한 다양한 접근점과 네트워크 주소를 포함하고 있는 기록을 의미한다.

KEM은 더블린 코어(Dublin Core)[6] 메타데이터의 15개 요소를 수용하고, LOM이나 GEM 등 국제적인 표준으로 성격을 갖는 교육관련 메타데이터 형식을 종합 검토하여 교육 자료에 공통으로 적용되는 기술요소들을 수용하였다.

한국 교육 자료에 적용된 기존의 형식과 교육자료의 분석을 통해 한국 교육실정에 적합한 기술요소들을 선별적으로 적용하고, 추가로 필요한 기술 요소를 새롭게 정의하였다. KEM은 IEEE에서 2002년 7월에 발표한 LOM 1.0을 상당부분 수용하였으며, 더블린 코어(Dublin Core) 메타데이터와의 호환성을 고려하여 개발되었다. KEM 3.0은 115개 메타데이터 요소로 구성되며, XML로 바인딩하기 위해 XML DTD 구조를 사용하지 않고 XML 스키마를 사용하였다. KEM의 상위 9개 요소는 그림 1에서 보는 바와 같으며 이 상위 요소에 다시 여러 개의 하위 요소들로 구성되어 있다.

2.2 OWL(Web Ontology Language)

시맨틱 웹에서 정보는 명시적인 의미를 부여받게 되는데, 이를 통해 기계는 좀더 쉽게 웹 상에 존재하는 정보들을 자동으로 처리하고 통합할 수 있다. 시맨틱 웹은 사용자 정의 태그 스키마를 정의할 수 있는 XML과 유연하게 데이터를 표현할 수 있는 RDF를 바탕으로 구축된다. 시맨틱 웹의 구현에 있어 RDF 바로 윗 계층에 필요한 것이 웹 문서에 포함된 용어의 의미를 형식적으로 기술할 수 있는 온톨로지 언어이다.

기계를 이용하여 웹 문서를 대상으로 유용한 추론 기능을 수행하려면 RDF 스키마가 제공하는 기초적인 의미 표

현력을 뛰어넘는 언어가 필요하다. 이런 요구에 의해 등장한 언어가 OWL이다.

ID	상위 요소	설명
1	일반사항 범주 (General category)	자원을 설명하는 일반적 정보들을 포함한다.
2	생명주기 범주 (Lifecycle category)	자원의 히스토리, 현재의 상태, 발전에 영향을 미친 사람 등과 관련된 정보들을 포함한다.
3	메타-메타데이터 범주 (Meta-metadata category)	지정된 기록이 설명하고 있는 자원에 대한 정보가 아닌 메타데이터 기록 자체에 대한 정보들을 포함한다.
4	기술 범주 (Technical category)	자원의 기술적인 필요조건들과 특성들에 관련된 정보들을 포함한다.
5	교육 범주 (Educational category)	자원이 갖고 있는 교육적 특징과 교수법상의 특징들에 대한 정보들을 포함한다.
6	저작권 범주 (Rights category)	자원과 관련된 지식소유권 및 사용권에 대한 정보들을 포함한다.
7	관계 범주 (Relation category)	사용하고 있는 자원과 다른 목표 자원과의 관계를 정의하고 있는 정보들을 포함한다.
8	주석 범주 (Annotation category)	자원의 교육적 사용에 대한 주석과 연해, 추가 주석을 달았는지에 대한 정보들을 포함한다.
9	분류 범주 (Classification category)	자원이 특별한 분류시스템에서 어디에 속하는지에 대한 정보들을 포함한다.

그림 1 KEM 최상위 구성 요소[2]

OWL은 OWL-Lite, OWL-DL, OWL-FULL과 같은 3 가지 종류로 세분화되는데, 각각 지원하는 표현력이 차이가 난다. OWL-Lite가 가장 표현력이 약하며 OWL-Full이 가장 강력한 표현력을 가진다. 그러나 OWL-Full은 풍부한 표현을 장점으로 가지는 반면 이를 이용하여 표현된 온톨로지를 추론하는 경우 연산시간이 너무 많이 요소되거나 추론 불가능한 상황에 빠질 수 있다. 현재 OWL언어를 지원하는 대부분의 추론 엔진들은 OWL-DL 수준까지의 표현력을 추론할 수 있는 기능을 제공하고 있다. 본 연구에서도 온톨로지를 구축하기 위하여 현재 추론 시스템들을 활용할 수 있는 OWL-DL을 기본 온톨로지 언어로 사용하였다.

3. 설계

3.1 KEM의 OWL 바인딩

본 연구에서는 에듀넷의 검색 서비스를 개선하기 위하여 에듀넷에서 사용하는 KEM을 OWL-DL을 이용한 온톨로지 바인딩하였다. KEM 온톨로지를 구축하면서 사용한 온톨로지 엔지니어링 방법론은 기본적으로 OTK(On-To-Knowledge)[8] 방법론을 이용하였다. 그러나 본 연구에서는 다른 온톨로지 구축 상황과는 달리 기본적인 개념이 없는 상황에서 시작한 것이 아니라 KEM이라는 기본 개념에서 온톨로지 구축을 시작하는 것이므로 개념 도출과정은 생략되었다. 기존 연구들에 의하면 온톨로지의 구축 목표는 크게 두 가지로 나뉘어진다. 첫째는 해당 도메인의 지식전달 둘째는 온톨로지를 이용한 서비스 기능의 확장이다. KEM은 기본적으로 교육 콘텐츠의 의미에 대하여 설명하는 것이 아니며, 본 연구의 목적 또한 교육 콘텐츠의 의미적 정의를 세우려는 것이 아니므로 KEM을 온톨로지 바인딩하면서 세운 기본 전

제는 KEM 온톨로지를 이용한 서비스 기능 확장 즉, 검색 기능 확장에 초점을 두는 온톨로지를 구성한다는 것이다. 이런 대전제 하에 세부적으로 온톨로지를 구축하면서 세운 규칙들의 일부는 다음과 같다.

- Container로써 지칭되는 General, LifeCycle 등의 정보 모델은 Ontology의 설계 및 구축에 있어서 개념적인 부분이므로 검색에 활용되지 않는 경우 생략하였다.
- KEM 요소중에서 독립적으로 존재 가능한 개념 (concept)인 경우 class로 구축하고 그렇지 않은 경우는 property로 구축하였다. 예를 들어 KEM 요소중 '기관', 각종 '분류체계' 등의 요소는 독립적인 개념이므로 class로 구성하고 '제목' 요소는 개념상 종속적 개념이므로 property로 처리하였다.
- KEM 요소를 분석하여 요소간의 상하위 개념이 있는 경우 subClassOf, subPropertyOf를 이용하여 class 및 property의 개층 구조를 형성하였다. 예를 들어 KEM 요소중에 '제목'과 '부제목'이 있는데 이는 'title' property와 그것의 하위 property인 'maintitle'과 'subtitle'로 구성하였다.
- 'Class 명명은 첫 문자를 대문자로 한다.' 'Property의 명명은 ObjectProperty 인 경우 has~ 로 시작하며, DatatypeProperty인 경우 첫문자를 소문자로 한다' 등의 명명 규칙을 활용하였다.
- KEM 요소에는 각종 number restriction이 존재한다. 이러한 restriction들은 의미의 변형이 없는 한 사용하지 않았다. 그 이유는 추론엔진이 모든 종류의 restriction을 지원하지 않는 경우가 많으며 restriction을 주지 않아도 의미에 변화가 크게 없기 때문이다.

3.2 도메인 온톨로지

전술한 KEM 온톨로지는 KEM을 owl을 이용하여 온톨로지 로 바인딩 한 것이다. 이는 콘텐츠의 외형적 정보를 가지고는 있지만 콘텐츠 내용의 의미적 정보는 가지고 있지 않다. 서론에서 전술한 바와 같이 학습을 위하여 교육용 콘텐츠를 검색할 때는 해당 교과서의 의미 구조가 중요한 기능을 한다. 궁극적으로는 초, 중등분야 모든 교과서의 시맨틱 구조가 온톨로지 로 구축되어 활용되어야 하겠지만 이는 쉽지 않은 일이며 앞으로 많은 노력이 필요한 일이다. 본 연구에서는 의미 구조의 활용하면 검색 기능을 좀 더 고도화 할 수 있음을 보이기 위하여 중학교 수학 교과서의 도형 분야의 의미 구조를 교육 인적 자원부 교과 분류 체제를 참조하여 수학 교육 전문가와 협력하여 구축하였다.

3.3 온톨로지의 통합(Merge)

3.2절에서 언급한 도메인 온톨로지 와 3.1절에서 언급한 KEM 온톨로지는 별도의 namespace를 이용하여 통합하는 작업을 수행 하였다. 본 연구에서와 같이 도메인 온톨

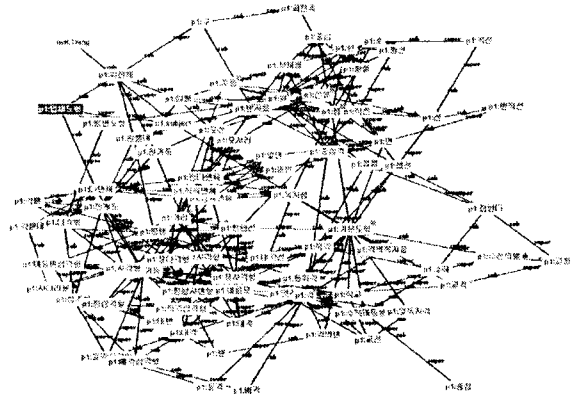


그림 2 도형 영역 도메인 온톨로지 일부

로지의 크기가 크지 않은 경우에는 온톨로지를 개발하는 초기 단계에서 도메인 온톨로지 와 KEM 온톨로지를 통합하여 개발할 수도 있으나 궁극적으로는 각종 도메인 온톨로지들과 KEM 온톨로지가 자연스러운 방식으로 통합되어야 하므로 namespace를 이용하는 통합 방식을 취하였다. 도메인 온톨로지 와 KEM 온톨로지 내에서 의미적으로 충돌이 발생할 수 있는 class와 property들은 서로의 관계를 owl:SameAs, owl:equivalentClasses 등의 owl property를 이용하여 관계를 정의하여 주었다.

3.4 온톨로지 활용 검색 시스템 프레임 워크

owl을 이용하여 구축한 온톨로지는 추론 엔진과 온톨로지 쿼리 언어를 사용하여 검색 할 수 있도록 하였다. 검색 시스템 프레임워크의 구조는 최하부에는 온톨로지가 위치하고 그 온톨로지를 추론 시스템이 읽어들이고, 추론하며 쿼리 인터페이스를 통하여 입력된 사용자 쿼리를 쿼리처리를 통하여 추론 시스템에 전달하면 추론 시스템은 사용자의 쿼리에 따라 추론한 결과를 사용자에게 돌려주는 형태로 되어 있다. 그림 3은 본 연구의 온톨로지 활용 검색 시스템 프레임 워크이다.

4. 구현 및 실험

본 연구의 KEM 온톨로지 와 도메인 온톨로지를 통합한 통합 온톨로지를 활용한 검색 시스템 다음과 같은 툴들을 이용하여 구현하였다.

- 온톨로지 구현 언어 : OWL-DL
- 온톨로지 에디터 : Protege 3.2[8]
- 온톨로지 추론 엔진 : KAON2, Bossam
- 온톨로지 질의 언어 : SPARQL, Buchingae

사용자 인터페이스는 JAVA Beans를 이용하였으며

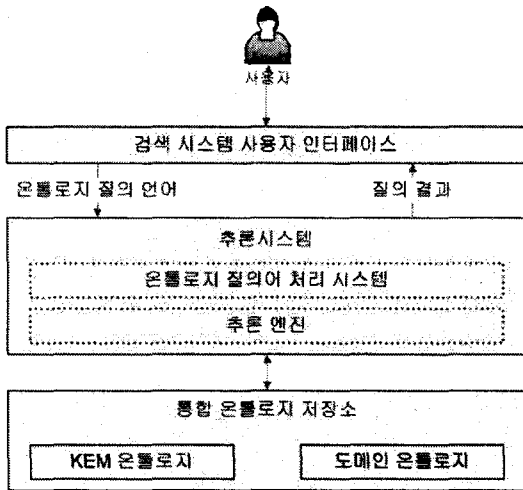


그림 3 온톨로지 활용 검색 시스템 프레임 워크

본 연구의 검색 시스템은 Pentium4 2.4Ghz, 1MB Main Memory의 linux 2.6 시스템에 구축 되었다.

기존의 키워드 기반 시스템의 사용자 인터페이스는 단순 키워드만 입력하면 검색 할 수 있는 구조이다. 그러나 온톨로지 검색 시스템은 단일 키워드 만으로는 검색하기가 어렵다. 따라서 본연구의 사용자 인터페이스는 기본적으로 SPARQL을 이용한 질의를 입력하는 것으로 하고 이에 익숙하지 않은 사용자를 위해 웹상에서 펼침 메뉴를 통한 항목별 입력을 통하여 쿼리를 생성하는 서비스를 웹상에 구축하였다. 단, 이 경우 SPARQL을 직접 입력하는 것 보다 표현력에 있어서 제한을 가진다.

실험을 위하여 한국교육학술정보원의 협조를 얻어 중학교 수학 콘텐츠 2,300여개와 그것의 KEM 정보를 확보 하였다. 그 다음 수학 전문가를 통하여 해당 콘텐츠가 도메인 온톨로지의 어느 class에 해당되는지 분류하고 해당 클래스의 인스턴스로 콘텐츠의 정보를 입력하였다.

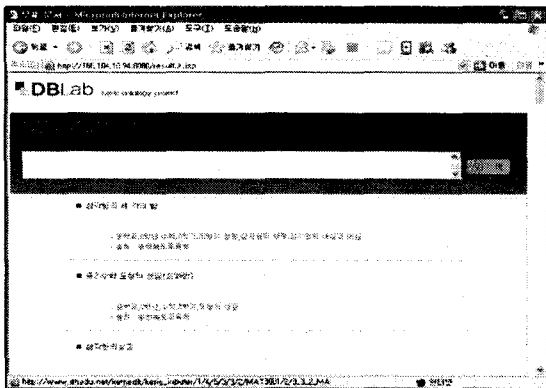


그림 4 사용자 질의 처리 결과

그림 4는 본 연구의 시스템을 실험한 결과로 사용자가 입력한 질의에 따라 결과를 보여주는 화면이다. 본 연구에서 개발한 시스템과 기존의 에듀넷 콘텐츠 검색 시스템을 비교하면, 기존의 시스템에서 등변사다리꼴에 대하여 학습을 할 목적으로 '등변사다리꼴'이라는 키워드를 입력하는 경우 에듀넷 검색 시스템에서는 단 하나의 결과를 리턴하는 반면 본 연구의 시스템에서는 등변사다리꼴의 기본 원리인 '대각', '대변'에 관한 콘텐츠도 검색 되었다.

5. 결론

본 연구에서는 한국교육학술정보원의 학습 콘텐츠 관리 시스템인 에듀넷 검색 시스템의 성능을 고도화하기 위하여 에듀넷 검색 시스템이 사용하는 KEM 온톨로지 기반인딩하였으며 시맨틱을 이용하여 검색 성능을 높이고자 실험적인 도메인 온톨로지를 구축하였다. 또한 이 온톨로지에 질의를 주고 추론하여 정보를 도출할 수 있는 프레임 워크를 구축하였다. 본 연구의 실험 결과에 의하면 본 연구의 시스템을 이용한 검색통하여 사용자는 기존의 키워드 검색 시스템보다 좀 더 만족스러운 결과를 얻을 수 있었다.

추후 풍부한 표현력을 가지면서도 사용자가 키워드 기반 검색 만큼 편리하게 질의를 입력할 수 있는 사용자 인터페이스에 대한 연구를 수행 할 예정이며, 온톨로지가 복잡해지고 인스턴스의 수가 많아 질 수록 추론의 성능이 떨어지는 문제를 해결하기 위하여 추론엔진의 성능 개선에 대한 연구를 지속적으로 수행할 예정이다.

참고문헌

- [1] 에듀넷, <http://edunet4u.net>
- [2] 손진근 외, "고등교육정보 및 저작, 관리를 위한 메타데이터(KEM v3.0)연구", 한국교육학술정보원 연구보고서KR-2005-27, 2005.
- [3] 서대원 외 (2003). 교수·학습 자료 분류 체계화 연구. 한국교육학술정보원 연구보고서
- [4] Web Ontology Language, <http://www.w3.org/2004/OWL>
- [5] KAON2, <http://kaon2.semanticweb.org>
- [6] Bossam, <http://mknows.etri.re.kr/bossam/documents>
- [7] Dublin Core, <http://dublincore.org>
- [8] York Sure, "On To Knowledge Technicla Fact Sheet for the OTK Tool Suite", Institute AIFB, 2002.
- [9] protege, <http://protege.stanford.edu>
- [10] Sure, Y., & Studer, R., "A methodology for ontology-based knowledge management", *Towards The Semantic Web (33-46)*. Wile.