

온톨로지 추론을 이용한 논문 검색에 관한 연구

강현민*, 위다현*, 김석동**, 선복근**, 한광록**

*호서대학교 메카트로닉스공학과

**호서대학교 컴퓨터공학과

e-mail:{pmalc,mayaki83}@hclab.net,{sdkim,bksun,krhan}@hoseo.edu

A Study on Paper Search using Ontology Inference

Hyun-Min Kang*, Da-Hyun We*, Suk-Dong Kim**, Bok-Keun Sun**, Kwang-Rok Han**

*Dept of Mechatronics Engineering, Hoseo University

**Dept of Computer Engineering, Hoseo University

요 약

사용자의 의도에 맞는 정보를 검색하기 위해서는 인터넷 환경에서 웹 문서 자원 사이의 의미 정보를 온톨로지로 표현하고, 이 온톨로지를 컴퓨터가 이해할 수 있게 하는 시맨틱 웹의 도입이 필요하다. 본 논문에서는 OWL 온톨로지 기반의 추론을 통한 논문 정보 검색시스템에 대하여 논한다. 시맨틱 웹의 새로운 온톨로지 언어로 부상한 OWL 기반의 논문 온톨로지를 구축하고, 논문 속성들 간의 다양한 상관관계를 서술논리 쿼리로 작성한다. 검색시스템은 이 쿼리를 기반으로 논문 온톨로지에 대하여 추론함으로써 지능적인 정보 검색이 가능하도록 하였다.

1. 서론

온톨로지(ontology)는 사람과 컴퓨터 간의 공유되는 지식을 개념적으로 표현한 것으로서, 지식의 개념들간의 구조와 여타 관계 및 그들의 제약을 표현하는 일종의 개념형 데이터베이스이다. 특히, 시맨틱 웹의 온톨로지는 기존의 온톨로지를 웹을 통해 확장시키는 방법을 제시하고 있다[1-5].

본 논문에서는 기존의 논문 검색 서비스가 가지는 단점을 극복하기 위하여 논문에 대한 OWL(Ontology Web Language) 온톨로지를 설계하였다[6]. 이를 기반으로 웹에 산재한 논문들에 대한 요약정보를 XML형태의 논문 주석 정보 데이터로 정의하여, 사용자 질의에 대한 OWL 온톨로지를 생성하여 추론을 통하여 논문 정보를 검색 가능한 시스템을 제안한다. 사용자로부터 질의를 입력받았을 때, 구축된 OWL 온톨로지와 추론 기법을 적용하여, 사용자가 찾고자 하는 검색 결과를 보다 정확하게 제공하고자 한다.

2. 온톨로지기반 기반의 검색 시스템

2.1. 시스템 구성

(1) 사용자 인터페이스

사용자 인터페이스를 통해 입력받은 데이터를 검색기의 질의 관리기로 전달하는 일과 추론 엔진으로부터 추론 결과를 전달받아 웹 상에서 원본 논문 정보자원을 다운로드할 수 있다.

(2) 전처리기

웹에 산재한 HTML 형식의 정보 데이터를 수집하여 저장하고, OWL 온톨로지 생성을 위한 메타 데이터가 되는 XML 형식의 논문 주석 정보 데이터를 생성하고 저장한다.

(3) 검색기

① 질의어 관리기

· 질의어들 중에서 제 1 키워드만을 추출하여 키워드 검색을 함으로써 1차로 이 키워드와 관련된 논문들의 주석 정보 데이터를 검색해낸다.

· DL(Description Logic) 쿼리를 생성하기 위하여 제 1 키워드를 제외한 나머지 질의어들을 추출하여 DL 쿼리 생성기에 전달한다.

② OWL 온톨로지 생성기

개념 온톨로지를 참조하여 키워드 검색에 의해 1차로 검색된 논문들의 주석 정보 데이터들에 대응하는 OWL 온톨로지를 생성한다.

③ DL 쿼리 생성기

질의어 관리기에서 전달받은 질의어들 중에서 제 1 키워드를 제외한 나머지 질의어들을 대상으로 추론을 위한 DL 쿼리를 생성하여 추론 엔진에 전달한다.

④ 추론엔진

온톨로지 생성기에서 작성된 논문의 OWL 온톨로지를 로드하고, 생성된 DL 쿼리를 기반으로 OWL 온톨로지에 대한 추론을 한다. 추론 결과로 해당하는 논문의 Identifier를 반환한다. 본 논문의 시스템은 Pallet 추론 엔진을 사용한다[7].

2.2 검색 프로세스

본 절에서 지금까지 설명한 시스템의 데이터 흐름을 종합하여 기술한다. 그림 1은 검색데이터의 종합적인 흐름을 보여주고 있으며, 이 흐름을 크게 분류하면 전처리과정과 사용자 질의에 대한 검색 처리과정으로 분류할 수 있다. 그림 1에서 점선은 온톨로지

Fig. 1. Data process of paper retrieval system based on ontology

을 확인하였다.

(2 개의 질의어 실험)

사용자 질의	키워드1+ 키워드2		키워드1+ 발행년도		키워드1+ 발행정보	
\backslash	A	B	A	B	A	B
η	9	17	31	40	16	16
ϑ	11	13	19	10	7	6
ϵ	14	6	13	13	16	14
Precision(%)	39	73	68	75	50	53
Recall(%)	45	56	62	80	69	72

(3 개의 질의어 실험)

사용자 질의	키워드1+키워드2+ 발행정보		키워드1+키워드2+ 학술분야	
	A	B	A	B
η	5	10	7	14
ϑ	13	7	13	6
ϵ	4	1	4	5
Precision(%)	55	90	63	73
Recall(%)	27	58	35	70

4. 결론

본 논문에서는 기존의 논문 검색 서비스가 갖는 단점을 극복하기 위하여 시맨틱 웹의 새로운 온톨로지 언어로 부상한 OWL 기반의 논문 온톨로지를 구축하고, 이 온톨로지에 대한 추론 기능을 부여하여 사용자의 의도에 적합한 논문을 검색해 낼 수 있는 OWL 온톨로지 기반의 논문정보 검색시스템에 대하여 기술하였다. 이를 위하여 웹에 존재하는 논문들의 요약 정보를 XML 형식의 메타 데이터인 논문 주석 정보 데이터로 작성하고, 이것을 기반으로 OWL 논문 온톨로지를 생성한 후에 사용자 질의어들로부터 DL(Description Logic) 쿼리를 작성하여 이를 근거로 논문 온톨로지에 대하여 추론함으로써 지능적인 정보 검색이 가능하도록 하였다.

논문의 검증을 위하여 논문 검색 사이트 KSI KISS에서 제공하는 논문 100 여건을 대상으로 검색 사이트에서 검색한 결과와 본 논문의 온톨로지 기반의 검색을 수행한 결과를 비교하였으며, 결과적으로 본 논문의 시스템이 사용자의 질의에 대해 보다 정확한 검색 결과를 제공하는 것을 확인하였다.

참고문헌

- [1]최중민, 조성정, 김진형, 이재호, 양정진, 김인철, 강민구, 박영택, "<특집 : 시맨틱 웹>", 정보과학회지, 21(3), pp.3-50, 2003.
- [2]정도현, "시맨틱웹을 위한 온톨로지 언어와 구현사례 연구, 정보관리연구", vol. 34, no. 3, pp.87-109, 2003,
- [3]김이란, 온톨로지의 의미정보를 이용한 RDF 스키마 생성에 관한 연구. 석사학위논문, 연세대학교 대학원 문헌정보학과, 2001.
- [4]김중태, 웹 2.0 시대의 기회, 시맨틱웹, 디지털미디어리서치, 2005.
- [5]오삼균, "시맨틱 웹 기술과 활용방안", 정보관리학회지, 19(4) : pp.297-319, 2003.
- [6]신복근, 위다현, 한광록, "OWL 온톨로지를 기반으로 하는 논문 검색 시스템에 관한 연구", 컴퓨터정보학회논문지 14권 2호, 2009.
- [7]Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, "Pellet: A Practical OWL-DL Reasoner, ", Web Semantics: Science, Services and Agents on the World Wide Web, Volume 5, Issue 2, 2007.