

# 점진적 전방향 추론을 이용한 DQL 검색 인터페이스

김제민\*, 박영택  
송실 대학교 컴퓨터학과

kimjemins@hotmail.com park@ssu.ac.kr

## DQL Search Interface Using Incremental Forward Chaining

Je min Kim\*, Young Tack Park  
Dept of Computer Science, Soongsil University

### 요 약

인터넷의 등장 이후 폭발적으로 증대되는 웹 정보를 효율적으로 사용하기 위한 방안이 연구 되고 있다. 현재, 웹 서비스에서 사용하고 있는 키워드 기반 검색은 syntactic 정보만을 제공하므로 원하는 서비스를 받고자하는 사용자의 의도와 상관없는 정보를 탐색하여 서비스한다. 그러나 시멘틱 웹을 기반으로 만들어진 시스템들은 서비스 정보의 질을 향상시키기 위하여 DQL(DamI Query Language)을 이용한 시멘틱 서치(Semantic search)기법을 사용하고 있다. 시멘틱 서치는 사용자가 입력한 질의들을 의미상으로 해석하고, 이를 통하여 사용자의 원하는 정보와 의미가 같은 해답을 얻어내는 것이다. 그러나 기존의 시멘틱 서치는 사용자가 복잡한 질의들을 수작업으로 처리하여 원하는 정보를 탐색해야하고 DQL(DamI Query Language)은 자체 추론 기능을 갖지 않기 때문에 DQL엔진에서 각각의 메타데이터들을 비교하여 사용자의 질의에 맞는 해답을 찾아내야 함으로 시스템 과부하 현상이 일어나게 된다. 본 논문은 이러한 기존의 시멘틱 서치 방식의 효율성과 속도를 높이기 위하여, 사용자 중심의 키워드 형태 질의를 시스템 중심의 DQL로 변환하는 DQL 컨버터 시스템과 추론엔진을 불러내어 전방향 추론과 단일화를 실행하는 DQL 엔진을 제안한다.

### 1. 서 론

현재 웹 문서의 대부분은 HTML로 작성되어 있기 때문에 이러한 문서들은 웹 브라우저를 통해서 사람이 보고 그 의미를 이해하기는 쉽지만 에이전트 같은 프로그램이 자동으로 문서의 의미를 파악하기 어려우며, 문서에 대한 정보를 검색 시 검색 결과의 정확도 저하와 정보 공유의 어려움 등이 존재한다. 시멘틱 웹은 이러한 문제점을 개선하고자 제시된 언어, 기반구조, 온톨로지 등 차세대 웹 관련 기술들을 총칭한다[1]. 시멘틱 서치의 궁극적인 목적은 시멘틱 웹 기술을 바탕으로 사용자가 입력한 질의들을 의미상으로 해석하고, 이를 통하여 사용자의 원하는 정보와 의미가 같은 해답을 검색하는 것이다. 시멘틱 서치의 성능은 사용자가 원하는 정보의 의미를 파악하는데 좌우되는데, 시멘틱 서치는 사용자가 원하는 정보의 의미를 파악하기 위한 방법으로 RDF 트리플 문을 기반으로 사용되는 DQL을 이용한다. 이는 웹 상의 문서들의 의미를 표현하기 위하여 DAML+OIL과 OWL을 이용하여 문서의 의미를 메타 데이터로 생성하고 이렇게 만들어진 메타데이터는 검색 엔진에서 검색을 용이하게 하기 위해 RDF 트리플 문으로 변환되어, 이를 기반으로 사용자가 입력하는 질의의 목적과 같은 의미를 지니는 정보를 검색해주기 때문이다.

기존의 시멘틱 서치기술은 사용자가 복잡한 질의들을 수작업으로 처리하여 원하는 정보를 탐색하기 때문에 키워드 검색에 비해서 일반 사용자가 이용하기에는 많은 어려움이 따랐다. 또 DQL(DamI Query Language)은 자체 추론 기능을 갖지 않기 때문에 DQL엔진에서 각각의 메타데이터를 매칭하여 사용자의 질의에 맞는 해답을 찾아내야 함으로 시스템 과부하 현상이 일어나 검색 속도가 저하된다. 본 논문에서는 사용자의 질의를 단순한 키워드로 입력받아서, 에이전트가 복잡한 DQL 질의를 자동으로 생성하여, 사용자가 원하는 정보를 탐색하는데 도움을 주고, DQL 엔진이 추론 엔진을 불러내어 전방향 추론을 실행하고 단일화함으로써 시스템 과부하를 최소한으로 줄임으로서 탐색속도를 높이기 위한 시스템을 구현하는 것을 목적으로 한다. 이에 기존의 시멘틱 서치 방식의 사용 효율성과 속도를 높이기 위하여 점진적 전방향 추론을 이용한 DQL 검색 인터페이스 시스템을 제안한다.

터페이스 시스템을 제안한다.

### 2. 관련 연구

#### 2.1 DQL search

United States/European Union ad hoc Agent Markup Language Committee에 의해 개발된 DQL은 시멘틱 웹을 위한 질의 언어로서 DAML+OIL/OWL로 표현된 지식을 기반으로 에이전트 간에 질의를 던져서 응답을 받기 위해 만들어진 언어다. 즉 응답을 도출해내기 위한 시멘틱 웹상의 지식 베이스를 기반으로 여러 질의들과 응답들 간의 의미적인 연관 관계(Semantic Relationship)를 정밀하게 분류한다. DQL이 데이터 베이스 시스템을 위한 표준 질의 언어(SQL)나 웹에서 정보를 검색하는 언어(XML query)와의 가장 큰 차이점은 JTP와 같은 자동 추론 엔진과 연결하여 시멘틱 웹상에 존재하는 다중 지식 베이스에서 질의에 대한 답을 도출할 수 있다는 것이다.

#### 2.2 RDF

RDF (Resource Description Framework)는 W3C의 가장 기본적인 시멘틱 웹 언어로서 시멘틱 웹의 구현을 위한 최하위 계층의 역할을 담당하며 웹에 있는 자원에 관한 메타정보를 표현하기 위한 언어다. 또한 프로그램이 처리할 수 있는 간단한 데이터 모델과 컴퓨터가 처리할 수 있는 표준화된 문법을 제공한다. RDF가 구조화된 메타 데이터를 생성해줌으로써 검색 엔진이 웹상에서 보다 정확히 정보를 검색할 수 있는 기능을 제공한다. 또 의미가 자원(Resource)과 그 특성 값으로 표현되므로 같은 내용에 대해서는 해석이 하나로만 정의되기 때문에 기존의 XML이 가지는 문제점을 해결하고 있다.

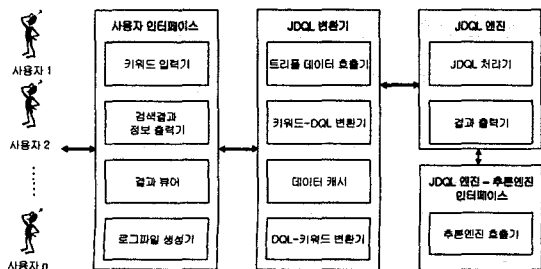
#### 2.3 JTP

JTP(Java Theorem Prover)는 스탠포드 대학에 소속되어 있는 지식시스템 연구소의 Gleb Frank에 의해 개발된 객체 지향 추론 시스템이다. JTP는 DAML+OIL로 만들어진 메타 데이터를 KIF(Knowledge Interchange Format) 형식으로 읽어 들여 전방향 추론을 할 수 있도록 개발되었다. 또 KIF형태로 저장된 DAML문서의 질의를 보내면 검색된 결과를 얻을 수 있는 인터페이스를 제공한다.

### 3. 전체적인 시스템 구조

본 논문에서는 기존 시멘틱 서치 방식의 효율성과 속도를 높이기 위하여 정진적 전방향 추론을 이용한 DQL 검색 인터페이스 시스템을 구축하기 위한 방안을 제시한다. DQL 검색 인터페이스 시스템의 핵심은 입력된 키워드를 통해 사용자의 의도를 이해하여 이러한 의도를 바탕으로 정확한 질의문을 생성하고, 추론 엔진의 전방향 추론기능을 을 적절히 활용하여 검색 속도를 높이는데 있다. 본 논문에서 제안하는 정진적 전방향 추론을 이용한 DQL 검색 인터페이스 시스템은 복잡한 DQL형식의 질의문을 사용자가 키워드 형태로 쉽게 입력하기위한 사용자 인터페이스, 사용자 중심의 키워드 형태의 질의를 시스템 중심의 DQL로 변환하는 JDQL 변환기, 전방향 추론을 실행하는 추론 엔진을 인터랙티브하게 호출하고 입력된 질의를 처리하여 정보를 검색한 후 출력해주는 JDQL 엔진으로 구성하였다. 첫째, 사용자 인터페이스는 시멘틱 서치 시스템을 이용하는 사용자의 의도를 키워드 입력기를 통해 키워드로 입력 받고, 검색된 결과를 사용자에게 보여준다. 이를 위해 온톨로지 구조와 의미가 일치하도록 입력 메뉴를 구성하고, 입력된 키워드는 서치 시스템에서 처리할 수 있는 데이터로 추상화 한다. 또 검색 과정을 로그 파일로 만들어서 각 사용자가 이전에 검색했던 정보를 정확히 기록하는 기능을 수행한다. 이렇게 생성한 로그파일을 바탕으로 사용자는 같은 검색을 반복하지 않는다. 둘째, JDQL 변환기는 사용자 인터페이스에서 생성한 추상화된 데이터로부터 정확한 DQL 질의문을 만드는 기능을 한다.

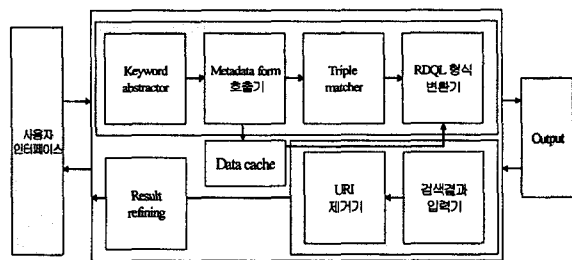
본 논문에서는 온톨로지 질의를 위해서 SQL과 유사한 문법을 가진 RDQL질의 언어를 사용한다. RDQL은 DAML API가 내장되어 DAML 온톨로지 작업이 가능한 Jena에서 제공된다. RDQL 변환기는 질의문 생성의 정확성을 극대화시키기 위하여 메타데이터의 트리플 데이터들을 호출하여 비교하고, 트리플 구문의 형식에 맞는 질의문을 생성하는 동시에 데이터 캐시에 호출된 트리플 데이터 형식을 저장하여 반복된 호출을 예방한다. 또 검색을 마치고 출력된 정보 역시 트리플 문이기 때문에 사용자가 인식할 수 있는 키워드 형태의 데이터로 변환한다. 셋째, JDQL 엔진은 JDQL 변환기가 만들어준 트리플 형태로 구성된 JDQL 질의문을 받아들인다. 질의문을 처리하기 위해 쓰이는 메타데이터는 그 양이 방대하기 때문에 먼저 트리플 형태의 질의 문장을 규칙이 적용된 추론 엔진에 보내어 전방향 추론을 실행한 후, 추론을 통해 정제된 데이터만을 바탕으로 검색을 실행하고 결과를 출력한다. 본 논문에서는 DAML+OIL로 만들어진 메타 데이터를 KIF(Knowledge Interchange Format) 형식으로 읽어 들여 전방향 추론을 할 수 있도록 개발된 JTP를 적극 활용하였다.



[그림 1] 전체적인 시스템 구조도

### 4. JDQL 변환기

사용자 측면을 고려하여 설계된 사용자 인터페이스에서 입력되어진 키워드 형태의 정보는 시멘틱 서치에서 사용하기에는 어려움이 따른다. 이에 본 논문에서는 사용자가 손쉽게 입력한 키워드 형태의 질의를 검색 시스템에서 사용 가능한 DQL형태로 적절히 변환하는 JDQL 컨버터를 연구한다. 본 논문의 서치 엔진에서 검색을 위해 사용하는 DQL은 온톨로지 질의를 위해서 SQL과 유사한 문법을 가진 RDQL 질의 언어를 사용했다. Jena는 DQL을 이용하여 검색을 하기위해 DAML+OIL문서를 트리플 형태로 변환하고, RDQL의 WHERE Clause는 이러한 트리플문과 같은 형태로 자신의 조건을 표현함으로써 검색을 용이하게 한다. 아래의 그림은 본 논문에서 제안하는 JDQL 컨버터의 구조를 나타낸 것이다. 사용자 인터페이스에서 입력된 키워드 데이터들은 JDQL 컨버터에서 처리할 수 있는 데이터 형태로 정제된 후 전송된다. 전송된 데이터는 RDQL의 문법의 맞게 순서(SELECT-WHERE-AND)를 정해주는 키워드 추상화 과정을 거친다. 앞서 말했듯이 RDQL은 트리플문과 같은 형태로 자신의 조건을 표현하는데, 이러한 트리플문은 <object>, <property>,<subject>로 구성되며,<object>,<property>는 URI 형식을 갖기 때문에 JDQL 컨버터가 변환한 질의문의 트리플 역시 URI형식으로 작성되어야 한다. 키워드로 입력된 데이터를 URI로 표현하기위해 메타데이터 형식이 담겨있는 파일을 호출하여 트리플 매칭기에 의해 각각의 데이터에 맞는 URI로 매치시킨다. 이렇게 입력된 데이터와 매칭된 URI는 RDQL형식 변환기에 의해 RDQL형식의 질의문으로 변환된다.



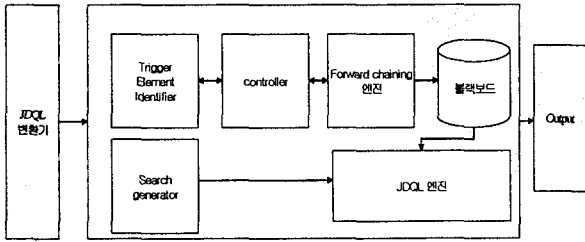
[그림 2] JDQL 컨버터

JDQL 컨버터는 생성된 질의문을 JDQL 엔진에 전송하고, 검색된 결과물 들을 받는데 검색된 결과물 역시 URI 형식으로 표현되기 때문에 사용자 인터페이스에서 사용할 수 있는 짧은 스트링 형태의 키워드로 변환하여, 변환된 데이터를 정제과정을 통해 정리한 후 사용자 인터페이스에 출력한다.

### 5. JDQL 엔진

JDQL 엔진은 JDQL 컨버터에서 생성한 RDQL 형태의 질의문에 대한 검색을 실행한다. JDQL 엔진은 질의문을 처리하기 위해 쓰이는 메타데이터는 그 양이 방대하기 때문에 규칙이 적용된 추론 엔진에 보내어 전방향 추론을 실행한 후, 방대한 양의 메타 데이터로부터 실제 검색엔진에서 사용하는 데이터만을 추출하는 단계와 추출된 메타데이터를 바탕으로 검색을 실행하여 결과를 찾는 단계로 구성된다. 시멘틱 서치를 위한 JDQL 엔진을 구축하기 위한 가장 중요한 단계인 전방향 추론 엔진 시스템은 입력된 메타 데이터를 여러 사실들로 구성하고 추론을 실행할 수 있는 규칙들이 삽입된 지식 베이스와 지식 베이스내의 규칙을 이용하여 사실들을 추론하여 단일화하는 추론 엔진으로

구성된다. 본 논문에서 제시한 귀납적 추론방식 중의 하나인 전방향 추론은 사용자가 추론하고자 하는 문제와 관련된 모든 사실들이 제공되면, 추론 엔진이 각각의 사실(fact)과 지식베이스(knowledge Base)에 있는 규칙(Rule)들의 전제조건을 매치하여 일치하는 규칙을 찾아 해당 규칙을 수행하고 그 규칙의 결론부분을 새로운 사실로서 추가한다. 본 논문에서의 작업메모리에는 규칙들의 조건문과 매치하기위해 트리플 형태의 매타데이터가 저장되어 있고, 규칙 베이스에는 전방향 추론을 하기 위한 IF-THEN 형태의 규칙들이 구성되어 있다.



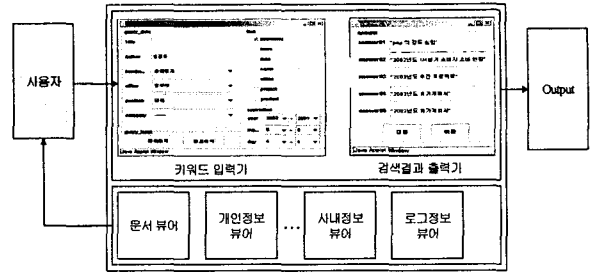
[그림 3] JDQL 엔진 구조도

위의 그림 3은 본 논문에서 제안하는 JDQL 엔진의 구조를 표현한 것이다. JDQL 변환기에서 생성되어 전달된 질의문의 전제 조건(WHERE문, AND문)들은 trigger element identifier에 의해 subject, property, object 구조를 지니는 각각의 트리플 문으로 분리된다. 분리된 트리플문은 전방향추론 엔진이 추론하기 위한 규칙들의 전제 조건과 비교하여, 올바르게 매치되는 규칙들이 존재하면 그 규칙을 실행하고, 실행된 결과를 블랙보드에 저장한다. 질의문들을 분석하여 전방향 추론을 실행하는 일련의 과정은 controller에서 제어한다. 이렇게 블랙보드에 저장되어 있는 추론 결과물들이 실제 검색 엔진에서 검색하기위한 유용한 데이터가 된다. JDQL 검색엔진은 블랙보드에 저장된 데이터를 바탕으로 시맨틱 서치를 실행하고 그 결과를 JDQL 변환기에 전송한다.

6. 사용자 인터페이스

사용자들 입장에서 검색에 쓰일 질의문을 보다 간편하게 사용하기를 원한다. 그래서 본 논문에서는 사용자 중심의 인터페이스를 설계하여 각 사용자가 간편하게 질의문을 입력할 수 있게 한다. 본 논문에서 사용자가 간단히 질의를 입력하는 방법으로서 텍스트 필드와 폼 박스를 이용한 키워드 입력을 사용하였다. 키워드 입력은 일반적으로 웹 페이지를 통해서 사용자가 컴퓨터에 매우 작은 숫자나 스트링 형태를 입력하는 방식이다. 이러한 인터페이스 시스템은 폼 박스와 같이 사용자가 사용하기 편하게 구성되었고, 단순한 키워드 입력만으로 통해서 사용자가 원하는 정보의 의미를 파악할 수 있는 근거를 획득하게 된다. 또한 시스템에서는 검색된 결과를 사용자가 이해하기 쉽게 보여주는 역할을 수행한다. 사용자 인터페이스를 통해서 얻어진 키워드들은 JDQL 컨버터가 쉽게 사용자가 원하는 질의문으로 변환위한 형식화된 기록정보를 구축하기 위해서 필요한 정보만을 골라서 저장하는 데이터 정제과정을 거친 후 정제된 데이터를 JDQL 컨버터에 넘겨지게 된다. 아래의 그림은 본 논문에서 제안하는 사용자 인터페이스의 구조를 설명한 것이다. 사용자는 자신이 원하는 의미의 정보를 검색하기 위해서 키워드 입력기에 질의문을 간단한 키워드를 입력하게 된다. 이렇게 입력된 키워드들은 JDQL컨버터에서 사용가능 한 데이터

로 정제된다. 한편 검색 엔진에서 검색된 결과 정보들은 검색 결과 출력기에 의해 사용자가 이해할 수 있는 형태로 출력되며, 사용자가 각각의 정보를 선택하면 여러 가지 정보의 세부적인 내용을 자세히 보여주는 정보 뷰어 창을 통해 정보의 상세한 부분들을 볼 수 있다.



[그림 4] 사용자 인터페이스 구조

7. 결과

본 논문에서 제안하는 전방향 추론을 이용한 DQL 검색 인터페이스 시스템은 기존의 시맨틱 서치 방식의 사용 효율성과 속도를 높이기 위하여 설계되었다. 즉, 사용자의 질의를 단순한 키워드로 입력받아서, 에이전트가 복잡한 DQL 질의를 자동으로 생성하여, 사용자가 원하는 정보를 탐색하는데 도움을 주고, DQL 엔진이 추론 엔진을 불러내어 전방향 추론을 실행하고 단일화함으로써 시스템 과부하를 최소한으로 줄임으로서 탐색 속도를 높이기 위한 시스템을 이용함으로써 사용자가 시맨틱 웹 서치를 보다 쉽고 효율적으로 실행할 수 있다. 시맨틱 웹 환경으로부터 본 논문에서 제안한 전방향 추론을 이용한 DQL 검색 인터페이스 시스템을 활용함으로써 HTML로 작성되어 있는 웹 문서들을 에이전트 같은 프로그램이 자동으로 문서의 의미를 파악하여, 문서에 대한 정보를 검색 시 검색 결과의 정확도를 향상시키고 정보 공유의 어려움을 해결하는데 도움이 된다.

참고 문헌

- [1] 최종민, "시맨틱 웹 개요와 연구동향", 정보과학회지, 2003년 3월
- [2] 이재호, "시맨틱 웹의 온톨로지 언어", 정보과학회지, 2003년 3월
- [3] Richard Fikes, Pat Hayes, Ian Horrocks, "DQL-A Query Language for the Semantic Web", WWW 2003, May 20-24, 2003, Budapest, Hungary
- [4] Gleb Frank, "JTP user manual", KSL, Stanford University, <http://www.stanford.edu/~gkfrank/jtp/#proof>
- [5] Larry Kerschberg, Wooju Kim, Anthony Scime, k "Intelligent Web Search via Personalizable Meta-Search Agents", E-Center for E-Business, George Mason University, Chonbuk National University, SUNY-Brockport
- [6] Andy Seaborne, "Jena Tutorial : A Programmer's Introduction to RDQL", April 2002.
- [7] John Davies, Richard Weeks, Uwe Krohn, "QuizRDF: Search Technology for the Semantic Web", BTexact Technologies, Orion 5/12, Adastral Park