

질의 추론을 통한 온톨로지 기반

시맨틱 검색 시스템의 성능 향상

하상범^o 박영택

송실대학교 컴퓨터학과

terrie@ailab.ssu.ac.kr^o, park@computing.ssu.ac.kr

Performance Enhancement of A Ontology-based Semantic Search System with Query Inference

Sangbum Ha^o YoungTack Park

Dept. of Computer Science, SoongSil University

요 약

시맨틱 웹 기술을 활용한 시맨틱 검색은 문서의 의미를 온톨로지의 메타데이터로 생성하여 이를 바탕으로 검색을 수행하게 된다. 이와같은 온톨로지 기반의 시맨틱 검색은 논리를 바탕으로 추론을 적용할 수 있다. 본 논문에서는 온톨로지 기반의 추론을 적용한 시맨틱 검색 시스템을 언급하고 시맨틱 검색 시스템에서의 성능향상을 위해 추론엔진의 작업메모리 영역의 부하를 줄여 기존의 시스템보다 빠른 성능의 시맨틱 검색 시스템을 제안한다. 본 논문에서 시맨틱 검색 시스템의 성능향상을 위한 방법론으로는 다음과 같다. 첫째, 추론엔진이 검색 도메인내의 전체 메타데이터를 가지고 추론을 수행하지 않고 메타데이터의 온톨로지부분만을 사용하여 사용자가 원하는 질의문을 추론하여 검색에 사용하게 한다. 둘째, 시맨틱 검색 방법에서 Directly 매칭 검색과 시맨틱 추론검색을 병행하여 수행하게 한다. 이를 위해 본 논문에서는 메타데이터의 온톨로지부분과 인스턴스부분을 분리하는 단계와 분리된 온톨로지부분에서 사용자가 원하는 질의를 추론하는 단계, 추론된 질의문을 검색시스템에서 매칭하는 단계를 수행하게 된다. 이러한 방법은 메타데이터의 양이 증가하여도 온톨로지부분은 증가하지 않으므로 추론엔진에서 전방향 추론단계의 수행시간을 단축과 추론엔진의 호출 횟수를 단축 시키는 결과를 가져온다.

1. 서 론

현재 인터넷의 발달로 사용자가 원하는 정보를 찾기 위한 수단 중 많은 부분을 문서검색 시스템에 의존하고 있다. 시맨틱 웹 기술을 활용한 시맨틱 검색 시스템은 온톨로지 기반의 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. 하지만 시맨틱 검색 시스템에서는 검색수행이 추론엔진에서 전방향 추론을 통해 이루어짐으로써 메타데이터의 양이 방대해지면 검색이 느린 단점이 있다. 본 논문에서는 시맨틱 검색 시스템에서 추론엔진의 속도를 향상시키는 방법으로 질의문을 추론하여 검색하는 방법을 제안한다.

2. 관련 연구

2.1 TAP 시맨틱 검색 시스템

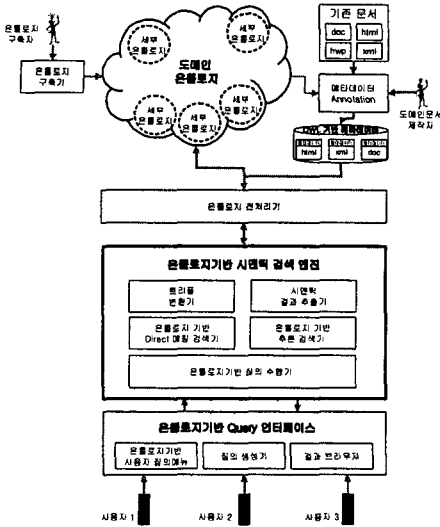
미국의 stanford 대학의 TAP 시스템은 시맨틱 웹 기술을 사용하여 문서의 검색 영역을 확장한다. TAP 시스템의 시맨틱 검색 방법은 사용자가 입력한 질의의 의미를 파악하여 이를 기반으로 웹 환경에서 분산되어 있는 정보들을 통합하여 답을 제시하는 시스템이다. 즉, 시맨틱 웹 기술을 통해서 정보들을 온톨로지처럼 계층구조와 링크된 정보들로 표현하고 사용자가 입력한 질의와 관련된 정보들을 TAP 시스템이 답변하는 것이다. TAP 시스템은 시맨틱 웹 기반에서 정보의 검색을 수행하지만 온톨로지에 표현된 속성을 위주로 검색결과를 확장하는 방식이

다. 이러한 방식은 온톨로지의 속성만을 이용한다는 점에서 본 논문에서의 제안하는 온톨로지 추론을 이용한 시맨틱 검색과 차이점을 갖는다.

본 논문에서 제안하는 시맨틱 검색시스템을 설명하기 전에 기존의 연구들을 살펴보았다. 앞에서 살펴본 방법들은 완전한 시맨틱 검색이라 말하기 힘들다. 접근방법은 다르지만 시맨틱 검색을 위해서는 문서의 의미를 정의할 수 있는 온톨로지가 필요하다. 본 논문에서 제안하는 시맨틱 검색은 온톨로지를 기반으로 하기 때문에 일반 태그가 아닌 키워드에 의미를 부여하는 태그를 사용하며, 태그의 계층적 구조도 활용한다. 또한 추론을 통하여 매칭으로 검색되지 않는 부분도 해결할 수 있다.

3. 온톨로지 기반 시맨틱 검색시스템

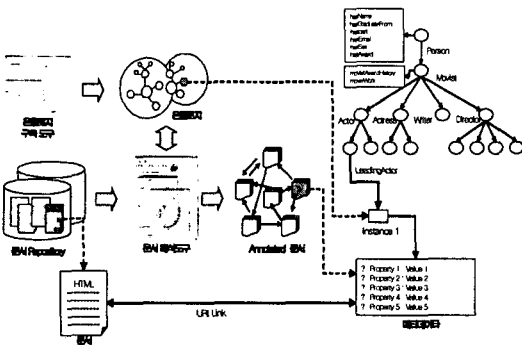
본 논문에서는 온톨로지 기반의 추론을 적용한 시맨틱 검색 시스템을 언급하고 시맨틱 검색 시스템에서의 성능향상을 위해 추론엔진의 작업메모리 영역의 부하를 줄여 기존의 시스템보다 빠른 성능의 시맨틱 검색시스템을 제안한다. 다음 [그림 1]은 본 논문에서 제안하는 온톨로지 기반 시맨틱 검색 시스템의 구조이다. 온톨로지 기반 시맨틱 검색 시스템은 크게 세부분으로 나누어져 있는데 첫 번째로 도메인 온톨로지를 구축하고 기존 문서를 해석하는 부분과 두 번째로 작성된 메타데이터와 추론엔진을 사용하여 시맨틱 검색을 수행하는 부분 마지막으로 온톨로지 기반의 시맨틱 검색을 사용자로 하여금 무리 없이 수행할 수 있게 도와주는 사용자 인터페이스 세부분으로 나누어진 다.



[그림 1] 제안하는 시스템 구조

본 논문에서는 이러한 온톨로지 기반 시맨틱 검색 시스템에서 성능향상을 위한 방법으로 온톨로지 전처리기를 추가하여 구축하게 된다. 온톨로지 전처리는 온톨로지에서 온톨로지 부분과 인스턴스부분을 분리하는 역할을 하게 되는데 이것은 추론 엔진이 검색 도메인내의 전체 메타데이터를 가지고 추론을 수행하지 않고 메타데이터의 온톨로지부분만을 가지고 추론하게 하여 추론 엔진의 부하를 줄여서 보다 빠르게 사용자가 원하는 질의문을 추론하게 해준다. 이와 같은 방법은 메타데이터의 양이 증가하여도 온톨로지부분은 증가하지 않으므로 추론 엔진에 전방향 추론 단계의 수행 시간을 단축시키는 결과를 가져온다.

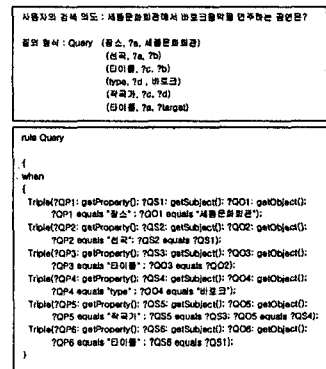
본 논문에서는 온톨로지 기반의 시맨틱 검색을 수행하기 위해서 OWL (Ontology Web Language) 기반의 온톨로지를 구축한다. OWL은 다양한 공리와 규칙 기반의 추론 엔진에 적용이 가능한 N-Triple 형태의 구문을 제공하므로 OWL을 사용하게 된다. 본 시스템을 위해 구축된 OWL 온톨로지를 사용하여 기존의 문서들을 메타데이터로 해석하고 시맨틱 검색 시스템의 지식 베이스를 구축한다. 이와 같이 구축된 지식 베이스는 온톨로지 기반의 추론을 적용하여 검색이 가능하게 된다. 다음 [그림 2]는 온톨로지 기반 시맨틱 검색의 흐름도이다.



[그림 2] 온톨로지 기반 시맨틱 검색 흐름도

3.1 시맨틱 검색 메카니즘

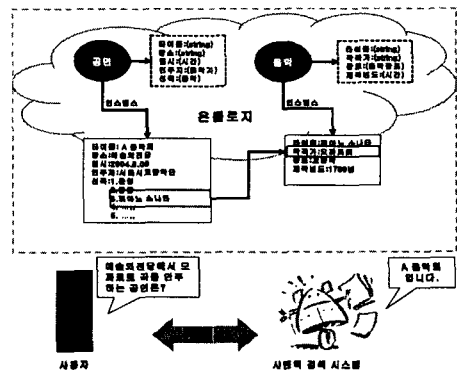
본 논문에서 온톨로지 기반 시맨틱 검색을 수행하기 위해서 다음과 같은 단계를 수행하게 된다. 첫 번째 단계는 온톨로지 기반의 속성들을 이용하여 Directly 매칭을 시도하게 된다. 온톨로지의 다양한 속성을 사용하여 메타데이터를 정의하여 사용자의 질의로부터 이러한 메타데이터를 매칭시켜서 검색을 수행하게 된다. 첫 번째 단계에서 검색결과가 없거나 사용자로부터 보다 많은 정보를 요구받을 때 두 번째 단계에서는 온톨로지 기반의 공리들과 규칙으로 이루어진 메타데이터를 사용자 질의를 바탕으로 추론하여 시맨틱 검색을 수행하게 된다. 본 논문에서 검색의 방법을 이와 같이 두 단계로 나누는 점은 사용자가 원하는 정보가 Directly 매칭 결과에 있을 때 추론 엔진의 호출 시간을 줄여서 검색의 효율성을 높이기 위함이다. 다음 [그림 3]은 사용자의 검색 의도를 반영하여 본 시스템에서 추론을 통해 생성된 질의문이다.



[그림 3] 추론된 질의문

3.1.1 온톨로지 기반 Directly 매칭 검색

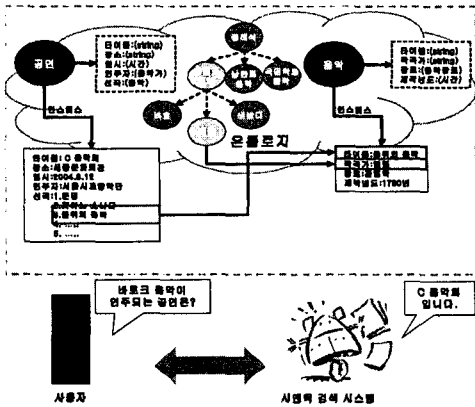
온톨로지가 도메인내의 다양한 정보를 속성과 계층구조로 정확히 표현할 수 있다면 이를 기반으로 기존 문서를 해석한 메타데이터도 정확하고 의미있는 정보를 포함하게 된다. 이러한 메타데이터의 속성을 사용하여 Directly 매칭 검색을 수행하면 [그림 4]의 검색 시나리오처럼 메타데이터의 의미를 파악할 수 있고 이를 검색에 반영할 수 있게 된다.



[그림 4] 온톨로지 기반 Directly 매칭 검색

3.1.2 온톨로지 기반 시맨틱 추론 검색

본 논문에서는 온톨로지 기반의 시맨틱 추론을 수행하여 검색을 하게 된다. 여기서 말하는 시맨틱 추론은 앞서 언급한 온톨로지와 인스턴스부분을 분리하여 추론하는 것을 말한다. 본 논문의 시맨틱 추론엔진은 사용자의 의도를 반영하여 검색 시스템에 실제 적용되는 질의문을 추론하게 된다. 시맨틱 검색 엔진에서 추론된 질의문으로 검색을 수행한다는 것은 일종의 추천과 같은 개념으로써 지식베이스에는 존재하지만 사용자가 파악하지 못한 정보를 추론된 질의문으로 찾아 주는 것과 같은 의미를 갖는다.

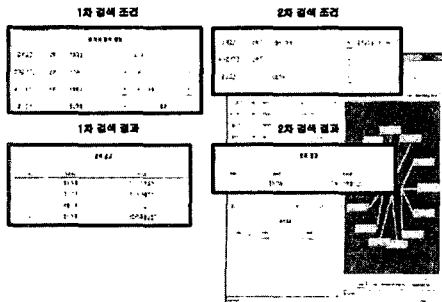


[그림 5] 온톨로지 기반 시맨틱 추론 검색

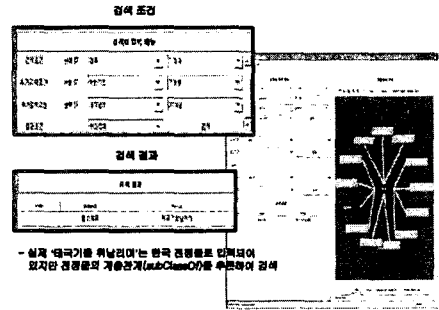
이러한 방식은 전체 메타데이터를 추론할 필요가 없고 메타데이터의 온톨로지부분만을 추론에 사용하기 때문에 시맨틱 검색 시스템의 성능향상에 도움이 된다. 다음 [그림 5]는 온톨로지 기반 시맨틱 추론 검색시나리오이다. 온톨로지 기반 메타데이터의 속성들만을 이용하는 Directly 매칭 검색을 사용해서는 검색이 안되는 부분까지 검색이 가능함을 보인다.

4. 실험

본 논문에서 온톨로지 기반의 시맨틱 검색 시스템을 구현하기 위해 다음과 같은 연구를 수행하였다. OWL기반의 온톨로지를 구축하여 기존문서를 메타데이터화 시켜 지식베이스를 구축하였고 본 시스템에서의 시맨틱 추론엔진을 구축을 위해서 일반적인 규칙기반의 추론을 제공하는 JRules를 이용하여 OWL 공리를 JRules형태로 생성하여 시맨틱 추론에 사용하였다.



다음은 앞서 언급한 두가지 형태의 시맨틱 검색 시나리오를 본 논문의 시스템에 적용한 결과이다. 첫 번째를 온톨로지 기반의 다양한 속성들을 갖고 사용자가 원하는 속성들을 메타데이터 지식베이스에서 Directly 매칭 검색을 수행하는 화면이다. 두 번째를 Directly 매칭으로 해결될 수 없는 부분에 대한 시나리오를 본 시스템에 적용하여 해결한 결과 화면이다. 이것은 온톨로지로부터 정의된 subClassOf 공리를 가지고 시맨틱 추론 엔진에서 메타데이터와 융합하여 사용자가 의도한 검색결과보다 좋은 검색 결과를 나타낸다.



[그림 5] 시맨틱 검색 결과 화면

4. 결론

본 논문에서는 온톨로지 기반의 시맨틱 검색 시스템의 성능 향상을 위한 두가지 방안을 제안하였다. 첫 번째로 온톨로지로부터 생성된 메타데이터를 온톨로지부분과 인스턴스부분으로 분리하여 추론엔진에 적용시키는 방안과 검색시에 온톨로지 기반의 속성을 이용하여 Directly 매칭과 시맨틱 추론 검색단계를 병행하는 방안이다. 검색 시스템에서 처리하는 데이터의 양이 많아질수록 추론엔진의 전방향 추론이 시스템의 많은 부하를 가져오게 되는데 본 논문에서 제안하는 방안을 사용하여 추론엔진의 부하를 줄여서 시맨틱 검색 시스템의 성능 향상을 유도하였다.

6. 참고문헌

[1] Deborah L. McGuinness and Richard Fikes, James Hendler, Lynn Andrea Stein, "DAML+OIL: An Ontology Language for the Semantic Web", IEEE 2002

[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] Richard Fikes, Deborah McGuinness, "An Axiomatic Semantics for RDF, RDF-S, and DAML+OIL", W3C, December 2001.

[4] Andy Seaborne, "Jena Tutorial : A Programmer's Introduction to RDQL", April 2002.

[5] Richard Fikes, Pat Hayes, Ian Horrocks, "DQL-A Query Language for the Semantic Web", WWW 2003, May 20-24, 2003, Budapest, Hungary.