

# 온톨로지 기반의 정보검색

정은경<sup>o</sup> 김영민 변영철 이상준  
제주대학교 공과대학 컴퓨터공학과  
{rainbow0707<sup>o</sup>, mincando, ycb, sjlee}@cheju.ac.kr

## The Ontology based Resource Discovery

Eun Gyoung Joung<sup>o</sup>, Young Min Kim, Young Choi Byun, Sang Joon Lee  
Dept. of Computer Engineering, Cheju National University

### 요 약

본 연구에서는 현재 웹 환경에서의 단순한 string matching 검색에 대한 한계점을 해결하기 위해서 온톨로지 기반의 자원검색 방안을 논한다. 테스트 베드로서 제주도의 숙박, 관광정보에 따른 온톨로지를 DAML+OIL언어로 생성하고 Jena에서 지원하는 API를 이용하여 사용자가 원하는 정보검색을 수행할 수 있는 테스트 베드 구축 방안도 제시한다.

## 1. 서 론

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 문제점을 해결하기 위해서 1990년대 말에 W3C(World Wide Web Consortium)에서 시맨틱 웹(Semantic Web)을 제안하였다.<sup>1-2)</sup>

Tim Berners Lee에 의해 제안된 시맨틱 웹은 컴퓨터(Intelligent Soft Agents)가 웹상의 정보를 이해하고, 정보를 참조할 수 있는 웹 환경으로서, 정보의 탐색과 의사결정이 인간이 아닌 컴퓨터가 할 수 있도록 만들어진 Web 환경이다<sup>3)</sup>. 즉, 메타데이터의 개념을 통하여 웹 문서에 시맨틱 정보를 덧붙이고, 이를 이용하여 에이전트가 의미 정보를 자동으로 추출할 수 있는 패러다임을 조성하는 것이다. 부수적으로 의미 정보의 자동 추출뿐 아니라 정보의 확장이나 공유 등도 가능하다.

Semantic Web의 목적은 웹에 있는 정보를 컴퓨터가 쉽게 이해 할 수 있도록 도와주는 표준과 기술을 개발하여 시맨틱 검색, 데이터 통합, 네비게이션, 타스크의 자동화 등을 지원하는 것이다.

Semantic Web에서 이러한 기능을 지원하기 위해서는 컴퓨터의 지능적인 정보처리가 가능도록 웹 문서 내에 지식 표현을 위한 온톨로지를 삽입하고, 지식간의 관계를 설정하며 추론 규칙을 포함 시켜야한다. 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 가능하도록 한다<sup>4-7)</sup>.

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<sup>8)</sup>.

본 연구에서는 기존 웹 응용으로 제공되었던 여행정보를 Semantic Web 환경에서 제공할 수 있도록 숙박, 관광 정보에 대한 온톨로지를 온톨로지 언어인 DAML+OIL을 이용하여 생성하고, 그에 따른 검색을 제안한다.

대부분의 여행정보는 현재 웹 응용으로 가장 많이 제공되는 데이터이다. 또한 데이터베이스에 저장되어 가장 빈번히 갱신되는 데이터중의 하나이다.

이러한 데이터에 온톨로지의 사용은 자동화된 기계처리를 가능하게 하며 여행자에게는 좀더 정확한 데이터를 제공하는 이점을 줄 수 있다.

제안한 방식은 기존의 웹 검색에서 string matching를 이용한 검색보다 온톨로지 기반의 검색으로 사용자가 원하는 좀더 정확한 정보를 사용자들에게 제공할 수 있도록 한다.

## 2. 관련 연구

현재 여러 분야에서 온톨로지를 생성하고 온톨로지를 이용하는 여러 프로젝트들이 연구되고 있다. 예를 들어, 바이오 온톨로지를 이용한 검색시스템<sup>9)</sup>, 온톨로지의 라이프 사이클을 다루는 온톨로지 공학<sup>10)</sup>, RDQL을 이용한 검색, DAML, OIL, OWL 등 W3C의 시맨틱 웹 제안이후로 수많은 연구가 시작되고 있다.

현재까지 개발된 바이오 온톨로지는 다음과 같다.

- RiboWeb ontology  
<http://smi-web.stanford.edu/project/helix/riboweb.html>
- EcoCyc ontology  
<http://ecocyc.PangeaSstems.com/ecocyc/ecocyc.html>
- Schulze-Kremer ontology for molecular biology (MBO)  
<http://igd.rz-berlin.mpg.de/~www/oe/mbo.html>
- The Gene ontology  
<http://genome-www.stanford.edu/GO/>
- TAMBIS ontology  
<http://img.cs.man.ac.uk/tambis>

이외에도 사용 가능한 온톨로지들의 예를 들자면, 온톨로지 라이브러리는 Ontolingua ontology library



그림 5. 온톨로지를 이용한 자원 검색

그림 5에서 보면 사용자가 검색을 원하는 분야를 선택하면 그에 따른 온톨로지가 선택된다. 사용자가 원하는 자원을 검색할 때 사용할 수 있는 어휘들을 알 수 있도록 하기위해 먼저 선택된 온톨로지에서 검색 가능한 속성들을 보여준다. 사용자는 속성들을 사용하여 주어진 쿼리 형식으로 원하는 자원을 검색함으로써 기존의 웹 검색보다 좀 더 정확한 검색을 할 수 있게 되는 것이다.

4. 구현

기존의 웹 검색에서는 단순히 검색어에 입력된 문자열이 존재하는 페이지만을 검색해 왔었다. 검색어에 '한라산 위치'라고 입력를 하고 검색을 하면 사용자가 원하는 것은 한라산의 위치였는데 그림 6처럼 원하지 않는 결과들이 검색되어 나온다.

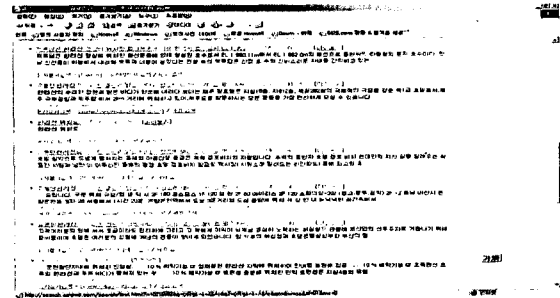


그림 6. 기존 포털사이트에서의 웹 검색

그림 6처럼 기존 포털 사이트에서의 검색 결과를 보면 한라산의 위치뿐만 아니라 '한라산'과 '위치' 두 단어가 모두 들어간 사이트를 검색하게 된다. 그러나, 본 연구에서는 사용자가 원하는 의미가 관광정보 중 한라산의 위치를 컴퓨터가 온톨로지를 이용하여 이해하고 처리하여 검색된 결과에서 다시 사용자가 확인해야하는 처리를 줄일 수 있다.

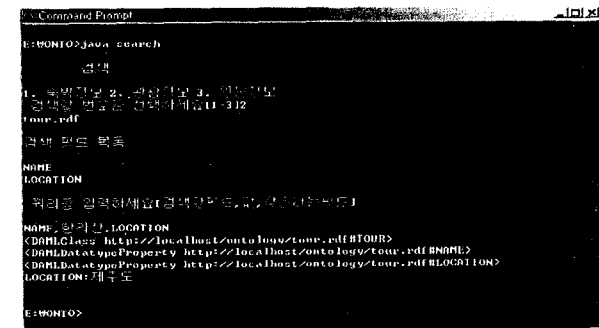


그림 7. 온톨로지 기반의 자원 검색

또한 같은 의미를 갖는 태그 즉, 예를 들어 숙박정보에서 호텔의 이름을 NAME과 TRADE라는 태그를 이용하여 정보가 저장되어 있다면 같은 의미로 이해하여 정보를 검색할 수 있다. 이것은 태그를 다르게 표시하더라도 같은 의미인 것을 정의해 주면 원하는 정보를 찾을 수 있다는 것이 가능하다는 것을 볼 수 있다.

5. 결론

온톨로지를 이용하여 사용자가 원하는 서비스 자원 발견을 수행할 수 있는 제시를 해보았다. 구현 결과 기존의 HTML 방

식의 Web 보다 좀 더 정확한 결과의 추출이 가능하다는 것을 알 수 있다. 진정한 의미의 Semantic Web은 자동화된 자원 발견과 수행, 구성과 상호 운영이라 볼 수 있다. 이와 같은 Semantic Web을 구축하기 위해 프로토타입을 기반으로 하여 소프트웨어 agent가 문서 내용을 파악하여 사용자 제약 조건을 이용하여 자동적인 자원의 발견, 검색 실행, 작업 실행등을 할 수 있도록 해야 할 것이다.

물론 현재 자원들에 대해 메타 데이터를 이용하여 자원의 정보를 annotation되어야 할 것이고, 그에 해당하는 검색을 자동화 할 수 있는 시스템이 제시 되어야 할 것이다.

시맨틱 웹의 출현과 더불어 핵심인 온톨로지의 중요성이 계속 대두되고 있다. 이러한 온톨로지는 새로운 연구 분야라기 보다는 새로운 응용분야라 봐야 타당할 것이다. 끊임없는 연구가 계속 이루어져 온톨로지를 이용한 시스템이 구현되길 바란다.

6. 참고문헌

- 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, 시맨틱 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) 정희준, 유명환, 이강찬, 민재홍, 정인정, 시맨틱웹을 위한 온톨로지 시스템의 설계, 정보처리학회 추계학술대회 논문집 제 9권 제 2호(2002,11)
- 8) 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
- 9) 김인철, 시맨틱 웹과 바이오 온톨로지, 정보과학회지 제 21 권 제3호, 2003.3
- 10) 양정진, 시맨틱 웹에서의 온톨로지 공학, 정보과학회지 제 21제 3호, 2003. 3권
- 11) Robert Stevens, Carole A, Goble and Sean Bechhofer, "Ontology-based knowledge representation bioinformatics", 2001, Briefings Bioinformatics
- 12) W3C RDF Validation Service, <http://www.w3.org/RDF/Validator/>