

온톨로지 기반 추론을 이용한 시맨틱 검색 시스템 (Semantic Search System using Ontology-based Inference)

하 상 범 [†] 박 영 택 ^{**}
(Sang-Bum Ha) (Yong-Tack Park)

요 약 시맨틱 웹은 단순한 문서들의 링크가 아닌 문서들의 의미와 관계를 표현하는 웹으로 소프트웨어 에이전트가 이해할 수 있도록 구성되어 있다. 본 논문에서 제안하는 검색방식은 온톨로지 기반의 추론을 통한 시맨틱 검색방법으로 다음과 같은 특징을 갖는다. 첫째, 검색키워드와 문서의 키워드가 다르더라도 의미적으로 같으면 온톨로지의 추론을 통해 검색이 가능하게 한다. 둘째, 규칙기반의 변환기가 서로 다른 온톨로지의 컨셉을 정확한 매치(exact match)가 아니어도 유사한 컨셉으로 추론할 수 있게 한다. 셋째, 온톨로지가 검색 키워드의 의미를 뚜렷하게 정의할 수 있으므로 단순한 키워드 매칭과 빈도만으로 검색하는 것 보다 정확한 검색이 가능하도록 한다. 넷째, 최적화된 질의문 자동 생성이 도메인 온톨로지를 통해 가능하므로 자연어와 유사한 검색영역과 정확성을 갖게 한다. 다섯째, 에이전트가 단순히 키워드가 포함된 문서만을 찾는 것이 아니라 온톨로지에 표현되어 있는 정보를 토대로 사용자가 원하는 정보와 지식을 자동적으로 찾게 한다. 이러한 방식은 데이터베이스의 질의문을 사용하거나 일반적인 키워드기반의 정보검색 기법을 사용하여 자료를 검색하는 기존의 검색 시스템보다 정확한 검색을 가능하게 한다. 본 논문에서는 온톨로지를 기반으로 추론을 적용한 시맨틱 검색시스템에 대하여 문서검색에 초점을 맞추어 연구 결과를 제안한다.

키워드 : 시맨틱 웹, 시맨틱 검색, 온톨로지, 추론

Abstract The semantic web is the web paradigm that represents not general link of documents but semantics and relation of document. In addition it enables software agents to understand semantics of documents. We propose a semantic search based on inference with ontologies, which has the following characteristics. First, our search engine enables retrieval using explicit ontologies to reason though a search keyword is different from that of documents. Second, although the concept of two ontologies does not match exactly, can be found out similar results from a rule based translator and ontological reasoning. Third, our approach enables search engine to increase accuracy and precision by using explicit ontologies to reason about meanings of documents rather than guessing meanings of documents just by keyword. Fourth, domain ontology enables users to use more detailed queries based on ontology-based automated query generator that has search area and accuracy similar to NLP. Fifth, it enables agents to do automated search not only documents with keyword but also user-preferable information and knowledge from ontologies. It can perform search more accurately than current retrieval systems which use query to databases or keyword matching. We demonstrate our system, which use ontologies and inference based on explicit ontologies, can perform better than keyword matching approach.

Key words : Semantic Web, Semantic Search, Ontology, Inference

1. 서 론

시맨틱 웹의 등장으로 온톨로지를 통하여 에이전트가

이해할 수 있는 의미(semantic)를 갖는 문서를 생성하는 것이 가능해졌다[1]. 이러한 시맨틱 웹의 영역은 비즈니스 업무 효율을 증가시키고 이를 통해 이윤을 극대화 시키는 방법으로 시맨틱 검색을 통한 정보 검색시스템으로 확대적용 될 수 있다. 이러한 정보 검색시스템에서의 목표는 저장되어있는 다량의 정보들 중에서 사용자가 요구하는 정보와 문서를 어떻게 하면 사용자의 의도를 정확히 파악하여 효율적인 검색으로 누락되지 않

· 본 연구는 숭실대학교의 교내 연구비 지원으로 이루어졌습니다.

[†] 비 회 원 : 숭실대학교 컴퓨터학과
terrie@ailab.ssu.ac.kr

^{**} 종신회원 : 숭실대학교 컴퓨터학과 교수
park@computing.soongsil.ac.kr

논문접수 : 2004년 4월 20일

심사완료 : 2004년 12월 23일

고 요구문서를 사용자에게 제대로 전달할 수 있는가이다. 기존의 검색 시스템의 방식은 문서의 키워드를 추출하여 사용자가 요구하는 질의문과 단순한 매칭을 시도하고 키워드의 빈도수로써 결과를 보여준다. 하지만 정보의 양이 무한대로 늘어나는 가운데 이러한 방식으로는 사용자가 요구하는 올바른 정보와 지식을 찾아주는 것이 매우 어렵게 되었다.

이에 본 논문에서는 정보 검색시스템에 시맨틱 웹의 개념을 적용시켜 기존의 정보 검색시스템의 문제를 개선하여, 보다 정확하고 효율성이 높은 시스템을 제안하고자 한다. 본 논문에서 제안하는 검색방식은 온톨로지를 기반으로 추론을 통한 시맨틱 검색방법으로 다음과 같은 특징을 갖는다. 첫째, 검색키워드와 문서의 키워드가 다르더라도 의미적으로 같으면 온톨로지의 추론을 통해 검색이 가능하게 한다. 둘째, 규칙기반의 변환기로서 다른 온톨로지의 컨셉을 정확한 매치(exact match)가 아니어도 유사한 컨셉으로 추론할 수 있게 한다. 셋째, 온톨로지가 검색 키워드의 의미를 뚜렷하게 정의할 수 있으므로 단순한 키워드 매칭과 빈도만으로 검색하는 것 보다 정확한 검색이 가능하도록 한다. 넷째, 최적화된 질의문 자동 생성이도메인 온톨로지를 통해 가능하므로 자연어와 유사한 검색영역과 정확성을 갖게 한다. 다섯째, 에이전트가 단순히 키워드가 포함된 문서만을 찾는 것이 아니라 온톨로지에 표현되어 있는 정보를 토대로 사용자가 원하는 정보와 지식을 자동적으로 찾게 한다. 이렇게 시맨틱 웹 기술을 적극 활용하므로써 다량의 데이터에서 발견하기 힘든 정보와 지식을 자동적이고 정확하게 추출할 수 있다. 예를 들어 사용자의 요구가 "바로크시대 작곡가의 음악이 연주되는 공연은?" 이와 같은 복잡한 질의문일 경우 별도의 자연어처리가 없다면 기존의 검색방법에서는 복잡한 질의문의 조합으로 여러 번의 재검색을 통하여 한다. 이것 또한 검색 키워드와 원하는 문서의 키워드가 의미적으로 유사하지만 키워드 자체가 다를 경우 원하는 결과를 찾을 수 없다. 하지만 본 검색 시스템은 온톨로지 추론을 통해 검색 키워드가 달라도 의미적 맥락이 유사한 컨셉까지 찾아주게 되며 검색에이전트를 사용하여 자동적이고 정확한 결과를 낼 수 있다.

본 논문의 구성으로는 2장에서는 시맨틱 검색방식에 대한 기존 연구에 대하여 살펴보고, 3장에서는 본 논문에서 제안하는 온톨로지 기반 시맨틱 검색시스템의 구현방법과 시맨틱 검색방법에 대하여 설명한다. 마지막으로 4장에서는 본 논문의 구현된 시맨틱 검색시스템에 다양한 질의문을 적용한 실험과 기존의 시스템과 성능비교를 통하여 제안하는 시스템의 효율성을 입증하도록 한다.

2. 관련 연구

정보의 양이 기하급수적으로 늘어나는 가운데 사용자의 요구에 해당하는 문서를 효율적으로 검색하는 방법에 대한 많은 연구가 진행되고 있다. 기존의 정보 검색 분야에서도 많은 연구가 진행되고 있지만 본 논문에서는 기존의 정보검색과는 차별화되는 시맨틱 검색분야에 대한 관련된 연구방법을 설명한다. 이 번장에서는 시맨틱 검색분야에서 연구되어온 두 가지 방법에 대해 설명한다[2,3].

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 계층 구조와 링크된 정보들로 표현하고 사용자가 입력한 질의와 관련된 정보들을 TAP 시스템이 답변하는 것이다. 다음 그림 1과 같이 사용자가 Yo-Yo-Ma에 대한 정보를 원할 때 시스템은 Yo-Yo-Ma에 대한 구성된 정보들 즉, 온톨로지를 기반으로 링크된 정보를 통합하여 사용자에게 알려준다.

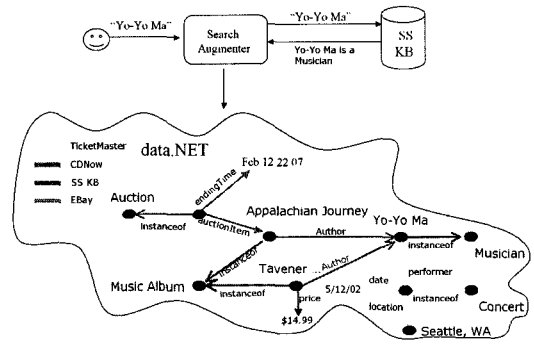


그림 1 TAP 시맨틱 검색 시스템

TAP 시스템은 시맨틱 웹 기반에서 정보의 검색을 수행하지만 온톨로지에 표현된 속성을 위주로 검색결과를 확장하는 방식이다. 이러한 방식은 온톨로지의 속성만을 이용한다는 점에서 본 논문에서의 제안하는 온톨로지 추론을 이용한 시맨틱 검색과 차이점을 갖는다.

2.2 DQL(DAML Query Language)

United States/European Union ad hoc Agent Markup Language Committee에 의해 개발된 DQL (Daml Query Language)은 시맨틱 웹을 위한 질의 언어로서 DAML+OIL/OWL로 표현된 지식을 기반으로

에이전트 간에 질의를 던져서 응답을 받기위해 만들어진 언어다. 즉 응답을 도출해내기 위해 시맨틱 웹상의 지식 베이스를 기반으로 여러 질의들과 응답들 간의 의미적인 연관 관계(Semantic Relationship)를 정밀하게 분류한다. DQL은 온톨로지에 정의된 속성들을 기반으로 기존의 데이터베이스 시스템을 위한 구조적 질의어(SQL)와 유사한 형태의 문답형식을 지원한다. DQL은 기본적으로 질의문과 지식베이스의 정보가 정확한 매치일 때 지식 베이스에서 질의에 대한 답을 도출할 수 있다는 것이다.

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 필요하며 온톨로지에 대한 추론이 필요하다. TAP 검색 시스템은 검색에 있어서 온톨로지를 사용하지만 검색결과에 있어 온톨로지의 속성을 추가적으로 첨부하는 성격이 강하다. DQL 시스템은 기본적으로 정확한 매치만을 지원하기 때문에 온톨로지의 다양한 속성을 이용하지만 효율성에 있어서 공리를 이용한 추론방법과 비교해 볼 때 낮은 효율성이 문제될 수 있다. 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 추론 검색

본 논문은 효과적인 검색시스템의 구축을 위해 서술 논리(Description Logic)바탕의 온톨로지를 사용하여 메타데이터를 생성하고 추론엔진을 접목시켜 온톨로지의 계층구조와 규칙, 공리(axiom)를 사용하여 추론을 함으로써, 기존의 지식베이스에 없는 지식을 검색영역에 포함하므로 모호한 질의에 대하여 해당하는 문서의 검색이 가능한 시맨틱 검색을 사용하는 방법을 제안하였다. 먼저 본 논문에서 제안하는 정보 검색시스템의 전체적인 구조를 살펴보고 세부적인 부분을 언급하여 설명하도록 한다. 다음은 시맨틱 검색시스템의 전체적인 구조이다. 그림 2에서 보듯이 시맨틱 검색시스템은 크게 세 부분으로 나뉘어져 있다.

첫 번째로, 정보 검색시스템 내에서 사용할 정보들을 정의하고 있는 온톨로지를 구축하고 온톨로지를 통해 해석(annotation)한 메타데이터 지식베이스부분, 두 번째로, 지식베이스를 정제하기 위해 메타데이터를 추론적용이 가능하게 변환하고 추론엔진을 통해 추론을 적용시키는 부분, 마지막으로 사용자의 의도로부터 질의문을

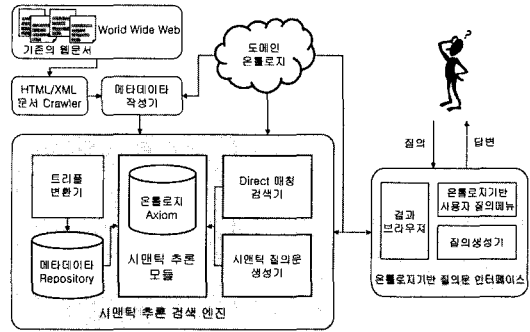


그림 2 시맨틱 추론 검색 시스템 구조

자동 생성하여 검색을 수행하고 결과를 출력하는 인터페이스부분으로 나눌 수 있다. 본 검색시스템은 단순한 키워드가 포함되어 있는 문서를 찾는 것이 아니라 사용자가 원하는 정보와 지식을 찾는 것에 중점을 두었다. 즉, 일반적인 문서가 아니라 온톨로지로부터 생성된 메타데이터를 바탕으로 이를 추론하고 온톨로지를 기반으로 질의문을 자동 생성하여 검색을 수행하는 방법을 사용한다는 것이다. 이러한 방법은 검색키워드와 문서의 키워드가 달라도 의미적으로 일치하면 온톨로지의 추론을 통해 검색이 가능하고 서로 다른 온톨로지의 컨셉을 규칙기반 추론을 통해 유사한 컨셉으로 추출하여 검색 영역의 확장을 이끌어낼 수 있다. 또한 검색 키워드의 뚜렷한 의미를 온톨로지를 통해 정의 할 수 있으므로 단순한 키워드 매칭 방법보다 정확한 검색이 가능하다. 이렇게 구축된 검색시스템의 정확성과 효율성은 잘 정의된 온톨로지와 빠른 속도의 추론엔진에 의해 많은 영향을 받게 된다.

본 논문은 정보 검색시스템의 구축에 있어서 온톨로지를 활용하여 문서들을 해석하여 정보검색시스템의 메타데이터로 사용함으로써 시맨틱 검색을 유도하고 사용자 요구에 해당하는 질의문을 자동 생성하여 검색의 정확성과 효율성을 높이는 방법을 집중 연구하였다. 다음 절에서는 전체 검색시스템의 구조에서 온톨로지의 구축 방안과 메타데이터 생성부분에 대해 구체적으로 설명하도록 한다.

3.1 온톨로지 구축 및 메타데이터 생성

본 논문에서 온톨로지의 역할은 도메인 내에서 사용되는 검색 키워드의 의미를 정의함으로써, 시맨틱 검색을 수행하여 검색의 정확도를 향상시키는데 있다. 이러한 온톨로지는 시맨틱 웹 환경에서 도메인내의 정보공유와 재사용을 하기 위한 중요요소로 사용된다. 본 논문에서의 온톨로지는 의미적 연관성을 통한 소프트웨어 에이전트로 하여금 이해할 수 있는 문서를 만들기 위한 시맨틱 웹 환경에서의 온톨로지이다[4]. 온톨로지는 기

본적으로 클래스(class)와 서브클래스(subClass) 그리고 프로퍼티(property), 도메인(domain), 레인지(range) 등의 구성요소를 가지고 도메인내의 정보를 표현하게 된다. 예를 들어 공연에 대한 온톨로지를 구성할 때 공연은 클래스로 선언이 되고 공연클래스의 프로퍼티중 이 공연의 지휘자라는 프로퍼티의 도메인은 공연클래스가 되고, 레인지는 지휘자의 형(type)이 되는 음악인 클래스가 된다.

본 검색시스템에서 자연어처리에 가까운 질의문을 처리하고 정확한 검색을 수행하기 위해서는 검색키워드의 의미를 온톨로지를 통해 뚜렷하게 정의해야 한다. 이는 검색의 키워드로 온톨로지의 다양한 클래스와 프로퍼티가 사용되기 때문이다. 따라서 시맨틱 검색시스템에서 보다 정확한 검색과 효율성의 문제는 도메인 온톨로지를 얼마만큼 정확하고 올바르게 모델링할 수 있는가에 대한 화두로 연결될 수 있다[5]. 앞서 언급한대로 예를 들어 "바로크시대 작곡가의 음악이 연주되는 공연은?"이라는 질의문을 처리한다고 하면 검색키워드가 되는 클래스에는 클래식음악 클래스와 작곡가 클래스가 된다. 각각의 클래스의 프로퍼티를 작곡가의 시대분류는 바로크시대로 선택하고 공연의 레퍼토리 속성에서는 클래식음악의 작곡가는 해당 작곡가로 설정하고 검색을 수행하게 된다. 여기서 온톨로지를 모델링할 때 있어 작곡가의 수상경력처럼 특정 클래스의 프로퍼티가 방대해질 수가 있다. 하지만 온톨로지를 모델링할 때 있어서 모든 상황을 고려할 수는 없다. 도메인 내에서 시맨틱 검색에 사용될 중요하고 필요한 부분만을 선택하고 나머지 부분은 qualification problem으로 생각할 수 있다. 또한 온톨로지의 클래스와 프로퍼티를 설정할 때 있어서는 시맨틱 검색시스템이 특정 도메인에 대하여 구축되기 때문에 해당 도메인 전문가의 의해 주요검색 키워드가 되는 속성을 특정 도메인 사용자에게 대한 프로파일 학습 기법이나 키워드의 빈도를 가중치로 계산하여 중요 속성을 선택하는 것과 같은 전처리과정을 수행하여 클래스와 프로퍼티를 설정하게 된다. 본 논문에서는 이러한 전처리 과정은 온톨로지 모델링에 관한 부분이므로 다루지 않도록 한다.

본 논문에서의 실제적인 검색시스템의 구현을 위하여 온톨로지는 온톨로지 구축도구를 이용하여 OWL(Ontology Web Language)의 형태로 구현하였다[6]. 본 논문에서 사용되는 온톨로지 표현 언어로 OWL 온톨로지 언어를 이용하는 이유는 OWL의 의미적 연관성과 미리 정의된 공리를 활용하면 다양한 추론이 가능하기 때문이다. 또한 시맨틱 검색시스템에서 시맨틱 추론 검색 엔진과의 연동부분과 에이전트가 시맨틱 정보 제공 기능을 효과적으로 수행할 수 있는 추론기관을 지원하는 장

점이 있다[6,7]. 따라서 본 논문에서는 OWL 온톨로지 표현 언어를 이용하여 온톨로지와 온톨로지 메타데이터를 표현한다. 다음 표 1은 OWL을 이용하여 온톨로지를 표현하는 예제이다.

표 1 OWL로 표현된 온톨로지

```
<owl:Class rdf:ID="Performance"/>
<owl:DatatypeProperty rdf:ID="hasTitle">
  <rdf:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  <rdf:domain rdf:resource="#Performance"/>
</owl:DatatypeProperty>
<owl:ObjectProperty rdf:ID="conductedBY">
  <rdf:domain rdf:resource="#Performance"/>
  <rdf:range rdf:resource="http://ailab.ssu.ac.kr/onto/Musician owl#Conductor"/>
</owl:ObjectProperty>
```

즉 공연이라는 컨셉을 온톨로지에서 표현할 때 공연이라는 클래스는 속성으로 공연 제목과 지휘자를 갖게 되고 공연 제목의 형태는 문자열이며 지휘자의 형태는 지휘자라는 클래스로 표현하게 된다. 표 1은 이와 같은 도메인 지식을 OWL로 표현하는 예제이다. 본 논문에서는 온톨로지 표현에 있어서 OWL에서 제공하는 subClassOf와 allValuesFrom 등 50여개의 공리를 사용하여 도메인 지식을 계층구조로 표현하게 된다.

이렇게 구축된 온톨로지를 바탕으로 기존의 문서와 생성된 문서들에 대해서 해석을 수행한다. 즉, 시맨틱 검색 시스템에서 사용하게 될 메타데이터를 생성하게 된다. 본 논문의 시맨틱 검색시스템내의 정보를 효율적으로 활용하기 위해서는 각각의 문서의 메타데이터 구축하고 이를 원본문서와 연결하여야한다. 시맨틱 정보를 효과적으로 해석하기 위해서는 다음 그림 3과 같은 개념적 구조의 방식을 취한다.

앞에서 언급한 방법으로 온톨로지를 구축하고 검색시스템의 도메인에서 사용될 기존의 문서와 새로 생성된 문서들을 수집하고 이를 해석도구를 사용하여 메타데이터로 생성한다. 온톨로지 구축과 메타데이터의 생성에 관한 부분은 현재까지는 수동적으로 수행되게 된다. 온톨로지의 경우 온톨로지 구축자가 구축도구를 사용하여 정의하게 되며 메타데이터의 경우도 해석도구를 사용하여 수동적으로 생성하게 된다. 이렇게 생성된 메타데이터에는 온톨로지의 다양한 클래스와 프로퍼티가 포함되어 정확한 의미를 표현하게 되어 원본문서의 의미적 정의가 가능하고, 기존의 문서들과 의미적 연관성을 갖게 된다. 즉, 온톨로지의 메타데이터에 도메인 정보들이 구체적인 내용으로 명시되어 있으므로 소프트웨어 에이전트가 효과적으로 추론하여 시맨틱 검색을 가능하게 하여 보다 정확한 검색과 풍부한 검색영역을 갖게 된다. 이러한 결과를 갖기 위해 앞에서 설명한 온톨로지를 추론하는 방식을 다음절에서 설명하도록 한다.

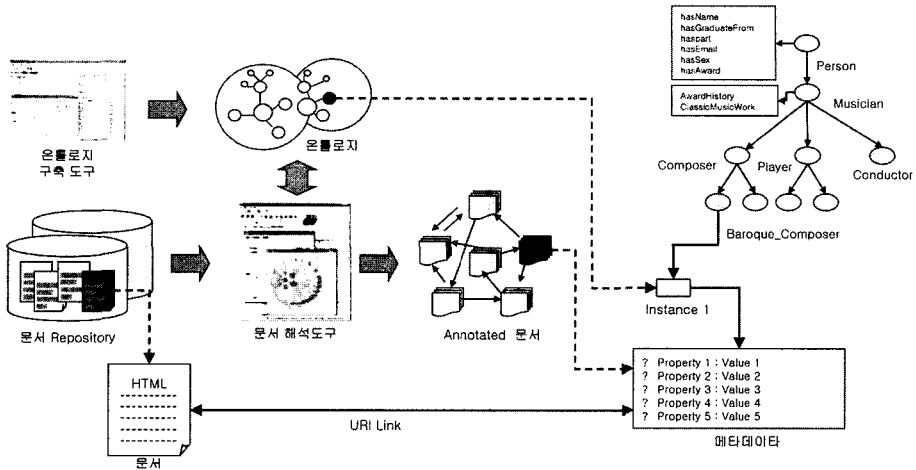


그림 3 Annotation을 통한 메타데이터 생성

3.2 추론을 사용한 시맨틱 검색

본 논문에서 제안하는 시맨틱 검색시스템에서의 검색 방법은 추론을 적용하여 전체 자원을 정제하고 시맨틱 추론 검색 엔진을 사용하여 사용자의 질의문에 응답을 하는 방식이다. 이러한 시맨틱 검색은 사용자의 의도를 정확히 파악하여 일반적인 키워드 매칭에 의해 나오지 않는 결과를 추론을 통해 풍부한 검색영역의 확장으로 이끌어낼 수 있게 된다. 앞에서 언급한대로 시맨틱 검색은 온톨로지를 바탕으로 생성된 메타데이터와 추론엔진을 통한 전방향 추론기법을 사용하여 검색하기 때문에 적용이 가능하다. 본 논문에서 제안한 시스템은 프롤로그(Prolog) 언어를 기반으로 시맨틱 추론 검색 엔진을 구축하였다. 시맨틱 추론 검색 엔진에서는 OWL로 생성된 메타데이터를 KIF(Knowledge Interchange Format) 형태로 읽어 들여 전방향을 추론을 수행할 수 있도록 술어(Predicate)형태로 변환과정을 거쳐서 추론 및 검색을 수행하게 된다[8-10]. 다음 그림 4는 본 시스템에서 추론엔진을 사용한 검색방법에 대한 개념도이다.

본 논문에서는 OWL로 작성된 온톨로지를 그대로 사용하여 추론을 할 수 없기 때문에 온톨로지의 변환이 필요하다. 기존의 RDF와 DAML+OIL에서 정의된 공리는 일차논리(First Order Logic) 표현방식을 따르고 있다. 즉, OWL 언어를 추론엔진을 통해 일차논리형식의 추론을 하기 위해서는 온톨로지를 <property> <subject><object> 형태의 트리플 형식으로 변환해야 한다. 본 시맨틱 검색 시스템의 트리플 변환기는 Jena의 RDF 파서를 사용하여 구현하였다[11,12]. 이렇게 온톨로지와 메타데이터를 트리플 변환기를 통하여 PSO형태의 트리플로 자동적으로 변환을 하고 다시 시맨틱 추론 검색엔진의 형태에 맞게 술어 형태로 변환 후 추론엔진

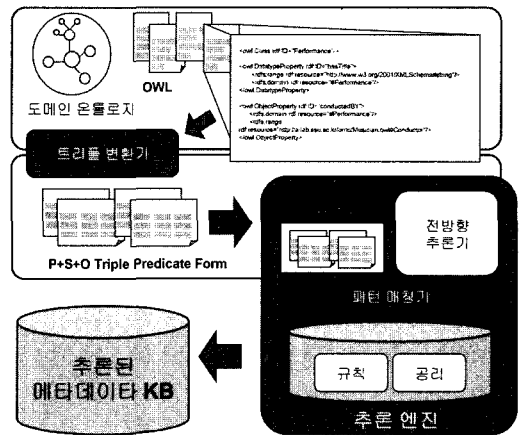


그림 4 시맨틱 검색시스템의 추론방식 개념도

의 작업메모리에 넣어 전방향 추론기를 통해 OWL의 공리와 사용자 요구 질의문을 바탕으로 구성된 규칙을 적용하여 전방향 추론을 수행한다.

본 논문에서 사용되는 지식베이스 정제방식은 기존의 구축된 메타데이터를 사용하여 시맨틱 검색을 수행할 때 추론엔진은 OWL 정의하고 있는 공리를 사용하여 기존의 메타데이터를 풍부하게 하거나 또는 축약하는 방식을 사용한다.

예를 들어 공연에 있어 지휘자에 관한 검색을 수행할 때 지휘자인 Conductor 클래스는 음악인인 Musician 클래스를 subClassOf의 관계로 정의하게 되고 추론엔진을 통해서 Conductor의 형도 Musician이라는 사실이 지식베이스에 추가되게 된다. 이렇게 OWL에서 미리 정의된 50여개의 공리를 사용하여 트리플로 변환된 메타데이터를 추론하게 되면 기존의 메타데이터보다 더욱

풍부한 자원을 확보하게 되고 결과적으로 시맨틱 검색의 영역이 확대되는 결과를 가져온다[13]. 또 다른 방식은 트리플로 변환된 메타데이터의 지식베이스에 사용자가 요구하는 질의문의 합당한 규칙을 자동 생성하여 추론을 적용시켜 방대한 지식베이스를 검색조건에 맞는 최적의 상태로 정제가 가능하다. 이러한 추론을 사용한 검색방법은 기본적으로 대용량의 지식베이스에서 검색을 수행하는 것 보다 수행시간을 감소시킬 수 있다. 다음 그림 5는 지식베이스의 메타데이터를 추론엔진으로 정제하는 과정을 나타낸다. 이렇게 정제된 지식베이스는 결과적으로 시맨틱 검색을 수행하는 시맨틱 추론 검색 엔진 처리부분의 자원이 된다. 본 논문에서 온톨로지를 추론하여 시맨틱 검색을 수행하는 방식은 새로운 자원의 추가와 방대한 자원의 축약이라는 두 가지 접근 방법으로 보다 정확한 검색과 풍부한 검색 영역을 갖게 된다. 본 시스템은 이러한 두 가지 방식을 사용자의 의도를 정확히 판단하여 동적으로 수행하였다.

3.3 시맨틱 검색 메커니즘

본 논문에서는 사용자의 입력 질의문에서 정확하게 매칭 되는 부분이 지식베이스에 없을 경우에 시맨틱 질의문을 생성하여 시맨틱 검색을 수행하는 시맨틱 검색 메커니즘을 제안한다. 본 논문에서는 시맨틱 검색을 수행하기 위해서 두 단계의 검색 방법을 거친다. 첫 번째는 온톨로지 기반의 직접 검색 방법으로 사용자의 질의문과 메타데이터의 속성이 정확하게 매칭 되는 부분을 검색하는 방법이다. 두 번째는 온톨로지 기반의 추론 검색방법으로 메타데이터에서 온톨로지의 계층구조와 공리로 표현된 부분을 규칙기반의 추론을 통해서 사용자의 질의문을 확장시켜서 검색하는 방법이다. 본 검색

시스템에 있어 시간적인 효율성을 감안할 때 매번 비교적 오랜 시간이 걸리는 추론형태의 검색을 사용하는 것이 아니라 정확히 매칭될 때에는 온톨로지 기반의 직접 검색방법을 사용하고 정확한 매칭이 없을 때에 추론 검색방법으로 질의를 확장하여 검색하는 것이 검색 시스템의 효율성을 고려한 수행방법이다. 다음 표 2는 시맨틱 검색 메커니즘을 적용하여 사용자가 입력한 질의문이 온톨로지의 계층구조상에서 어느 곳에 위치하느냐에 따라서 시맨틱 질의를 확장하는 방안을 나타낸다.

사용자가 질의문의 입력을 통해서 알고자하는 클래스가 온톨로지의 계층구조상에서 어디에 위치하는가에 따라서 각각 다른 시맨틱 질의문을 생성하여 검색을 수행하게 된다. 즉, 자식 노드가 있는 경우에는 자식 노드를 검색할 수 있는 질의문을 생성하고 단말 노드인 경우에는 부모 노드를 검색할 수 있는 질의문과 형제 노드를 검색할 수 있는 질의문을 생성하게 된다.

본 논문에서는 시맨틱 검색을 수행하기 위한 시맨틱 검색 메커니즘에 대해 제안한다. 앞서 언급한 시맨틱 질의문을 생성한 후에는 시맨틱 검색 알고리즘에 따라서 검색을 수행한다. 본 논문의 시맨틱 추론 검색 엔진은 다음 그림 6과 같은 알고리즘에 따라서 시맨틱 검색을 수행한다. 초기에 사용자로부터 입력된 질의문의 검색 조건은 단일 검색 조건일 수도 있지만 일반적인 경우에 다중 검색 조건의 형태를 갖는다. 초기에 시맨틱 추론 검색 엔진은 첫 번째 검색 조건부터 순차적으로 지식베이스와 단일화(Unification)를 시도한다. 이때 매칭이 성공되면 다음 검색 조건으로 넘어가게 되고 매칭이 실패했을 때 기존의 입력 질의문을 확장하여 시맨틱 질의문을 생성하고 다시 매칭을 시도하게 된다. 매칭의 시도

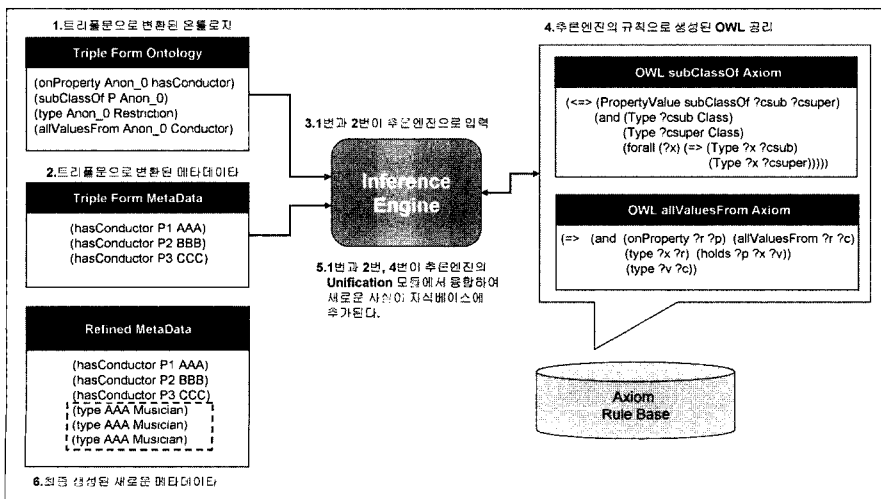


그림 5 추론엔진을 사용한 메타데이터 정제

표 2 시맨틱 검색 메커니즘

	검색노드 위치	입력 질의문	시맨틱 검색 질의문
child노드가 있을 경우 (childMatch)		$Q(?a, predicate, ClassA)$	$Q(?b, subclassOf, ClassA)$ $Q(?a, predicate, ?b)$
leaf노드일 경우 parent (parentMatch)		$Q(?a, predicate, ClassA)$	$Q(ClassA, subclassOf, ?b)$ $Q(?a, predicate, ?b)$
leaf노드일 경우 sibling (siblingMatch)		$Q(?a, predicate, ClassA)$	$Q(ClassA, subclassOf, ?b)$ $Q(?c, subclassOf, ?b)$ $Q(?a, predicate, ?c)$ <i>/* ClassA != ?c */</i>

Q_k : Set of query condition
 match: Direct match
 semanticMatch: Query inference by ontology
 childMatch: child node match
 parentMatch: parent node match
 siblingMatch: sibling node match

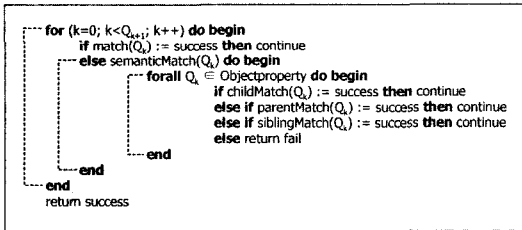


그림 6 시맨틱 검색 알고리즘

순서는 사용자의 입력 클래스에서부터 지식과 부모, 형제 노드의 차례로 매칭을 시도한다.

3.4 온톨로지기반의 질의문 자동생성

본 논문의 시맨틱 검색시스템에서 자연어와 유사한 검색영역을 갖기 위해서는 온톨로지를 기반으로 사용자의 요구에 맞는 질의문을 시맨틱 추론 검색엔진의 형태에 맞는 정확한 질의문으로 자동 생성하여야 한다. 다음 그림 7은 온톨로지기반 질의문 자동 생성방식에 대한 구조이다. 시맨틱 검색의 사용자 인터페이스는 사용자의 요구를 다양하게 포함할 수 있는 메뉴방식으로 구현하고 이를 통해 사용자의 요구로부터 검색 키워드를 추출하게 된다. 추출된 검색 키워드는 메타데이터 자원매칭기로 전송 되어 시맨틱 추론 검색 엔진의 형태에 맞게 재구성하여 메타데이터 자원의 해당하는 트리플문의 URI 형식으로 변환된다. 즉, 메타데이터 자원매칭기는 메타데이터 자원의 트리플문과 재구성된 검색키워드에 연결된 URI를 매치시킨다. 키워드 추출기와 메타데이터

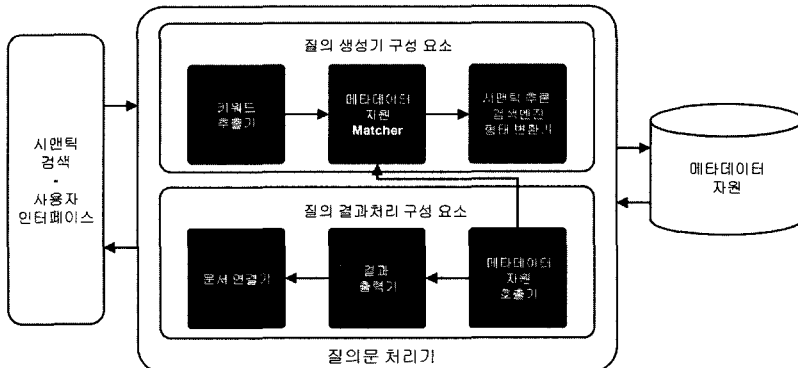


그림 7 온톨로지기반 질의문 자동 생성구조

자원매칭기로부터 재구성된 검색키워드는 온톨로지 기반 질의문 변환기를 통해 시맨틱 추론 검색엔진의 질의문으로 자동 생성되게 된다. 그리고 자동 생성된 질의문은 추론을 통해 메타데이터정제를 거친 메타데이터 자원으로부터 검색을 수행하게 된다. 질의문 처리기에 의한 검색 결과는 URI 같은 불필요한 요소를 제거하는 일반화 과정을 거쳐서 원본문서와 연결하는 과정을 수행하게 된다. 마지막으로 사용자는 검색 결과를 사용자 인터페이스를 통해 원본문서를 얻게 된다. 이러한 기능은 전체 시스템의 질의문 처리기에서 관장하도록 한다.

자동 생성된 온톨로지 기반의 질의문은 검색 조건절과 결과 조건절로 구성된다. 검색 조건절은 검색을 수행하기 위한 다양한 조건들로서 단일 검색 조건과 다중 검색 조건을 동시에 처리하게 된다. 결과 조건절은 사용자가 검색 조건을 이용하여 찾고자하는 결과를 의미하여 검색 조건절과 결과 조건절은 <property>, <subject>, <object> 형태로 구성되며 프롤로그의 술어 형태를 갖는다. 이러한 구성을 가지는 온톨로지 기반의 질의문을 사용하여 본 시맨틱 검색시스템의 질의문 처리기는 자연어처리에 가까운 다양한 요구를 처리하기 위해 온톨로지를 통해 정의된 클래스와 프로퍼티를 사용자가 자유롭게 선택함으로써 자동적으로 질의문을 생성하는 방식을 사용하였다. 앞서 언급한대로 “바로크시대 작곡가의 음악이 연주되는 공연은?” 이라는 질의문을 사용자가 검색인터페이스를 통하여 선택하여 질의문 처리기로부터 자동 생성하게 된다. 사용자 인터페이스로부터 사용자가 입력한 질의문의 변수들은 검색변수와 목표변수로 구분되게 된다. 검색변수는 사용자의 검색조건이 되며, 목표변수는 사용자가 찾기 원하는 정보가 된다. 이렇게 입력받은 검색조건은 질의문 처리기에 의해 시맨틱 추론 검색 엔진에 맞는 형태로 자동 생성되게 되어 최종적으로 검색을 수행하게 된다.

4. 실험 및 평가

지금까지 시맨틱 검색 시스템에 관한 정량적인 성능 평가에 관한 연구는 거의 보고되지 않았다. 따라서 본 논문에서의 제안하는 시맨틱 검색시스템의 정확성과 효율성을 검증하기 위해서 검색 시나리오로부터 결과를 도출하는 실험을 수행하고 온톨로지 기반의 질의응답을 지원하는 DQL과 제안한 시스템의 결과를 비교함으로써 시스템의 성능 평가를 수행하였다.

실제적인 검색시스템 구현을 위하여 시스템의 도메인은 문화컨텐츠로 한정하여 구축하도록 한다. 구체적인 검색 도메인은 영화와 클래식 음악 공연에 관한 것으로서 2004년도에 개봉한 영화와 클래식 공연을 중심으로 지식 엔지니어링을 통해 충분한 시맨틱 검색상황을 고

려하여 온톨로지를 구축하였다. 그리고 각 해당 메타데이터는 실제 존재하는 영화와 공연에 대한 웹문서들을 대상으로 문서해석을 통해 메타데이터로 생성하였다. 그리고 본 논문에서 제안하는 시스템의 각 구성별로 트리플 변환기, 시맨틱 추론 엔진, 질의문 처리기를 포함하는 사용자 인터페이스를 구현하였다. 본 논문에서는 시맨틱 검색을 실험하기 위하여 웹 기반으로 시맨틱 검색 시스템을 구현하였다. 본 논문의 시맨틱 검색 시스템에서 사용자는 원하는 검색 키워드 정보를 온톨로지를 기반으로 메뉴형식으로 표현되는 검색 콤보박스에서 쉽게 도메인 정보를 선택해서 검색을 수행할 수 있다.

4.1 온톨로지 추론을 통한 검색의 경우

본 논문에서 온톨로지 추론을 통하여 메타데이터를 검색하는 방법을 사용한다. 다음 그림 8은 기존 시스템에서 불가능한 검색 질의를 온톨로지 추론을 사용하여 검색이 가능한 시나리오이다. 다음과 같이 공연에 대한 온톨로지가 정의되어 있고 공연에 대한 인스턴스로서 C음악회가 메타데이터로 존재하게 된다. 여기서 사용자가 원하는 질의로 “세종문화 회관에서 바로크 음악이 연주되는 공연은 무엇인가?” 라고 입력하였을 때 기존의 검색 시스템에서 정확한 답변을 하기 힘들다. 왜냐하면 문서에 포함되는 키워드에 바로크 음악이라는 부분으로 매칭 되는 부분이 없기 때문이다. 본 논문의 시스템에서는 작곡가 ‘헨델’에 대한 정보가 계층구조로서 ‘바로크 음악’에 속한다는 것이 온톨로지에 존재하고 음악 온톨로지의 인스턴스로서 ‘물위의 음악’에 대한 메타데이터와 ‘물위의 음악’이 내포된 C공연의 메타데이터가 존재하기 때문에 사용자의 질의에 대해서 C음악회라고 답변할 수 있게 된다.

4.2 검색키워드와 대상 문서의 키워드가 다를 경우

본 논문의 방식은 온톨로지 기반의 추론을 사용하여 검색을 수행하므로 기존의 검색방법에서 movie라는 키워드와 cinema라는 키워드는 엄연히 다른 것처럼 검색 키워드와 사용자가 찾고자하는 문서의 키워드가 불일치하면 의미적 맥락이 같아도 검색이 되지 않는 점을 표 3과 그림 9에서 보듯이 본 검색 시스템에서는 서로 다른 검색키워드를 온톨로지 기반 추론엔진을 통해 규칙을 적용시켜서 정확한 일치가 아니어도 유사한 컨셉을 찾아주는 검색이 가능하다. 온톨로지 규칙에 의해 movie 클래스와 cinema 클래스는 sameClassAs라는 공리로 정의되어 있다. 즉, movie 클래스와 cinema 클래스로부터 생성된 각각의 메타데이터는 규칙에 의해 유사한 컨셉으로 추론이 가능하다. 그러므로 사용자가 cinema 클래스로부터 검색을 수행했을 때 본 시스템은 첫 번째로 cinema 클래스의 메타데이터 이하다 찾고 온톨로지 기반의 공리를 이용한 추론에 의해 movie 클래스의 메타데

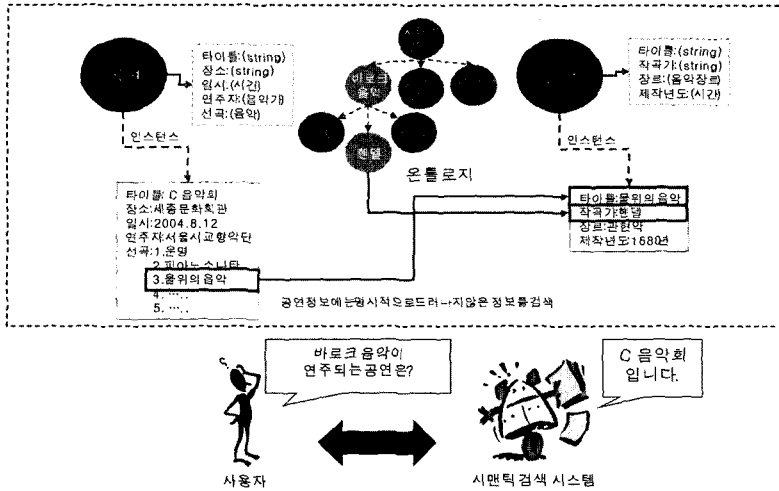


그림 8 온톨로지 추론을 이용한 검색 시나리오

표 3 검색 시나리오 1

<p><사용자의 검색 의도 : 영화배우 안성기가 주연한 영화는 ?></p> <ul style="list-style-type: none"> - 일반적인 검색 키워드 영화배우, 안성기, 영화(cinema) - 시맨틱 검색키워드 영화(cinema sameClassAs movie) (주연배우 : 안성기) + 영화 (영화제목 : A) - 규칙 기반 변환기에 의해 자동 생성된 시맨틱 질의문 Q(?a. sameClassAs cinema) Q(?b. type, ?a) Q(?b. <N#hasActor>, <N#안성기>) Q(?b. <N#movieTitle>, ?c)
<p>- 질의 시나리오 1 검색결과</p> <p>Result ?c ?c = 실미도</p>

이하다 자동으로 검색하게 된다. 이는 사용자가 원하는 정보가 의미적으로 유사하지만 단지 키워드가 다르기 때문에 검색이 불가능함을 가능하게 하고 사용자로 하여금 불필요하게 movie와 cinema로 두 번 검색을 하는 번거로움을 감소시켜준다.

4.3 온톨로지 추론을 이용한 유사한 컨셉의 검색의 경우

본 논문에서는 온톨로지의 계층구조를 이용한 추론을 검색에 적용시킨다. 다음의 표 4와 그림 11의 검색 시나리오 보듯이 사용자가 원하는 질의의 의미를 이해하여 A-1에 위치하는 T-1 극장에서 영화X를 상영하지는 않지만 Location 온톨로지의 계층구조를 추론하여 A-1과 장소가 유사한 A-2에 위치하는 T-2를 검색하여 사용자에게 T-2극장에서 영화X가 상영한다는 정보를 자동적으로 알려주는 것이 가능하다. 이처럼 유사한 정보가

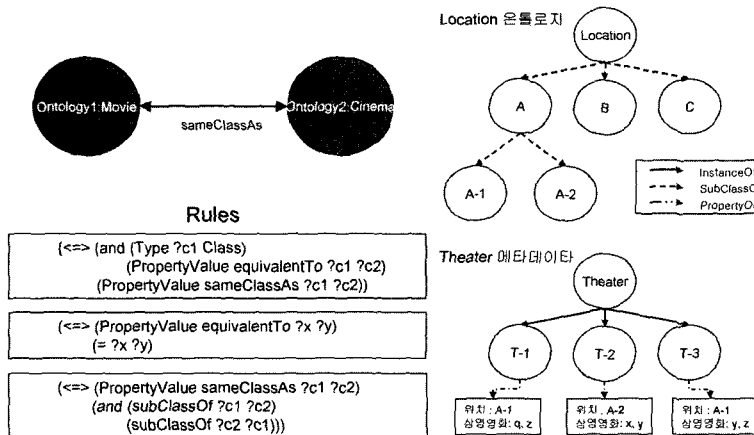


그림 9 온톨로지 추론 시나리오

표 4 검색 시나리오 2

<p><사용자의 검색 의도 : 위치 A-1의 극장에서 X영화를 상영하는 극장은 ?></p> <p>- 일반적인 검색 키워드</p> <p style="padding-left: 40px;">A-1, X영화, 극장</p> <p>- 시맨틱 검색 키워드</p> <p style="padding-left: 40px;">극장 (위치 : A-1) + 극장 (상영작 : X) + 극장 (극장이름 : ?)</p> <p>- 시맨틱 검색 시스템에 의해 자동 생성된 시맨틱 질의문</p> <p style="padding-left: 40px;">Q(A-1, subClassOf ?b) Q(?c, subClassOf, ?b) (: : ?c not equal A-1) Q(?a, <N#theaterLocation>, ?c) Q(?a, <N#theaterOnMovie>, <N#X>) Q(?a, <N#theaterName>, ?n)</p>
<p>- 질의 시나리오 2 검색 결과</p> <p style="padding-left: 40px;">Result ?n ?n = 1-2</p>

있지만 단순한 검색키워드의 빈도수와 키워드와 문서의 유사도만으로 찾을 수 없는 정보를 온톨로지 기반 검색 시스템은 사용자의 의도를 파악하여 자동으로 검색을 수행하게 되는 것이다. 이것은 온톨로지 기반으로 도메인의 정보들이 의미적인 연관성을 갖고 표현되어 있기 때문에 추론을 통하여 가능하다.

4.4 시맨틱 검색의 다양성

본 논문의 방식은 도메인 정보를 온톨로지를 통해 구체적으로 정의하고 있으므로 사용자가 선택하는 검색키워드의 의미를 검색시스템으로 하여금 정확히 이해할 수 있게 하여 표 3과 같이 보다 정확한 검색을 할 수 있다. 다음 표 5의 검색시나리오에서 보듯이 영화의 제목을 모르는 상태에서 남자주연배우의 출신학교를 검색하려면 몇 번의 검색을 거쳐야 한다. 또한 일반적인 검색에서 질의문으로써 검색 키워드를 나열할 때 출신학교라는 키워드는 영화의 구성요소의 어떤 부분에서도 매칭이 될 수 있다. 하지만 온톨로지기반으로 검색 키워드를 정의 할 수 있으므로 사용자는 정확히 남자주연배우의 출신학교라는 질의문을 시스템에서 선택할 수 있고 검색 시스템도 그 의미를 정확히 이해하여 검색을 수행할 수 있다.

이처럼 본 검색 시스템은 온톨로지를 통해 사용자가 자연어 검색과 유사한 복잡한 질의문을 쉽게 자동 생성하여 검색의 영역을 넓혀주는 효과를 가져다준다. 이와 더불어 본 검색 시스템은 시맨틱 웹 기술을 사용하여 사용자가 원하는 정보와 지식을 찾아주게 된다. 예를 들어 사용자가 표 6과 같은 질의문을 요청했을 때 검색 시스템은 단순히 사용자가 원하는 키워드가 포함된 문서, 즉 영화제목만을 포함하는 문서만을 찾는 것이 아니라 실제영화가 상영하는 극장과 주연배우와 감독에 관한 정보 등등 관련된 정보를 자동적으로 찾아줄 수 있다.

표 5 검색 시나리오 3

<p><사용자의 검색 의도 : 2002년 베니스 영화제에서 감독상을 수상한 영화의 남자주연배우의 출신학교는 ?></p> <p>- 일반적인 검색 키워드</p> <p style="padding-left: 40px;">2002년, 베니스영화제, 감독상, 남자주연배우, 출신학교</p> <p>- 시맨틱 검색키워드</p> <p style="padding-left: 40px;">영화제(영화제이름 : 베니스) + 영화제(수상부문 : 감독상) + 영화제(수상년도 : 2002년) + 영화(영화제목 : A) + 영화(A의 남자주연배우 : B) + 배우(배우이름 : B)(출신학교 : C)</p> <p>- 시맨틱 검색 시스템에 의해 자동 생성된 시맨틱 질의문</p> <p style="padding-left: 40px;">Q(?n, <N#AwardName>, <N#Venice>) Q(?n, <N#hasYear>, <N#2002>) Q(?n, <N#hasPosition>, <N#감독상>) Q(?a, <N#movieAwardHistory>, ?n) Q(?a, <N#hasLeadingActor>, ?b) Q(?b, <N#graduateFrom>, ?c)</p>
<p>- 질의 시나리오 3 검색 결과</p> <p style="padding-left: 40px;">Result ?c ?c = 한양대학교</p>

표 6 검색 시나리오 4

<p><사용자의 검색 의도 : 한국 영화중에서 흥행 성적이 1000만인 영화는 ?></p> <p>- 일반적인 검색 키워드</p> <p style="padding-left: 40px;">한국, 영화, 흥행성적, 1000만</p> <p>- 시맨틱 검색키워드</p> <p style="padding-left: 40px;">영화 (제작국가 : 한국) + 영화 (흥행성적 : 1000만) + 영화 (영화제목 : A)</p> <p>- 시맨틱 검색 시스템에 의해 자동 생성된 시맨틱 질의문</p> <p style="padding-left: 40px;">Q(?n, <N#madeFrom>, <N#Korea>) Q(?n, <N#score>, <N#1000만>) Q(?n, <N#movieTitle>, ?a)</p>
<p>- 질의 시나리오 4 검색결과</p> <p style="padding-left: 40px;">Result ?a ?a = 대극기 휘날리며</p> <p style="padding-left: 40px;">?a (상영극장) (JCinema) ?a (남우주연배우) (장동건) ?a (남우주연배우) (연민) ?a (감독) (김지규)</p>

이와 같이 사용자는 시맨틱 검색 사용자 인터페이스를 통해서 찾고자 하는 지식과 정보를 정확히 찾게 된다. 즉, 사용자 인터페이스로부터 검색 시나리오에 따른 지식과 정보를 결과 키워드 또는 검색결과에 연결된 원본 문서들에게서 찾을 수 있게 된다.

4.5 DQL 시스템과의 비교

DQL 시스템과 제안한 시스템의 성능분석을 위하여 본 논문에서는 다음과 같은 실험을 하였다. 즉, 동일한 지식베이스를 공유하는 DQL 시스템과 시맨틱 검색 시스템을 상대로 동일한 질의문을 입력하여 각각 출력된 검색결과에 대한 평균적인 정확도와 검색조건의 수에 따른 정확도를 살펴보았다. 다음 그림 10은 검색 조건에 대한 사용자들의 검색 만족도 평균값을 그래프로 나타낸 모습이다. 그림 10의 그래프에서 알 수 있듯이 본 논문에서 제안하는 시스템은 동일한 지식베이스를 공유하는 기존의 DQL 검색 시스템보다 좀 더 나은 검색의 정

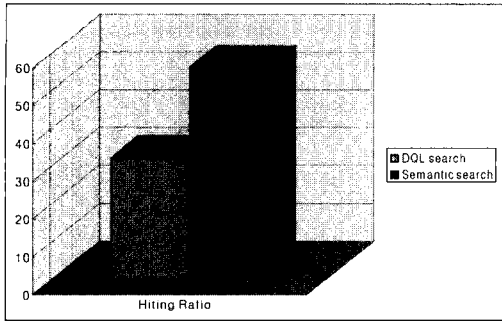


그림 10 질의문에 따른 검색 만족도

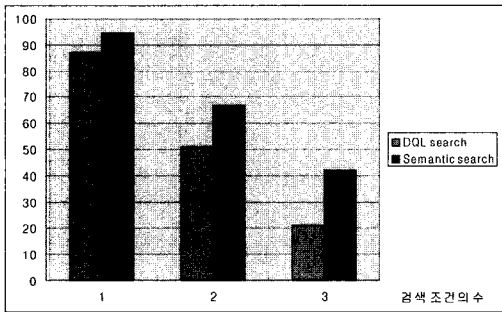


그림 11 검색 조건수에 따른 검색 정확도

확도를 보였다. 이것은 DQL이 사용자의 질의문과 지식 베이스의 메타데이터가 정확한 일치치를 보일 때만 검색을 수행할 수 있는 반면에 본 논문의 시스템은 정확한 일치치를 찾아내지 못했을 때 온톨로지를 추론하여 유사한 컨셉을 검색할 수 있는 질의문으로 확장시켜서 재검색을 수행하기 때문에 결과 그래프에서 보듯이 더 나은 정확도를 보인다. 다음 그림 11은 사용자가 입력한 검색 조건의 수에 따른 검색의 정확도를 나타낸 그래프이다. 본 그래프에서는 다수의 검색조건을 갖는 질의문이 검색의 정확도가 떨어지는 사실을 알 수 있다. 이것은 다수의 검색조건을 만족하는 메타데이터가 부족한 현상으로 해석할 수도 있지만 기존의 DQL 시스템과 본 논문의 시스템의 정확도를 비교해보면 검색조건이 늘어날 때 본 시스템이 낮은 감소율을 갖는다는 것을 알 수 있다. 이것은 검색조건이 많이 질수록 온톨로지로부터 추론될 수 있는 가능성이 증가하는 것과 다수의 검색조건을 만족하는 메타데이터가 없을 때 오히려 온톨로지를 추론하여 나올 수 있는 메타데이터가 늘어날 수 있다는 사실을 보여준다. 본 논문에서 제안하는 시스템은 정확한 매칭이 이루어지지 않을 때에 온톨로지를 추론하여 질의문을 확장하는 방식이다. 이는 기존의 DQL 검색 시스템보다 동일하거나 그 보다 많은 검색결과를 출력한다. 단순히 검색된 결과의 양만으로 검색 시스템의 정

확성과 효율성을 평가하기는 어렵다. 그러므로 본 논문에서 제안하는 시스템의 효율성을 증명하기 위해서 온톨로지를 추론하여 나온 검색결과에 대한 타당성을 검증할 실험을 추가적으로 수행하였다. 아직까지 시맨틱 검색 시스템에 관한 정량적인 성능평가에 관한 연구는 거의 보고되지 않았고 추가적으로 검색된 부분에 대한 평가가 이루어지기 힘들기 때문에 본 논문에서는 제안하는 시맨틱 검색시스템의 정확성과 효율성을 검증하기 위해서 다양한 검색 시나리오를 바탕으로 동일한 질의문을 DQL 시스템과 본 시스템에 적용한 후 사용자로부터 검색결과에 대한 만족도를 평가받는 실험을 수행하였다. 다양한 검색조건을 상대로 사용자에게

검색 만족도 평가를 5단계로 나누어 입력하도록 하였다. 여기서 검색 만족도의 5단계 중에서는 원치 않은 결과에 대해서는 음수를 입력하여 의도하지 않은 결과에 대해서는 가중치를 높게 주었다. 실험은 숭실대학교 컴퓨터학과에 재학 중인 대학원생과 일반인으로 구성된 피실험자들로부터 100개의 실험 데이터를 수집하였다. 다음 그림 12는 검색 조건에 따른 사용자의 만족도를 수치화하여 나타낸 그래프이다. 일반적으로 사용자의 질의문과 정확한 매칭을 이루는 검색조건과 같은 검색조건 1-4 는 기존의 DQL 시스템과 시맨틱 검색 시스템이 동일한 사용자 만족도를 얻었다. 또한 검색조건에 대해 DQL 시스템의 정확한 매칭의 결과가 없고 시맨틱 검색 시스템의 온톨로지를 추론한 결과도 없을 때인 검색조건 9-12 는 동일하게 사용자 만족도를 얻지 못하였다. 여기서 검색조건 5-8은 DQL 시스템이 결과를 내지 못한 부분에 대해 본 논문의 시스템이 온톨로지를 추론하여 유사한 컨셉에 대한 출력을 해 주었을 때 사용자에게 대한 만족도를 나타낸다. 여기서 검색조건 7에 해당하는 사항은 온톨로지를 추론하여 나온 검색 결과에 대해 사용자들은 오히려 낮은 점수를 부여함으로써 의도하지 않은 결과를 표현하였다. 그 외의 결과들은 DQL 시스템보다 나은 검색 만족도를 기록하였다. 즉, 본 시스템

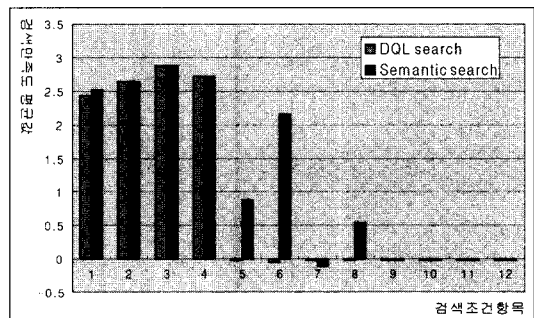


그림 12 검색 조건에 따른 사용자 만족도

에 의하여 온톨로지 기반의 유사한 컨셉이 검색되었을 때 DQL 시스템보다 검색 만족도 부분에서 상대적으로 높은 점수를 얻은 점을 알 수 있다.

5. 결론 및 향후 연구

정보의 양이 방대해질수록 정작 사용자가 원하는 정보와 지식은 찾기 힘들다. 이러한 정보의 홍수 속에서 시맨틱 웹 기술을 사용한 시맨틱 검색방식은 사용자에게 유용한 정보와 지식을 보다 정확하고 쉽게 찾게 해준다. 본 연구는 시맨틱 웹 기술과 온톨로지를 활용하여 효율적인 검색 시스템을 만들고자 하였다. 본 논문에서 제안하는 방법은 시맨틱 검색시스템으로 복잡한 질의문을 수행하기 위하여 온톨로지를 구축하고 온톨로지바탕의 클래스와 다양한 프로퍼티를 사용하여 온톨로지 기반 질의문으로 자동생성하고 트리플 컨버터를 구축하여 추론엔진으로 하여금 OWL에서 정의된 공리로 온톨로지와 메타데이터 지식베이스의 추론을 통해 별도의 자연어 처리 없이 복잡한 질의문의 검색을 보다 정확하게 수행하는 방식이다. 이러한 추론 검색방식은 질의문을 온톨로지를 기반으로 확장시키는 형태로써 키워드가 달라도 의미적으로 유사한 컨셉을 찾아주게 된다. 본 논문은 실제 도메인 온톨로지를 작성하여 이를 기반으로 검색 시스템에서 사용자로 하여금 쉽게 복잡한 질의문에 해당하는 검색결과를 얻게 하고 정확한 매칭으로 찾을 수 없는 결과를 온톨로지 추론으로 얻을 수 있게 하였다. 이는 사용자로 하여금 불필요한 재검색의 불편을 없애고 자동적인 정보와 지식을 제공받을 수 있게 하는 장점을 갖는다.

본 논문은 시맨틱 웹 환경에서의 검색 기법을 제안하였다. 본 검색 시스템은 온톨로지의 공리를 충분히 고려한 검색 시스템으로 기존의 온톨로지만의 계층구조를 이용하는 검색방법보다 효율성을 높일 수 있었고 정확한 매칭의 한계를 보완하는 검색 방법으로 시맨틱 검색 기법을 제안하였다. 이는 온톨로지를 근거로 하여 유사한 의미의 판별을 제안함과 동시에 온톨로지 공리를 사용함으로써 검색의 유연성을 높일 수 있는 방법이다.

향후 연구에서는 온톨로지 모델링 시에 클래스와 프로퍼티를 적절하게 설정하는 전처리 부분에 대한 연구가 필요하겠다. 현재 온톨로지 모델링과 온톨로지 합병, 온톨로지 변환 등의 관련된 연구가 진행되고 있다. 이와 더불어, 구축된 온톨로지로 하여금 동적으로 메타데이터를 생성하는 부분에 대한 연구가 진행된다면 좀 더 효율적인 시맨틱 검색시스템 구축에 큰 역할을 할 것으로 예상된다.

참고 문헌

- [1] Tim Berners-Lee, James Hendler and Ora Lassila, "The Semantic Web," Scientific American, 2001.
- [2] R. Guha, Rob McCool, Eric Miller "Semantic Search," WWW2003 conference paper submission.
- [3] Richard Fikes, Pat Hayes, Ian Horrocks, "DQL-A Query Language for the Semantic Web," WWW 2003, May 20-24, 2003, Budapest, Hungary.
- [4] Deborah L. McGuinness and Richard Fikes, James Hendler, Lynn Andrea Stein, "DAML+OIL: An Ontology Language for the Semantic Web," IEEE 2002.
- [5] Thomas R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," Stanford Knowledge Systems Laboratory, 23 August 1993.
- [6] Michael K. Smith, Chris Welty, Deborah L. McGuinness, "OWL Web Ontology Language Guide," W3C Recommendation 10 February 2004.
- [7] Dan Connolly, Frank van Harmelen, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, Lynn Andrea Stein, "DAML+OIL Reference Description," W3C Note 18 December 2001.
- [8] Fikes, Richard, Jessica Jenkins, and Gleb Frank. "JTP: A System Architecture and Component Library for Hybrid Reasoning," Proceedings of the Seventh World Multiconference on Systemics, Cybernetics, and Informatics. Orlando, Florida, USA. July 27-30, 2003.
- [9] "Knowledge Interchange Format," draft proposed American National Standard (dpANS), <http://logic.stanford.edu/kif/dpans.html>
- [10] Fikes, Richard, Jessica Jenkins, and Qing Zhou. "Including Domain-Specific Reasoners with Reusable Ontologies," Proceedings of the 2003 International Conference on Information and Knowledge Engineering. Las Vegas, Nevada, USA. June 23-26, 2003.
- [11] Frank Manola, Eric Miller, "RDF Primer," W3C Working Draft 23 January 2003.
- [12] Andy Seaborne, "Jena Tutorial : A Programmer's Introduction to RDQL," April 2002.
- [13] Richard Fikes, Deborah McGuinness, "An Axiomatic Semantics for RDF, RDF-S, and DAML+OIL," W3C, December 2001.



하 상 범

2003년 2월 성결대학교 컴퓨터공학과 졸업(학사). 2005년 2월 숭실대학교대학원 컴퓨터학과 졸업(석사). 관심분야는 시맨틱 웹, 온톨로지 추론, 개인화 에이전트, 유비쿼터스 컴퓨팅 등

**박 영 택**

1978년 서울대학교 전자공학과 졸업(학사). 1980년 KAIST 전산학 졸업(석사) 1992년 Univ. of Illinois at Urbana-Champaign 졸업(박사). 1981년~현재 숭실대학교 컴퓨터학과 교수. 관심분야는 개인화 에이전트, 시맨틱 웹, 유비쿼터스

컴퓨팅, 기계학습 등