

碩士學位論文

시맨틱 웹 환경에서의
온톨로지 기반 정보검색 시스템



指導教授 李尙俊 서관
JEJU NATIONAL UNIVERSITY LIBRARY

濟州大學校 大學院

컴퓨터 工 學 科

鄭 恩 京

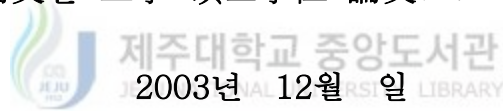
2003年 12月

시맨틱 웹 환경에서의 온톨로지 기반 정보검색 시스템

指導教授 李 尙 俊

鄭 恩 京

이 論文을 工學 碩士學位 論文으로 提出함.



康眞榮의 工學 碩士學位 論文을 認准함.

審査委員長 郭 鎬 榮 印

委 員 李 尙 俊 印

委 員 邊 暎 哲 印

濟州大學校 大學院

2003 年 12 月

A study of ontology-based information retrieval on Semantic Web

Eun-Gyoung Joung

(Supervised by Professor Sang-Joon Lee)



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING GRADUATE
SCHOOL
CHEJU NATIONAL UNIVERSITY

2003. 12.

[목 차]

I. 서론	1
II. 관련이론 및 연구	3
1. 시맨틱 웹.....	3
2. 온톨로지.....	9
3. 정보검색 연구.....	13
III. 온톨로지 기반 정보검색 시스템	15
1. 시스템 구조.....	15
2. 연구방법	17
IV. 시스템 구현 및 평가.....	18
1. 구현시스템 설계	18
2. 시스템 구현	25
3. 평가	30
V. 결론	31
참고문헌	32

[그림 목차]

Fig. 1 The development of web 4

Fig. 2 The Semantic Web "layer cake" presented by Tim Berners Lee at the XML 2000 conference 5

Fig. 3 The step of Ontology development 10

Fig. 4 OntoEdit 11

Fig. 5 OilEd 12

Fig. 6 protege-2000 12

Fig. 7 System structure 16

Fig. 8 System design 18

Fig. 9 The table of jeju lodge 19

Fig. 10 The table of jeju tour 20

Fig. 11 The tree of lodge class 22

Fig. 12 The tree of tour class 22

Fig. 13 The tree of ontology 23

Fig. 14 The screen of information searching for client 27

Fig. 15 The lodge information searching 27

Fig. 16 The searching for user word 28

Fig. 17 The result of searching for Fig.15 and Fig.16 28

Fig. 18 The case of searching for used a wordwritteninChinesecharacters
29

Fig. 19 The result for Fig. 18 29

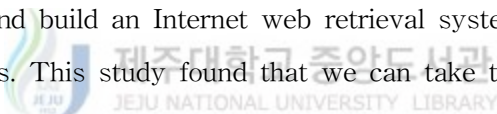
Fig. 20 The comparing of searching time and file loading for using demon
thread 30

[표 목차]

Table. 1 The list of lodge property 21
Table. 2 The list of tour property 21

Abstract

Ontologies are formal theories that are suitable for implementing the semantic web. Which is a new technology that attempts to achieve effective retrieval, integration, and reuse of web resources. Ontologies provide a way of sharing and reusing the knowledge among the people and the heterogeneous application systems. The role of ontologies is that of making explicit specified conceptualizations. In this context, domain and generic ontologies can be shared, reused, and integrated in the analysis and design stage of information and knowledge systems. This study aims to design an ontology for jeju lodges and tour. and build an Internet web retrieval system based on the proposed ontologies. This study found that we can take the same result for [synonym](#) that cilent use.



I. 서론

Tim Berners-Lee에 의해 1989년에 처음 제안된 월드와이드웹은 널리 알려진 클라이언트-서버 개념과 쉽게 익힐 수 있는 간단한 HTML 언어를 이용하여 편리성을 추구한 덕분에 일반 사용자 누구나 쉽게 정보를 접근하거나 게시할 수 있게 되었고, 결과적으로 폭발적인 정보의 증가를 가져왔다. 사용자들은 이러한 정보를 이용하기 위해 브라우저에 URL 주소를 직접 입력하거나 하이퍼링크를 따라가기도 하고, 검색엔진의 도움을 받기도 한다.

이처럼 사용하기 편리하다는 점이 주요 성장 요인이기도 하지만 현재 많은 량의 정보가 폭발적으로 생성되는 시점에서는 문제점으로도 인식되고 있다.

저장 측면에서 보면 현재의 웹 사이트에서는 데이터베이스를 이용하여 자료를 저장하고 스키마 구조에 의해 자료를 찾아 사용자들에게 단순하게 제공하고 있다. 이러한 시스템은 사용자가 자주 사용하는 용어와는 관계없이 웹 사이트 개발자가 일방적으로 정해놓은 검색필드만을 검색할 수 있다는 문제점을 가지고 있다. 현재의 웹은 데이터들의 의미를 부여하여 사용자가 원하는 정보를 자동으로 추출하고 정보와 그 정보의 속성을 나타낼 수 없기 때문이다. 예를 들어, 현재 웹 검색엔진은 주로 단어의 빈도수나 어휘 정보를 이용하여 문서의 유사도를 측정하고 순위를 부여하기 때문에 사용자의 질의와는 관계없는 많은 문서를 결과로 가져올 수 있고 이로 인해 사용자는 불필요한 정보를 걸러내느라 시간을 낭비하게 된다.

이러한 한계점을 해결하기 위해서 1990년대 말에 W3C(World Wide Web Consortium)에서 시맨틱 웹(Semantic Web)을 제안하였다(McIlraith.S.A 2001).

시맨틱 웹의 목적은 웹에 있는 정보를 컴퓨터가 좀 더 이해 할 수 있도록 도와주는 표준과 기술을 개발하여 시맨틱 검색, 데이터 통합, 네비게이션, 타스크의 자동화 등을 지원하는 것이다. 시맨틱 웹에서 이러한 기능을 지원하기 위해서는 컴퓨터의 지능적인 정보처리가 가능토록 웹 문서 내에

지식 표현을 위한 온톨로지를 삽입하고 지식간의 관계를 설정하며 추론 규칙을 포함 시킨다. 이를 통해서 이 기종간의 상호 운용성을 보장하고 사용자가 원하는 웹 서비스의 발견, 자동적인 웹 서비스의 실행과 동시에 웹 서비스들의 통합과 상호작용을 하여 사용자가 원하는 정보를 찾고 더 나아가 추론이 가능토록 한다(Hendler J.등 2001, 정희준 등 2002).

온톨로지를 이용하면 인터넷 웹 검색에서 검색 시간을 줄일 수 있다는 것은 기존 논문 중 김현희, 안태경등,온톨로지를 이용한 인터넷 웹 검색에 관한 실험적 연구(김현희 등 2003)에서 검증된 바가 있다. 이 논문에서는 국제기구에 대한 온톨로지를 작성하여 기존 웹 검색과 온톨로지 기반의 검색의 시간비교를 하였다. 그러나 온톨로지 적용 대상이 특정 분야에만 한정되어 연구되기 때문에 사용자에게 친숙한 분야의 적용 방안을 찾아보기 힘들다. 따라서 본 논문에서는 온톨로지를 일반 사용자에게 친숙한 분야에 적용하여 좀 더 사용자에게 빠른 정보와 정확한 정보를 제공하려고 한다.

특히 온톨로지는 용어들 간의 동일 의미 관계를 처리할 수 있다. 그렇기 때문에 온톨로지의 적용은 현재 웹 사이트에서 동일 의미여도 서로 다른 데이터로 취급되어 검색한다는 단점을 해결할 수 있는 장점가지고 있다. 따라서 본 논문에서는 온톨로지 개념을 한 특정 분야에 적용시켜 서로 다른 데이터베이스를 갖더라도 온톨로지에 의해 동일한 의미의 키워드에 대한 검색이 가능하기 때문에 현재 웹보다 효율적인 검색이 가능하다는 것을 보이는 데 본 논문의 목적이 있다.

본 논문의 구성은 다음과 같다. II장에서는 시맨틱 웹과 온톨로지에 관련된 이론 및 현재 연구되고 있는 동향에 대하여 언급하고, III장에서는 온톨로지 기반의 정보 검색 시스템 개요와 시나리오를 설명하였다. IV장에서 특정 영역을 설정하여 온톨로지 기반의 정보검색 구현하였다. 마지막으로 결론으로 앞으로의 연구동향을 제시한다.

II. 관련 이론 및 연구

본 장에서는 시맨틱 웹과 온톨로지에 대한 이론적 배경과 온톨로지를 이용한 정보검색에 대한 연구들을 살펴보겠다.

1. 시맨틱 웹(Semantic Web)

현재의 웹은 사람들이 읽고 이해하도록 정보의 외형적인 모습에만 치중되어 설계된 '사람들의 웹'이다. 사람들의 웹에서 컴퓨터의 역할은 웹 문서를 신속하게 파싱하여 문서의 틀을 잡아주고 문서를 다른 웹 페이지와 연결시켜 주는 등 매우 제한적이며, 대부분 정보처리는 사람들의 인지에 의해 이루어진다.

이러한 현재 웹의 문제점을 보완하기 위해 1990년대 말에 Tim Berners Lee에 의해 제안된 것이 바로 시맨틱 웹이다(Berners-Lee T. 등 2001, McIlraith S.A. 등 2001).

시맨틱 웹은 웹상에 존재하는 자료에 의미를 부가하고 사람이 관여하지 않아도 컴퓨터가 자동으로 처리할 수 있는 차세대 지능적인 웹을 말한다. 또한 웹상에 존재하는 정보들은 사람뿐만 아니라 컴퓨터 프로그램 같은 기계들이 해독하고 처리하기 쉽게 표현이 되어 다른 이 기종들 간의 상호 운용성을 갖을 수 있다.

이러한 시맨틱 웹을 실현하기 위한 핵심 기술이 바로 온톨로지이다. 온톨로지는 특정 도메인에 맞는 지식을 개념화(conceptualization)하고 이를 명세화(specification)한다. 즉, 시맨틱 웹에서 온톨로지란 사람과 컴퓨터간의 공유되는 지식을 개념화한 구체적인 형식이며, 개념화와 개념화간의 관계를 표현하는 것으로 정의된다(S.Decker 등 2000).

Fig. 1 에서 보이는 것과 같이 Tim Berners-Lee는 시맨틱 웹이 기존의 웹과 완전히 구별되는 새로운 웹의 개념이 아니라 현재 웹을 확장하여 웹에 올라오는 정보에 잘 정의된 의미를 부여하고 이를 통해 컴퓨터와 사람이 협동적으로 작업을 수행할 수 있도록 하는 패러다임이라고 그 역할을 정의하였다(Berners-Lee T. 등 2001).

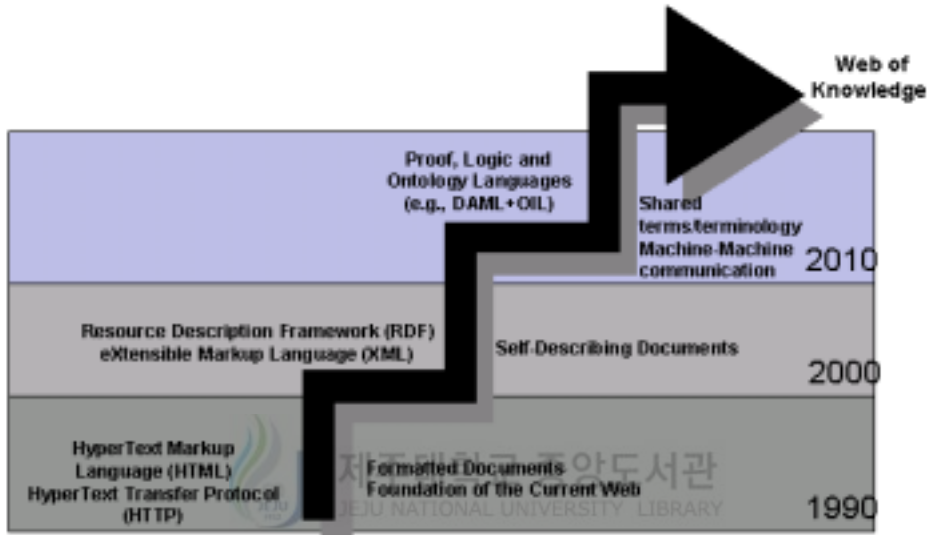


Fig. 1 The development of web

그리고 웹에 의미(semantic)을 부여한다는 것은 사용자 인터페이스를 위한 자연어 처리 기능을 부가하고자 하는 것이 아니라 컴퓨터가 처리하기 용이하게 하고자 선언적인 추가 정보를 부여하는 것을 뜻한다. 즉, 웹 문서에 시맨틱 정보를 덧붙이고 이를 이용하여 소프트웨어 에이전트가 이 의미 정보를 자동으로 추출할 수 있는 패러다임을 조성하는 것이다.

시맨틱 웹을 구현하기 위해서는 다음 Fig. 2 에서 보는 바와 같이 다양한 기반 기술의 발전이 필수적이다.

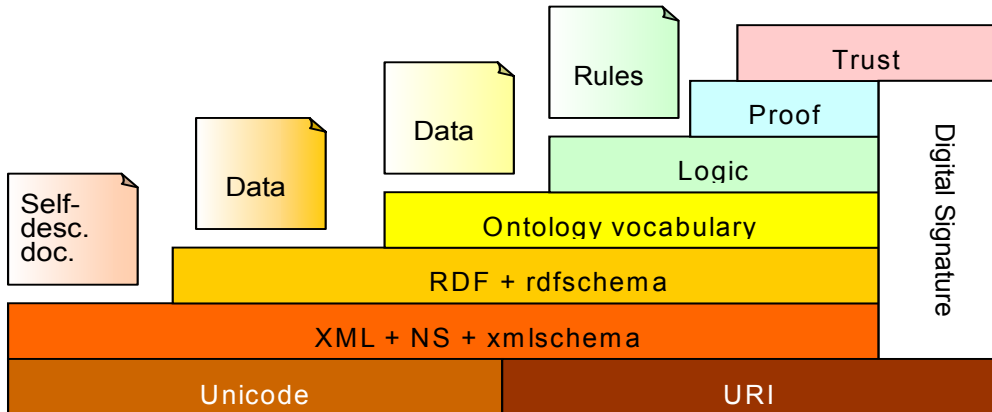


Fig. 2 The Semantic Web "layer cake" presented by Tim Berners Lee at the XML 2000 conference

Fig. 2의 계층구조에서 보면 가장 하위 레벨에서 웹 프로토콜에서 자원을 지칭하기 위한 주소지정(addressing) 방법인 URI가 밑받침되고 이를 기반으로 XML과 Namespace, RDF와 RDF 스키마, 온톨로지의 순서로 연구가 진행되고 있으며 그 위의 계층인 Logic에 대해서는 인공지능의 추론연구를 밑받침으로 일부 연구가 시작되었다. 또한 보다 더 상위 계층인 Proof와 Trust는 시맨틱 웹 정보의 신뢰성과 보안에 관한 내용으로서 아직 개념 정도만 얘기되고 있으며 차후 연구과제로 제시되고 있다.

1) XML/XMLSchema

XML은 SGML의 subset으로 구성된 마크업 언어로서 시맨틱 웹이라는 개념과는 별개로 HTML의 비구조성을 극복하기 위해 이전부터 제시되었던 것이다. HTML에 비해서 XML은 잘 정의된 구조화 문서(well-structured documents)를 작성할 수 있도록 해준다. 즉, 요소(element)라고 불리는 시작 태그와 종료 태그가 반드시 쌍으로 존재해야 한다는 것과 중첩 구조가 반드시 지켜져야 한다는 등의 제약조건이 반드시 만족되어야 한다.

시맨틱 웹과 관련된 XML의 역할은 이러한 구조화된 문서의 생성을 이끌어낸

다는 것도 있지만 태그의 이름을 사용자가 자유롭게 정의할 수 있기 때문에 의미정보를 나타낼 수 있는 태그 이름을 사용할 수 있다는 것이 더 비중을 차지한다.

그렇기 때문에 시맨틱 웹은 XML을 기반으로 하고 있다. 즉, 다양한 정보를 일정한 형식으로 정의하고, 표준화된 방식으로 상호 교환할 수 있는 기술 기반을 제공함으로써, 컴퓨터가 처리할 수 있도록 한다는 것이다. 이러한 특성은 시맨틱 웹의 구현에서 대단히 중요한 기반이다.

그러나 XML은 시맨틱 웹에서 요구하는 수준의 데이터와 의미, 즉 객체 형태의 데이터 구조, 데이터 구조 사이의 관계, 그리고 추론을 위한 규칙을 표현하기 위한 직접적인 방법은 없는 상태이다. 태그의 사용에 대한 규칙은 DTD나 XML Schema에 의해 정의될 수 있으나, 이들 사이의 관계를 컴퓨터에게 알려줄 수 없다. 이는 XML이 문서 구조를 지정할 수 있으나, 문서 구조가 갖는 의미에 대해서는 아무것도 정의하지 못하기 때문이다. 객체 형태의 데이터 구조의 정의 및 데이터 구조간의 관계의 표현을 위해 XML을 기반으로 하여 제안된 언어가 RDF이다.



2) RDF/RDF Schema

RDF는 XML의 문제점을 해결하고 시맨틱(의미)에 초점을 맞추기 위해 제시된 기반구조이다. RDF의 근본을 이루는 개념은 메타데이터이다(S.Decker 등 2000). 메타데이터는 데이터에 대한 데이터, 즉 어떤 객체나 리소스에 대한 서술적인 정보를 말한다.

RDF는 구조화된 메타데이터의 생성, 교환, 재사용 등을 가능하게 해주는 기반 구조이다.

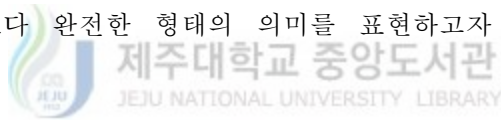
RDF 모델은 리소스(Resource), 특성(Property), 서술문(Statement)의 개념으로 구성된다. 웹 페이지나 웹 사이트 등의 모든 사물(thing)은 리소스로 표현되고, 각 리소스의 특성이나 다른 리소스와의 관계 등을 특성으로 나타낸다.

RDF는 메타 데이터를 처리 기반이 되는 표준으로서, 정보를 교환하는 응용 시

스텝 간의 상호 운용성을 제공한다. RDF 데이터 모델은 웹 자원간의 상호관계를 특성과 값으로 기술한다.

그러나 RDF는 자원에 대한 의미가 리소스와 그 특성값으로 표현되므로 같은 내용(의미)에 대해서는 해석이 하나로만 귀결된다. 그리고 문제점이었던 태그이름의 중첩성과 모호성은 여전히 존재한다. 이러한 문제점은 RDF 스키마로 해결할 수는 있다.

RDF Schema는 RDF 데이터 모델에 유형을 선언할 수 있는 메카니즘을 제공하게 된다. RDF Schema는 XML DTD나 XML Schema와 비교될 수 있다. DTD나 XML Schema는 XML문서의 구조에 대한 제약을 정의하는 반면에, RDF Schema는 RDF 데이터 모델로 표현된 문장의 해석을 위한 정보를 제공하는 것이다. 물론 RDF Schema는 이러한 제약을 표현할 수도 있다. 그러나, XML의 DTD와 마찬가지로 정의할 수 있는 속성의 범위에 많은 제약을 갖고 있다. 이를 더욱 확장하여 보다 완전한 형태의 의미를 표현하고자 한 언어 중 하나가 DAML이다.



3) DAML+OIL

DAML(DAML, 2002)은 웹 페이지에 존재하는 정보를 컴퓨터가 읽고 이해할 수 있도록 하기 위해서 XML 기술을 기반으로 고안된 의미론적 언어로서, 문맥 정의 언어(Context Definition Language)라고 정의된다. DAML이 서로 관련 있는 웹 페이지가 동일한 의미의 태그를 상이하게 표현하거나 또는 한 태그가 서로 다른 의미로 사용되어 발생하는 의미론적 장벽(Semantic Barrier)을 해결하기 위한 수단으로 온톨로지에 의한 연결(Ontological Connection)을 제공한다는 점에서, 평범한 메타 태그와 구별된다.

DAML 이전에 이미 OIL이라고 하는 온톨로지를 정의하기 위한 언어가 개발되었으나, W3C와 DARPA(Defense Advanced Research Projects Agency)에 속한 많은 연구원들은 OIL만으로는 온톨로지를 표현하기에 부족하며 이를 보완할 새

로운 언어가 필요하다는 사실을 인식하였다. 이로 인해 OIL이 RDF나 RDFS와 가지고 있었던 관계보다 의미론적으로 더욱 밀접하게 결합된 DAML+OIL이라는 언어가 탄생하게 되었다.

DAML+OIL은 컴퓨터가 쉽게 읽을 수 있고, 쉽게 이해할 수 있도록 정보를 표시할 수 있으며, 온톨로지를 생성할 수 있다. 의미론적 웹(Web)을 만들고, 이를 프로그램이 자유롭게 사용하기 위해서는 DAML+OIL은 그 핵심 기술이라고 할 수 있다. DAML+OIL의 목적은 웹 페이지, 데이터베이스, 프로그램들간의 의미론적 상호운영성을 가능하게 하는데 있다. DAML+OIL은 다른 메타 태그와는 다르게 서로 관련이 있는 웹 페이지들 간의 온톨로지를 제공함으로써, 기존에 장애였던 의미론적 장벽을 극복할 수 있도록 하고 있다. 그 결과로서 단순히 웹 페이지를 가져와 디스플레이 하던 기존의 방식과는 다르게, 컴퓨터에 의해 읽혀지고, 처리될 수 있는 웹 페이지를 가질 수 있다.

온톨로지를 표현하기 위해 스키마와 구문구조 등을 정의한 언어가 온톨로지 언어(ontology language)인데 현재 DAML+OIL, OWL, Ontolingua 같은 온톨로지 언어가 정의되었다. 이 중에서 W3C에서 표준안으로 제시한 DAML+OIL은 웹 리소스에 대한 시맨틱 마크업 언어이며 W3C의 RDF와 RDF 스키마 표준에 기반을 두고 이들을 확장한 프레임 기반의 온톨로지 표현 언어이다. 기본적으로 DAML+OIL로 표현된 온톨로지는 크게 클래스 요소(class element)와 특성 요소(property element)로 구성된다.

4) OWL

OWL은 DAML+OIL에 기반을 둔 온톨로지 구축 경험을 토대로 개념의 일관성을 확보하여 CLASS와 PROPERTY의 개념 및 그들 사이의 관계가 보다 명료하게 정의 되도록 정리한 온톨로지 언어이다(양정진 등 2003, 정희준 등 2002).

즉, OWL은 시맨틱 웹 형성의 기적적인 온톨로지 개발에 필요한 언어 구조를 재정립한 결과로 볼 수 있다. 따라서 현재 초안 단계인 OWL의 공식 스펙이 국제 표준으로 완성되는 과정에서 다소의 보안 작업을 거칠 가능성을 감안하더라도 차세대 온톨로지 언어로 광범위하게 활용될 것으로 예측된다.

2. 온톨로지(Ontology)

시맨틱 웹의 이상을 실현하기 위해서 반드시 선행되어야 할 기술 기반이 RDF 층 상단에 위치하는 온톨로지(Ontology)이다.

전통적인 철학분야에서 사용되는 온톨로지의 어의는 그리스어인 ontos [being: 존재하는 것]와 logos[word: 단어]에서 유래하였다. 그러므로, 온톨로지는 존재하는 것과 그것의 기본적인 범주를 연구하는 학문이라고 할 수 있다. 온톨로지는 지식 베이스에서 개념 유형이나 범주를 분류하는 분류체계(taxonomy)와 동의어로 쓰이기도 한다. 인공지능(AI) 및 지식 표현 연구 분야에서 오랫동안 사용되어 왔던 온톨로지는 최근 객체 모델링과 XML 등의 분야에서 넓은 관심을 모으고 있다. Tom Gruber는 온톨로지를 “개념화(conceptualization)에 대한 명시적인 명세사항(specification)”이라고 정의하고 있다. 다른 곳에서는 온톨로지를 “a meta-level description of a knowledge representation”이라고 정의 하고 있다. 이는 특정분야에서 사용되는 표준 어휘들의 모음이라 할 수 있다. 에이전트의 수행을 위해서는 도메인의 개념화를 해야 한다. 즉, 도메인 내의 지식을 표현하고 의사소통하기 위한 단어를 제공함으로써 이기종 간의 상호 운용성을 보장하고 지식의 통일화를 시킨다(Hendler J.등 2001, McIlraith.S.A 등 2001).

온톨로지는 어휘 사전의 역할 이외에 지식을 효과적으로 표현하기 위해 정보의 의미를 부여하고, 또 정보들 간의 관계를 설정해 준다. 즉, 온톨로지는 광범위한 도메인에 적용이 가능하도록 표준을 제시함으로써 웹 문서에 나타난 지식을 표현, 공유와 재사용을 그 목적으로 하고 있다. 더 나아가 시맨틱 웹의 목적인 자동 실행과 추론을 하기 위해 그 중요성이 부각되고 있다.

이러한 온톨로지를 독립적인 하나의 중심 구성요소로 보고 이를 개발과 운영의 중심에 놓는 것이 온톨로지 기반의 시스템(ontology-driven system) 또는 시스템 개발이다.

특히 애플리케이션 개발에 있어 주로 데이터 저장소로 많이 이용하는 것이 데이터베이스인데 데이터베이스 스키마의 경우는 개발자에 따라 다르게 개발되어

오직 하나의 애플리케이션에만 적용된다. 그러나 온톨로지는 정보의 의미를 부여하고, 정보들 간의 관계를 설정함으로써 다양한 애플리케이션에 적용할 수 있다.

이러한 장점으로 인해 정보검색, 의료정보와 바이오정보, 인공지능 및 에이전트, 전자상거래, 지능형 인터넷 등 다양한 기술 분야에 적용되며, 이미 분야별로 이에 대한 연구가 활발히 진행되고 있다.

1) 온톨로지 개발 방법

온톨로지 공학에서는 원칙적인 디자인, 수정, 응용, 평가단계를 거쳐 온톨로지를 개발한다(양정진 등 2003).

온톨로지 공학에서 소개하는 온톨로지를 개발하는 단계는 다음과 같다.

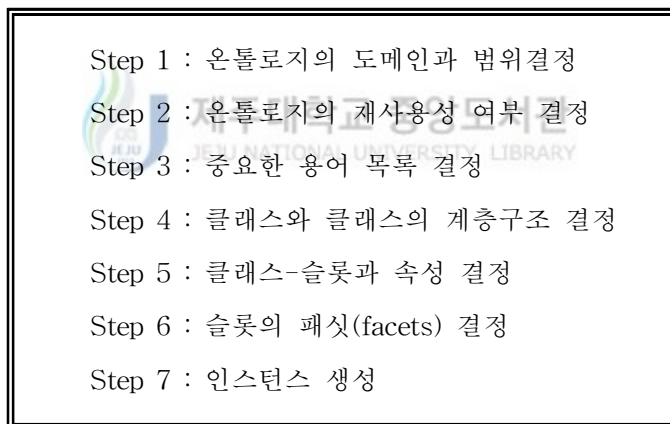


Fig. 3 The step of Ontology development

Fig. 3 에서 보는 온톨로지 개발에 앞서 고려해야 할 점들이 있다. 온톨로지는 도메인의 특성과 온톨로지 사용용도를 고려하여 온톨로지 구성영역을 정하여야 한다. 온톨로지 영역은 온톨로지가 응답해야 하는 질의들을 나열하여 온톨로지 내에 포함되어야 할 내용의 세부사항들을 정한다. 기존의 온톨로지 활용을 경제적으로 하기 위해서 그리고 특정 온톨로지를 이미 사용하고 있는 애플리케이션과 상호적인 처리를 하기 위해서 온톨로지 재사용여부를 확인하여야 한다. 또한

언어학적인 분석과 기계학습 관계를 이해하여 멀티미디어와 같은 다양한 자료에서 자동적으로 학습될 수 있는 부분을 고려한다. 온톨로지 개발 목적에 따라서 메타데이터의 주석이나, 찾는 대상의 구조, 개인별 특징도 고려해야 할 것이다.

2) 개발 도구

온톨로지를 개발하는 툴킷으로 OntoEdit(<http://www.ontoprise.de>), OilEd(<http://oiled.man.ac.uk>), Protege-2000(<http://protege.stanford.edu>), OntoWeb 등이 있다.

OntoEdit

Ontoprise사에서 개발된 OntoEdit는 각 concept 와 instance 를 연결하여 그래프로 만들어준다. 그러나 인터페이스가 사용하기에 그 다지 쉽지 않다.



Fig. 4 OntoEdit

OilEd

영국의 Manchester 대학에서 개발된 OilEd는 Oil 기반의 단순한 온톨로지 편집기이다.

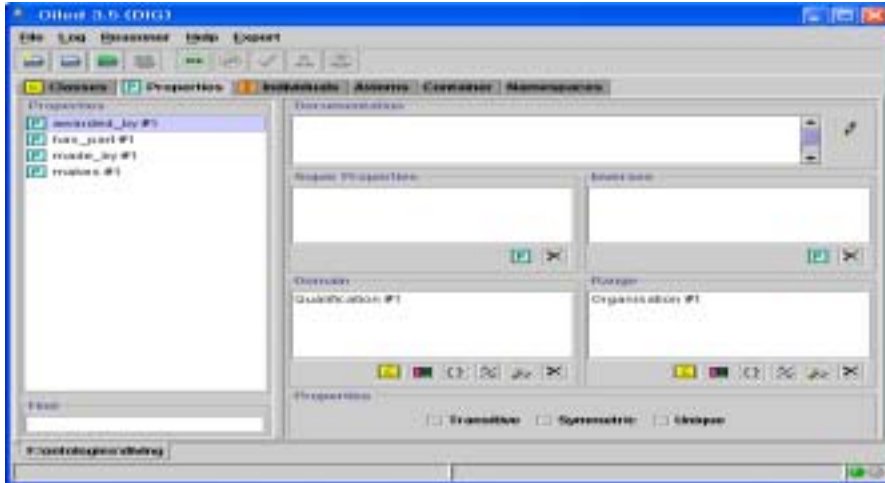


Fig. 5 OilEd

protege-2000



미국의 stanford 대학에서 개발한 protege-2000은 쉽게 subclass 나 instance 를 추가할 수 있는 등 사용하기가 쉽다.

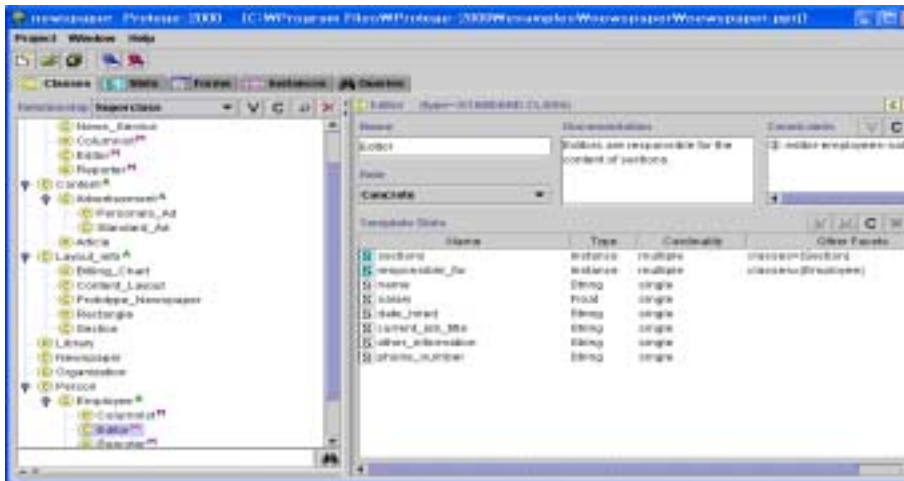


Fig. 6 protege-2000

이외에도 온톨로지에 관련된 프로젝트들이 연구되고 있다.

(<http://peercom.co.kr/contents/semantic.php>)

그러나 각 시스템들은 현재 연구 개발 단계여서 실용화되지 못하고 있고 또한 온톨로지의 모든 용어들을 지원하지 못하는 경우도 있다.

3. 정보검색 연구

현재 온톨로지에 대한 연구는 온톨로지 개발 방법, 이론적 이슈, 전략적 온톨로지 필요성 인식 및 개발, 향상된 톨의 개발등에 방향이 맞추어져 있다. 특히 온톨로지는 특정 분야에 대해서 온톨로지에 대한 연구가 많이 되고 있는데 그중에서 바이오 온톨로지에 관한 연구가 많이 진행되고 있다.

바이오 온톨로지의 용도는 크게 두가지로 나누어 볼수 있다. 첫째는 온톨로지를 하나의 데이터 저장소 혹은 서로 다른 데이터 저장소들 간에 공유할 수 있는 통제어휘(Controlled vocabulary)로 이용하는 것이다. 일반적으로 온톨로지는 데이터 항목을 나타내는 용어들의 의미를 정의함으로써 각 용어들이 하나의 공통된 의미를 가지고 일관되게 쓰일 수 있도록 한다. 또 온톨로지 구조는 질의 조정(query refinement)에도 도움이 된다. GO(Gene Ontology)와 MBO(Molecular Biology Ontology)는 이러한 연구의 대표적인 예들이다.

둘째는 생물정보학 응용시스템들에 지식 서비스를 제공하기 위해 온톨로지들을 사용하는 것이다. TAMBIS, RiboWeb, EcoCyc과 같은 시스템들에서는 데이터베이스 스키마를 제공하기 위해 온톨로지를 이용하고 있다.

생물학분야뿐만 아니라 의학, 자동차 부품 등 여러 분야에서 온톨로지를 적용하는 시스템에 대한 연구가 활발하여 이러한 온톨로지의 장점을 본 논문에서는 전문지식이 필요한 부분이 아닌 일반 사용자를 위한 정보검색에 적용하려 한다.

온톨로지를 기반으로하는 정보검색에 대한 연구도 진행 중이다.

(1) Semantic 환경에서의 자원발견(박재홍 등 2002)

Semantic Web 환경 프로토타입의 도메인을 제주도 여행으로 설정하였고 OWL 형식으로 제주도 숙박 시설과 렌트카, 항공정보에 대한 온톨로지를 구축하고 인스턴스들은 RDF파일로 작성을 하였다. 자원 발견 툴킷은 JENA에서 지원하는 RDQL API를 사용하였다.

RDQL API는 RDF query 이기 때문에 자원들의 관계에 의미를 부여하는 것뿐이지 모든 관계에 대한 의미를 부여하지 않는다. 이 논문에서는 온톨로지를 생성하였지만 온톨로지를 기반으로 검색이 이루어지지 않았다. 즉, 제주도 여행정보를 이용하고 있지만 OWL은 현재 W3C에서 초안인 상태이고 이 논문에서는 온톨로지를 생성했지만 실제 온톨로지 기반의 정보검색은 이루어지지 않고 있다.

따라서 본 논문에서는 온톨로지를 사용하여 온톨로지 기반의 정보검색을 하고 또한 웹 사이트에서 각각의 데이터베이스의 스키마 구조가 같은 의미인 것이어도 서로 다르게 취급하여 검색할 수 없다는 단점을 보완하여 사용자가 효과적으로 정보 검색을 할 수 있도록 하였다.



III. 온톨로지 기반의 정보검색 시스템

1. 시스템 구조

기존의 웹은 분산된 환경 하에서 자료를 공유한다는 목적으로 만들어져 웹 개발자들에 의해 많은 량의 정보가 제공되었다.

그러나, 폭주된 정보량은 사용자로 하여금 한 번 더 필요한 정보만을 걸러내야 하는 번거러움을 주기도 한다. 이것은 HTML은 의미정보를 제공하여 의미적인 검색이 이루어지지 않기 때문이다. 따라서 시맨틱 웹 환경에서 온톨로지를 기반으로 의미적인 검색을 한다면 이와 같은 불편함을 줄일 수 있을 것이다.

그러나 지금 현재 온톨로지는 전문적인 지식을 필요로 하는 분야에만 작성되어 연구되고 있다. 따라서 일반사용자들이 필요로한 정보에 대한 온톨로지 활용 연구도 필요하다. 특히 주 5일 근무가 활성화되면서 주말을 계획되게 보낼려는 사람들이 늘어가면서 기존 포탈 사이트에서는 여행 쪽의 검색이 늘어가고 있다. 그러나 각 사이트의 웹 개발자 마다 서로 다른 구조로 데이터를 저장하기 때문에 같은 의미의 데이터도 서로 다른 의미로 처리할 수 있다. 또한 현재 웹에서는 의미 정보를 표현 해주지 못하기 때문에 같은 필드이름에 해당하는 서로 다른 의미의 값을 가질 수 있는 것을 의미적으로 구분할 수 가 없다 . 예를 들어 ‘초롱이’라는 이름이 사람이름인지 강아지 이름인지 구별할 수 없다.

따라서, 현재 웹에서는 의미적인 검색을 할 수 없다. 그래서 기존 웹 환경에 온톨로지를 적용을 한다면 더 효과적인 검색을 할 수 있다는 것이 본 논문에서의 취지이다.



Fig. 7 System structure

Fig. 7에서 보면 기존 웹에서 웹 사이트를 제공하는 웹 서버는 특정 분야의 지식을 기반으로 의미 구조를 정리하는 온톨로지를 생성하여 자신의 온톨로지를 갖고 있다. 물론 온톨로지는 재사용이 가능하여 기존의 온톨로지를 상속받아 사용할 수도 있다. 기존 웹에 분산되어 있는 온톨로지는 온톨로지 사상(mapping)에 의해 연결되어 지능적인 에이전트들에 의해 의사소통하여 더 나아가 스케줄링이나 가격 협상과 같은 웹 서비스를 제공할 수 있다.

2. 연구 방법

인터넷 정보 검색시 여행정보는 앞에서 언급한바와 같이 많이 검색되는 분야 중 하나이다. 여행정보로는 숙박, 숙박지의 연락처, 호텔정보 등이 있다.

기존 관련 연구에서는 온톨로지를 OWL을 이용하여 작성하지만 실제 검색 시스템은 온톨로지를 기반으로 하여 검색하지 않았고 정보의 의미 관계만을 따져서 검색하는 RDQL을 이용하였다. 본 논문에서는 온톨로지 언어인 DAML+OIL을 이용하여 온톨로지를 작성하고 온톨로지를 기반으로 검색을 가능하게 하기 위해 JENA API를 이용한다.

물론 온톨로지 언어인 OWL도 있지만 실제 OWL을 지원하는 검색 API는 테스트 중이다. 따라서 DAML+OIL을 이용한다.

먼저 온톨로지 작성 단계에 따라 온톨로지를 작성하고, 그에 해당하는 인스턴스 파일을 작성하여 웹 환경에서 사용자가 직접 질의를 입력을 하였을때 사용자가 원하는 정보가 빠르고 정확하게 검색되도록 한다.

IV. 시스템 구현 및 평가

1. 구현 시스템 설계

현재 웹 환경에 온톨로지를 생성하여 테스트 할 수 있는 툴이 아직 연구 중에 있기 때문에 본 논문에서는 테스트베드로 기존 데이터베이스에 저장되어 있는 정보를 이용하여 온톨로지 언어인 DAML+OIL을 이용하여 온톨로지를 작성하였다.

본 논문에서 구현하는 시스템 구조도는 다음과 같다.

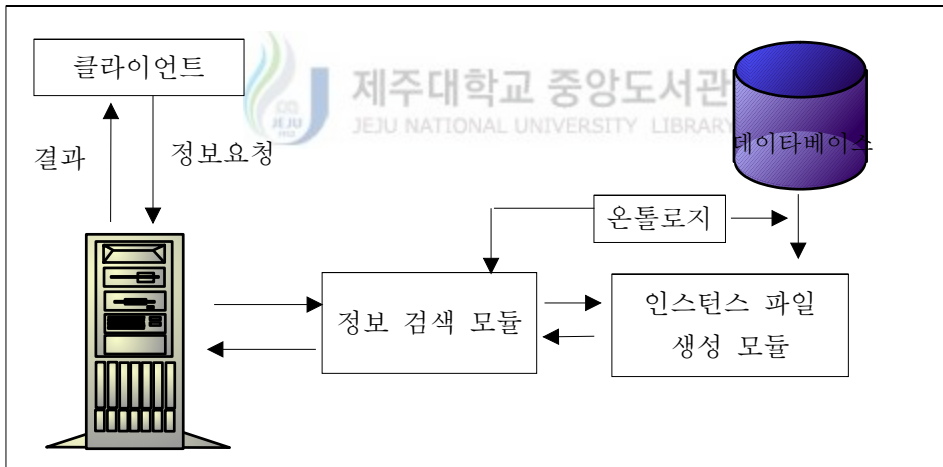


Fig. 8 System design

기존 연구된 논문들에서 보면 온톨로지를 생성했지만 실제 정보 검색할 때는 온톨로지를 이용하지 않고 있다[12].

그래서 본 논문에서는 온톨로지를 직접 정보 검색을 이용하려고 한다.

Fig. 7에서 보면 웹 서버는 온톨로지를 생성하고 그 온톨로지를 기반으로 데이터 베이스에서 정보를 검색하여 인스턴스 파일을 생성하여 저장하고 있다.

인터넷 망에 연결된 클라이언트는 알고 싶은 정보를 클라이언트가 자주 사용 하던 키워드를 이용하여 입력한다. 입력받은 정보를 이용하여 서버내의 정보검색 모듈에서는 온톨로지를 이용하여 키워드의 의미를 파악하고 그에 해당하는 정보를 검색한다. 웹 서버는 그 결과를 클라이언트에게 출력한다.

예를 들어 클라이언트가 신라호텔의 전화번호를 알고 싶다면 정보를 입력할 때 "NAME이 신라호텔인 TEL을 찾아라" 이렇게 입력하여 사용하지는 않을 것이다. 이것은 정보가 어떤 구조로 저장되어 있는지 알 경우이다. 보통 일반 사용자라면 "호텔이름이 신라호텔인 전화번호를 찾아라" 이렇게 입력할 것이다. 또는 한자를 사용하는 사람이라면 "名稱이 신라호텔인 電話番號를 찾아라" 이렇게도 입력할 수 있을 것이다. 두 질의에 대해 모두 처리할 수 있도록 온톨로지에는 호텔이름, NAME, 명칭과 같이 사용자가 사용할 수 있는 단어에 대해 관계를 정의 해 둔다. 정보 검색 모듈은 그 용어에 대하여 온톨로지에 동일 의미인 속성들을 찾고 그에 해당하는 정보를 반환한다.

본 논문에서 다루는 데이터베이스는 제주도청에서 제공하는 여행정보이다.

호텔명	전화번호	주소	홈페이지		
제주도관광호텔	064-14-4971	55, 599	100,000	제주도	www.visitjeju.com
제주도관광호텔	064-14-2525	75	30,000	제주도	www.visitjeju.com
신라호텔	064-799-6222	28	80,000	제주도	www.visitjeju.com
제주도관광호텔	064-799-6999	58	40,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-9911	21	21,000	제주도	www.visitjeju.com
제주도관광호텔	064-741-7908	52	45,000	제주도	www.visitjeju.com
도산호텔	064-746-9154	50	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-741-7914	39	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-741-7909	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-9181	50	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6811	50	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6809	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6808	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6807	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6806	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6805	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6804	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6803	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6802	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6801	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6800	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6799	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6798	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6797	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6796	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6795	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6794	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6793	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6792	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6791	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6790	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6789	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6788	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6787	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6786	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6785	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6784	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6783	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6782	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6781	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6780	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6779	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6778	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6777	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6776	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6775	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6774	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6773	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6772	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6771	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6770	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6769	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6768	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6767	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6766	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6765	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6764	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6763	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6762	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6761	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6760	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6759	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6758	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6757	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6756	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6755	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6754	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6753	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6752	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6751	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6750	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6749	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6748	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6747	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6746	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6745	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6744	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6743	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6742	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6741	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6740	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6739	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6738	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6737	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6736	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6735	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6734	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6733	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6732	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6731	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6730	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6729	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6728	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6727	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6726	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6725	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6724	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6723	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6722	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6721	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6720	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6719	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6718	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6717	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6716	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6715	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6714	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6713	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6712	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6711	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6710	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6709	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6708	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6707	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6706	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6705	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6704	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6703	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6702	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6701	48	30,000	제주도	www.visitjeju.com
제주도관광호텔	064-792-6700	48	30,000	제주도	www.visitjeju.com

Fig. 9 The table of jeju lodge

No	Name	Address	Phone	Fax	E-mail	Website	Category	Level	Price	Status	Remarks
1	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔
2	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔
3	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔
4	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔
5	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔
6	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔
7	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔
8	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔
9	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔
10	신라호텔	제주도 제주시	064-733-1218				호텔	5성급	1,000,000	정상	신라호텔

Fig. 10 The table of jeju tour

테이블에 대상으로 가능한 질의는 다음과 같다.

- 신라호텔의 전화번호는 무엇인가?
- 신라호텔의 TELEPHONE은 몇 번인가?
- 신라호텔의 PHONE은 몇 번인가?
- 신라호텔의 電話番号는 몇 번인가?
- 등급이 특급호텔은 어떤 것들이 있는가?
- 여미지 식물원의 어른 입장료는 얼마인가?
- 어떤 호텔들이 있는가?
- 동굴인 관광지는 어떤 것들인가?

이러한 질의를 처리할 수 있도록 Fig. 10과 Fig.11 같이 온톨로지를 Fig. 3의 단계에 따라 설계하였다.

1) 온톨로지 설계

Step 1 : 온톨로지의 도메인과 범위결정

온톨로지의 도메인은 각각의 숙박과 관광지로 나눈다.

Step 2 : 온톨로지의 재사용성 여부 결정

DAML+OIL에 관련된 온톨로지는 IMPORT받아서 사용하고 숙박과 관광지에 해당하는 온톨로지를 작성한다.

Step 3 : 중요한 용어 목록 결정

다음 Table. 1와 Table. 2 에 보이는 것과 같이 온톨로지에 각 property의 동일한 의미에 해당하는 용어들을 정의해 두었다.

Table. 1 The list of lodge property

PROPERTY	동일의미
NAME	이름, 명칭, 名稱
TEL	TELEPHONE, PHONE, 전화번호, 電話番號
GRADE	등급
AREA	지역, 地役
CITY	위치, 位置
HOMEPAGE	홈페이지, URL

Table. 2 The list of tour property

PROPERTY	동일의미
NAME	이름, 관광지, 명칭, 名稱
TEL	TELEPHONE, PHONE, 전화번호, 電話番號
ADDR	ADDRESS, 주소, 住所
ENTERANCE_ELDER	노인입장료, 老人入場料
ENTERANCE_MAN	어른입장료, 성인입장료, 成人入場料
ENTERANCE_YOUNG	청소년입장료, 青少年入場料
ENTERANCE_CHILD	어린이입장료, 小人入場料
ENTERANCE_SOLDIER	군인입장료, 軍人入場料
CITY	위치, 位置
AREA	지역, 地役

Step 4 : 클래스와 클래스의 계층구조 결정

먼저 LODGE 와 TOUR를 각각의 도메인으로써 CLASS로 설정하였다. LODGE의 SUB_CLASS로 HOTEL, CONDO, MINBAK, MOTEL이 있다. HOTEL의 SUB_CLASS로 SPECIAL, TOURIST, NORMAL이 있고 MINBAK의 SUB_CLASS로 FOREIGN으로 구성하였다.

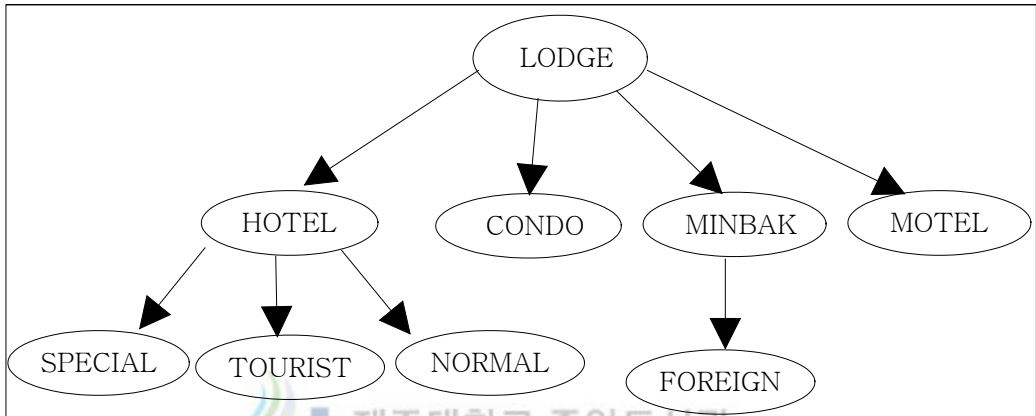


Fig. 11 The tree of lodge class

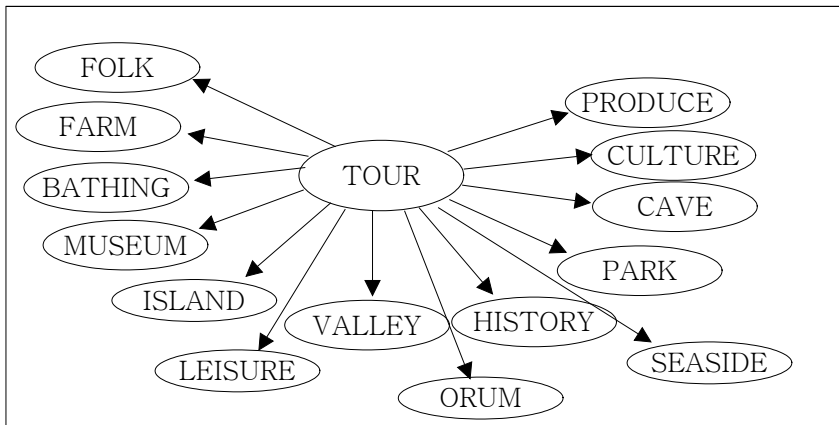


Fig. 12 The tree of tour class

TOUR는 각각의 공통된 분류로 각 관광지를 SUBCLASS로 정의하였다.

Fig. 13은 Fig. 11과 Fig. 12에 대한 클래스도를 OilEd에서의 온톨로지를 구현한 것이다.

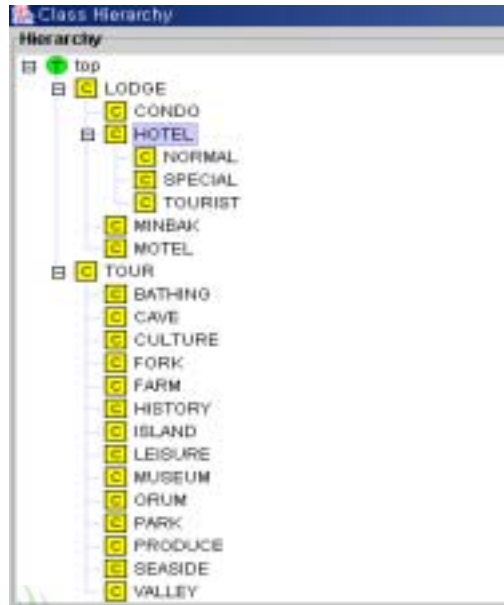


Fig. 13 The tree of ontology at OilEd

LODGE의 SUB_CLASS로 HOTEL, CONDO, MINBAK, MOTEL이 있다. HOTEL의 SUB_CLASS로 SPECIAL, TOURIST, NORMAL이 있고 MINBAK의 SUB_CLASS로 FOREIGN으로 구성하였다.

Step 5 : 클래스-슬롯과 속성 결정

Step 6 : 슬롯의 패싯(facets) 결정

다음은 위의 Table. 1 에서 정리한 property 중 전화번호에 대한 DAML+OIL로 작성된 온톨로지 부분이다.

```

<daml:DatatypeProperty rdf:ID='전화번호'>
  <daml:range

```

```

    rdf:resource='http://www.w3.org/2000/10/XMLSchema#string'/>
  <daml:equivalentTo rdf:resource='#TEL'/>
</daml:DatatypeProperty>
  <daml:DatatypeProperty rdf:ID='PHONE'>
    <daml:range
      rdf:resource='http://www.w3.org/2000/10/XMLSchema#string'/>
    <daml:equivalentTo rdf:resource='#TEL'/>
  </daml:DatatypeProperty>
  <daml:DatatypeProperty rdf:ID='TELEPHONE'>
    <daml:range
      rdf:resource='http://www.w3.org/2000/10/XMLSchema#string'/>
    <daml:equivalentTo rdf:resource='#TEL'/>
  </daml:DatatypeProperty>
  <daml:DatatypeProperty rdf:ID='電話番號'>
    <daml:range
      rdf:resource='http://www.w3.org/2000/10/XMLSchema#string'/>
    <daml:equivalentTo rdf:resource='#TEL'/>
  </daml:DatatypeProperty>

```

이처럼 ‘전화번호’, ‘TELEPHONE’, ‘電話番號’, ‘TEL’, ‘PHONE’을 DAML+OIL의 equivalentTo를 이용하여 동일 의미를 나타낼 수 있다.

Table. 1 와 Table. 2 에서 보이는 것처럼 각 프로퍼티들은 하나의 프로퍼티만 존재하는 것이 아니라 같은 의미를 나타내는 프로퍼티들의 관계를 온톨로지에 작성한다. 따라서 전화번호를 찾고 싶으면 영어권의 관광객은 ‘TELEPHONE’이 나 ‘PHONE’등의 단어를 이용하여 물어볼 것이고 한자권의 관광객은 ‘電話番號’를 입력할 것이다. 이처럼 특별한 언어번역 처리 없이도 의미관계에 의하여 같은 의미를 나타내는 단어들을 지원할 수 있다.

Step 7 : 인스턴스 생성

인스턴스는 RDBMS에 저장되어 있는 테이블에서 STEP 6단계까지 작성된 온톨로지 파일을 이용하여 자동 생성모듈에 의해 생성한다.

2. 시스템 구현

1) 시스템 환경

본 논문에서 구현한 온톨로지 기반의 여행 정보 검색 시스템의 구현 환경은 Pentium3-300MHz 컴퓨터에서 linux9.0 운영체제를 사용하였다. 프로그램 구현언어로는 Java-servlet를 사용하였고 JENA API를 이용하였다. 웹 서버로 jakarta-tomcat-4.0을 사용하였다. 데이터는 제주도 도청에서 Web Application으로 제공되는 여행정보 데이터베이스를 사용하였다.

2) 정보 검색

클라이언트가 질의를 입력하면 서버는 온톨로지와 인스턴스들을 T-BOX와 A-BOX 개념으로 사용하여 CLASS와 속성을 찾고 그에 해당하는 자원을 검색하여 그 결과를 보여준다.

T-BOX는 사용할 수 있는 용어들의 terminology box를 말하고 A-Box는 assertion box로 T-Box에서 정의된 용어들을 사용하여 기술한 사실들을 말한다.

차후에 데이터베이스에 저장되는 정보의 관리는 웹에서 XML/RDF형식의 메타 정보가 들어 있는 페이지들이 활성화 된다면 온톨로지에 작성된 용어가 들어있는 링크정보를 이용하여 데이터는 자동 입력/관리 될 수 있다.

온톨로지

웹에서 제공되는 여행정보 데이터베이스를 기준으로 작성되었다. 작성된 온톨로지의 구문적인 타당성을 검증하기 위해

<http://www.w3.org/RDF/Validator>를 이용하여 검증하였다.

인스턴스 생성 모듈

JDBC를 이용하여 온톨로지에 해당하는 클래스와 속성에 대한 질의 결과를 RDF파일로 저장을 하였다.

온톨로지 기반 검색 모듈

검색 모듈에서는 DAML+OIL을 지원하는 JENA API를 이용하였다.

Fig. 14 화면에서 클라이언트는 원하는 검색 분류를 선택하고 무엇에 대해 어떤 것을 알고 싶은지에 대해서 입력한다. 여기서 중점은 바로 클라이언트가 자주 사용하는 용어를 입력하더라도 검색이 될 수 있다는 것이다. 예를 들어 '신라호텔'의 주소를 알고 싶다고 하면 '무엇에 대해' 란에는 '신라호텔'을 '어떤 것을' 란에는 '전화번호' 또는 'telephone' 또는 'phone'을 입력하더라도 검색 시스템은 모두 전화번호를 찾아서 출력한다.

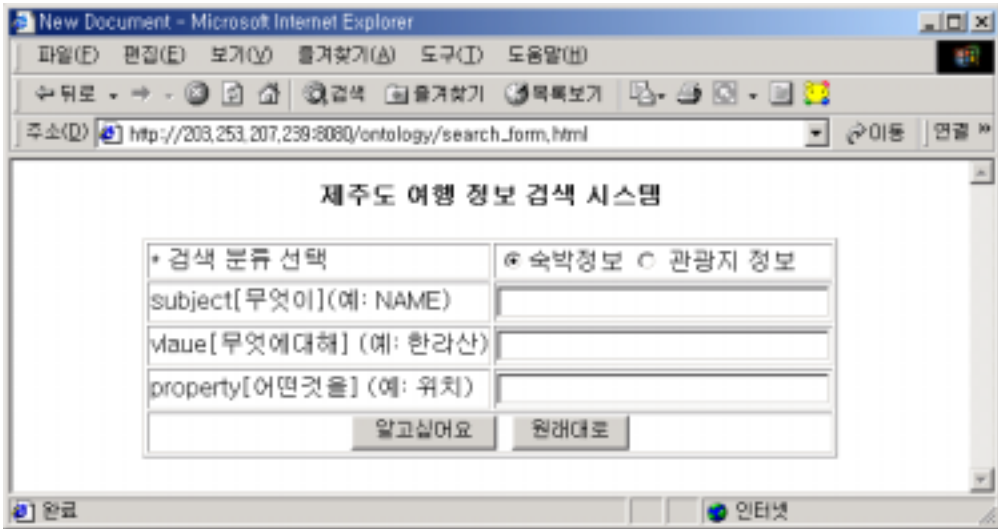


Fig. 14 The screen of information searching for client

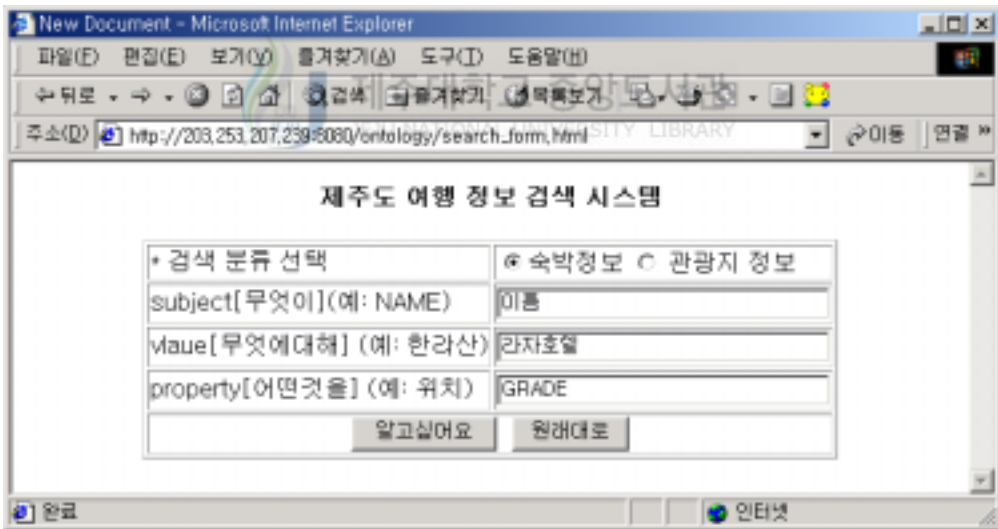


Fig. 15 The lodge information searching

Fig. 15 은 숙박정보 중 '이름'이 '라자호텔'의 'GRADE'를 찾는 화면이다. 기존 웹에서의 검색 방식이라면 검색할 수 있는 필드를 사용자에게 모두 알려 주어서 그 필드를 이용하여 검색해야 할 것이다. 또한 의미적으로 '이름'이 숙박정보에

서의 이름인지 관광지정보의 '이름'인지를 구별할 수 가 없어 의미적인 검색이 불가능하였다. 그러나 위 화면에서는 검색분류에서 검색분야를 선택하여 검색분야에서 쓰일 것 같은 용어를 사용자 중심으로 입력하여 그 용어가 온톨로지에 정의되어 있다면 해당하는 정보를 찾아 보여준다. 만약 그 용어가 온톨로지에 정의 되어 있지 않다면 알 수 없는 정보라고 결과가 나올 것이다.

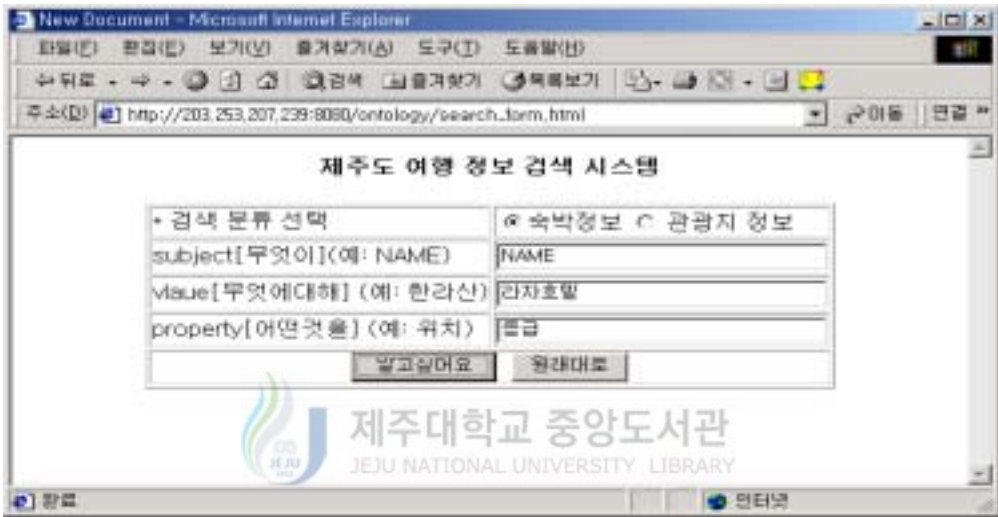


Fig. 16 The searching for user word

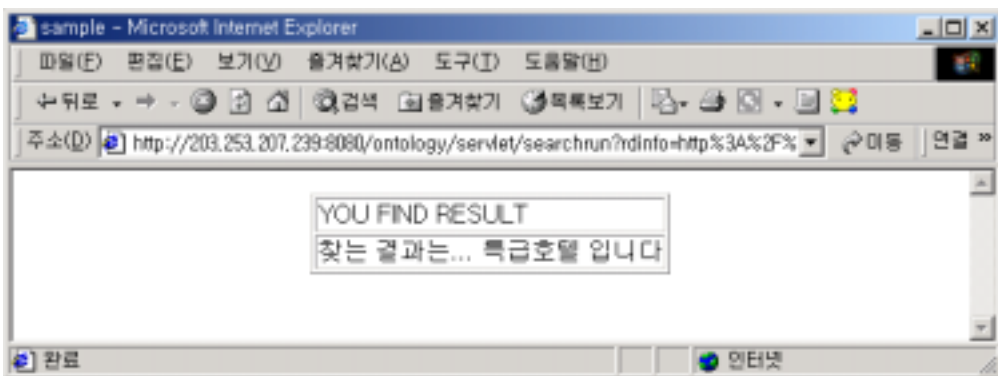


Fig. 17 The result of searching for Fig.15 and Fig.16

Fig. 17에서와 같이 Fig.15와 Fig.16처럼 서로 다른 용어이지만 같은 의미를 나타내기 때문에 모두 같은 결과를 출력하는 것을 볼 수 있다. 또한 Fig. 17에서처럼 한자어에 대한 단어가 온톨로지에 정의되어 있을 경우에 별도의 언어 번역 없이 검색 처리를 할 수 있다.

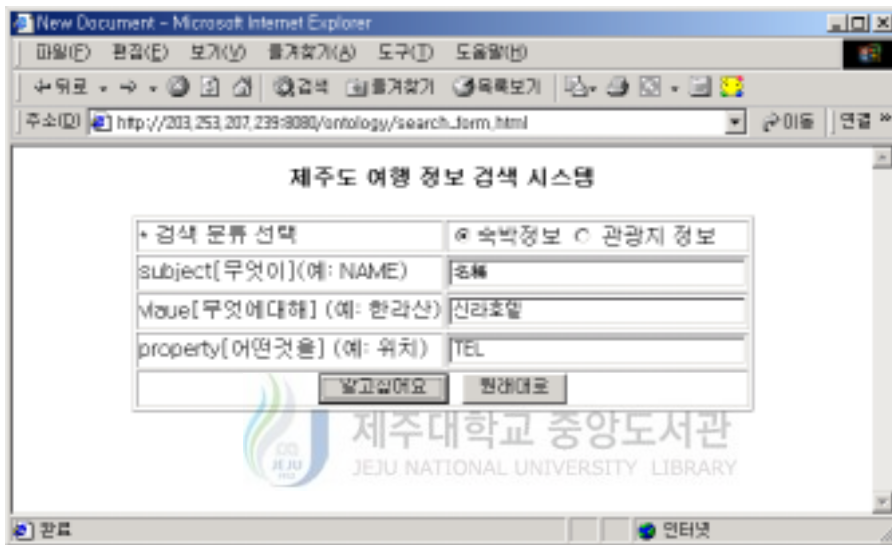


Fig. 18 The case of searching for used a [wordwritteninChinese characters](#)

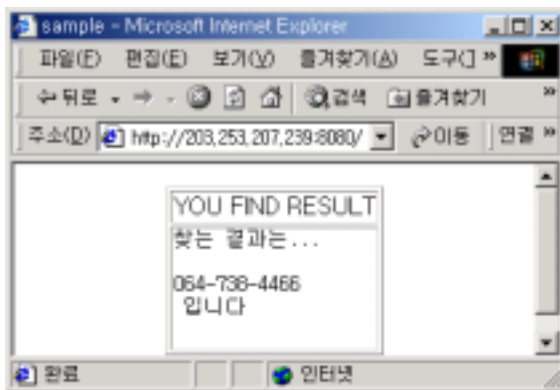
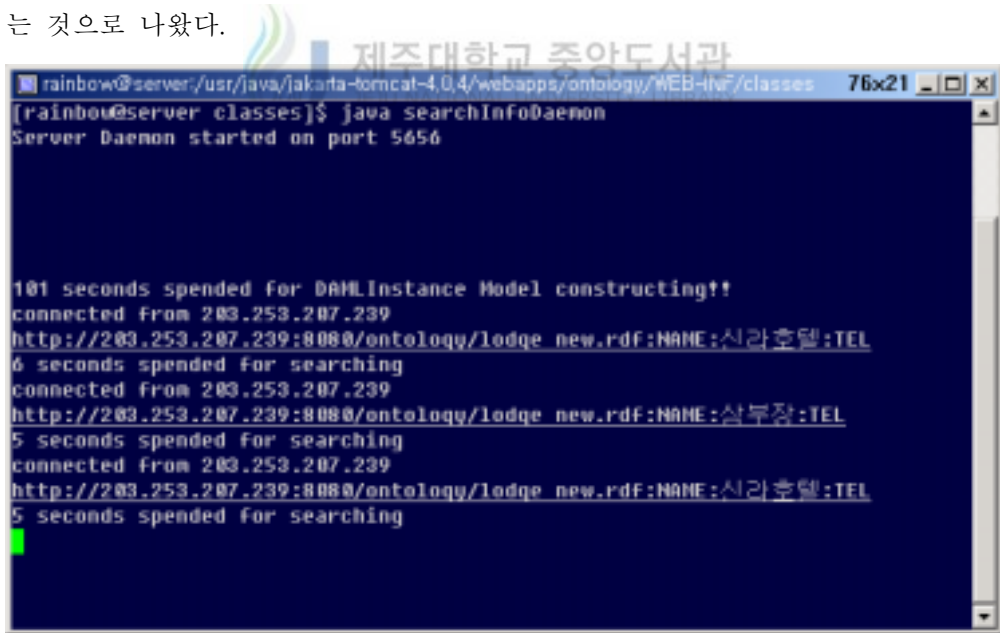


Fig. 19 The result for Fig. 18

3. 평가

의도한 바대로 구현된 온톨로지 기반의 검색 시스템에서 사용자의 용어로 질의를 주었을때 같은 의미를 가지고 있는 서로 다른 표현의 단어에 대하여 모두 같은 처리 결과를 검색하여 출력하였다. 그러나, 10000줄 이상이 되는 온톨로지와 인스턴스 파일을 모두 읽어오면 시스템 부하가 cpu 점유율과 메모리 점유율이 99%이상인 되는 문제점이 발생하였다. 이 문제점은 검색 처리에 대한 시간이 아니라 10000줄 이상 되는 온톨로지와 인스턴스 파일을 메모리에 읽어오는데 시스템 부하가 걸리고 있는 것이어서 읽어오는 부분만 별도의 쓰레드를 생성하여 읽어오게 하였다. 그 결과 다음 Fig. 20에서 보이는 것과 같이 온톨로지와 인스턴스 파일을 읽어오는 시간은 101초가 걸리고 검색을 하는 데는 6초 정도 소요되는 것으로 나왔다.



```
rainbow@server:/usr/java/jakarta-tomcat-4.0.4/webapps/ontology/WEB-INF/classes 76x21 [rainbow@server classes]$ java searchInfoDaemon
Server Daemon started on port 5656

101 seconds spended for DAHLInstance Model constructing!!
connected from 203.253.207.239
http://203.253.207.239:8080/ontology/lodqe_new.rdf:NAME:신라호텔:TEL
6 seconds spended for searching
connected from 203.253.207.239
http://203.253.207.239:8080/ontology/lodqe_new.rdf:NAME:삼부장:TEL
5 seconds spended for searching
connected from 203.253.207.239
http://203.253.207.239:8080/ontology/lodqe_new.rdf:NAME:신라호텔:TEL
5 seconds spended for searching
```

Fig. 20 The comparing of searching time and file loading for using demon thread

V. 결론

현재까지의 인터넷 환경에서는 컴퓨터가 웹 데이터의 의미를 파악하고 인식하는 능력이 부족하기 때문에 사용자의 도움 없이 자동으로 정보를 추출하고 처리하여 저장하고 합성하는 것이 힘들었다. 뿐만 아니라 같은 의미인 데이터도 서로 다른 데이터로 취급하여 개발자들에게는 웹 사이트를 개발할 때 마다 데이터를 스키마 구조에 따라서 다시 저장해야 하는 번거로움을 주고 사용자에게는 어떤 용어로 찾아야 하는지에 대한 혼란을 주었다. 따라서, 본 논문에서는 온톨로지를 이용하여 같은 의미에 해당하는 용어들을 정리하고 전문가가 아닌 사용자도 쉽게 정보를 검색할 수 있는 정보 검색 시스템을 구현하였다. 구현된 시스템은 사용자가 자주 사용하는 용어로 정보를 검색하더라도 같은 의미를 나타내는 용어인 경우에는 같은 정보 검색 결과를 사용자에게 제공할 수 있도록 구현하였다.

온톨로지에 각 속성들의 의미와 관계를 정리하여 일반 사용자가 빠르고 정확한 검색을 할 수 있게 구현하였다. 또한 제주도 여행정보를 이용하여 설계한 온톨로지에 다국어 지원도 가능하도록 같은 의미에 해당하는 다국적인 표현에 대해서도 관계를 정의하였다. 그리하여 외국인 관광객들을 위한 데이터 검색도 가능하게 하였다.

참고문헌

Berners-Lee T., Hendler J. and Lassila, O., 2001, The Semantic Web, *Scientific American*.

Cost R., Finin T., Joshi A., Yun P., Nicholas C., Soboroff I., Chen H., Kagal L., Perich F., Youyong Z. and Tolia S., 2002, ITtalks: a case study in the Semantic Web and DAML+OIL, *IEEE Intelligent Systems*, Vol. 17, No. 1, pp.40-47.

Hendler J., 2001, Agents and the Semantic Web, *IEEE Intelligent Systems*, Volume:16 Issue:2, pp.30-37, March-April.

McIlraith.S.A, son T.C, and Honglei Zeng, 2001, Semantic web services, *IEEE Intelligent Systems*, Volume:16 Issue:2, pp.46-53, March-April.

Introduction to Ontologies on Semantic Web

<http://www.cs.umd.edu/users/hendler/ontologies.html>

Lassila O., 1998, Web metadata: a matter of semantics, *IEEE Internet Computing*, Vol. 2, No. 4, pp.30-37.

McGuinness D., Fikes R., Hendler J. and Stein, L., 2002, DAML+OIL: an ontology language for the Semantic Web, *IEEE Intelligent Systems*, Vol. 17, No. 5, pp. 72-80.

S.Decker, P. Mitra, S.Melnik, 2000, Framework for the Semantic Web: an RDF tutorial, *IEEE Internet Computing*, Vol.4 Issue:6, pp.68-73, Nov.-Dec.

Swartz A., MusicBrainz, 2002, a semantic Web service, *IEEE Intelligent Systems*, Vol. 17, No. 1, pp.76-77.

김현희, 안태경, 2003, 온톨로지를 이용한 인터넷 웹 검색에 관한 실험적 연구, 한국정보관리학회지, V.20, n.1, pp417-455 1013-0799.

박재홍, 임유정, 김도완, 박찬규, 조현규, 2002 .11, Semantic Web 환경에서의 자원발견, 한국정보처리학회 추계학술대회 논문집 제 9권 제 2호.

양정진, 2003.3, 시맨틱 웹에서 온톨로지 공학, 한국정보과학회.

정희준, 유명환, 이강찬, 김성한, 민재홍, 정인정, 2002, 시맨틱 웹 기반의 바이오 온톨로지 시스템 설계, 정보과학회.

정희준, 유명환, 이강찬, 민재홍, 정인정, 2002.11, 시맨틱웹을 위한 온톨로지 시스템의 설계, 정보처리학회 추계학술대회 논문집 제 9권 제 2호.

<http://www.ontoprise.de>

<http://oiled.man.ac.uk>

<http://protege.stanford.edu>



감사의 글

지금 제 옆에서 이 글을 같이 쓰는 예림이가 벌써 5살이 됩니다. 그동안 함께 있지 못하고 제 일에만 바쁘게 살았습니다. 많이 미안한 마음과 저와 예림이를 돌봐주신 어머니께 고마운 마음이 앞섭니다.

물론 제가 대학원 과정을 마칠 수 있도록 힘이 되어주신 모든 분들께도 감사하다는 말씀을 전하고 싶습니다.

우선 저에게 대학원을 권유해 주시고 뒷바라지 해주시는 어머님께 너무 감사합니다. 어머니의 도움이 아니었다면 지금의 저는 없었을 것입니다. 그리고 같은 분야를 먼저 공부하면서 저를 이끌어주고 격려를 해준 하나 뿐인 동생 윤경이에게 감사합니다.

항상 아버지처럼 너그러이 이끌어주시고, 격려해주신 이상준 교수님께 감사합니다. 그리고 논문 심사에 애써주신 곽호영 교수님과 변영철 교수님께도 감사합니다.

항상 옆에서 모든일을 꼼꼼히 챙겨주시며 힘이 되어 준 영민이 오빠께 정말 감사합니다. 그리고 많은 격려를 해주시며 할 수 있다는 자신감을 주신 박충희 선생님과 김휴찬 선생님께 감사합니다. 또한 저에게 자상한 배려를 해주는 의남매인 경복이 오빠께도 감사합니다. 그리고 많은 도움을 준 대학원 동료, 선배님과 후배님 들에게도 감사합니다.

이밖에 미처 감사드리지 못한, 저와 연이 닿았던 모든 분들께 진심으로 감사드립니다.

못난 딸이 이렇게 제 길을 가는 모습을 보고 싶어 하신 하늘에 계신 우리 아빠.... 아빠 영전에 이 논문을 받칩니다.