
온톨로지 활용 감귤 자원 검색시스템 개발

김민철* · 김영익**

Constructing Search System on Botanic resources (Mandarine) by using Ontology

Kim, min cheol* · Kim, Young Ick**

요 약

정보통신 인프라의 급속한 발달 및 인터넷 사용 환경의 편의성 제고 등으로 인하여 웹의 사용이 일반화되고 있을 뿐만 아니라 현대인의 삶에 있어 필수 불가결한 요소로서 자리잡아가고 있다. 하지만 너무나 방대한 정보의 범람으로 인하여 필요한 정보를 정확하고 편리하게 획득하는 기술을 확보하는 것이 인터넷 활용에 있어서 가장 중요한 요소로 떠오르고 있다. 이러한 문제의 해결책의 일환으로서 현재의 웹기술에 시맨틱웹 기술을 접합한 새로운 시스템이 활발히 개발되고 있으며, 시맨틱웹 개발의 현실화 기반은 온톨로지를 구축 함으로서 이루어지게 된다. 이 연구는 식물자원의 온톨로지 기술을 활용한 에이전트 개발에 목표를 두고 있으며, 향후 다양한 자원 정보의 개발 및 축적에 연계하여 효율적으로 활용이 가능할 것이다.

ABSTRACT

Because of the rapid development of information technology and easiness in using internet, Web becomes indispensable utility in every day life and various information has been shared in internet. However, caused by the flood of information in the Web, the skill of acquiring required information with easy becomes one of the most urgent on utilizing internet. As a solution for the requirement, Semantic Web, which upgrades present web technology by using semantic skills, has been researched actively and Ontology is regarded as the basis of constructing semantic web system. The aim of this research is on developing agent using ontology skills for botanic resources. Moreover, the system can be used for developing and accumulating various resources effectively.

키워드

시맨틱웹, 온톨로지, 지식, 인터넷, 에이전트 Semantic Web, Ontology, Knowledge, Internet, Agent

I. 서 론

1990년대부터의 정보통신의 발달 및 인터넷 사용 환경의 편의성 제고 등으로 인하여 웹의 사용이 일반화되고 있을 뿐만 아니라 각종 다양한 정보들이 웹을 통해서 공유되고 있다. 즉, 각종 멀티미디어 정보들이 디지털 화되

고 대용량의 데이터 교환이 가능해졌으며 단순한 텍스트 뿐만 아니라 이미지, 음성, 동영상 등 다양한 형태의 자료를 공유할 수 있게 됨에 따라서 다양한 정보의 공유와 그 방대한 양으로 인하여 웹은 정보의 바다로 부각되었다 [1]. 하지만 일반화된 인터넷의 사용으로 너무나도 방대한 정보가 웹상에 제공이 되고, 그로 인하여 필요 정보의

* 제주대학교 경영정보학과 조교수

** 제주대학교 관광과 경영경제연구소

적절한 선택 및 추출의 어려움이 대두 되어있는 상황이다. 이 문제의 해결이 인터넷 활용의 비약적 발전을 이끄는 하나의 화두가 되고 있다. 인터넷 활용 효율성 증대 측면의 대안으로 여겨지는 것 중 하나가 바로 온톨로지를 활용한 시맨틱 웹이며, 본 연구 개발에서는 식물 분야의 특정 도메인의 지식체계 온톨로지 서버에서 시스템 통합 기술 개발 및 응용을 위한 에이전트 프로토타입 시스템 개발 목표에 초점을 두고 있다.

II. 시맨틱웹 및 온톨로지 이해

2.1. 시맨틱웹의 필요성

현재의 웹은 사이트 개발자가 일방적으로 정해놓은 검색필드만을 검색할 수 있다는 문제점과 함께 데이터에 원하는 의미를 부여하여 사용자가 원하는 정보를 자동으로 추출하고 정보의 속성을 나타낼 수 없다는 단점을 가지고 있다[2]. 이러한 문제점을 해결하기 위한 노력의 일환으로서, '시맨틱 웹'(Semantic Web)'이라는 개념에 이르게 되었다. 기존 인터넷은 정보과부하의 상태에 도달 했으며 [3] 이의 해결책으로서 팀 버너스리(2001)는 정보의 탐색과 의사결정의 인간이 아닌 컴퓨터가 할 수 있도록 만들어진 웹 환경인 시맨틱웹을 제창하게 된 것이다. 즉 시맨틱 웹은 기계를 이용한 정보의 탐색과 의사 결정을 지원하는 진정한 의미의 인터넷 환경이라 할 수 있는 것이다 [4].

2.2. 시맨틱웹 개요

시맨틱웹을 한마디로 정의 하자면, '컴퓨터가 정보의 의미를 이해하고 의미를 조작할 수 있는 웹'이라 말할 수 있다. 검색엔진과 에이전트 프로그램 등이 데이터 처리수준을 넘어 웹에 있는 정보의 의미를 이해한다는 것은 정보의 리소스들 사이의 연관성을 표현해 주는 것을 말한다. 하지만 시맨틱 웹이라는 것이 기존의 웹 시스템을 완전히 바꾸어 놓는 개념은 아니다. 팀 버너스 리는 시맨틱 웹이 기존의 웹과 완전히 구별되는 개념이 아니라 현재 웹을 확장하여 웹에 올라오는 정보에 잘 정의된 의미를 부여하고 이를 통해 컴퓨터와 사람이 협동적으로 작업을 수행 할 수 있도록 하는 패러다임이라 그 역할을 정의 하였다[5].

2.3. 온톨로지 개요

온톨로지란 원래 철학용어로서 존재론으로 번역이 된다. 하지만 전산학에서 온톨로지라는 말이 사용이 되기 시작하였고 Gruber는 온톨로지를 "공유된 개념화(Shared conceptualization)에 대한 정형화되고 명시적인 명세(formal and explicit specification)"라고 정의 하였다[6]. 다시 말하자면 "온톨로지는 공유하기 위한 개념들의 개념화를 형식적이고 명백하게 설명해 놓은 명세서라고 할 수 있다. 따라서 온톨로지의 역할은 해당 분야에서 사용되는 어휘의 해석과 의미의 한계를 정하는데 있다. 이러한 의미에서 온톨로지는 사용되는 어휘사이의 논리적 형식보다는 어휘의 내용을 중심으로 창조된다고 할 수 있다. 즉 온톨로지란 시맨틱웹을 구축하기 위한 필수적 기반인 것이다.

표 1. 중요용어 목록의 예

Table 1. The Example List of Important Terms

Property	동일의미
Admission-Fee	입장료, 관망료, 入場料
Adult	성인입장료, 어른입장료, 일반입장료, 성인입장료
Youth	청소년입장료, 중고생입장료, 청소년입장료
Child	소인입장료, 어린이입장료, 소인입장료
Soldier	군인입장료, 軍人入場料, 군경입장료
Elder	노년입장료, 老人入場料, 경로입장료, 노인입장료
Name	이름, 名稱, 명칭
Tel	전화, 電話番號, Telephone, Phone, Tel
Price	가격, 價格, 요금, 料金

2.4. 온톨로지의 동향

South China Morning Post(2002.5.28) 기사에 따르면 온톨로지 관련 웹 시장의 규모도 2001년 21.8억 달러에서 2006년 63.5억 달러로 증가가 예상되고 있다. 이러한 웹 기술 시장 규모의 증가 추세를 고려한다면 이 분야 시장은 향후 지속적 증가가 예상된다. 특히 2003년 웹 콘텐츠 제작에 XML(Extensible Markup Language) 비중이 50%에서 2007년에는 웹 온톨로지 언어 비중이 오히려 더 앞서갈 전망이다.

III. 개발 방법

3.1. 시스템 구축모델

3.1.1. 온톨로지 구축 방법론

Noy & McGuinness(2001)가 제안한 방법론(Ontology Development 101)에 근거하여 구축 한다.

3.1.2. 온톨로지 구축단계

- step 1 : 온톨로지 도메인과 범위 결정
- step 2 : 온톨로지 재사용 여부 결정
- step 3 : 중요한 용어 목록 결정
- step 4 : 클래스와 클래스의 계층 구조 결정
- step 5 : 클래스-슬롯과 속성 결정
- step 6 : 슬롯의 패시(facets) 결정
- step 7 : 인스턴스 생성

3.1.3. 온톨로지 저작 툴

Protege 2000(Stanford Medical Informatics)

3.1.4. 시스템 구성도

기본적으로는 정해진 감귤 온톨로지(RDF)에서 정보검색 모듈인 Jena 시스템에 의해 여러 지식 등을 검색하도록 설계하였다. Jena는 RDF를 제어하기 위해 오픈소스 기반의 JAVA API로서 RDF의 생성, 파싱, 검색, 추론이 가능하며 사용방법이 손쉬운 장점이 있다.

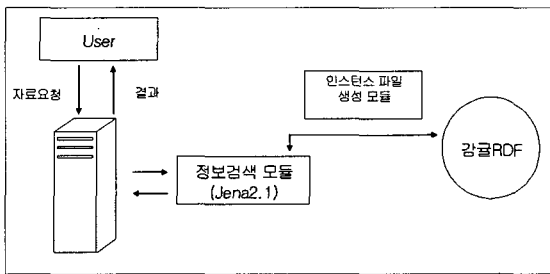


그림 1. 시스템 구성도
Fig. 1. Structure of System

3.2. 에이전트 개발

3.2.1. 개발계획 수립

본 연구 및 개발의 원활한 수행을 위해서 연구의 내용, 연구 개발의 절차, 프로그램의 설계 등을 구체적으로 기술하는 연구 계획서를 작성하였고, 이를 기초로 하여 에이전트 소프트웨어를 효과적으로 개발하기 위한 개발 계획 및 전략을 수립하였다.

3.2.2. 기존 에이전트 분석

최근 에이전트시스템은 랭킹 시스템, 중복검색 결과 제거, 전문 검색 등 다양한 분야의 정보들 중 사용자에게

더욱더 빠르고 정확하게 의미 있는 정보를 전달하는 기술로 계속 발전하고 있다. 특히 최근에는 자연언어처리 기술을 이용한 의미기반 정보검색 기술이나 질의응답시스템 등이 개발되어있다. 기존의 정보검색 기술 중의 하나였던 디렉터리 서비스, 키워드 기반 검색 등은 문서의 의미를 판단할 수 있는 기술은 아니었으므로, 그만큼 사용자의 의도에 맞는 정답 문서를 제공하지는 못하였다. 따라서 사용자의 질의를 확장하기 위한 색인어 확장용으로 시소러스를 이용하여 사용자에게 의미 있는 문서를 제공하거나, 자연어 질의를 형태소 분석, 구문 분석, 의미 분석 단계를 거쳐 사용자의 의도를 분석하여 지식베이스와 추론 기법을 이용하여 의미 있는 문서를 제공하는 정보검색 기술이 이용되고 있다. 그러나 이러한 정보검색 기술 또한 아직까지 사용자의 질의 분석에서의 오류, 문서에서의 정확한 의미 분석 처리 문제, 단어 의미 중의성(word sense ambiguity) 문제 등 여러 가지 문제 때문에 실생활에서 사용하기에는 더 많은 시간이 필요할 것으로 보인다.

3.2.3. 구조 분석 및 화면설계

프로그램 구조는 일반적인 에이전트 프로그램의 검색 형식이 아닌 RDF 문서 검색을 기반으로 하였다. 처음 사용자도 쉽게 검색을 할 수 있도록 인터페이스 부분을 강화 시켰다.

3.2.4. 프로그램 개발

프로그램 개발은 웹 서버가 여러 플랫폼에서 작동하므로 프로그램 도구로 대부분 자바를 사용하였다. 그리고 RDF 검색을 위하여 Jena가 이용이 되었다. 프로그램 개발에 사용된 프로그래밍 도구로는 Java를 중심으로 하였고 인터페이스 부분은 JSP를 이용하였다. 그리고 개발에 사용된 컴퓨터 시스템은 웹 서버 환경은 인텔 펜티엄 4를 사용하였고 서버시스템은 Windows2000 서버를 설치하였으며, 웹 서버로는 톰캣을 사용하였다.

3.3. 시스템 구축 방법 및 환경

표 2. 개발환경 및 도구 내역
Table 2. Development Environment and Tools

개발환경	개발 S/W
Windows2000server	JDK 1.4.2_02
Java,Servlet,Jsp	Tomcat 4.0.6
자바 MVC모델링 적용	Jena 2.1

IV. 프로세스 구성

4.1. 온톨로지 구축

본 과제에서는 제주 지역 내 특산물이라는 특수한 환경을 고려하여 감귤이라는 한정된 도메인 내에서 한방 효능에 초점을 맞추어 개발하였다. 따라서 <그림 2>는 감귤 온톨로지 화면으로서 감귤의 기본적인 자료는 공식적인 감귤 웹 사이트(감귤박물관)[7] 및 관련 책자 등을 활용하였다.

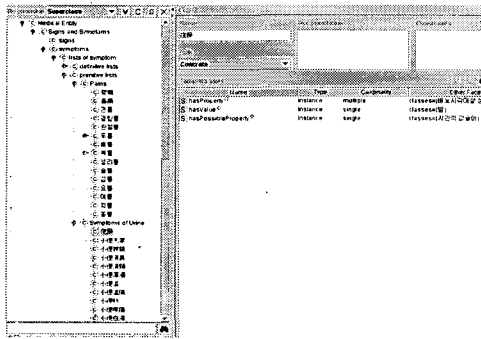


그림 2. 상위 카테고리인 한방 온톨로지
Fig. 2. Oriental Medicine Ontology of Upper Category

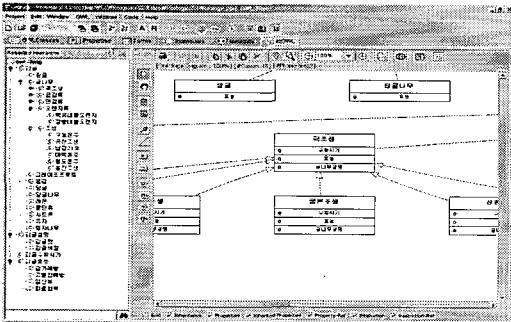


그림 3. 감귤 온톨로지
Fig. 3. Mandarin Ontology

온톨로지 구축은 Protege-2000을 이용하여 구축하였고 이를 토대로 <그림 4>와 같이 감귤의 인스턴스가 생성되었다. 감귤의 맛(신맛, 아주신맛, 약간신맛) 그리고 감귤의 색(노란, 진한노란)등의 특징 별로 분류하였다.

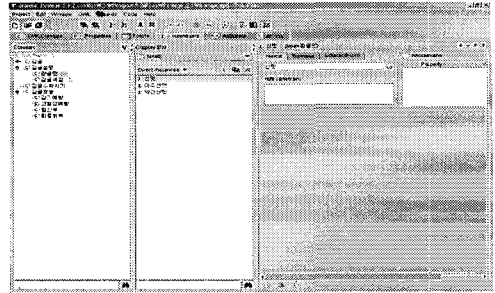


그림 4. 인스턴트의 생성
Fig. 4. Generated Instance

4.2. RDF 생성

온톨로지 기반으로 하여 감귤의 종류 및 감귤의 효능에 대한 RDF 문을 만들었다. 이는 OWL로 표현된 온톨로지의 예로서 감귤과 만감류가 감귤의 서브클래스이며, 만감류는 수확시기와 껍질무설명, 효능의 특성을 갖는다는 것을 정의한다.

```

<?xml version="1.0" ?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/unnamed.owl#"
  xml:base="http://www.owl-ontologies.com/unnamed.owl#"
  >
  <owl:Ontology rdf:about="" />
  <owl:Class rdf:ID="감귤예발" />
  <rdf:subClassOf />
  <owl:Class rdf:ID="감귤효능" />
  </rdf:subClassOf />
  </owl:Class />
  <owl:Class rdf:ID="만감나무" />
  <rdf:subClassOf />
  <owl:Class rdf:ID="감귤" />
  </rdf:subClassOf />
  </owl:Class />
  <owl:Class rdf:ID="극조생" />
  </rdf:subClassOf />
  </owl:Class />
  </rdf:RDF />
  
```

그림 5. OWL 표현
Fig. 5. Expressed with OWL

4.3. 에이전트

이 파일 프로그램의 결과인 <그림 5>는 사용자가 원하는 정보를 입력할 수 있는 화면이다. 인터페이스 예 <그림 6>에서 볼 수 있는 것처럼, 사용자에게 친밀감을 줄 수 있는 인터페이스가 많이 사용이 되고 있다. 본 화면 또한 사용자의 친밀감을 높이고, 감귤 온톨로지 검색을 위한 인터페이스 에이전트로서 도메인의 주제를 명확히 인식시키기 위하여 감귤 닥터라는 이미지화된 모습으로 구성하였다.

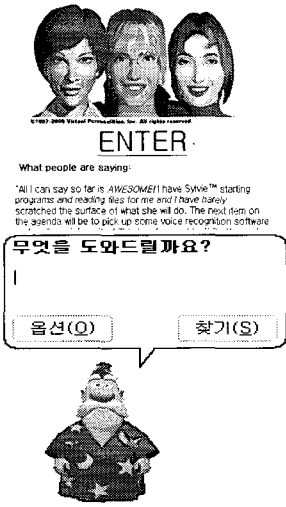


그림 6. 아래 에이전트 인터페이스에
Fig. 6. Below. Example of Agent Inter-pace

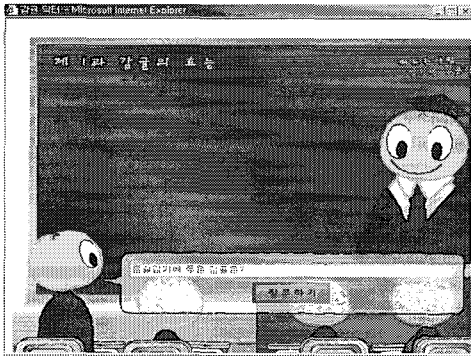


그림 7. 검색 질문 화면
Fig. 7. Screen of Searching Question

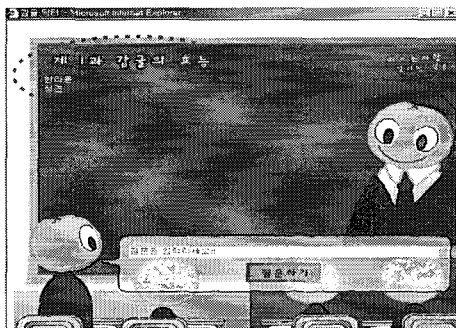


그림 8. 검색 답변 화면
Fig. 8. Screen of Searching Answer

구조 분석 및 화면설계 프로그램 구조는 일반적인 에이전트 프로그램의 검색 형식이 아닌 RDF 문서 검색을 기반으로 하였다. 처음 사용자도 쉽게 검색을 할 수 있도록 인터페이스 부분을 강화시켰다.

V. 개발 효과

5.1. 기술적 측면

에이전트시스템은 사용자가 원하는 지식 정보를 얼마나 정확하고 빠르게 검색하여 의미 있는 지식 정보를 제공할 수 있는가에 따라 시스템의 성능과 평가가 좌우된다고 할 수 있다. 개발된 에이전트시스템은 사용자 측면에서 다양한 정보 검색을 통해 의미 있는 지식 정보를 제공할 수 있도록 하는 매개체 역할을 담당하며, 정보 검색에서의 질의 처리 성능, 대용량 정보 처리 방법, 색인 기술 등과 같이 지식 정보에 대한 표현, 저장, 조직, 접근 등의 기술적 처리의 효율성에 초점을 맞추고 있다.

따라서 많은 기업들이 웹 기술의 진보를 위한 최근 기술 사항 이전(특히 온톨로지 기술의 적용 가능성 탐색)을 요청하고 있는 상황에서 이번 연구는 기술적으로 여러 효과를 지닌다. 더욱이 온톨로지 분야는 바이오 지식을 구축하는 데 있어서 매우 효율적인 틀이 되므로 본 연구 개발은 동종 또는 유사 산업의 효율성 제고의 의의를 지닌다.

5.2. 경제·산업적 측면

현재 일부 에이전트는 사용자가 원하는 정보를 정확히 전달하지 못하고 그로인하여 사용자에게 정보검색 및 활용에 있어서의 부담을 주고 있는 상황이다. 온톨로지가 이용한 에이전트 시스템은 이러한 문제점이 중복검색과 의미론적인 검색을 활용하여 극소화시킬 수 있으므로, 사용자가 원하는 정보를 신속하고 정확하게 전달할 수 있다.

최근 시맨틱 웹 개발에 대한 필요성이 가중되고 있으며 시맨틱 웹의 핵심 기술인 온톨로지 분야를 개발할 전문 인력이 없는 상황이다. 따라서 이번 연구 및 개발을 활용함으로써, 온톨로지 기술을 적용해야 할 에이전트 시스템을 개발할 수 있는 관련 전문가의 양성 효과를 기대할 수 있다. 또한 향후 의료의 원격 서비스 및 각 병원의 웹 서비스 활용 측면에서 유용성을 지니며 의료정보의 비대칭 해소 등의 긍정적 측면을 지니고 있다.

VI. 결 론

시맨틱 웹에서의 온톨로지 활용 기술은 웹 서비스 관련 분야뿐만 아니라 생물정보학, 자연언어처리, 데이터베이스, 인공지능, 정보검색, 기계번역, 분산시스템 등 다양한 분야에서 개발대상으로 삼고 있으며, 현재 온톨로지에 대한 개발은 국내외적으로 지식기반·지능형 시스템 개발에서는 제외할 수 없는 핵심 기술이자 자원으로 인식되고 있다. 따라서 정보검색시스템뿐만 아니라 다른 온톨로지 활용 시스템 개발을 위해서는 이러한 기반 연구가 선행되어야 하며, 국가적 차원의 연구 기관과 대규모 프로젝트를 통해 국내외 온톨로지 표준화 작업의 참여뿐만 아니라, 시맨틱 웹 기술의 국내 기반 확충과 성장, 국제 표준에의 참여를 위해 많은 연구자들이 노력해야 할 것이다. 본 연구 개발은 식물자원이라는 카테고리 안에서도 감귤의 한방화 측면이라는 특정 도메인을 정리하였다. 물론 현재의 도메인 범위가 작은 관계로 현실적인 사용에 있어서는 그리 큰 의미를 갖지 못한다는 한계를 가진 것은 사실이다. 하지만, 온톨로지 체계의 기반을 준비함으로써, 향후 제주 관광산업, 바이오 산업 등에서 지식 체계화 시킬 수 있는 프레임틀을 제시하는 것에 의미를 부여할 수 있으며, 궁극적으로는 유비쿼터스 환경의 기반이 될 수 있을 것이다.

참고문헌

- [1] 정인영, 박기식, 민재홍, 김성찬, 이강찬. 2003. “시맨틱 웹 기반의 검색 시스템 구조” 주간기술 동향 통권 1094호. pp15-17)
- [2] S.A.McIlraith, T.C. Son, and H. Zeng. 2001. “Semantic Web Services”. IEEE Intelligent Systems. vol16. no2. pp46-53.
- [3] 권혁철. 2004. “시맨틱웹의 가능성과 한계”. 지식정보 인프라 통권15호. pp15-19.
- [4] 박재홍, 임유전, 김도완, 박찬규, 조현규. 2002. “Semantic Web 환경에서의 자원발견”. 정보처리학회 추계학술발표대회 논문집 제9권 제2호. pp1-4.
- [5] Benjamins Lee, T., Hendler, J., Lassila. 2001. “the Semantic Web”. Scientific American. May 2001. p32.
- [6] Gruber, T.R. 1993. “Towards Principles for the design of

- Ontologies used for Knowledge Sharing” [online].
 [cited:2002.12.20] <<http://www.cise.ufl.edu/~jhammer/classes/6930/XML-FA02/papers/gruber93ontology.pdf>>.
 [7] <http://www.citrusmuseum.com/> 서귀포 감귤박물관 홈페이지

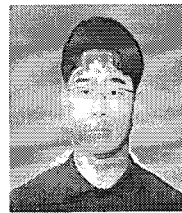
저자소개



김민철 (Min-cheol, Kim)

고려대학교 박사(경영정보 전공)
 제주대학교 경영정보학과 조교수

※ 관심분야: MIS, 의료정보, 지식경영



김영익 (Young-ick, Kim)

제주대학교 관광과 경영경제 연구
 소 연구원.
 중앙대학교 철학과 졸업.
 University of Technology, Sydney.
 MBA.

※ 관심분야: 온톨로지, Culture Technology.