

사용자에 따라 검색 결과의 순위를 적용하는 DQL 검색 시스템

김제민^o 박영택
송실대학교 컴퓨터학과

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

DQL Search System That Apply Ranking Of Search Result

JeMin Kim^o YongTack Park
Dept. of Computer Science, Soongsil University

요 약

현재 웹 서비스에서 사용하고 있는 키워드 기반 검색은 syntactic한 정보만을 제공한다. 웹 서비스의 이러한 단점을 보완하고자 시멘틱 웹 기술이 제안되고 있다. DQL 검색 시스템은 검색 결과의 질을 향상시키기 위해서 시멘틱 웹 기반의 검색 시스템 중의 하나다. 시멘틱 검색은 입력된 질의들을 통해서 사용자가 원하는 정보와 의미가 일치하는 결과를 얻어내는 것이다. 본 논문은 시스템을 이용하는 각 사용자에 따라 검색 되는 결과들에 우선순위를 적용하는 DQL 검색 시스템을 제안하고 다음 3가지 부분에 중점을 두어 시스템을 설계하였다. 첫째, 각 사용자들의 성향을 검색 결과에 반영하기 위해 개인화 프로파일과 온톨로지를 생성하는 개인화 온톨로지 생성기를 설계한다. 둘째, 검색 정확도를 높이기 위해 검색 대상이 되는 정보들의 속성과 관계를 추론해주는 검색 엔진을 설계한다. 마지막으로 DQL 검색 시스템의 사용 효율성을 높이기 위해 사용자가 입력한 키워드를 DQL로 자동 변환하는 DQL 컨버터를 설계한다.

1. 서 론

현재 웹 문서의 대부분은 HTML로 작성되어 있기 때문에 웹 브라우저를 통해서 사람이 문서의 의미를 이해하기는 쉽지만 에이전트 같은 프로그램이 자동으로 문서의 의미를 파악하기 어렵다[1]. 이 때문에 문서에 대한 정보를 검색한 후 결과의 정확도 저하와 정보 공유의 어려움 등이 존재한다. 웹 서비스의 이러한 단점을 보완 하고자 시멘틱 웹(semantic web)기술이 제안되고 있는데, 시멘틱 웹의 기반구조를 바탕으로 하는 시멘틱 검색은 사용자가 원하는 의미를 가지는 정보를 찾아줄 수 있다. 그러나 웹 페이지의 급격한 증가로 인하여 인터넷 사용자들은 필요한 정보를 빠른 시간 내에 찾기가 점점 어려워지고 있다. 이 때문에 사용자가 원하는 정보를 검색하는 것과 필요한 정보를 찾아내기 위해 검색 시간을 단축하는 것이 주요한 사안으로 떠오르고 있다. 사용자가 단시간 내에 유용한 정보를 획득하기 위해서는 개인적인 정보 선호도를 기록한 프로파일을 이용하여 개인화 온톨로지를 만들 필요가 있다.

이에 본 논문에서는 DQL(DamL Query Language)을 이용한 시멘틱 검색(Semantic search)기법과 프로파일 기반의 개인화 온톨로지 자동 생성기법을 이용하여 사용자에 따라 검색 되는 결과들에 우선순위를 적용하는 DQL 검색 시스템을 제안한다. 본 논문에서 제안하는 시스템은 사용자 프로파일을 구축하는 온톨로지 기반의 사용자 프로파일 에이전트, 사용자가 입력한 키워드 질의를 DQL로 변환하는 DQL 컨버터, 입력된 질의를 처리하여 정보를 검색한 후 출력해주는 DQL 엔진, DQL 엔진이 전방향 추론을 실행하는 추론엔진을 호출하기 위한 DQL-추론엔진 인터페이스 부분으로 구성된다.

2. 기초 연구

시멘틱 웹이 제안된 이후, 여러 가지 관련 연구가 활발히 진행되고 있다. 본 논문에서 제안하고 있는 DQL 검색 시스템은 개인화 웹 서비스와 시멘틱 웹에서 기본이 되는 두 가지 컨셉을 바탕으로 연구되었다.

2.1 개인화 (Personalization)

검색 서비스를 이용하는 사용자들의 측면에서 개인화는 현재의 정보 과잉과 관련이 있다. 현재 사용자가 검색 사이트를 이용하여 정보를 검색하면 사용자가 원하는 것과 관련 없는 정보들이 쏟아진다. 이러한 단점을 보완하기 위해 여러 검색 사이트들은 사용자들에게 개인화된 정보 검색 서비스를 시행하고 있다. Yahoo의 My Yahoo와 같은 기능은 기본 화면을 사용자의 취향에 맞게 편집할 수 있는 기능과 사용자의 스타일에 맞는 정보를 선별해준다. 웹 검색에서의 개인화는 가치 교환과 정이다. 사용자가 자신이 선호하는 검색 정보의 결과를 검색 사이트에 제공하면 검색 사이트는 사용자가 제공한 자료를 기초로 사용자에게 가장 알맞은 정보를 제공한다.[2]

2.2 온톨로지 (Ontology)

온톨로지는 개념(Concept)과 관계(Relation)들로 구성된 사전으로서 특정 도메인에 관련된 객체들을 계층적 구조로 표현하고 추가적으로 이를 확장할 수 있는 추론 규칙을 포함한다. 온톨로지의 역할은 서로 다른 데이터베이스가 같은 개념에 대해서 서로 다른 단어나 식별자를 사용할 경우에 이를 해결해주는 데 있다. 예를 들어, 주소를 포함하는 두 데이터베이스에서 postal code와 zip code는 같은 것을 의미한다. 이 두 데이터베이스의 정보를 비교하거나 통합하려는 프로그램이 있다면 온톨로지의 추론을 통해서 이루어진다. 이러한 역할로 인하여 온톨로지는 웹 기반의 지식을 처리하거나 응용 프로그램 사이의 지식공유, 재사용등을 가능하게 하는 아주 중요한 요소로 자리 잡고 있다.[3]

2.3 DQL (DamL Query Language)

DQL(DamL Query Language)은 시멘틱 웹을 위한 질의 언어로서 DAML+OIL/OWL로 표현된 지식을 기반으로 에이전트 간에 질의를 던져서 응답을 받기 위해 만들어진 언어다[4]. 즉 응답을 도출해내기 위해 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 JTP[5]와 같은 자동 추론 엔진과 연결하여 시멘틱 웹상에 존재하는 다중 지식 베이스에서 질의에 대한 답을 도출할 수 있다는 것이다.

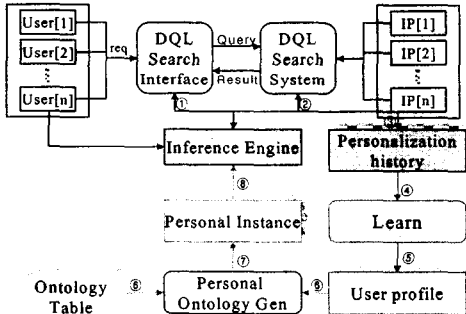
3. 시스템 구조

사용자에 따라 검색 결과의 순위를 적용하는 DQL 시스템의 핵심은 사용자가 어떠한 정보를 검색하는지 정확하게 모니터링하여, 관심을 가지는 정보들을 바탕으로 만들어진 사용자 프로파일의 정확도와 사용자가 입력한 키워드를 통해 사용자의 검색하고자하는 정보의 의미를 이해하여 정확한 DQL 질의문을 생성하며, 추론 엔진의 추론기능을 적절히 활용하여 정보들의 속성과 관계들을 추론하여 검색 정확도를 높이는데 있다. 본 논문은 시멘틱 웹상에서 정확하게 학습된 사용자 프로파일을 기반으로 각 사용자가 원하는 검색 정보를 제공할 수 있도록 다음과 같은 DQL 시스템 구조를 제안한다.

3.1 사용자 프로파일과 온톨로지 생성

사용자 온톨로지는 사용자 히스토리 파일 생성, 사용자의 검색 행위에 가중치가 적용된 사용자 프로파일 생성, 온톨로지 테이블을 기반으로 개인화 온톨로지 구축의 세 단계를 거쳐 생성된다. 다음 그림 1은 사용자 검색 인터페이스로부터 입력된 검색 행위에 대한 데이터를 기반으로 사용자의 개인 온톨로지 생성되는 과정을 보여주고 있다.

각 사용자의 개인화 온톨로지를 생성하기 위해서 가장 먼저 수행되어야 하는 것은 사용자의 검색 행위를 포괄적으로 탐지하는 것이다. 본 논문에서는 사용자 히스토리 파일 구축기를 사용하여 각 사용자가 검색하고자 하는 정보와 검색된 정보, 검색된 정보 중에서 사용자가 선택한 정보를 로그 파일로 생성한다.



[그림 1] 사용자 개인 온톨로지 생성 과정

사용자 히스토리 파일은 검색 정보, 출력 정보, 선택 정보로 구성된다. 검색 정보(search information)는 사용자가 검색하고자하는 검색 아이템을 나타내고, 출력 정보(output information)는 사용자가 검색하고자하는 검색 아이템에 관한 검색된 결과를 나타내며, 선택 정보(choice information)는 출력 정보 중에서 사용자가 일차적으로 선택한 정보를 나타낸다.

생성된 사용자 히스토리 정보는 사용자 개인 프로파일을 구축하기 위해 사용된다. 본 논문에서는 보다 정확한 개인 프로파일을 구축하기 위해 출력 정보의 가중치를 고려하여 사용자 관심을 측정한다. 각 사용자의 정보 선호도의 척도가 되는 가중치를 평가하는 평가 함수는 다음과 같다.

$$\begin{aligned}
 &Total\ Information\ Weight = Hit\ Ratio - Relative\ Hit\ Ratio \\
 &* Hit\ Ratio = Hit\ Number(HN) / Output\ Number(ON) \\
 &* Relative\ Hit\ Ratio = \sum Relation\ Weight, \\
 &* Relation\ Weight = \sum Co\ Relation\ Weight, \\
 &* Co\ Relation\ Weight = Non\ Hit\ Number(NHN) / \\
 &Output\ Number(ON) \times Relative\ Information\ Total\ Weight
 \end{aligned}$$

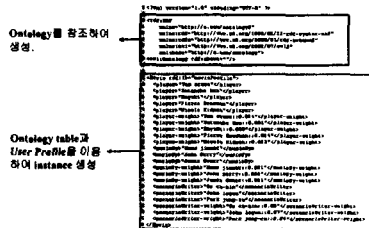
평가 함수에 의해 평가된 각 정보의 가중치는 각 사용자의 프로파일에 적용되어, 각 사용자에 대한 검색 결과의 선호도를 나타낸다. 사용자 프로파일은 정보 카테고리, 검색 대상, 검색 정보와 가중치로 구성 된다.

검색 정보들의 category	사용자가 검색한 정보의 Weight
Information Category - News	
Information Category - Music	
Information Category - Education	
Information Category - Entertainment	
Information Category - University	

[그림 2] 사용자 프로파일

개인 온톨로지 구축기는 각 사용자의 정보 선호도(weight)가 기록되어 있는 사용자 프로파일을 바탕으로 OWL 기반의 개인화 인스턴스 파일을 동적으로 생성한다. 본 논문에서는 정확한 개인화 인스턴스 파일을 구축하기 위해서 온톨로지의 각 클래스와 클래스의 속성이 정확하게 정리 되어있는 온톨로지 테이블을 제안한다. 온톨로지 테이블을 구축하는 방법으로 온톨로지 디자이너가 온톨로지를 구축한 후 온톨로지 테이블을 수작업으로 작성하는 방법, 온톨로지 and 온톨로지 테이블을 동시에 작성 할 수 있는 도구를 사용하는 방법, ontoEdit나 protege와 같은 도구로 작성된 온톨로지를 입력 받아서 온톨로지 테이블을 자동 생성해주는 plug-in을 구현하여 사용하는 방법 등을 생각해 볼 수 있다.

그림 3은 개인 인스턴스 구축기에 의해서 생성된 개인 인스턴스 파일을 보여준다. 생성된 개인 인스턴스 파일은 해당 사용자가 검색을 실행할 때마다 트리플 컨버터에 의해 트리플로 변환되어 검색된 결과에 순위를 적용하는데 구간이 된다.



[그림 3] 개인 인스턴스 파일

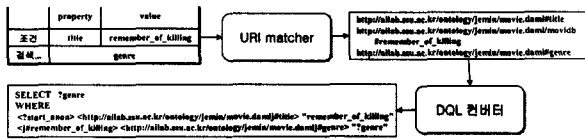
3.2 JDQL 컨버터

본 논문에서 제안한 검색 결과의 순위를 적용하는 DQL 검색 시스템은 사용자의 질의를 기존의 키워드 형태로만 입력받아 사용자의 의도를 파악하도록 처리 한다. 이렇게 사용자 측연을 고려하여 설계된 사용자 인터페이스에서 입력되어진 키워드 형태의 정보는 일반적으로 기존의 DQL을 기반으로 한 시멘틱 검색에서 사용하기에는 어려움이 따른다. 이에 본 논문에서는 사용자가 손쉽게 입력한 키워드 형태의 질의어를 DQL형태로 적절히 변환하여, 시멘틱 검색 시스템에서 인식하여 검색을 실행할 수 있도록 JDQL 컨버터를 연구한다.

본 논문의 JDQL 엔진에서 사용하는 DQL은 트리플(RDF-Triple) 형식의 메타데이터에 SQL과 유사한 문법의 질의문을 사용하여 정보를 검색하는 RDQL이다. RDQL은 SELECT문, FROM, WHERE, AND, USING문으로 구성된다.

아래의 그림 4는 JDQL 컨버터의 구조를 나타낸 것이다. 사용자 검색 인터페이스에서 입력된 키워드형태의 데이터들은 조건 벡터와 검색 벡터에 저장되어 전송된다. 전송된 데이터는

RDQL의 문법의 맞게 순서(SELECT-WHERE- AND)를 정해주는 키워드 추상화 과정을 거친다. 즉 검색 백터에 저장된 데이터들은 사용자가 검색을 원하는 정보로서 SELECT문에서 변수로 변환되고, 조건 백터에 저장된 데이터들은 사용자가 정보를 검색하기 위한 조건으로써 WHERE문에 트리플 구문의 형태로 변환된다. RDQL은 트리플과 같은 형태로 자신의 조건을 표현하기 때문에, WHERE문의 <subject>, <property> 또한 URI형식으로 작성되어야 한다. 키워드로 입력된 데이터를 URI로 표현하기 위해 메타데이터 형식을 호출하여 트리플 매칭기에 의해 각각의 데이터에 맞는 URI로 매치시킨다. 이러한 방법을 통해서 사용자 검색 인터페이스에서 입력된 키워드 데이터와 각 키워드에 매칭된 URI는 RDQL형식 변환기에 의해 RDQL형식의 질의문으로 변환된다.



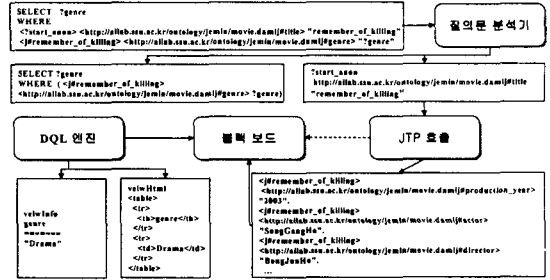
[그림 4] JDQL 컨버터 구조 및 작업 흐름

3.3 JDQL 엔진

JDQL 엔진은 JDQL 컨버터에서 생성한 RDQL 형태의 질의문에 대한 검색을 실행한다. 본 논문에서 제안한 검색 결과의 순위를 적용하는 DQL 검색 시스템은 기존 시멘틱 웹상에서 사용하는 검색 엔진보다 검색 효율성을 높이기 위해 2단계 구조를 가지는 JDQL 엔진을 연구한다. 먼저 정보를 검색하기 위해 사용되는 트리플 형태의 메타데이터는 그 양이 매우 방대하기 때문에 질의문의 조건부분에 해당하는 트리플 구문을 규칙이 적용된 추론 엔진에 전송하여 전방향 추론을 실행한 후, 매칭한 양의 메타 데이터로부터 실제 검색엔진에서 사용하는 데 필요한 데이터만을 추출하는 단계와 추출된 메타데이터를 바탕으로 검색을 실행하는 단계로 구성된다.

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 논문에서는 DAML+OIL로 만들어진 메타데이터를 KIF (Knowledge Interchange Format) 형식으로 읽어서 전방향 추론을 하도록 개발된 JTP를 추론 엔진으로 활용하였다.

그림 5는 본 논문에서 제안하는 JDQL 엔진의 구조를 표현한 것이다. JDQL 변환기에서 생성되어 전달된 질의문의 전제 조건(WHERE문, AND문)들은 질의문 분석기에 의해 각각의 트리플 구문으로 분리된다. 분리된 트리플 구문은 전방향 추론 엔진이 추론하기 위한 규칙들의 전제 조건과 비교하여, 올바른 매치되는 규칙들이 존재하면 그 규칙을 실행한다. 규칙이 실행되면 그 결과에 해당하는 모든 사실(트리플 데이터)들이 블랙보드에 저장한다. 질의문들의 조건부분을 분석하여 전방향 추론을 실행하는 일련의 과정은 controller에서 제어하는데, 이 과정은 재귀적으로 이루어진다. 즉 조건 부분이 여러 개 존재할 경우 처음 조건부분을 이용해서 전방향 추론을 실행한 결과 데이터를 바탕으로 그 다음 전제조건을 이용해서 전방향 추론을 실행한다. 이러한 과정을 거쳐서 블랙보드에 저장된 데이터들이 실제 검색 엔진에서 검색하기 위한 유용한 트리플 데이터가 되며, JDQL 검색엔진은 블랙보드에 저장된 데이터를 바탕으로 시멘틱 검색 실행하고, 실행된 결과에 순위를 적용하여 사용자 검색 인터페이스에 전송한다.



[그림 5] JDQL 엔진의 구조 및 작업 흐름

4. 결론

기존의 시멘틱 검색은 사용자가 복잡한 질의들을 수작업으로 처리하여 원하는 정보를 검색해야 함은 물론이고 사용자 개인의 마음에 드는 정보 역시 직접 선택해야 한다. 본 논문은 이러한 기존의 시멘틱 검색 방식의 효율성을 높이고 검색된 결과에 대해 사용자 개인의 정보 선호도를 적용하기 위하여 다음 세 가지 시스템을 제안하였다. 첫째 각 사용자들의 성향을 검색 결과에 반영하기 위한 사용자 프로파일과 온톨로지 생성 시스템, 둘째 사용자 중심의 키워드 형태 질의를 시스템 중심의 DQL로 변환하는 JDQL 컨버터 시스템, 셋째 인터랙티브하게 추론엔진을 불러내어 전방향 추론과 단일화를 실행하는 JDQL 엔진을 제안하였다. 사용자에게 따라 검색 결과의 순위를 적용하는 DQL 검색 시스템은 기존의 시멘틱 검색 방식의 사용 효율성을 높이고 사용자 개인에 맞는 정보를 제공하기 위해서 설계되었다. 즉, 사용자의 질의를 단순한 키워드로 입력받아서, 에이전트가 DQL 질의를 자동으로 생성하여, 사용자가 원하는 정보를 탐색하는데 도움을 주고, DQL 엔진이 추론 엔진을 사용하여 시스템 과부하를 최소한으로 줄이며, 사용자 프로파일과 인스턴스 데이터를 이용하여 사용자가 선호하는 정보들을 선별해냄으로써 사용자가 시멘틱 검색을 보다 쉽고 효율적으로 실행할 수 있다. 본 논문에서 제안한 사용자에 따라 검색 결과의 순위를 적용하는 DQL 검색 시스템을 활용함으로써 HTML로 작성되어 있는 웹 문서들을 소프트웨어 에이전트가 의미를 파악하여, 문서에 대한 정보 검색 결과의 정확도를 향상시키고 정보 선택의 어려움을 해결하는데 도움이 된다.

5. 참고 문헌

- [1] 최중민, "시멘틱 웹 개요와 연구동향", 정보과학회지, 2003년 3월
- [2] <http://www.personalization.co.kr>
- [3] 이재호, "시멘틱 웹의 온톨로지 언어", 정보과학회지, 2003년 3월
- [4] Richard Fikes, Pat Hayes, Ian Horrocks, "DQL-A Query Language for the Semantic Web", WWW 2003, May 20-24, 2003, Budapest, Hungary
- [5] Gleb Frank, "JTP user manual", KSL, Stanford University, <http://www.stanford.edu/~gkfrank/jtp/#proof>
- [6] Chen, L. and Sycara, K. A Personal Agent for Browsing and Searching. In Proc. 2nd Intl. Conf. on Autonomous Agents, pp. 132-139, New York, USA, 1998
- [7] Heflin, J., Hendler, J. and Luke, S. SHOE: A Knowledge Representation Language for internet Applications. Technical Paper, Institute for Advanced Computer Studies, University of Maryland, College Park.
- [8] Luke, S, Spector, L., Rager, D. and Hendler, J. Ontology-Based Web Agents. In Proc. First International Conference on Autonomous Agents(AA'97), 1997.
- [9] RDF Primer, W3C Working Draft 23 January 2003, <http://www.w3.org/TR/rdf-primer>