

온톨로지 기반 의미적 정보검색 시스템 설계

공현장^a, 황명권^a, 김판구^b

^a 조선대학교 전자정보공과대학 전자계산학과
광주 동구 서석동 375, 501-759

Tel: +82-62-230-7799, E-mail: kisofire@chosun.ac.kr

^b 조선대학교 전자정보공과대학 컴퓨터공학과
광주 동구 서석동 375, 501-759

Tel: +82-62-230-7636, Fax: +82-62-230-7636, E-mail: pkkim@chosun.ac.kr

Abstract

많은 정보가 저장되어 있는 웹은 더 이상 우리의 삶과 별개로 생각할 수 없는 존재가 되었다. 무수한 정보가 웹상에 존재하게 되어 웹은 커졌지만, 단순한 정보의 저장소로 전락할 위기에 놓여있다. 이러한 거대한 정보 저장소에서 사용자는 원하는 정보를 찾기 위해 많은 시간을 소비해야만 한다. 이러한 문제의 해결을 위해 현재까지 많은 연구가 진행중에 있으며, 본 연구에서는 온톨로지 기술에 기반한 정보검색 시스템을 그 해결방안으로 제안한다. 본 논문에서는 온톨로지 기반의 의미적 정보검색을 위해 요구되는 세부적인 요소 기술들을 소개하고, 실제로 이들을 설계하고, 유기적인 결합을 통한 의미적 정보검색 시스템을 설계하였다.

Keywords:

온톨로지, 정보검색, 시맨틱 웹

1. 서론

웹은 진화하고 있으며, 웹의 사용자 역시 변화하였다. 사용자는 검색엔진을 통해 자신이 원하는 정보를 찾기위해 자신이 알고 있는 많은 것을 표현하지 않더라도 기계가 스스로 원하는 정보를 찾아주길 바란다. 예를 들어, 사용자가 'car'라는 검색어로 웹상에서 검색을 할때, 사용자는 자동차에 대한 많은 지식을 이미 알고 있으며, 기계 역시 그러한 정보를 알고 있기를 희망한다. 그러나 현실적으로 그러하지 못하므로 사용자는 정확한 정보검색을 위해서 'car is same as motorvehicle' 또는 'BMW is a car'와 같은 많은 정보를 직접 기계에 표현해야만 현재의 웹에서는 그러한 모든 지식에 대한 대응 정보를 제공한다. 그렇지만 이것은

복잡하고 피곤한 일이다. 그러므로 그러한 표현의 부분을 대신 담당하는 기술을 필요로 하게 되었으며, 많은 연구에서 온톨로지가 이 부분을 담당하게 되었다. 그리하여 웹이 의미적으로 변하는 차세대 웹인 시맨틱 웹이 대두되었다.

본 논문에서는 현재의 웹에서 시맨틱 웹으로 진화하는 과도기적 단계에서의 온톨로지 기술을 현재 웹 기술에 접목하여 웹이 좀더 의미적으로 구성되도록 설계하였다. 웹에서 의미적 정보검색을 위해 많은 기술들이 요구되며, 특히 온톨로지에 관련된 기술이 큰 비중을 차지하고 있다. 본 논문에서는 의미적 인터넷 정보검색을 위한 전체적인 시스템 설계와 세부적으로 필요한 요소 기술들에 대하여 소개함으로써 의미적 인터넷 정보검색의 가능성을 보인다.

2장에서는 기존의 온톨로지 기반 정보검색 시스템의 접근 방법에 대하여 조사하고 분석하며, 3장에서는 본 시스템의 전체 구조와 세부 모듈들에 대하여 자세하게 소개한다. 그리고 4장에서는 기존 관련 시스템과의 비교 및 평가의 내용을 서술한다. 끝으로, 5장에서는 본 논문의 결론과 향후 연구 방향에 대하여 기술한다.

2. 관련 연구

온톨로지 기반 정보검색 시스템을 개발하기 위해서는 다음 [표 1]과 같은 여러가지 사항들을 고려해야 한다.

또한, 온톨로지 기반 정보검색 시스템 개발을 위해서 [표 1]에서 소개한 여러 가지 요소 기술들뿐만 아니라, 실질적인 온톨로지 구축, 구축 기관과의 연계, 기존 정보검색 기술과의 융합 기술

그리고 개발결과와 적극적 활용 방법 등에 대한 많은 부분들을 고려하여야 한다.

표 1- 온톨로지 기반 정보검색 시스템 구축시 고려사항

- 온톨로지를 이용한 애플리케이션의 성질 결정
- 사용자와 관련된 사항 확인
- 특정 조직의 정보와 지식에 대한 분석
- 소유권 문제
- 특정 조직 내 시스템 조사
- 추론 처리 결정
- 평가 기준과 측정 기준 결정
- 활용 가능한 범위 결정
- Data noise 고려, 존재하는 온톨로지들에 대한 연구
- 온톨로지 관리 도구와 처리 절차 분석

현재까지 연구 개발이 진행되어진 몇가지 온톨로지 기반 정보검색 시스템의 사례를 통해, 온톨로지를 어떻게 정보검색에서 활용하고 있는지를 살펴보도록 하자.

먼저, OntoWeb 프로젝트에서는 온톨로지 기반 정보검색 시스템을 연구 개발하였다.[1,2] OntoWeb 프로젝트에서 온톨로지는 사용자 질의에 대한 더 많은 관련 결과들을 검색하기 위한 가이드 역할을 담당하고 있으며, 이전의 정보검색 시스템에서 해결하지 못한 질의나 문맥에서의 의미에 대한 처리를 시도하였다.

또, 웹 문서를 분석하기 위하여 온톨로지를 이용함과 동시에 온톨로지 기반 질의 처리를 위한 시스템으로서 OntoBroker가 연구 개발되었다.[3] OntoBroker는 사용자들이 HTML 문서를 온톨로지 구조로 변환하기 위하여 제안되었으며, 사용자들은 OntoBroker 인터페이스를 통해 수집된 지식 정보를 검색하고 문헌의 내용을 이해한다. OntoBroker에서의 온톨로지는 정보 제공자와 검색자를 위한 공통의 언어이며, 개념, 의미관계 그리고 특정한 규칙으로 구성되어 있다.

그리고 MELISA(Medical Literature Search Agent)는 의학 분야 문헌 검색 시스템[26]으로서, 의학 분야에서의 온톨로지 기반 정보검색 에이전트 시스템의 프로토타입이다. MELISA는 사용자 질의 처리 문제와 의학 분야 문헌의 검색 정확률 향상을 위해, Medical 온톨로지를 사용하고 있다.[10]

다음 [그림 1]은 시맨틱 웹 정보기반 검색 시스템 구조를 설명한 것이다.[4] 이 시스템은 서버 시스템인 검색 엔진과 온톨로지로 구분할 수 있다. 온톨로지 저장소(ontology repository)를 중심으로 윗부분을 검색 엔진이 담당하고 있으며, 아랫부분은 온톨로지 관련 기술들이 차지하고 있다. 검색 엔진은 사용자가 시맨틱 웹의 사용을 위한 에이전트와 RDF 온톨로지를 이용한 추론 엔진의

연동에 대한 서브시스템들로 세분화할 수 있으며, 온톨로지 관리 시스템은 온톨로지를 생성하고 유지 및 관리하기 위한 시스템 구조이다[4].

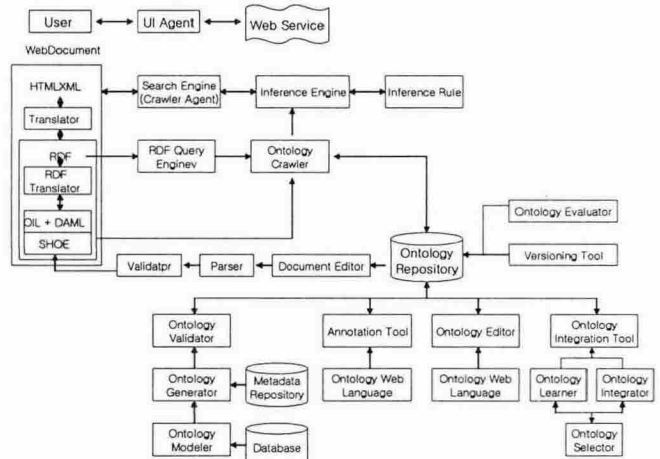


그림 1- 시맨틱 웹 기반 정보검색 시스템 구조

앞에서 소개한 지금까지의 온톨로지 기반의 정보검색 시스템은 대부분 온톨로지를 사용하여 의미적인 정보검색을 피하고 있다. 그렇지만 온톨로지 기반의 정보검색 시스템은 앞에서도 언급했듯이, 온톨로지의 구축에서부터 온톨로지의 관리 그리고 온톨로지를 이용한 추론을 비롯하여 온톨로지 기반의 질의 처리에 이르기까지 여러 가지 요소 기술들을 필요로 한다. 이러한 기술들이 모두 유기적으로 구성되어짐으로써, 이상적인 온톨로지 기반의 정보검색 시스템이 구축되어질 수 있다. 본 논문에서는 이러한 온톨로지 기반의 정보검색 시스템이 갖추어야 할 요소 기술들에 대하여 자세하게 소개하고, 각 요소 기술들을 설계하고 구현함으로써 온톨로지 기반의 의미적 정보검색의 가능성을 보인다.

3. 본론 - 온톨로지 기반 정보검색 시스템 요소 기술 및 시스템 구조

본 시스템에서 온톨로지는 가장 중요한 역할을 하고 있으며, 온톨로지를 효율적으로 관리하고 어떻게 활용할 것인가에 대한 문제가 가장 중요하다. 본 논문에서는 이전의 시스템과 비교하여 본 시스템의 특징을 소개하고, 본 시스템의 중요한 요소 기술들을 소개한다. 그리고 중요 요소들이 유기적으로 결합된 전체 시스템 구조를 소개한다. 온톨로지 기반의 의미적 정보검색 시스템의 구축을 위해 다음 4가지 온톨로지에 관련된 주요 기술들이 필요하다.

3.1 온톨로지 저작 도구(OntoMan)

온톨로지 저작도구는 온톨로지 구축의 일련 과정을 지원한다. 온톨로지 저작 도구를 통하여 온톨로지 생성이 가능하고, 또한 온톨로지 크롤러에 의해 수집된 온톨로지의 일관성을 체크하거나 수정 및 변환하여 이를 온톨로지 저장소에 저장하고 관리한다. 본 연구에서 연구 개발한 온톨로지 기반 정보검색 시스템을 위한 첫번째 구성 요소인 온톨로지 저작도구는 쉬운 인터페이스로 구성되어 있으며, 그 이외에도 몇가지의 주요 특징들을 가지고 있다. 그리하여 본 온톨로지 저작도구는 국내 온톨로지 개발자 및 사용자에게 큰 도움이 될 수 있을 것으로 기대된다.

뿐만 아니라 온톨로지 저작 도구가 갖추어야 할 모든 구성 요소 기술들을 갖추고 있다. 본 온톨로지 저작도구(OntoMan)의 주요 특징은 특정 도메인에 대한 자동 온톨로지 구축 지원과 온톨로지 언어에 대한 사용자 가이드를 제공하는 것이다. 앞에서 언급한 주요 특징들에 대해서 자세하게 설명하면 다음과 같다.

- 특정 도메인에 대한 자동 온톨로지 구축 지원: 워드넷에 기반하여 자동으로 특정 도메인에 대한 프레임을 구축하고, 도메인 전문가의 특정 입력 형식을 바탕으로 사용자 정의 도메인 대해 자동으로 온톨로지를 구축한다.
- 사용자 중심의 쉬운 인터페이스 지원: GUI를 사용하여 사용자가 그래픽적으로 쉽게 온톨로지를 구축할 수 있고, 특히 온톨로지의 구조를 트리의 형식으로 표현하여 개념간의 관계를 쉽게 파악할 수 있도록 설계하였다.
- 온톨로지 언어에 대한 사용자 가이드 지원: 본 시스템은 대표적 온톨로지 언어인 RDF(S)와 OWL에 대한 사용자 지침에 대한 한글 가이드를 제공하고 있다. RDF(S)와 OWL은 온톨로지 구축을 위해 가장 널리 사용되는 언어들이지만 국내 온톨로지 사용자가 사용하기에는 조금은 무리가 있다. 특히 많은 문서들이 영문 표준안으로 작성되어져 있으며, 문서의 내용도 이해하기 어려운 부분이 많기 때문이다. 이에 본 연구에서는 이러한 국내 사용자를 위한 RDF(S)&OWL 에 대한 한글 사용자 가이드를 제공하고, 이를 온톨로지 저작도구에 접목시켜 보다 효율적인 온톨로지 저작이 가능하도록 하였다.

온톨로지 기반 정보검색 시스템의 첫번째 구성요소인 온톨로지 저작 도구의 시스템 구성 화면은 다음 [그림 2]와 같다.

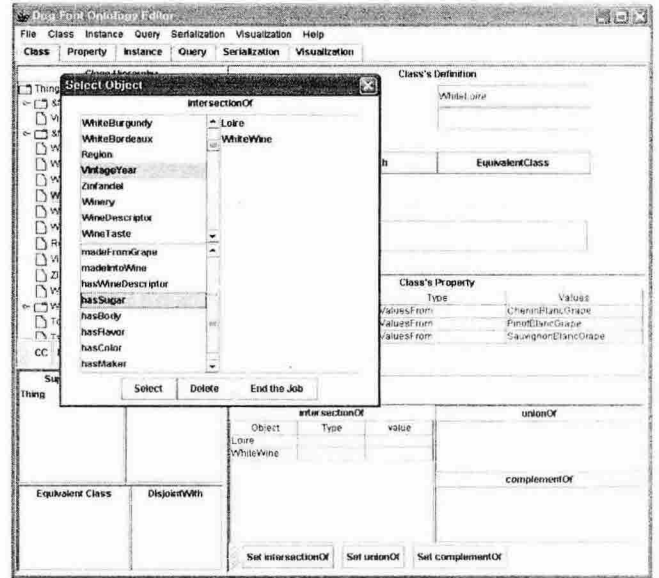


그림 2 - OntoMan

3.2 온톨로지 저장소

온톨로지 저장소는 온톨로지들을 한곳에 수집하고 관리하며, 모든 사용자들이 쉽게 온톨로지에 접근할 수 있도록 지원함으로써, 의미적 인터넷 정보검색을 가능하게 하는 두번째 기본 구성 요소이다. 온톨로지 저장소는 온톨로지 저작도구와 연결되어 온톨로지 저작도구에서 만들어지 온톨로지를 직접 저장하기도 하고, 또 온톨로지 크롤러에 의해 웹으로부터 수집된 온톨로지를 검증, 수정 그리고 저장하는 기능을 포함하고 있다. 온톨로지 저장소는 크게 온톨로지 파일(OWL, RDF)과 온톨로지서 추출된 사실들을 모아놓은 사실파일(FACT)들로 구성되어져 있으며, 온톨로지가 생성되거나 수집되면 미리 정의된 추론 규칙에 의해서 새롭게 사실 파일을 추론하고 생성하여 함께 저장한다.

3.3 온톨로지 크롤러

온톨로지 크롤러는 웹 상에 분산되어 존재하는 많은 온톨로지 파일을 찾아서 두번째 구성 요소인 온톨로지 저장소에 저장을 하는 작업을 담당하고 있으며, 본 시스템에서 제작한 크롤러는 웹상의 OWL, RDF 온톨로지 파일에 대해서만 수집을 한다. 또한 이렇게 수집된 온톨로지는 본 온톨로지 저작도구에 의해 온톨로지 일관성을 체크하거나 수정하여 새롭게 온톨로지를 재구성한다. 온톨로지 기반 정보검색 시스템의 세번째 구성 요소인 온톨로지 크롤러는 수집된 온톨로지를 도메인 별로 분류하는 분류모듈, 분류된 문서들에 대한 순서를 부여하는 랭킹모듈 그리고 사용자에게 편리한 인터페이스를 제공하여 원하는 도메인 온톨로지를 빠르게 검색할 수 있는 검색모듈로 구성되어 있다.

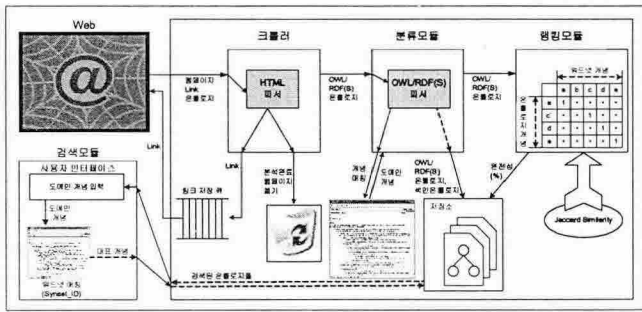


그림 3- 온톨로지 크롤러 구조

[그림 3]은 본 시스템의 온톨로지 크롤러 구조를 설명하고 있다. 온톨로지 내부에 정의된 개념들을 분석하고, 도메인 내에 정의된 모든 개념들을 이용하여 랭킹 정보를 갖기 때문에 각 모듈의 기능들은 기존의 검색 시스템들과는 확연히 구분된다.

3.4 온톨로지 추론 엔진

온톨로지 기반 정보검색 시스템의 마지막 구성 요소인 온톨로지 추론 엔진은 크게 두 부분으로 구성된다. 하나는 온톨로지 어휘에 대한 추론 규칙을 정의하여 온톨로지 파일을 새로운 사실 트리플의 형태로 변환하여 관리하는 추론 부분이고, 다른 하나는 이렇게 추론 규칙을 통하여 만들어진 새로운 사실 트리플에서 사용자가 원하는 정보를 추출하기 위한 질의 처리 엔진 부분이다. 본 추론 엔진에서의 추론 규칙은 기본적인 그래프 모델에 대한 규칙 30개와 온톨로지 언어 어휘에 대한 추론 규칙 20개 그리고 사실의 일관성을 체크하는 5개의 일관성 체크 규칙으로 구성되어 있다. 새로운 온톨로지가 구축되거나 수집되면 추론 엔진은 온톨로지 파일에 추론 규칙을 적용하여 트리플 형식의 사실 파일로 변환하여 온톨로지 저장소에 저장한다. 그런 다음 질의 처리엔진에서 사실 파일의 트리플에 기반한 질의 처리를 거쳐 의미적 정보를 추출하게 되며, 이렇게 추출된 의미적 정보가 실제 인터넷 정보 검색에 사용되어 의미적 인터넷 정보검색을 기대할 수 있다.

3.5 온톨로지 기반 정보검색 시스템 전체 구성도

앞에서 설명한 4가지의 온톨로지 관련 주요 기술을 사용하여 본 논문에서 최종적으로 제안한 전체 시스템의 구성도를 살펴보면 다음 [그림 4]와 같다.

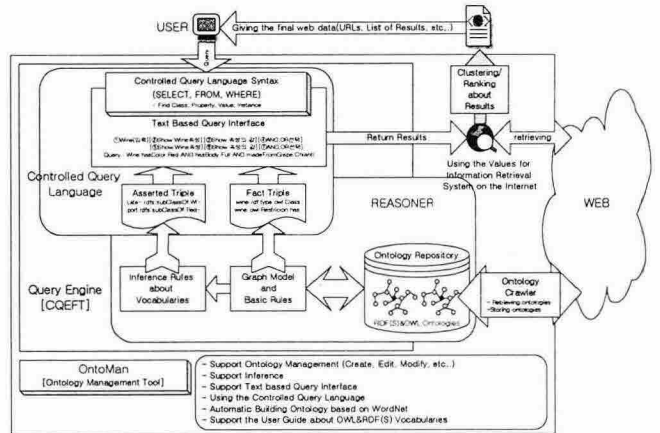


그림 4- 온톨로지 기반 정보검색 시스템 구조

[그림 4]는 본 연구에서 제안한 온톨로지 기반 정보검색 시스템의 흐름을 보여주고 있으며, 여기에는 앞에서 소개한 4가지의 요소 기술들이 유기적으로 결합되어 전체 시스템이 구성되어 있다. 시스템은 기존의 웹과 접목되어 웹의 정보에 대하여 의미적 정보검색이 가능하도록 설계되었다. 전체 시스템의 흐름을 순서적으로 살펴보면 다음과 같다.

- ① 사용자 : 정보검색 엔진에 접근
- ② 추론엔진에서 지원하는 Text Based Query Engine(CQEFT)을 통한 사용자 질의 입력
- ③ 질의는 온톨로지 저장소에 저장된 온톨로지를 사용하여 특정 추론 규칙을 적용한 추론 엔진의 Reasoner에 의해 추론된 사실들을 기반으로 하여 해당 결과값을 얻음
- ④ 온톨로지 저작도구는 온톨로지를 새롭게 구축하고 관리함
- ⑤ 온톨로지 크롤러는 웹에 산재되어 있는 온톨로지를 온톨로지 저장소에 저장
- ⑥ 온톨로지 기반의 검색 결과값을 웹에서 검색하여 해당 URI 또는 URL을 사용자에게 제공

위의 절차에 의해 의미적 인터넷 정보검색이 가능하다. 앞에서 소개한 정보검색 절차는 매우 정확한 검색을 위한 과정이다. 그렇지만 사용자가 좀더 많은 관련 결과를 얻기 위해서는 확장 검색을 통한 검색도 가능하다. 그러므로 사용자가 그 내용에 대하여 정확하게 알고 있지 않더라도 근접한 결과값을 쉽게 구할 수 있게 된다.

4. 비교평가

본 시스템은 성능 평가를 위한 단계는 아직 되지 않는다. 현재 추론 과정을 통한 정보의 추출과 전체적인 시스템의 구성 요소들은 완성되었지만, 실험 초기 단계인 본 시스템은 추론 엔진에서 더욱 강화되고 추가된 규칙과 웹 연동 부분의 미비한 점이 추가적으로 연구되어야 한다. 그렇지만 현재 온톨로지에 기반한 정보검색 시스템에 대한 이전의 기술들은 대부분 온톨로지의 구축과 관리등과 같은 온톨로지에 관련된 기술들은 대부분 배제하고, 온톨로지를 정보검색 시스템에 접목하고자 하는 연구에 치중하고 있다.

본 시스템은 이러한 시스템들에서 간과한 온톨로지에 대한 기본 기술에서부터 온톨로지에 기반한 추론 그리고 질의에 대한 기술까지 모두 다룰 수 있는 시스템을 설계하고 구축하였다는데 연구의 의의가 있다.

5. 결론

본 연구는 온톨로지에 기반한 의미적 정보검색의 가능성을 보이기 위해 연구의 초점을 맞추고 있으며, 현재까지의 모든 연구에서는 이러한 온톨로지 기반의 의미적 정보검색을 이론적으로만 표현하고 있는 부분을 본 연구를 통해서 극복하고자 시도 하였으며, 앞으로 시맨틱 웹의 실현에 본 연구가 초석이 될 것으로 기대한다. 아직은 전체적인 구성이 완벽하지 못하여 새롭게 추가하여 진행하여야 할 연구 부분이 많지만, 이러한 부분은 온톨로지 기반의 의미적 정보검색의 실현을 위한 향후 연구과제로 남아있다.

References

- [1] Uschold, M., King, M., Moralee, S., Zorgios, Y., "The Enterprise Ontology", AIAI-TR-195, Artificial Intelligence Applications Institute (AIAI), the University of Edinburgh, 1997
- [2] <http://www.cs.umd.edu/projects/plus/SHOE/>
- [3] <http://ontoweb.aifb.uni-karlsruhe.de>
- [4] 이재호, "시맨틱 웹의 온톨로지 언어", 정보과학회지, 제21권, 제3호, pp. 18-27, 2003
- [5] 정도현, "시맨틱웹을 위한 온톨로지 언어와 구현사례 연구", 정보관리연구, 제34권, 제3호, pp. 87-109, 2003
- [6] 김현희, 안태경, "온톨로지를 이용한 인터넷 웹

검색에 관한 실험적 연구", 정보관리학회지, 제20권, 제1호, pp. 417-455, 2003

- [7] 이강찬, 김성한, 민재홍, 박기식, 정인정, "시맨틱 웹 기반의 검색 시스템 구조", 주간기술동향, 제1094호, IITA IT정보단, 2003, <http://www.itfinder.or.kr>
- [8] Aitken, S., Reid, S., "Evaluation of an Ontology-Based Information Retrieval Tool", 12th European conference on Artificial Intelligene(ECAI'00) Workshop on Applications of Ontologies and Problem-Solving Method, 2000
- [9] Gruber, T., "Toward Principles for the design of ontologies used for knowledge sharing", International Journal of Human-Computer Studies, vol.43, no.5/6, pp. 907-928, 1995
- [10] Abasolo, J.M., Gómez, M., "MELISA. An Ontology-based agent for information retrieval in medicine", ECDL 2000 Workshop on the Semantic Web(SemWeb2000), pp. 73-82, 2000