

온톨로지 기반 논문정보 검색 시스템 설계

*위다현, 강현민, 손석원, 한광록

호서대학교 메카트로닉스공학과, 메카트로닉스공학과, 뉴미디어학과, 컴퓨터공학과
e-mail : mayaki83@hcilab.net, pmalc@hcilab.net, sohn@hoseo.edu, krhan@hoseo.edu

Design of Treatise Retrival System based on Ontology

*We-Da Hyun, Kang-Hyun Min, Sohn-Surg Won, Han-Kwang Rok
Dept. of Mechatronics, Mechatronics, Newmedia, Computer Engineering
Hoseo University

Abstract

Nowadays world wide web has been abundant in quantity. However, the quality decreaed. Due to the much information, we need to select search some key words and review the search results using keywords related search methods in order to obtain information users want. In this paper, we propose the system design of treatise retrieval through metadata reasoning using ontology. In the process of this design, we express a particular treatise information as semantic-based metadata using ontology instead of using simple keywords relationship which has been used conventionally.

I. 서론

하이퍼텍스트에 기반을 둔 기존의 웹은 어휘 정보를 이용하여 문서를 검색하기 때문에 유의어, 동음이의어 등의 문제를 해결하지 못하며, 상업 및 학술정보 등의 정보의 특징을 분별해낼 수 없다[1].

또한, 지금까지의 웹 기술은 사람과 컴퓨터간의 정보 교환에 치중하여 컴퓨터와 웹 데이터간의 의미적 해석 및 처리의 문제를 가지고 있다. 시맨틱 웹은 기존의

웹을 확장한 형태로, 정보에 정형화된 의미를 부여하여 소프트웨어 에이전트가 이 의미정보를 자동으로 추출할 수있는 환경을 제공한다[2]. 온톨로지는 사람과 컴퓨터 간의 공유되는 지식을 개념적으로 표현 한 것으로써, 지식의 개념들간의 구조와 여타 관계 및 그들의 제약을 표현하는 일종의 개념형 데이터베이스이다. 특히, 시맨틱 웹의 온톨로지는 기존의 온톨로지를 웹을 통해 확장시키는 방법을 제시한다[3].

본 논문에서는 사용자에게 맞는 논문정보를 제공하기 위하여 OWL 온톨로지를 통해 논문속성에 연결되는 의미정보를 적용하여 시맨틱 웹 기반의 논문 검색 시스템을 설계를 제안한다.

II. OWL 온톨로지 기반의 시맨틱 논문 검색 시스템

2.1 시맨틱 추론 엔진

시맨틱 웹 환경에서는 추론 엔진에서 사용가능한 지식[사실(Fact)과 규칙(Rule)]이 시맨틱 웹 언어인 OWL로 기술되어 제공되며, 추론 엔진은 시맨틱 웹상에 존재하는 지식들을 검색하여 추론에 사용할 수 있다[4]. 시맨틱 추론엔진은크게 온톨로지 일관성 검사, 클래스들에 대한 계층관계 추론, 개체와 클래스의 관계 추론과 같은 3가지 기능을 제공한다[5]. 온톨로지 추론은 규칙을통해 검색조건에 해당하는 단어들의 의미와 관계를 확장해 나가는 방식으로 수행되며, OWL로 표현

된 지식 베이스를 이용하는 질의언어이자 프로토콜인 OWL-QL(OWL Query Language)을 사용한다.

2.2 논문 검색 시스템

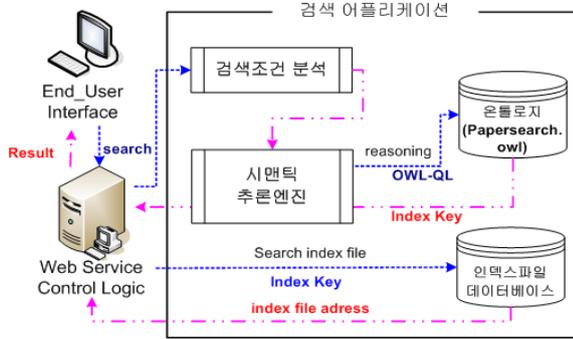


그림 1. 전체적인 시스템 구조

본 논문의 시스템에서 최종 사용자는 논문 정보 검색을 위해 키워드, 분야, 저자 등의 다양한 검색 조건을 적용할 수 있다. 그림 1은 전체적인 시스템 구조를 보여준다. 검색 어플리케이션은 java와 같은 객체 지향 언어로 구현되며, 최종 사용자가 웹 서비스 컨트롤 로직에서 제공하는 웹 브라우저 인터페이스를 통해 상기 시스템을 통한 논문정보의 검색을 가능하도록 한다. 시맨틱 추론엔진의 입력과 출력 데이터 구조는 OWL 온톨로지를 통해 표시되며, 웹 서비스 컨트롤 로직은 출력 데이터를 이용해 XML구조화한 논문정보가 있는 인덱스 파일 데이터베이스로부터 링크주소를 넘겨받아 검색결과를 출력한다.

III. OWL 온톨로지 설계 및 구축

본 논문에서의 실제적인 검색시스템 구현을 위하여 도메인은 공학 관련 분야로 한정하여 구축한다. 온톨로지 에디터인 Protege를 이용하여 OWL 형태로 구현하였다. 그림2는 시맨틱 논문 정보 검색 시스템 내에서 사용될 정보들을 표현하기 위해서 구축할 온톨로지의 전반적인 구조를 보여준다. 온톨로지는 계층적 구조를 가지고 있으며, 's'는 'rdfs:subClassof'로써 하위 클래스 개념을 표현하고 있다. 'p'는 요소에 해당하는 프로퍼티 값을 가진다. 'Paper'는 최상위 클래스로서 'Domain'과 논문 정보를 묘사하는 'Descriptor' 클래스를 서브 클래스로 가진다. 또한 논문의 속성을 상세히 정의함으로써 발행된 논문의 등급과 국내외 학회 정보를 이용한 검색이 가능하도록 설계하였다.

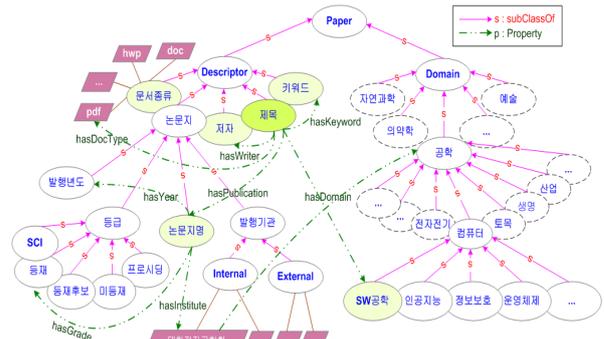


그림 2. 계층적인 온톨로지 구조

IV. 결론 및 향후 연구 방향

본 논문에서는 시맨틱 웹 환경에서 OWL 온톨로지를 이용한 논문 정보 검색 시스템을 설계하였다. 기존 인터넷 환경에서 사용자가 원하는 논문을 검색하는 방법은 두 가지로 검색엔진을 통한 논문의 키워드검색 방법과 학회지별, 논문지별로 제공되는 검색 방법이다. OWL 온톨로지를 이용하여 논문에 대한 클래스 계층 관계를 정의하고, 공리를 적용한 추론기에 로드하여 보다 정확한 검색결과를 도출하고자 하며, 앞서 말한 두 가지 검색방식을 함께 적용하도록 설계하였다. 향후 과제로는 시나리오에 따른 검색 어플리케이션을 구현하여 질의 검색 결과를 타 논문 검색엔진의 결과와 비교하여 성과를 측정해 보는 것이 필요하다. 또한 보다 유연한 웹 서비스를 제공하기 위하여 전문가에 의한 온톨로지 구조의 보완과 기존의 논문 정보를 온톨로지 공리에 따라 자동으로 맵핑시켜 주는 기술이 연구되어 져야 한다.

참고문헌

- [1] Berners-Lee, T.,Hendler, J. and Lassila, O., "The Semantic Web", Scientific American, 2001.
- [2] 최중민, "시맨틱 웹의 개요와 연구동향", 정보과학회지, 제21권, 제3호, pp.4-10, 2003. 3.
- [3] 신병호 외, "온톨로지 기반의 소프트웨어 프로세스 개념 모델 설계", 한국컴퓨터정보학회 논문지, 제13권, 제2호, 2008. 3.
- [4] 한국정보통신기술협회, "의미 웹 추론 엔진의 표준 사양", 정보통신단체표준, 2004.
- [5] 김제민, 박영택, "온톨로지 기반 웹 콘텐츠 추천 기법", 한국정보과학회 가을 학술발표논문집 제33권, 2호(B), pp.294-299, 2006.
- [6] 오삼균, "Web Ontology Language와 그 활용에 관한 고찰", 데이터베이스연구, 제18권 3호, pp.63-79, 2002