

# 온톨로지 확장을 이용한 교육자료 관리 기능의 개선

최한웅<sup>0</sup>, 도현오, 이윤수, 강현상, 차재혁  
 한양대학교 전자컴퓨터통신공학과

{hw23, lunar0m, eclipse, coreafirst, [chajh](mailto:chajh@hanyang.ac.kr)}@hanyang.ac.kr

## Enhancing the Education Resource management with ontology population

Hanwoong Choi<sup>0</sup>, Hyunoh Doh, Yoonsoo Lee, Hyunsang Kang, Jaehyuk Cha  
 Dept. of Electronics and Computer Engineering, Hanyang Univ.

### 요 약

본 연구에서는 한국교육학술정보원의 교육 콘텐츠 관리 시스템인 에듀넷의 키워드 기반 콘텐츠 검색 기능을 고도화하기 위해 온톨로지 확장을 활용한 검색 프레임워크를 설계, 구현하였다. 의미 기반 교육 콘텐츠 검색이 가능하도록 중학교 수학의 일부 영역을 도메인 온톨로지로 구축하였으며 콘텐츠와 온톨로지의 클래스 이름간의 유사도를 계산하여 자동으로 인스턴스로 추가시키는 시스템을 설계, 구현하였다. 도메인 온톨로지를 확장하여 풍성해진 온톨로지를 활용하여 콘텐츠를 검색할 수 있는 시스템을 구축하였다. 실험을 통하여 본 연구에서 구축한 시스템이 키워드 매칭을 통한 검색 보다 사용자에게 의미 있고 유용한 결과를 도출함을 보였다.

### 1. 서 론

최근의 이러닝 기술 동향을 살펴보면 협력, 에듀테인먼트, 맞춤형 학습과 교육적 3D 콘텐츠 가시화 기술, 교육용 콘텐츠 적응화 기술 등이 개발되고 있다. [1] 위와 같이 이러닝 기술이 점점 더 발전하고 여러 분야에 광범위하게 활용되면서 이러닝 콘텐츠의 종류가 점점 더 증가하고 있다. 이에 따라 이러닝 콘텐츠들을 체계적으로 관리, 검색, 활용하는 시스템의 필요성이 커졌다.

현재 이러닝 콘텐츠를 검색하는 시스템의 경우 대부분 사용자가 원하는 키워드 중심의 검색 기능을 제공 하고 있다. 초, 중, 고등학교용 이러닝 콘텐츠를 가장 많이 보유하고 있는 국가 수준의 시스템인 한국교육학술정보원의 에듀넷(Edunet)[2] 역시 키워드를 기반으로 한 검색 서비스를 제공하고 있다. 이 시스템은 KEM(Korea Education Metadata)[3]이라는 메타데이터 구조를 이용하여 학습 콘텐츠의 체계적인 메타 정보를 기록하고 있는데 이 메타 정보를 활용하여 통합 키워드 검색에 추가하여 세부 검색을 할 수 있는 기능을 추가적으로 제공하고 있다. 그림 1은 에듀넷의 서비스 화면을 나타내고 있다.

그러나 에듀넷의 이와 같은 검색 서비스는 여러 가지 문제점을 가지고 있다. [4] 첫째로 검색의 불편함이다. 사용자 요구 조사에 따르면 ‘원하지 않는 결과가 도출된다’, ‘너무 많은 결과가 나온다’, ‘필요 없는

데이터가 너무 많다’, ‘정확한 검색을 하려면 비슷한 단어라고 판단되는 단어를 모두 입력 하여야 한다’ 등의



그림 1 에듀넷 서비스 화면

불편함을 표현하였다. 둘째로 지식의 구조를 검색하거나 콘텐츠의 상관 관계를 검색하는 것이 어렵다는 점이다. 대부분의 이러닝 콘텐츠 사용자들은 콘텐츠를 검색할 때 콘텐츠 서로간의 상관관계를 고려하여 관련 있는 것들이 동시에 검색 되기를 원하는 경우가 많다. 그러나 현재 에듀넷을 비롯한 여러 콘텐츠 검색 업체들이 제공하는 서비스의 경우에는 이러한 지식의 구조를 검색하거나 콘텐츠의 상관관계를 검색하는 것이

불가능하다. 예를 들어 중학교 2학년 수학의 도형 분야에서 평행 사변형의 성질에 대하여 공부하려고 이와 관련된 콘텐츠를 검색하려고 '평행 사변형'이라는 키워드로 검색을 할 경우 기본적으로 '평행 사변형'이라는 키워드에 매칭되는 콘텐츠만 검색이 된다. 하지만 '평행 사변형'을 학습 하기 위해서는 관련 학습 단원인 대각, 엇각 등의 개념을 설명하는 콘텐츠도 검색이 되어야 학습자가 체계적이고 효과적인 학습을 할 수 있다. 이런 목적으로 지식의 구조를 검색하거나 콘텐츠의 상관관계를 검색하는 시스템으로는 [4]에서 제안 되었다.

[4]에서 제안된 시스템은 KEM(Korea Education Metadata)[3]을 OWL[5]을 이용하여 온톨로지로 바인딩하고 사용자에게 지식의 구조에 따른 체계적인 검색 결과를 제공하기 위하여 중학교 수학의 도형 영역의 일부 개념을 온톨로지화 했다. 그리고 두 개의 온톨로지를 통합하여 온톨로지 추론을 통해 콘텐츠를 검색하는 시스템이다. 그 결과 기존의 검색보다 더 의미 있는 검색 결과가 도출되었으나 몇 가지 문제점이 존재했다. 첫째, 콘텐츠의 메타데이터를 이용하기 때문에 콘텐츠를 제작할 때 이러한 메타데이터를 모두 기록해야 하는 불편함이 존재했고 메타데이터가 존재하지 않는 콘텐츠는 제대로 활용할 수가 없었다. 둘째, 두 개의 온톨로지를 통합해야 하기 때문에 통합하는 과정의 불편함과 추론 과정에서의 속도저하의 문제가 발생했다.

이와 같은 문제점을 해결하기 위해 본 논문에서는 자동화 된 온톨로지 확장을 통해 에듀넷을 이용하여 학습 콘텐츠를 검색하는 사용자에게 사용자가 원하는 정확한 콘텐츠 검색 결과를 제공하여 교육 콘텐츠 재활용과 콘텐츠 공유기능의 개선을 위한 콘텐츠 검색시스템을 설계하고 구현하였다. [4]에서 사용한 중학교 수학의 도형영역의 일부 개념을 온톨로지화한 온톨로지의 클래스 명과 콘텐츠의 내용을 분석하여 Vector Space Model[6]로 모델링 하였다. 이것을 바탕으로 각각의 콘텐츠가 유사도가 가장 높은 클래스에 자동으로 인스턴스화 되도록 하였다. 이렇게 확장된 온톨로지를 이용하여 콘텐츠를 검색하는 프레임워크를 제시, 구축하였다.

## 2. 관련 연구

### 2.1 Vector Space Model

Vector Space Model은 정보 필터링, 문서 내에서의 정보 검색, 색인과 유사도를 계산하기 위한 수학모델이다. 다차원 선형공간에서의 Vector 정보를 이용해서 자연어를 포함한 문서의 중요도를 분석하기 위한 방법을 제시한다. 문서의 유사도를 다음과 같은 간단한 식(1)로 나타낼 수 있다.

$$\cos \theta = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|} \quad (1)$$

### 2.2 온톨로지 확장

Ontology Population으로 불리는 온톨로지 확장은 온톨로지를 구성하고 있는 각각의 클래스와 관련이 있는 인스턴스를 해당 클래스의 인스턴스로 추가함으로써 온톨로지의 크기를 증가시키는 것을 의미한다. 이것은 개체명 인식(Named-Entity Recognition)과 정보추출과 밀접한 관련이 있다. [7]

### 2.3 OWL(Web Ontology Language)

시맨틱 웹에서 정보는 명시적인 의미를 부여 받게 되는데, 이를 통해 기계는 좀 더 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 위의 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 표현력을 뛰어 넘는 언어가 필요하다. 이런 요구에 의해 등장한 언어가 OWL이다.

OWL은 OWL-Lite, OWL-DL, OWL-FULL과 같은 3가지 종류로 세분화되는데, 각각 지원하는 표현력이 차이가 난다. OWL-Lite가 가장 표현력이 약하며 OWL-FULL은 풍부한 표현을 장점으로 가지는 반면 이를 이용하여 표현된 온톨로지를 추론하는 경우 연산시간이 너무 많이 소요되거나 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 도메인 온톨로지

서론에서 밝힌 바와 같이 학습을 위하여 교육용 콘텐츠를 검색할 때는 해당 교과 의미 구조가 중요한 기능을 한다. 궁극적으로는 모든 학년과 교과의 시맨틱 구조가 온톨로지로 구축되어 활용되어야 하겠지만 이는 많은 비용이 드는 쉽지 않은 일이며 앞으로 많은 노력이 필요한 일이다. 본 연구에서는 의미 구조를 활용하면 검색 기능을 좀 더 고도화 할 수 있음을 보이기 위하여 중학교 수학 교과의 도형 분야의 의미 구조를 교육 인적 자원부 교과 분류 체제를 참조하여 수학 교육 전문가와 협력하여 구축하였다. 그림 2는

도메인 온톨로지를 나타낸다.

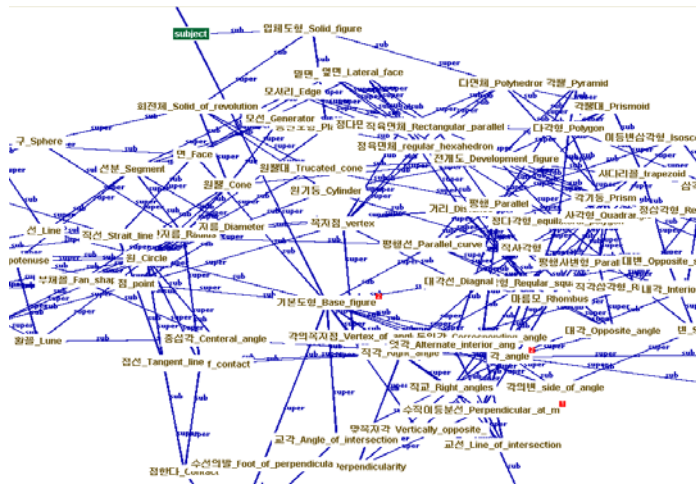


그림 2 도형 영역 도메인 온톨로지

같은 틀들을 이용하여 구현하였다.

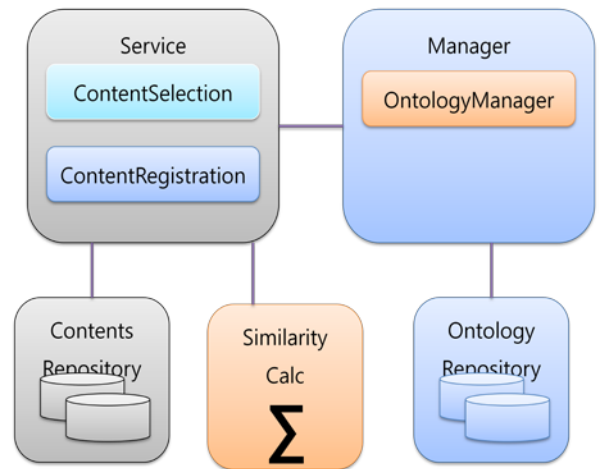


그림 3 온톨로지 활용 검색 시스템 프레임 워크

### 3.2 Vector Space Modeling

본 연구에서는 학습 콘텐츠와 온톨로지의 클래스명 간의 유사도 비교를 위해서 Vector Space Model을 이용하였다. 각각의 콘텐츠에서 도형 영역 도메인 온톨로지의 클래스명들의 발생 빈도를 조사하여 이를 Vector로 표현하였다. 각각의 콘텐츠의 가중치를 표현하기 위해 Vector Space Model의 TF-IDF 모델을 사용하였다. 식(2)은 본 연구에서 사용한 TF-IDF 모델을 나타낸다.

$$W_{i,j} = \frac{\text{freq}_{i,j}}{\max_k \text{freq}_{k,j}} \times \log \frac{N}{n_i} \quad (2)$$

$\text{freq}_{i,j}$  = 특정 콘텐츠에서 나타나는 문자  $k_i$ 의 횟수  
 $N$  = 전체 콘텐츠의 개수  
 $n_i$  = 문자  $k$ 를 포함하고 있는 콘텐츠의 개수

각각의 콘텐츠에 대한 가중치를 표현한 Vector를 식(2)의 유사도 공식을 사용하여 콘텐츠들이 어떤 클래스와 가장 밀접한 관련이 있는지 계산한다. 가장 유사도가 높은 클래스의 인스턴스로 콘텐츠를 추가하는 방법으로 도형 영역 도메인 온톨로지의 크기를 확장시키는 방법을 택하였다.

### 3.3 온톨로지 활용 검색 시스템 프레임 워크

owl[5]을 이용하여 구축한 온톨로지는 추론 엔진과 온톨로지 쿼리 언어를 사용하여 검색 할 수 있도록 하였다. 그림 3은 본 연구의 온톨로지 활용 검색 시스템 프레임 워크이다.

## 4. 구현 및 실험

본 연구의 온톨로지를 활용한 검색 시스템은 다음과

- 온톨로지 구현 언어 : OWL-DL
- 온톨로지 에디터 : Protégé 3.4.4[8]
- 온톨로지 추론 엔진 : KAON2
- 온톨로지 질의 언어 : SPARQL

Java를 이용하여 모든 시스템을 통합하였으며 사용자 인터페이스는 [4]의 연구에서 구현한 시스템을 활용하였다.

실험을 위하여 한국교육학술정보원의 협조를 얻어 중학교 수학 콘텐츠 2,300여개를 확보하였다. 그 다음 수학 전문가를 통하여 해당 콘텐츠가 도메인 온톨로지의 어느 class에 포함되는지 수동으로 분류하고 해당 클래스의 인스턴스로 콘텐츠의 정보를 입력하였다. 실험 방법은 수학 전문가가 수동으로 분류한 목록과 논문에서 제안된 시스템이 자동으로 분류한 목록의 비교를 통해서 논문에서 제안된 시스템의 실효성을 검증하였다. 실험 결과는 그림 4와 같다.

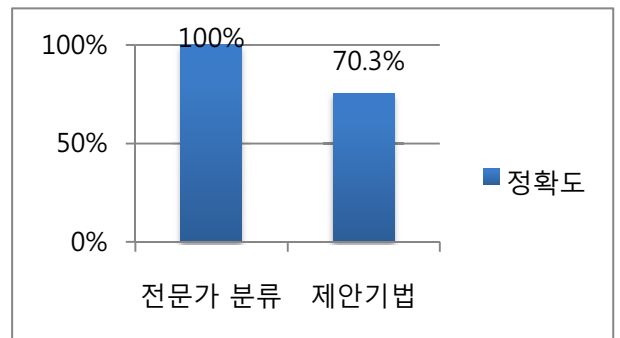


그림 4 정확도 비교

전문가가 수동으로 분류한 목록의 정확도를 100%라고 가정하였을 경우에 본 논문에서 제안된 시스템의 정확도는 70.3%였다.

## 5. 결 론

본 연구에서는 자동화 된 온톨로지 확장을 이용하여 교육 콘텐츠를 효과적으로 검색하는 시스템을 구축하였다. 이러한 시스템의 효과는 [4]의 연구에서 잘 나타나 있다. 사용자는 본 연구의 검색 시스템을 통하여 기존의 키워드 검색 시스템보다 좀 더 만족스러운 결과를 얻을 수 있었다.

이번 연구는 기존의 연구를 자동화 된 방법을 이용하여 결과를 얻었다는데 의의가 있다.

정확도의 향상을 위해서 클래스들간의 연관 관계를 가중치에 반영 할 수 있는 방법을 계속하여 연구해 나갈 예정이다.

## 6. 참고 문헌

- [1]지형근 외, "e-러닝 기술 동향", 한국정보과학회지 제26권 제12호, 2008.12
- [2]에듀넷, <http://edunet4u.net>
- [3]서대원 외(2003). 교수-학습 자료 분류 체계화 연구, 한국교육학술정보원 연구보고서
- [4]B. Chang et al. Educational Information Search Service Using Ontology. ICALT 2007. pages 414-415, July 2007.
- [5]Web Ontology Language, <http://www.w3.org/2004/OWL>
- [6]G Salton, A Wong, CS Yang, "A vector space model for information retrieval", Journal of the American Society for Information Science, Volume 18, Issue 11, 1975
- [7]Philipp Cimiano, "Ontology Learning and Population from Text", pages 30-31, Springer.
- [8]Protege, <http://protege.stanford.edu>