

시맨틱 웹 환경에서의 온톨로지 기반의 정보검색

정은경* · 김영민* · 변영철** · 이상준** · 박경린***

*제주대학교 대학원, **제주대학교 통신·컴퓨터공학부, *** 제주대학교 전산통계학과

ABSTRACT

온톨로지는 시맨틱 웹을 구현하기 위한 이론이다. 온톨로지는 웹 자원의 효과적인 검색, 통합, 재사용을 위한 새로운 기술이고 사람과 이질의 응용 시스템 간에 지식을 공유하고 재사용하기 위한 방법을 제공한다. 온톨로지의 역할은 특별한 개념을 명세화하는 것이다. 도메인과 일반적인 온톨로지는 정보와 지식 시스템을 분석하고 디자인하는 측면에서 공유, 통합, 재사용될 수 있다.

본 연구는 제주도의 숙박과 관광정보를 위한 온톨로지를 디자인하는데 목적이 있다. 그리고 제안된 온톨로지 기반으로 인터넷 웹 검색 시스템을 만드는데 목적이 있다. 본 연구에 의해 클라이언트가 사용하는 표현은 다르지만 뜻이 같은 동의어에 의해 같은 결과를 얻을 수 있다는 것을 알게 되었다.

1. 서론

현재의 웹은 사람이 보고 잘 이해할 수 있도록 하기 위한 브라우저의 디스플레이 또는 레이아웃 기술에 초점을 맞추고 있다. HTML을 이용한 이러한 표현방식은 문서의 내용과 의미를 나타내는 시맨틱 정보를 표현하기가 어려우며, 따라서 사람이 아닌 프로그램 또는 소프트웨어 에이전트가 자동으로 문서로부터 의미를 추출하기가 어렵다.

이러한 한계점을 해결하기 위해서 1990년대 말에 W3C(World Wide Web Consortium)에서 시맨틱 웹(Semantic Web)을 제안하였다[1,2].

Tim Berners Lee에 의해 제안된 시맨틱 웹은 컴퓨터(Intelligent Soft Agents)가 웹상의 정보를 이해하고, 정보를 창출할 수 있는 웹 환경으로서, 정보의 탐색과 의사결정이 인간이 아닌 컴퓨터가 할 수 있도록 만들어진 Web 환경이다[3]. 즉, 메타데이터의 개념을 통하여 웹 문서에 시맨틱 정보를 덧붙이고, 이를 이용하여 에이전트가 의미 정보를 자동으로 추출할 수 있는 패러다임을 조성하는 것이

다. 부수적으로 의미 정보의 자동 추출뿐 아니라 정보의 확장이나 공유 등도 가능하다.

시맨틱 웹에서 이러한 기능을 지원하기 위해서는 컴퓨터의 지능적인 정보처리가 가능토록 웹 문서 내에 지식 표현을 위한 온톨로지를 삽입하고, 지식 간의 관계를 설정하며 추론 규칙을 포함 시켜야한다[4-7].

시맨틱 웹에서 온톨로지는 특정 도메인에 맞는 지식을 개념화(conceptization)하고 이를 명세화(specification)한다. 즉, 시맨틱 웹에서 온톨로지란 사람과 컴퓨터간의 공유되는 지식을 개념화한 구체적인 형식이며, 개념화와 개념화간의 관계를 표현하는 것으로 정의된다[8].

본 논문에서는 온톨로지를 일반 사용자에게 친숙한 분야에 적용하여 좀 더 사용자에게 빠른 정보와 정확한 정보를 제공하려고 한다.

또한, 각 용어들간의 관계를 온톨로지에 표현하면 동일한 의미의 키워드에 대한 검색이 가능하기 때문에 현재 웹보다 효율적인 검색이 가능하다는 것을 보이는 데 본 연구의 목적이 있다.

II. 관련 연구

2.1 시맨틱 웹

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[8].

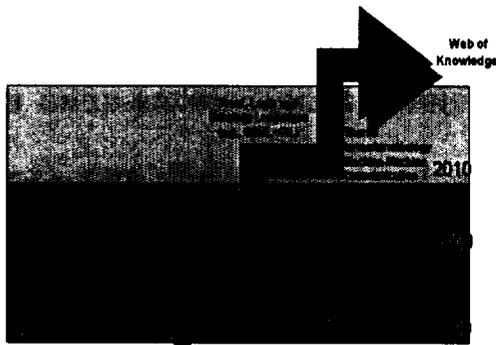


Fig. 1. 진화하는 웹.

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.

2.2 온톨로지(Ontology)

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

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 관심

을 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 정의하고 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.

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.

III. 제안하는 방법의 특징

본 논문에서 제안하는 방법은 Fig. 2에서 보이는 바와 같이 특정 영역의 온톨로지를 생성하고 온톨로지에 정의된 용어를 이용하여 데이터베이스에 저장되어 있는 데이터를 인스턴스 파일로 생성하여 클라이언트가 정보를 원하였을 때 온톨로지기반의 의미적 정보검색이 가능토록 한다.

3.1 온톨로지 생성

시멘틱 웹 환경의 제주도 여행 정보에 대해 크게 숙박, 관광정보로 한정하여 나누었으며 DAML+OIL을 이용하여 온톨로지를 생성하였다.

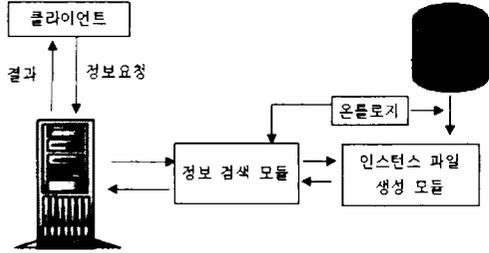


Fig. 2. 시스템 설계.

Table. 1 LODGE의 PROPERTY 목록

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 TOUR의 PROPERTY 목록

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, 靑少年入場料
ENTERANCE_CHILD	어린이입장료, 小人入場料
ENTERANCE_SOLDIER	군인입장료, 軍人入場料
CITY	위치, 位置
AREA	지역, 地役

숙박 정보를 도메인으로 작성된 온톨로지에는 호텔, 콘도, 펜션으로 나누어 클래스와 속성을 정의하였다. 관광정보도 숙박정보와 같이 클래스와 속성을 정의하였다.

온톨로지 생성은 Jena API를 이용하여 DAML+OIL의 온톨로지 생성하는 프로그램을 구현하여 별도로 작성하였다.

Table. 1과 Table. 2은 각 숙박과 여행정보를 도메인으로 하는 온톨로지의 용어들을 정리한 것이다.

다음은 Table. 1의 각 속성의 동의 관계를 DAML+OIL언어에 의해 정의한 온톨로지이다.

```

<daml:DatatypeProperty rdf:ID = '전화번호'>
<daml:range
rdf:resource = 'http://www.w3.org/2000/10/XMLSch
ema#string'/>
<daml:equivalentTo rdf:resource = '#TEL'/>
</daml:DatatypeProperty>
<daml:DatatypeProperty rdf:ID = 'PHONE'>
<daml:range
rdf:resource = 'http://www.w3.org/2000/10/XMLSch
ema#string'/>
<daml:equivalentTo rdf:resource = '#TEL'/>
</daml:DatatypeProperty>
<daml:DatatypeProperty rdf:ID = 'TELEPHONE'>
<daml:range
rdf:resource = 'http://www.w3.org/2000/10/XMLSch
ema#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/XMLSch
ema#string'/>
<daml:equivalentTo rdf:resource = '#TEL'/>
</daml:DatatypeProperty>
    
```

3.2 온톨로지 기반의 정보검색

본 논문에서 구현한 온톨로지 기반의 여행 정보 검색 시스템의 구현 환경은Pentium3-300MHz 컴퓨

터에서 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.

IV. 구현

본 연구에서는 온톨로지 생성, 인스턴스 생성, 온톨로지 기반의 정보검색 이렇게 세가지 부분에 대하여 구현하고 제주도 여행정보검색에 적용하였다.

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 저장하였다.

온톨로지 기반 검색 모듈

검색 모듈에서는 DAML+OIL을 지원하는 JENA API를 이용하였다.

Fig. 3에서 보이는 바와 같이 클라이언트가 자주 사용하는 용어들을 이용하여 정보검색을 하게 된다. 여기서 중점은 바로 클라이언트가 자주 사용하는 용어를 입력하더라도 검색이 될 수 있다는 것이다.

Fig. 6에서와 같이 Fig. 4와 Fig. 5처럼 서로 다른

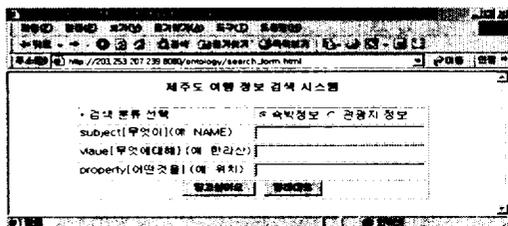


Fig. 3. 클라이언트 정보 검색 화면.

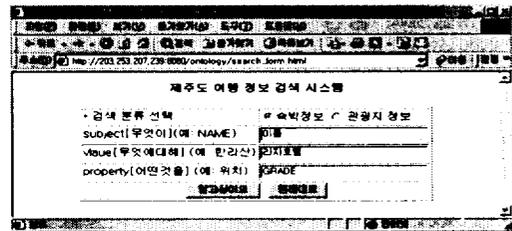


Fig. 4. 숙박정보에서 특정 정보 검색.

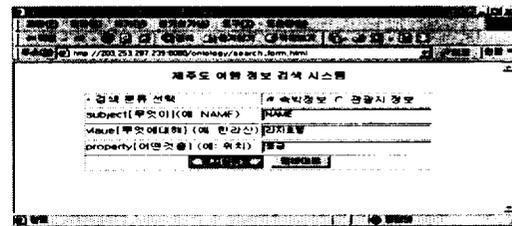


Fig. 5. 사용자 중심의 용어로 검색.

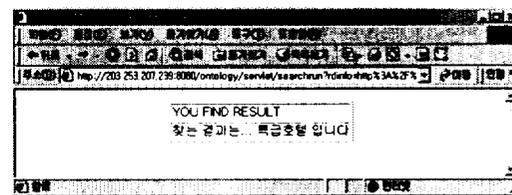


Fig. 6. Fig. 4와 Fig. 5의 검색 결과.

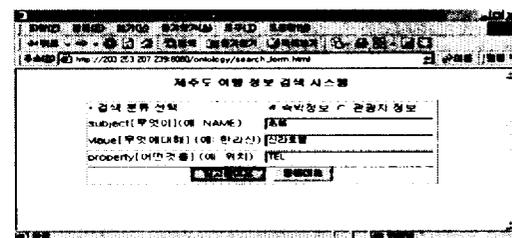


Fig. 7. 한자어를 사용하여 검색한 경우.

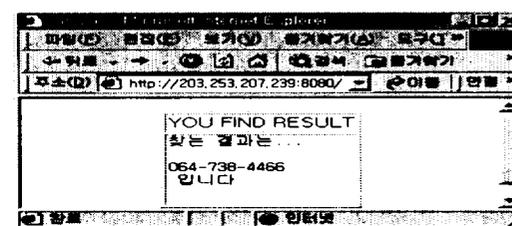


Fig. 8. 한자어 사용에 대한 검색 결과

용어이지만 같은 의미를 나타내기 때문에 모두 같은 결과를 출력하는 것을 볼 수 있다. 또한 Fig. 7에서처럼 한자어에 대한 단어가 온톨로지에 정의되어 있을 경우에 별도의 언어 번역 없이 검색 처리를 할 수 있다.

V. 결론

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.

물론 앞으로 이러한 각 분야의 온톨로지를 작성하는데 자동적으로 생성할 수 있는 온톨로지 시스템이 만들어져야 한다. 또한 분산환경에서 각각 온톨로지가 생성이 되었을때 온톨로지와 온톨로지를 mapping 할 수 있는 방법과 온톨로지를 이용하여 자동 추론을 할 수 있는 Agent 개발도 앞으로의 연구과제로 남아있다.

참고 문헌

- [1] S.A. McIlraith, T.C. Son, H. Zeng. Semantic Web Service. IEEE Intelligent Systems, March/April 2001.
- [2] 정희준, 유명환, 이강찬, 김성한, 민재홍, 정인정. 시멘틱 웹 기반의 바이오 온톨로지 시스템 설계. 정보과학회, 2002.
- [3] 박재홍, 임유정, 김도완, 박찬규, 조현규. Semantic Web 환경에서의 자원발견. 정보처리학회, 2002.
- [4] McIlraith.S.A, son T.C, Honglei Zeng. "Semantic web services", IEEE Intelligent Systems, Volume: 16 Issue:2, pp.46-53.March-April.2001
- [5] J.Hendler, Agents and the Semantic Web. IEEE Intelligent Systems, Volume:16 Issue:2, pp.30-37.March-April.2001
- [6] Introduction to Ontologies on Semantic Web <http://www.cs.umd.edu/users/hendler/ontologies.html>
- [7] S.Decker, P. Mitra, S.Melnik. Framework for the Semantic Web: an RDF tutorial. IEEE Internet Computing, Vol.4 Issue:6, pp.68-73. Nov.-Dec., 2000
- [8] Berners-Lee T., Hendler J. and Lassila, O., 2001. The Semantic Web. *Scientific American*
- [9] DAML+ OIL ontology <http://www.daml.org/ontologies/>
- [10] Jena API <http://www.hpl.hp.com/semweb/>

A study of ontology-based information retrieval on Semantic Web

Eun-Gyoung Joung*, Young-Min Kim*, Yung-Cheol Byun*, Sang-Joon Lee*, and
Gyung-Leen Park**

**Faculty of Telecommunication & Computer Engineering., Cheju Nationall University.*

***Department of Computer science and statistics, Cheju Nationall University.*

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 client use.