

시맨틱 웹을 이용한 온톨로지 기반의 정보검색 시스템 설계 및 구현

서우진^{1*}, 유경택²

¹강동대학교 컴퓨터정보과 외래교수, ²강동대학교 컴퓨터정보과 부교수

Design and Implementation of Information Retrieval System Based on Ontology Using Semantic Web

Woo-Jin Seo^{1*}, Kyeong-Taek Rhyu²

¹Dept. of Computer Information, Gangdong Univ. adjunct professor

²Dept. of Computer Information, Gangdong Univ. associate professor

요 약 본 논문에서는 시맨틱 검색 수행을 위해 검색 도메인에 알맞은 온톨로지를 이용, 구축하고 정보에 관한 검색, 변환, 통합, 공유가 가능한 검색 엔진을 구현하여 검색 시스템의 기반을 마련하는 것을 목적으로 하였다. 기존 방식에서 벗어나 온톨로지를 활용하여 계층 관계를 추론하고, 그 계층을 근거로 개체를 추론한 다음 속성을 추출하여 사용자가 원하는 자료와 관련있는 분야를 검색하는 것이다. 이러한 방식으로 정보를 검색할 수 있도록 정보검색 시스템을 '자격증'과 관련된 키워드를 입력하여 구현하였다. 구현된 시스템은 온톨로지에서 각 속성들의 의미와 관계를 정리하여 일반인 정보검색을 사용자가 빠르고 쉽게, 정확한 검색을 할 수 있도록 하였다. 또한, 구현 결과를 2개의 다른 검색엔진과 비교하였다. 비교한 검색엔진은 대표적인 검색엔진인 '네이버'와 '다음'이다. 시맨틱 웹을 이용한 검색을 수행하기 위해 검색 도메인에 맞는 온톨로지를 이용하여 구축한 본 연구의 검색 엔진은 상당히 우수한 결과를 보여주는 것으로 평가되었다. 그러나 검색 엔진의 정확성과 신뢰성을 높이고 좀 더 포괄적인 범주의 검색어 포함하기 위해서는 더욱 정형화된 온톨로지가 필요하다고 사료된다.

주제어 : 온톨로지, 시맨틱 웹, 검색시스템, 기반정보 검색시스템, 검색엔진

Abstract In this paper, the purpose of this paper is to lay the foundation for the search system by using and building an online search engine suitable for the search domain and enabling search, conversion, integration and sharing of information. It is to use the ontology to infer hierarchical relationships, deduce objects based on that layer, and extract attributes to search areas that are relevant to the data that the user wants. In order to search for information in this way, the information search system was implemented by entering key words related to 'qualifications'. The implemented system arranged the meaning and relationship of each attribute online so that the general public can search information quickly, easily, and accurately. In addition, the implementation results were compared with two different search engines. Comparable search engines are Naver and Daum, the two major search engines. The search engine of this study, which was built using an ontology suitable for the search domain to perform searches using the semantic web, was evaluated to have excellent results. However, it is thought that a more formalized online location is necessary to increase the accuracy and reliability of search engines and to include more comprehensive categories of search terms.

Key Words : Ontology, Semantic web, Search system, Information retrieval system, Search engines

*Corresponding Author : Woo-Jin Seo(youiwin1@naver.com)

Received November 21, 2018

Accepted January 20, 2019

Revised January 2, 2019

Published January 28, 2019

1. 서론

Tim Berners-Lee가 1989년 처음 제안된 웹(World Wide Web)은 클라이언트-서버 개념과 쉽게 학습할 수 있는 HTML(HyperText Markup Language) 언어를 사용하여 누구나 쉽게 정보에 접근하거나 정보를 검색할 수 있게 되었고, 결과적으로 방대한 정보의 증가를 가져왔다. 웹 사용 횟수가 증가함에 따라 정보의 양이 적을 때에 비해 사용자가 원하는 정보를 검색하거나 구성 및 통합하는 작업이 복잡해지고 상당한 시간을 요구하였다.

이를 개선하기 위해 1999년 제안된 차세대 웹인 시맨틱 웹 기술 연구가 진행 중이다.

시맨틱 웹 기술은 기계를 통해 지식을 처리하고 웹을 통해 확장, 공유할 수 있는 핵심 기술이다. 이미 미국과 유럽을 중심으로 시맨틱 웹에 대한 연구가 진행되었고, 시맨틱 웹에서 핵심기술 중 하나인 온톨로지는 인간과 컴퓨터간의 연결되는 지식을 개념화하여 구체적인 형식으로 표현하는 것으로 정의된다. 이것은 특정한 분야에서 사용되는 표준 단어들의 모음이라 할 수 있다. 그러나 온톨로지는 표준 단어들의 개념 속에서 통제되는 어휘를 제공함으로써 웹에서 응용 프로그램이 이해할 수 있도록 다양한 데이터를 구조화한다. 더 나아가 의미까지 지원하고, 인간과 웹 프로그램 사이에서 간결한 의사소통을 가능하게 한다[1].

시맨틱 웹 기술은 현재 상업적인 시도가 이루어지고 있으며, 시맨틱 웹 기술 자체가 시장을 형성하기 보다는 관련된 응용분야의 요소적인 기술로 시장을 형성해 나아가 갈 가능성이 현재로서는 매우 높다. 관련된 응용분야로는 먼저 지식관리시스템(KMS), e-러닝 시장, 기업의 어플리케이션통합(EAI) 및 기업과 기업 간 통합된 시장을 볼 수 있으며, 그 수요가 늘어나고 이러한 시장을 바탕으로 시맨틱 웹에 대한 고유 기술과 응용 시장이 활성화될 것이며, 앞으로 지식을 기반으로 사회의 모든 산업 부문에 다양한 영향을 줄 것으로 예상된다[2].

2. 이론적 배경

2.1 온톨로지(Ontology)의 고찰

2.1.1 온톨로지의 개념

1999년 Berners-Lee가 주장했던 시맨틱 웹은 현재 웹이 가진 정보 자원에 정의된 의미를 부여함으로써 기존

웹의 범위를 확장할 수 있는 웹을 말한다. 이러한 시맨틱 웹의 이상을 구현하기 위해 반드시 선행되어야 할 기술이 RDF(Resource Description Framework)층 상단에 위치하는 온톨로지이다. 이 온톨로지는 지식을 처리하거나 인공지능, 컴퓨터적인 과학 등 많은 분야에서 사용되고 있으며, 특히 자연어 처리 및 지식관리, 전자상거래, 정보 검색, 지능형 정보 통합 시스템, 생물 정보과학, 데이터베이스 설계와 통합, 교육과 관련된 응용프로그램에서 많이 사용되고 있다.

또한 온톨로지는 넓은 의미로 ‘데이터베이스’라 표현할 수 있지만 실제로는 데이터보다 복잡한 형태로 많은 지식과 관련되어 있다는 의미에서 ‘지식베이스(knowledge base)’라 부르기도 한다. 하지만 엄밀히 표현하자면 ‘온톨로지’는 지식적인 내용과 절차적인 추론 과정을 포함하여 포괄적 의미의 지식보다 용어사이의 ‘개념’적 관계로 국한되어 있기 때문에 지식베이스와는 구별되어 다른 형태의 ‘데이터베이스’라 할 수 있다.

온톨로지를 사용함으로써 설계자는 높은 재활용성을 얻을 수 있게 되는데 마치 프로그램을 재사용하는 것과 비슷한 형태로 지식을 재사용할 수 있게 되는 것이다. 게다가 다른 기종의 시스템 간에 공통된 단어와 용어를 사용함으로써 도메인에 대한 지식을 재사용하고 공유할 수 있는 애플리케이션 프로그램을 설계할 수 있도록 도움을 준다. 또한, 시스템 설계자가 너무 많은 부분을 세밀하게 구현하지 않아도 되는 부담감을 덜어주고 도메인의 구조적인 부분이나 다양한 부분을 작업할 수 있도록 해준다[1].

온톨로지의 세부적인 특징을 살펴보면 아래 내용과 같다.

- Conceptualization : 사람들이 사물에 보고 생각하는 것을 나타내는 추상화된 모델이다.
- Explicit Specification : 사용상의 제약조건이나 개념의 형식들이 명시적으로 정의된다.
- Formal : 온톨로지는 정형화되어 기계도 이해할 수 있도록 해야 한다.
- Shared : 온톨로지는 그룹전체 구성원들이 합의하여 개인에게 국한되지 않도록 지식을 공유할 수 있어야 한다.

지금의 온톨로지는 관계와 단어들로 구성된 일종의 사전 처럼 생각할 수 있으며, 그 안에는 계층적으로 특정

도메인과 관련된 단어들 이 표현되어 있다. 이를 확장할 수 있도록 추론 규칙이 포함되어 있으며 웹 기반의 지식 또한 처리하거나 실제 응용 프로그램 사이의 지식을 공유 및 재사용이 가능하도록 되어 있다.

2.1.2 온톨로지 시스템 구조

기존 웹에서 사이트를 제공하는 서버는 특정 분야에 대한 지식을 기반으로 의미구조를 정리하고 하나의 온톨로지를 생성하여 자신이 가지고 있다. 물론 생성된 온톨로지는 재사용이 가능하며, 상속받은 기존의 온톨로지를 사용할 수도 있다. 분산되어 있는 웹 상의 온톨로지는 온톨로지 사상(mapping)과 연결되어 지능적인 에이전트들이 의사소통하며, 가격 협상과 스케줄링 같은 웹 서비스 등을 제공할 수 있다.

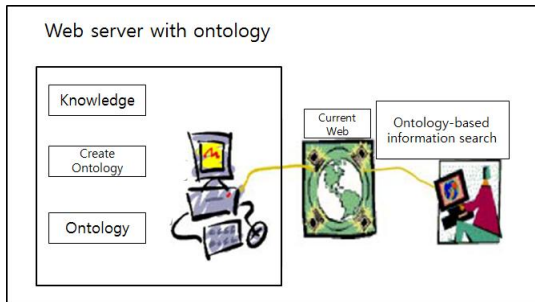


Fig. 1. Architecture of Ontology Systems
 <source : JungEunKyung, 2004 [3]>

2.1.3 온톨로지의 구축

온톨로지를 개발한다는 것은 생각만큼 가볍고 쉬운 일이 아닌 것으로 많이 알려져 있으며 데이터 모델링과 같은 일반적인 온톨로지 모델링 역시 폭넓은 지식과 풍부한 경험을 필요로 하는 분야이다[4].

특정 영역(Domain)의 일반적인 온톨로지 개발은 Fig. 2와 같은 단계를 거친다.

온톨로지 개발의 과정은 크게 분석, 수집, 구축, 설계의 4단계로 구분할 수 있다. 정보검색 시스템 중 학술 논문 검색 시스템의 경우, 세부적으로 논문제목의 수집, 구문 구조 분석을 통한 키워드에 대한 역할개념 및 관계 설계, 실제 온톨로지의 구축도구를 이용한 구축 단계로 구분 지을 수 있다. 대부분 지금까지 제시된 온톨로지 구축 방법들은 전체 순환 개발만 고려하여 분석단계에서 반복되는 순환 개발의 중요성이 간과되었다[6]. 검색 시스템 중 학술논문의 경우 온톨로지 구축을 위하여 분석 단계

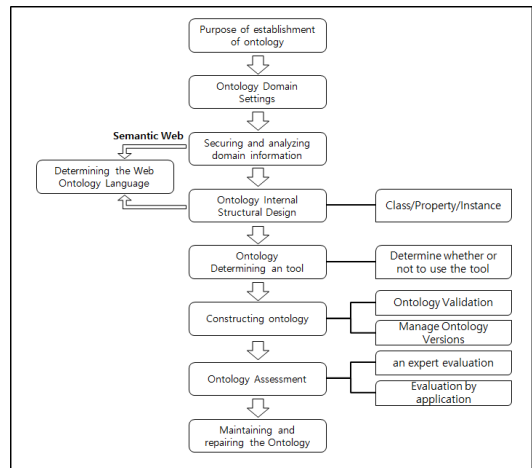


Fig. 2. Ontology Deployment Stage
 <source : HeoYeon, 2010 [5]>

가 앞으로 개발에 중요한 영향을 주기 때문에 다음 개발 단계로 이전하기 전 반복적인 정밀화를 거쳐서 효율적으로 온톨로지를 구축할 수 있다. 온톨로지를 구축하기 위해 논문제목의 구문구조 분석단계에서 산출된 결과는 관계속성 설계와 키워드(인스턴스)의 개념 뿐만 아니라 온톨로지를 실제 구축하는데 중요한 의미를 가진다.

2.2 시맨틱 웹에 대한 고찰

2.2.1 시맨틱 웹(Semantic Web)

SGML(Standard Generalized Markup Language)은 사용성의 편리함과 표현의 단순함으로 인하여 인터넷에서의 표준 웹페이지를 표현하는 언어로 자리 잡게 되었다[7]. HTML은 사용자와 개발자 그리고 제공자가 쉽게 웹의 자원을 접근, 생성 그리고 유지할 수 있도록 해주는 특성을 가지고 있었다. HTML의 영향으로 지난 10년 동안 기하급수적으로 웹 사용자 수가 늘어나면서 정보의 양도 늘어나고 있다. 웹 환경에서는 단순링크를 기반으로 색인어에 의해 주제에 따른 홈페이지 분류나 홈페이지를 검색하는 시스템에서 제공되는 정보를 통해 필요한 정보를 획득해야 한다. 이러한 정보검색 시스템에서 지나치게 많은 정보를 제시하고, 분류시스템 의한 분류방법은 사용자가 원하는 형태로 대부분 되어있지 않는 경우가 많다. 그래서 정보를 획득하는 데에는 어려움이 있을 수밖에 없다[8].

정보를 획득하는데 어려움을 해결하기 위한 대안은 1998년, W3C에서 차세대 웹으로 '시맨틱 웹'이 제안되었다[5]. 기존의 웹은 키워드 기반 검색으로 사용자가 원하는

는 정보의 문장이 의미는 같고 표현 방법이 다를 경우 정확하게 검색 결과를 나타내기 어려운 것과 달리 '시맨틱 웹'은 인터넷이라는 공간에서 존재하는 정보에 시맨틱 웹 기술을 사용하여 온톨로지와 같은 의미(Semantic)를 부여하고, 소프트웨어 에이전트가 문서의 상태 정보 및 의미를 인지한 후 각 사용자에게 적절한 경로의 검색을 판단하게 한다. 그러므로 시맨틱 에이전트를 사용하다 보면 지능적이고 자동적인 테스트가 가능하게 된다 [9-10]. 즉, 에이전트는 인간의 정보 활용 수행 방식을 위임함으로써 사용자의 간섭 없이도 정보를 검색, 관리, 통합, 추출 등 요구사항에 대해 자발적 수행을 할 수 있도록 하는 것이다.

2.2.2 시맨틱 웹(Semantic Web)과 관련된 연구

시맨틱 웹을 이용한 검색방법을 기존의 웹을 이용한 검색방법과 효과적인 면에서 서로 비교하기 위한 방법으로 검색결과와 평가요인에 초점을 맞추고 있다.

Guha et al.(2003)[11]의 선행연구에서는 기존 웹 검색 결과와 시맨틱 웹 검색 결과를 하나의 화면에 표시하고, 사용자가 결과 화면에서 선택한 문서가 어떠한 결과 화면에서 도출되었는지가 검색시스템의 평가기준으로 사용되었다.

Bonino et al.(2004)[12]의 선행연구에서는 기존 웹 검색이 용어 간의 상관관계가 높은 용어를 추천하는 방식인 반면, 시맨틱 웹 검색에서는 온톨로지 구조를 통해 사용자가 요구하는 검색 의도에 맞게 검색에 대한 질의를 확대할 수 있다고 가정하였다. 시맨틱 웹 검색을 위한 핵심은 질의 정제 과정이라 간주하고, 온톨로지 개념의 관계성과 도메인 온톨로지의 가용성의 이해력이 중요성이라고 강조하였다. 이를 토대로 시맨틱 웹 검색 엔진을 개발 하여, 이 연구에서 웹 페이지를 대상으로 제한된 범위 내에서의 정확율과 재현율에 대한 평가를 수행하였다. 정확율과 재현율은 기존 웹 검색엔진의 성능을 평가 요인이었는데, 시맨틱 웹 검색 엔진의 품질을 평가하는 요인에서도 이 연구에서는 검색 결과의 중요한 요인으로 간주하였다.

Makela et al.(2006)[13]의 선행연구에서도 화면 구성 관점과 시맨틱 웹의 UI(User Interface)관점에서 사용자 간의 질의 생성에 관련된 연구를 수행하였다. 시맨틱 웹 기술은 관계와 개념을 이용하여 그래프로 지식을 표현하기에 적합하지만 검색을 이용하는 사용자는 본인들의 지

식을 검색하는데 필요한 질의를 생성하는데 있어서 어려움이 있음을 파악하였다.

3. 검색시스템 설계

3.1 개발 개요 및 환경

3.1.1 개발 개요

기본적으로 필요한 크롤링(Crawling), 인덱싱(Indexing), 서치(Searching)를 이용하여 구현하는 source를 만들고 사용자가 원하는 적절한 결과물을 추출할 수 있도록 간단한 인터페이스를 활용 것을 목표로 하며, Perl을 이용해서 검색엔진 시스템을 구축한다.

3.1.2 개발 환경

CPU

- AMD processor
- Pantium processor

OS

- Red Hat Linux release 9 (Shrike), Kernel 2.4.20-8 on an i686
- Windows XP sp2

Developing program

- Perl v5.8.0 built for i386-linux-thread-multi
- Microsoft Visual C++ 6.0

3.2 사용된 함수 및 변수 설명

본 연구에서 통해 2개의 설계된 파일은 crawl.pl과 search.pl로 각기 클라이언트와 서버가 수행하는 파일이 된다. crawl.pl은 가능한 한도 내의 뉴스를 서버가 네트워크망으로부터 크롤링하고 그 결과물을 테마별로 인덱싱해서 파일로 저장한다. search.pl은 사용자가 검색어를 입력하면 저장된 파일의 인덱서들을 살펴보고 일치하는 검색어와 내용들이 들어 있는 뉴스들을 사용자가 원하는 수만큼 찾아 결과파일(results.txt)로 저장해서 남기게 된다.

3.2.1 Crawl.pl

주어진 조건으로부터 웹 서버의 입장으로 원하는 정

보를 크롤링하고 인덱싱해서 제목별로 인덱서들이 10개의 파일로 분리 저장한다.

Table 1. Important variable description

important variables
\$MAX = 100; The number of news that will be maximized by initializing to 100
\$newsnum = 1; Crawl news search from the initial value 00001 to 25000.
\$url= "http://news.empas.com/show.tsp/" . \$year . \$date . "n" . \$newsnum; This 'url' address is used until newsnum and date are changed and the desired number of news is repeatedly crawled. In the case of date, the localtime () function was used to retrieve the date in the reverse order and select the method to revert to the most recent article to improve the efficiency of the news.

Table 2. Used functions

Used functions
chmon(); Functions that return the month as a number with the localtime () function, \$month
indexing(); In case of Hangul, the first ASCII value is 176 ~ 200, and it is indexed with 3 differences. In case of special characters or English, it is placed at the first or last indexing. There were many cases where special characters were coming at the beginning of the title, so it tried to separate it as much as possible but it was not within the range. The first character value of the nouns when searching through the hash table generated is extracted as ASCII code, and the same storage space is searched first, and then the efficiency is pursued in terms of time. According to the title, it was classified so that it could be stored in 10 spaces.
StoreToFile(); After the indexing is done, the stored news is saved as a separate file.

3.2.2 Search.pl

사용자가 원하는 검색어를 입력하면 실행하여 검색어를 형태소 분리하여 일치되는 내용을 가진 문서들을 인덱서(Indexer)에서 사용자가 원하는 수만큼 찾아 결과물을 돌려준다.

Table 3. Important variable description

important variables
\$compare_value = 4; The frequency increases and searches for nouns that match the search term. Frequency comparison variable. The larger this value, the easier it is to search.
\$wanted_doc = 4; The maximum value is displayed. Find documents with more than the frequency. Finds up to four documents and ends the search.
@final_doc; Arrangement means. An array to store the found documents.

Table 4. Used functions

Used functions
search(@ARGV); Enter the search term. Search terms can be used freely and can be used to search up to 10 morphemes.
parse_search(); A function to store only nouns by morphemically separating input words.
parse_firstc(); Extract the ASCII value of the first character of the morpheme-separated query nouns to find the appropriate indexer.
FileToVariance(); A function that puts what was stored in a file back into a variable through indexing.
find_from_indexer(); Locate the appropriate indexing location from the indexer using ASCII values
spare_of_findnum(); A function that specifies the location of another indexer in advance if the desired document is not found. Find indexer according to the value stored in @findnum.
choose_to_use_from_indexer(); A function to search for articles that ultimately compares search terms with news.parse_storage_1 (); Used in conjunction with ~ parse_storage_10 () ; prase_sentence (\$ temp) ; get_frequency (@parse_storage);
parse_storage_1(); Staple documents taken from storage_1_1.txt.
parse_sentence(\$temp); Compare the contents of the morpheme news with get_frequency ();
get_frequency(@parse_storage);

4. 시스템 구현 및 평가

4.1 키워드 검색

본 논문에서 제공하는 검색 기법은 찾고자 하는 분야에 해당하는 영역을 설정하고, 원하는 정보와 관련한 키워드를 입력하면 해당 값이 파라미터로 엔진에 전달되고, 엔진에서 조건을 만족하는 문서를 찾아 사용자 화면에 출력하게 된다.

먼저, 검색어로서 Fig. 3에 제시한 바와 같이, '취업에 필요한 자격증'을 입력창에 입력하였고, 이와 관련한 cmd 화면은 Fig. 4와 같이 디스플레이 되었다.

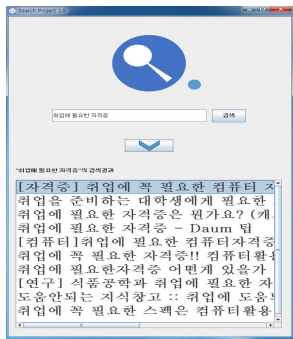


Fig. 3. Search project screen for search word 'Certificate required for employment'

입력된 값이 추론기에 전달된 후 추론기에서 상위 클래스인 '자격증' 클래스의 서브 클래스를 추출하는 그림이다. Fig. 4에서 보는 바와 같이 '자격증' 클래스의 서브 클래스로는 온톨로지에서 구축된 서브 클래스들이 존재하며 추론기는 상위 클래스인 '자격증' 클래스의 서브 클래스를 추론하여 개체를 추출한다.



Fig. 4. cmd screen for search word 'Certificate required for employment'

Fig. 5은 입력받은 검색 정보를 이용하여 '취업에 필요한 자격증' 객체의 link정보를 출력하는 GUI(Graphical User Interface) 실행 화면이다. Fig. 5에 제시한 바와 같이, 검색어와 일치된 문서의 'link'정보를 찾게 되면 이 문서정보를 데이터베이스에 저장된 link필드의 값으로 검색하여 원하는 문서의 파일과 함께 사용자의 결과 화면에 출력하게 된다. 실행 결과, 구현된 GUI 실행 화면은 Fig. 5에 나타난 바와 같다.

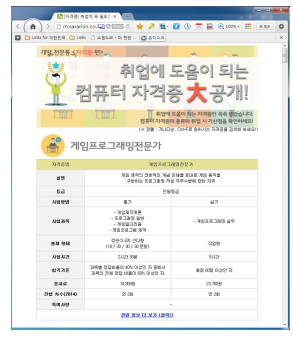


Fig. 5. GUI execution screen for the search word 'Certificate required for employment'

다음으로, 검색어로서 Fig. 6에 제시한 바와 같이, '정보처리산업기사'를 입력창에 입력하였고, 이와 관련한 cmd 화면은 Fig. 7와 같이 디스플레이 되었다.

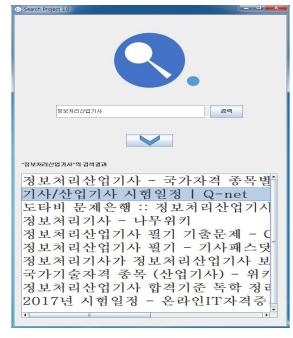


Fig. 6. Search project screen for search word 'An information processing industry engineer'

입력된 값이 추론기에 전달된 후 추론기에서 상위 클래스인 '자격증' 클래스의 서브 클래스를 추출하는 그림이다. Fig. 7에서 보는 바와 같이 '자격증' 클래스의 서브 클래스로는 온톨로지에서 구축된 서브 클래스들이 존재하며 추론기는 상위 클래스인 '자격증' 클래스의 서브 클래스를 추론하여 개체를 추출한다.

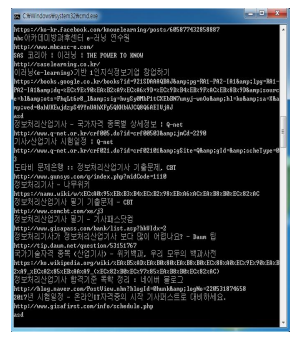


Fig. 7. cmd screen for search word 'An information processing industry engineer'

Fig. 8은 입력받은 검색 정보를 이용하여 ‘정보처리산업기사’ 객체의 link정보를 출력하는 GUI 실행 화면이다. Fig. 8에 제시한 바와 같이, 검색어와 일치된 문서의 ‘link’정보를 찾게 되면 이 문서정보를 데이터베이스에 저장된 link필드의 값으로 검색하여 원하는 문서의 파일과 함께 사용자의 결과 화면에 출력하게 된다. 실행 결과, 구현된 GUI 실행 화면은 Fig. 8에 나타난 바와 같다.

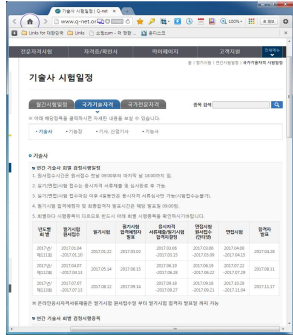


Fig. 8. GUI execution screen for the search word 'An information processing industry engineer'

마지막으로, 검색어로서 Fig. 9에 제시한 바와 같이, ‘극동대학교’를 입력창에 입력하였고, 이와 관련된 cmd 화면은 Fig. 10과 같이 디스플레이 되었다.

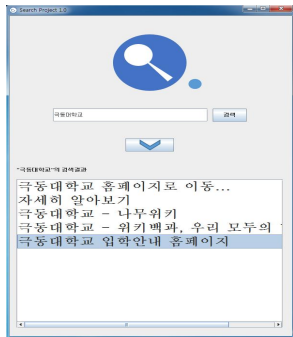


Fig. 9. Search project screen for search word 'FAR EAST UNIVERSITY'

입력된 값이 추천기에 전달된 후 추천기에서 상위 클래스인 ‘자격증’ 클래스의 서브 클래스를 추출하는 그림이다. Fig. 10에서 보는 바와 같이 ‘자격증’ 클래스의 서브 클래스로는 온톨로지에서 구축된 서브 클래스들이 존재하며 추천기는 상위 클래스인 ‘자격증’ 클래스의 서브 클래스를 추천하여 개체를 추출한다.

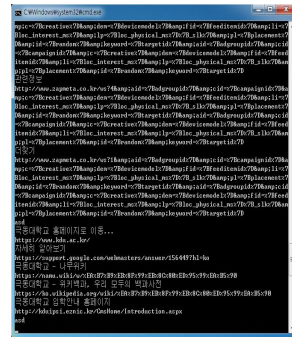


Fig. 10. cmd screen for search word 'FAR EAST UNIVERSITY'

Fig. 11은 입력받은 검색 정보를 이용하여 ‘극동대학교’ 객체의 link정보를 출력하는 GUI 실행 화면이다. Fig. 11에 제시한 바와 같이, 검색어와 일치된 문서의 ‘link’정보를 찾게 되면 이 문서정보를 데이터베이스에 저장된 link필드의 값으로 검색하여 원하는 문서의 파일과 함께 사용자의 결과 화면에 출력하게 된다. 실행 결과, 구현된 GUI 실행 화면은 Fig. 11에 나타난 바와 같다.

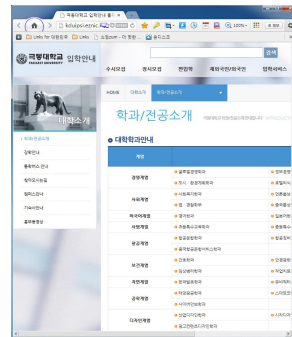


Fig. 11. GUI execution screen for the search word 'FAR EAST UNIVERSITY'

마지막으로 본 연구에서 온톨로지를 구축하고, 저장된 지식 정보에 대한 검색, 변환, 통합, 공유가 가능한 의미 검색 엔진을 구현하였고, 구현 결과를 다른 검색엔진과 비교하였다. 비교한 다른 검색엔진은 국내 대표적인 검색엔진인 ‘네이버’와 ‘다음’ 검색엔진으로 수행하였다. 검색어의 경우 앞서 구현하였던 ‘취업에 필요한 자격증’, ‘정보처리산업기사’, ‘극동대학교’ 이렇게 3개 키워드의 구현결과를 토대로 비교 분석하였다. 비교 결과 Table 5에 제시한 바와 같이 나타났다. 구현 결과의 정확도를 바탕으로 본 연구자가 판단한 결과, 본 연구 검색엔진의

경우 90%로 나타났으며 ‘네이버’ 검색엔진의 경우 85%, ‘다음’ 검색엔진이 80%로 나타남으로써 시맨틱 검색을 수행하기 위해 검색 도메인에 맞게 온톨로지를 이용하여 온톨로지를 구축한 본 연구의 검색 엔진은 매우 우수한 결과를 보여주는 것으로 평가된다.

Table 5. Implementation Results by Search Engine

A search engine of main thesis	Naver	Daum
90%	85%	80%

5. 결론

현재 인터넷 환경은 컴퓨터가 웹 데이터를 파악하고 인식하는 능력이 약간 부족하기 때문에 사용자의 작은 도움 없이 정보를 자동으로 추출하고 처리하여 저장 및 통합하는 것에 어려움이 존재하였다. 또한, 같은 의미의 데이터도 서로 다른 데이터인 것 처럼 취급하여 개발자들에게는 하나의 웹 사이트를 개발할 때 마다 각각의 데이터를 스키마 구조에 다시 저장해야 하는 번거로움과 사용자에게는 어떠한 용어로 찾아야 하는지에 혼란을 주었다.

이에 본 논문에서는 온톨로지를 기반으로 하여 논문에 대한 클래스 계층관계와 데이터에 대한 단어를 명세화하여 구성하고, 공리를 적용한 추론기에 입력하여 보다 정확한 검색을 가능하게 하는지를 확인하였다. 기존의 문서 및 단어 추출 알고리즘에서 벗어나 온톨로지를 기반으로하여 계층 관계를 형성하여 추론된 계층을 근거로 개체의 속성을 추출하고 정보검색을 이용하는 사용자가 원하는 자료와 정보를 검색하는 것이다. 온톨로지를 이용하여 서로 같은 의미의 용어들을 정리하고 정보 검색 전문가가 아닌 일반 사용자도 보다 쉽게 정보를 검색할 수 있는 시스템을 ‘자격증’과 관련 키워드로 구현하였다.

본 논문에서 구현된 시스템은 자주 사용하는 간단한 용어로 정보를 검색하더라도 서로 같은 의미를 나타내는 용어의 경우 검색 결과를 사용자에게 쉽게 제공할 수 있도록 구현하였다.

그러나 검색 엔진의 포괄적인 범주의 검색자료를 포함하고 정확성을 높이기 위해서는 정형화된 온톨로지가 필요하다.

이러한 온톨로지를 위해서는 향후 추가되는 논문과 전문가와 공학자의 참여가 이루어져 데이터 타입에 대한 충분한 보완이 필요할 것으로 사료된다.

REFERENCES

- [1] Y. H. Yang, (2007). *Design of Biological Curriculum based on Ontology Reasoning System and implementation*, Graduate School of Korean National University of Education.
- [2] J. K. Kim. (2007). *Implementing a system integration project ontology and information retrieval system*, Yonsei University Graduate School of Education.
- [3] Y. K. Jung. (2004). *Ontology-Based Information Retrieval System in a Semantic Web Environment*, Jeju University Graduate School.
- [4] H. S. Choi & J. H. Lim. (2006. 4). Methods and Examples of Ontology, *Information science journal*, 31-44.
- [5] Y. Huh. (2005). *How to Search Web Documents Based on Fuzzy Logic and Ontology*, Graduate School of Chonbuk National University.
- [6] H. S. Lee. (2003). *A Study on the Design of Ontology-Based Oriental Medicine Knowledge Management System*, Graduate School of Chung-Ang University.
- [7] S. Y. Park. (2006). Korea's S/W Industry and Semantic, *Information science journal*, 24(4), 5-10.
- [8] H. C. Kwon. (2006). Semantic Web and Ontology : Possibility and Limitations, *Information science journal*, 24(4), 11-16.
- [9] S. Y. Huh & Y. K. Kim. (2007). A Study on the Semantic Search System Based on Ontology, *Korea Information Processing Association*, 263-466.
- [10] D. L. Han. (2009). *A Study on the Semantic Search System for the Efficiency of Knowledge Services*, Graduate School of Chung-Ang University.
- [11] R. Guha, R. McCool & E. Miller. (2003). *SemanticSearch*, WWW 2003Conference, May20-24,ACM Press, Budapest, Hungary, 700-709.
- [12] D. Bonino et al. (2004). Ontology Driven Semantic Search. *WSEAS Transaction onInformationScienceandApplication*, 1, 1597-1605.
- [13] E. Makela, e tal. (2006). Ontogator - A SemanticView-Based Search Engine ServicesforWebApplications, *5th International Semantic Web Conference 2006*,

ISWC2006, Athens, GA, USA, 4273(2006), 847-860.

- [14] K. B. Kim. (2007). *Online newspaper article search system based on semantic Web Ontology*, Sogang University Graduate School of Information and Communication.
- [15] H. C. Park. (2006). *Implementing a system integration project ontology and information retrieval system*, Yonsei University Graduate School of Education.

서 우 진(Seo, Woo Jin)

[정회원]



- 2018년 2월 : 극동대학교 정보통신과(석사)
- 2011년 5월 : JS Company 대표
- 2017년 3월 ~ 현재 : 강동대학교 외래교수
- 관심분야 : SW코딩, 아두이노, 시

스템통합

· E-Mail : youiwin1@naver.com

유 경 택(Rhyu, Kyeong Taek)

[정회원]



- 1990년 8월 : 광운대학교 전산공학 학과(석사)
- 2006년 2월 : 원광대학교 컴퓨터공학 학과(박사)
- 1995년 3월 ~ 현재 : 강동대학교 부교수

· 관심분야 : 분산컴퓨팅, XML, 시스템통합

· E-Mail : rhyukt@gangdong.ac.kr