

OWL 기반의 SPARQL을 이용한 시맨틱 검색

하상범^o 한은영 최호준

KT 컨버전스 본부

{terrie^o, heyhey, chopchop}@kt.co.kr

OWL-Based Semantic Search using SPARQL

Sangbum Ha^o Eunyoung Han Hojun Choi

KT Convergence Center

요 약

시맨틱 웹의 등장으로 시맨틱 검색에 대한 관심이 높아졌다. 이에 본 논문에서는 시맨틱 웹 환경의 자원들을 이용하여 시맨틱 검색을 수행하는 검색방법을 제안한다. 기존의 SPARQL(Simple Protocol and RDF Query Language) 시맨틱 질의언어는 추론의 기능 없이 RDF(Resource Description Framework)에서 제공하는 SPO(subject, predicate, object) 형태의 트리플 패턴 매치만을 제공한다. 본 논문의 시스템은 기존의 SPARQL질의 시스템에 시맨틱 추론기능을 추가하여 검색 결과에 효율성을 증가 시키는 것을 목적으로 한다. 본 논문에서의 시스템은 다음과 같은 특징을 갖는다. 첫째, 시맨틱 웹 환경의 온톨로지 구축을 위해서 W3C에서 온톨로지 언어로 표준화된 OWL(Web Ontology Language)를 사용하여 검색 환경을 구축한다. 둘째, 온톨로지와 메타데이터를 추론하여 시맨틱 검색을 유도하는 OWL추론기를 사용한다. 시맨틱 추론은 온톨로지의 공리(Axiom)를 충분히 활용하는 온톨로지기반 시맨틱 추론과 검색 도메인에 맞는 규칙을 활용하는 사용자 컨텍스트 기반의 시맨틱 추론으로 이루어진다. 셋째, 다양한 시맨틱 검색을 위해 W3C에서 제안되어 차세대 시맨틱 검색 질의언어로 연구중인 SPARQL을 사용한다. 이와 같은 특징은 시맨틱 검색 시스템이 시맨틱 웹 환경의 자원을 충분히 활용하는 결과를 가져온다.

1. 서 론

현재 인터넷 환경에서는 사용자가 원하는 정보를 찾기 위한 수단 중 많은 부분을 문서검색 시스템에 의존하고 있다. 하지만 기존의 키워드중심의 검색형태는 사용자가 원 의도하는 검색을 수행하는데 많은 한계점을 갖는다. 시맨틱 웹 기술을 활용한 시맨틱 검색 시스템은 온톨로지기반의 메타데이터를 추론하여 검색에 사용함으로써 보다 정확하고 사용자가 발견하지 못하는 지식까지 검색할 수 있게 해준다. 이와 같이 시맨틱 검색 시스템은 기존의 검색 시스템이 갖는 한계점을 많은 부분에서 보완하고 향상시킨다. W3C에서 시맨틱 검색 질의문으로 제안한 SPARQL은 온톨로지 언어로부터 작성된 문서에 대해 질의응답을 가능하게 하는 질의 언어지만 추론의 기능이 없으므로 의미검색을 수행하기는 부족하다. 본 논문에서는 SPARQL 시맨틱 질의문에 온톨로지기반의 추론을 적용함으로써 단순한 매치에서 벗어나 시맨틱 웹 환경의 자원을 충분히 사용한 시맨틱 검색 시스템을 제안하고자 한다.

2. 관련 연구

2.1 SPARQL

W3C의 DQWG(RDF Data Access Working Group)에서

제안한 SPARQL 질의언어는 기존의 RDF 문서에 질의응답을 가능하게 했던 RDQL 질의언어의 발전적인 형태로 시맨틱 웹 환경에서의 질의응답을 수행한다. 시맨틱 웹의 모든 자원은 RDF에서 제공하는 서브젝트, 프레디케이트, 오브젝트 형태의 SPO 구조로 이루어져 있다. SPARQL의 트리플 패턴 매치는 이러한 SPO 형태의 모든 URI 주소와 리터럴, 노드에 대해 매칭을 제공한다. SPARQL 질의언어는 기존의 RDQL에서 지원하지 못하는 다중매칭 및 다양한 조건적 매칭, 검색 조건의 제약을 지원하고, SPO 형태의 추론에 용이한 구조를 지원함으로써 효과적인 시맨틱 검색을 유도한다.

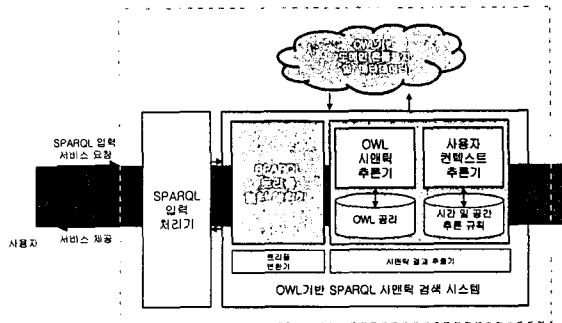
2.2 ARQ

ARQ(A SPARQL Processor for Jena)는 SPARQL을 지원하는 질의엔진으로 HP연구소에서 개발된 시맨틱 웹 라이브러리인 JENA를 기반으로 구성된 SPARQL 질의응답 시스템이다. ARQ 시스템은 서버와 클라이언트 형태로 구축된 RDF 자원에 대해 SPARQL 형태의 질의응답 구조를 지원한다. 하지만 아직까지는 SPO 형태의 트리플 패턴 매치만을 지원하기 때문에 온톨로지 기반의 추론을 적용하지 못한다. 이는 표현된 시맨틱 웹의 자원을 적극적으로 활용하지 못하는 경우이다. 단순히 트리플 패턴매치만으로는 시맨틱 검색의 한계점을 갖게 된다.

3. 본 론

3.1 제안하는 시맨틱 검색 시스템

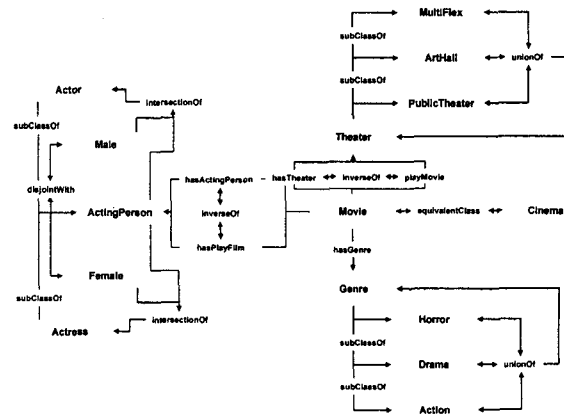
본 논문에서는 OWL 기반의 시맨틱 웹 환경에서 SPARQL 질의형식으로 시맨틱 질의문을 요청했을 때 OWL 기반의 자원을 추론하여 단순한 트리플 패턴 매칭이 아닌 새로운 지식정보를 찾게 해주는 시맨틱 검색 시스템을 제안한다. 다음 [그림 1]은 본 논문에서 제안하는 OWL기반 SPARQL 시맨틱 검색 시스템의 구조이다. 본 논문의 시스템은 크게 세부분으로 나누어져 있는데 첫 번째로 OWL 언어를 사용하여 도메인 온톨로지를 구축하고 기존문서를 해석하는 부분과 두 번째로 작성된 메타데이터와 OWL 공리, 사용자 컨텍스트를 대상으로 시맨틱 추론을 수행하는 시맨틱 추론기 부분, 마지막으로 SPARQL 시맨틱 질의문으로 시맨틱 검색을 수행하는 SPARQL 트리플 패턴 매칭기 부분으로 나누어진다.



[그림 1] 제안하는 시스템 구조

3.2 OWL기반의 온톨로지 구축

본 논문에서는 시맨틱 검색환경을 구축하기 위해서 OWL 기반의 온톨로지를 구축하고 메타데이터를 생성하였다. OWL은 W3C에서 시맨틱 웹의 온톨로지 언어로 표준화된 언어로써 다양한 OWL공리를 제공한다. 이는 도메인 온톨로지를 객체와 속성의 관계로 유기적으로 정의할 수 있게 한다. 또한 OWL구문을 추론하기 위해서 사용되는 n-Triple형태는 기존의 추론엔진에서 추론에 용이한 구조로써 다양한 온톨로지 기반의 추론을 가능하게 한다. 본 논문에서는 OWL Full, OWL DL, OWL Lite언어 중에서 OWL DL을 사용하며 이에 해당하는 공리를 사용하여 온톨로지를 구성하였다. 다음 [그림 2]는 본 논문에서 사용된 온톨로지의 일부분이다.



[그림 2] OWL 기반의 온톨로지

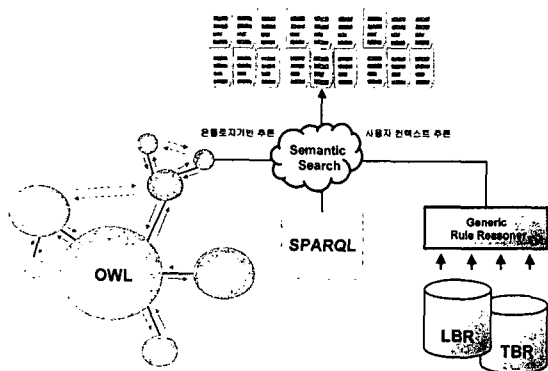
3.3 OWL기반 시맨틱 추론

본 논문에서 제안하는 시맨틱 검색은 검색방법에 추론을 적용하여 사용자의 질의문에 응답을 한다. 이러한 시맨틱 검색은 사용자의 의도를 정확히 파악하여 일반적인 키워드 매칭에 의해 나오지 않는 결과를 이끌어낼 수 있게 된다. 본 논문에서는 시맨틱 검색의 효율성과 정확성을 위해서 온톨로지기반의 시맨틱 추론과 사용자 컨텍스트기반의 시맨틱 추론으로 분리하여 시맨틱 추론을 수행한다.

3.3.1 온톨로지기반 시맨틱 추론

본 논문에서의 온톨로지 기반의 시맨틱 추론의 목적은 온톨로지에서 생성된 메타데이터에 대해 OWL에서 제공하는 공리를 사용하여 추론하고 이러한 추론된 사실을 바탕으로 사용자로부터 입력된 질의문으로 시맨틱 검색을 수행하기 위함이다. 본 논문에서의 시맨틱 검색은 Description Logic 기반의 온톨로지를 바탕으로 생성된 메타데이터에 대해 전방향 추론기법을 사용하여 검색하기 때문에 기존의 키워드검색이나 일반적인 SPARQL 질의응답 시스템에서 찾기 힘든 정보에 대한 검색이 가능하다. 본 논문에서 구축된 시스템의 추론엔진 부분은 HP 연구소에서 개발된 JENA 시맨틱 웹 라이브러리중에서 일반적인 추론 기능을 제공하는 추론 모듈(Generic Rule Reasoner)을 활용한다. JENA는 RDF형태의 문서의 변환 및 추론에 용이한 모델을 제공하므로, OWL 형태의 온톨로지를 JENA의 모델로 생성하고 일반적인 규칙 형태로

OWL 공리를 표현하여 온톨로지와 메타데이터를 융합하여 추론을 수행한다. 다음 [그림 3]은 본 시스템에서의 추론방식에 대한 개념도이다.



[그림 3] 시맨틱 검색을 위한 추론방식 개념도

3.3.2 사용자 컨텍스트기반 시맨틱 추론

본 논문에서는 온톨로지기반의 시맨틱 추론 외에 사용자 컨텍스트 기반의 시맨틱 추론을 수행한다. 이는 사용자로부터 입력되는 다양한 질의문 형태를 처리하기 위함이다. 즉, 온톨로지의 공리로 표현되는 부분은 온톨로지기반의 시맨틱 추론으로 처리하고 그 외에 검색에 필요한 부분은 규칙을 정의하고 추론하여 처리한다. 예를 들어 사용자의 검색 상황에 맞는 시간정보와 위치정보를 활용하고 이를 규칙으로 정의한 후 시간적 추론과 공간적 추론을 수행함으로써 시맨틱 검색시스템의 효율성을 증대시킬 수 있다. 다음 [표 1]은 본 논문에서 정의되고 있는 시간과 공간에 대한 규칙의 예를 보여준다.

[표 1] 시간적 추론 규칙 공간적 추론 규칙

<pre>[eventBefore: (?person hasSchedule ?schedule), (?schedule scheduleTime ?schTime), (?movie startTime ?sTime), (?movie runTime ?rTime), callBeforeTime(?schTime, ?sTime, ?rTime) -> (?sTime eventBefore ?schedule)]</pre>
<pre>[locationNear: (?person hasScheduleLocation ?schLocation), (?theater hasLocation ?theaLocation), (?schLocation hasXCoordinate ?schX), (?schLocation hasYCoordinate ?schY), (?theaLocation hasXCoordinate ?theaX), (?theaLocation hasYCoordinate ?theaY), callLocationNear(?schX, ?schY, ?theaX, ?theaY) -> (?theaLocation locationNear ?schLocation)]</pre>

4. 실험

OWL 기반의 SPARQL을 이용한 시맨틱 시스템을 구현하기 위해 다음과 같은 연구를 수행하였다. 첫째, OWL기반의 온톨로지를 구축하고 검색 도메인에 해당하는 메타데이터를 작성하여 시맨틱 검색 환경을 구축하였다. 둘

째, OWL 시맨틱 추론기는 온톨로지로부터 생성된 메타데이터를 OWL의 공리와 사용자 컨텍스트를 사용하여 추론하고 이를 시맨틱 검색결과에 반영하도록 하였다. 셋째, 사용자로부터 SPARQL 질의문을 구성하여 시맨틱 검색을 처리할 수 있게 하였다. 다음 [표 2]는 본 논문의 시스템에서 사용자의 질의문을 SPARQL로 처리하여 시맨틱 검색을 수행한 모습이다.

[표 2] 시맨틱 검색 수행

질의 Semantics	내일 오후 3시 전후로 종로구에서 코미디 장르를 상영하는 영화관과 상영시간은?
질의 SPARQL	<pre>PREFIX : <http://onto.kl.co.kr/movieSW.owl#> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> PREFIX rdf: <http://www.w3.org/2001/XMLSchema#> SELECT ?theater ?screen ?sTime WHERE { ?movie :hasgenre ?typeGenre . ?typeGenre rdf:type :Comedy . ?movie :startTime ?sTime . ?theater :hasScreen ?screen . ?screen :hasSchedule ?mSchedule . ?mSchedule :showingMovie ?movie . ?theater :hasLocation ?location . ?location :locationNear :종로구 . ?sTime eventBefore :2005-08-03T15:00:00.</pre>
답변 내보	?theater = CineCore ?screen = 3관 ?sTime = 2005-08-03T12:00:00

5. 결론

본 논문에서는 OWL기반에서의 시맨틱 검색을 수행하기 위해서 메타데이터를 추론하고 SPARQL 질의문을 이용하여 시맨틱 검색을 수행하는 방법에 대해 제안하였다. 이러한 시맨틱 추론을 검색에 사용하면 기존의 키워드기반의 검색이나 일반적인 SPARQL 질의응답보다 효율성과 정확성에서 나은 결과를 가져온다. 향후 연구로는 사용자 컨텍스트 기반 시맨틱 추론규칙에 대한 확장으로 사용자 검색 패턴에 대한 학습결과로 사용자 컨텍스트 규칙을 생성하는 것과 사용자로부터 친숙한 방법으로 SPARQL 질의문을 일반화 시키는 질의문장 형태의 추출에 관한 연구를 수행할 예정이다.

6. 참고문헌

- [1] Eric Prudhommeaux, Andy Seaborne, "SPARQL Query Language for RDF", W3C, October 2004
- [2] Sean Bechhofer, Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, Lynn Andrea Stein, "OWL Web Ontology Language Reference", W3C, February 2004.
- [3] Andy Seaborne, "Jena Tutorial : A Programmer's Introduction to RDQL", April 2002.