

시맨틱 웹 기반의 분산 시스템을 위한 인덱스 및 질의 변환 기법

채광주^o 김연희 임해철
홍익대학교 컴퓨터공학과
ckj4300@gmail.com

Index and Query Rewriting Strategy for Distributed based on Semantic Web

Kwangju Chae^o Younhee Kim HaeChull Lim
Dept. of Computer Engineering, Hongik Univ.

차세대 웹으로 인식되고 있는 시맨틱 웹 환경에서는 지능적인 소프트웨어 에이전트가 웹 자원간의 의미를 파악하여 자동으로 원하는 결과를 찾을 수 있다. 이를 위해, 웹 자원이 가지고 있는 의미를 컴퓨터가 처리할 수 있는 형태로 기술한 것을 온톨로지라고 한다. W3C에서는 온톨로지를 기술하기 위한 언어로서 RDF, RDFS, OWL등을 제안하였다. 이 중 OWL은 가장 최근에 시맨틱 웹 표준 언어로 권고되었고 RDF/S보다 다양한 표현들을 제공한다. 따라서 앞으로 시맨틱 웹을 보다 발전시키기 위해서는 웹 자원의 의미를 다양하게 기술할 수 있는 OWL에 기반한 연구가 필요하다.

한편, 웹 자원의 수는 계속 증가하고 있고 지역적으로도 점차 분산되고 있는 웹 환경을 고려할 때, 분산 시맨틱 웹에서의 데이터 관리에 대한 연구가 필요하다. 분산 환경에서는 데이터가 다양한 지역의 저장소에 존재하기 때문에, 각각의 저장소에 존재하는 데이터를 논리적으로 통합하여 사용자의 질의를 처리해줄 수 있는 미들웨어가 필요하다. 이 미들웨어를 통해 질의를 처리함으로써, 사용자는 물리적으로 분산되어 있지만 논리적으로는 마치 하나의 시스템에서 제공하는 것과 같은 서비스를 제공받을 수 있다. 이러한 시스템의 성능을 결정하는 요인 중에 중요한 것이 질의에 적합한 데이터를 포함하고 있는 지역 저장소를 빠르게 찾는 것이라 할 수 있다. 따라서 질의에 적합한 데이터를 포함하고 있는 지역 저장소를 빠르게 검색하기 위한 기법에 대한 연구가 필요하다.

또, OWL은 RDF/S와 달리 분산되어 있는 웹 자원간의 의미적 관계를 기술할 수 있는 추가적인 표현이 존재하기 때문에, 분산 환경에 적합한 언어라 할 수 있다. 따라서 분산 환경을 고려한 OWL의 연구는 시맨틱 웹을 발전시키고 현실의 웹 시스템에 적용시키기 위해 중요한 연구이다.

본 논문에서는 첫째, 분산된 시맨틱 웹 환경에서 질의에 적합한 데이터를 포함하고 있는 지역 저장소를 빠르게 찾기 위한 인덱스 기법을 제안한다. 둘째, 분산 환경을 지원할 수 있는 표현을 제공하는 OWL의 다양한 표현을 활용하여 사용자에게 보다 정확한 결과를 반환해주기 위해 사용자가 제시한 초기 질의를 변환하기 위한 기법을 제안한다.

OWL의 표현은 계층관계, 속성의 특징, 속성의 제약 사항, 온톨로지 매핑, 복잡한 클래스 등으로 분류할 수 있고, 본 논문에서 제안한 OWL 온톨로지를 위한 저장 구조는 각 분류마다 테이블을 생성함으로써 모든 표현에 대한 정보를 저장할 수 있으며, 각 표현에 대한 정보가 어느 테이블에 있는지를 알 수 있다. 하지만 위의 저장 구조만으로는 분산 환경에서 질의에 적합한 데이터를 포함하고 있는 지역 저장소를 찾을 수 없다. 따라서 본 논문에서는 첫 번째로 질의에 적합한 데이터를 포함하고 있는 지역 저장소를 빠르게 검색하기 위해 지역 저장소에 대한 정보를 저장하고 있는 중재 인덱스를 제안한다.

중재 인덱스는 클래스 인덱스와 속성 인덱스로 구분된다. 속성 인덱스는 속성이 질의 조건으로 명시된 질의에 대해 적합한 지역 저장소가 어느 곳인지를 알려준다. 하지만 속성 인덱스만으로는 클래스만 질의 조건으로 주어진 질의는 처리할 수 없다. 이 경우, 클래스 인덱스는 클래스만 질의 조건으로 주어진 질의에 대해 적합한 지역 저장소가 어느 곳인지를 알려줄 수 있다. 중재 인덱스의 노드는 클래스나 속성의 이름, 데이터가 있는 지역 저장소의 식별자와 함께 해당하는 OWL의 표현에 대한 정보를 포함한다. 중재 인덱스의 노드가 클래스 이름을 포함하면 클래스 인덱스로 분류되며, 속성 이름을 포함하면 속성 인덱스로 분류된다.

하지만 리소스만 질의 조건으로 주어진 질의는 중재 인덱스를 통해 적합한 지역 저장소를 검색할 수 없다. 예를 들어 사용자는 한 리소스에 대한 모든 정보를 원할 수 있지만, 저장 구조나 중재 인덱스에는 리소스가 어느 클래스

의 인스턴스인지에 대한 정보가 저장되어 있지 않기 때문에, 적합한 지역 저장소를 검색할 수 없다. 따라서 모든 경우의 질의 유형에 대해 적합한 지역 저장소를 검색하기 위해서는 리소스와 클래스간의 관계를 알 수 있는 리소스-클래스 인덱스가 추가적으로 요구된다. 리소스-클래스 인덱스의 노드는 리소스 이름과 클래스 이름을 포함한다. 따라서 리소스만 질의 조건으로 주어진 질의를 처리할 때, 리소스-클래스 인덱스를 통해 리소스가 어느 클래스의 인스턴스인지 검색할 수 있고, 구해진 클래스 이름으로 클래스 인덱스를 검색하면 질의에 적합한 지역 저장소를 파악할 수 있다. <그림 1>은 중재 인덱스의 노드 구조를 보여주며, <그림 2>는 리소스-클래스 인덱스의 노드 구조를 보여준다.

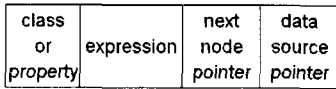


그림 1 중재 인덱스의 노드 구조

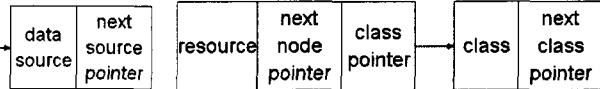


그림 2 리소스-클래스 인덱스의 노드 구조

두 번째, OWL의 다양한 표현을 활용하고 사용자에게 보다 정확한 검색 결과를 반환하기 위해서는 질의 변환 알고리즘이 필요하다. 트리플 구조로 표현된 질의는 OWL의 표현에 대한 정보를 포함하고 있지 않기 때문이다. 질의 변환을 수행하기 위해서는 OWL의 표현을 클래스 관련 표현과 속성 관련 표현으로 구분할 필요가 있다. 이것은 중재 인덱스가 클래스 인덱스와 속성 인덱스로 구성되어 있고 중재 인덱스의 각 노드는 OWL의 표현 정보를 포함하고 있기 때문이다. 즉, 클래스 인덱스의 노드는 클래스 관련 표현을 포함하고, 속성 인덱스의 노드는 속성 관련 표현을 포함하고 있다. 따라서 위와 같이 표현을 분류하면, 질의 변환을 수행할 때, 중재 인덱스의 노드에 대한 접근을 통해 관련된 OWL의 표현을 알 수 있다. 이로 인해, 질의 변환을 수행할 때, 중재 인덱스를 활용함으로써 빠르게 질의 변환을 수행할 수 있다. <그림 3>은 "subClassOf" 표현에 대한 질의 변환 알고리즘을 보여준다.

```

// q is (s,p,o) triple query
1  s = S(q); p = P(q); o = O(q);
2  // return expression of class
3  if(s) // s is resource
4      expr = getExpr(type(s)); type = type(s);
5  else if(!s && p == [type] && o) // s is variable
6      expr = getExpr(o); type = o;
7  switch(expr){
8      case(subClass):
9          // get superclass from Hierarchy table
10         c = GetSuperClass(type);
11         q' = generateQ(v, "type", c); // generate query
12         addQ(q', GQ(v, p, o), &); i++;
13         addQ(q, q', ||); // extend query
14         break;
15 }
    
```

그림 3 "subClassOf" 표현에 대한 질의 변환 알고리즘

지금까지 설명한 것처럼, 본 논문에서는 첫째, 분산된 온톨로지 환경에서 질의에 적합한 데이터가 위치하고 있는 지역 저장소를 빠르게 검색하기 위한 인덱스 구조를 제안하였고, 둘째, OWL의 다양한 표현을 이용하여 질의를 변환하기 위한 기법을 제안하였다. 본 논문에서 제안한 기법을 통해 OWL의 다양한 표현을 이용하여 사용자의 질의에 보다 정확한 검색 결과를 반환하는 것이 가능하였고, 모든 질의 유형에 대해 적합한 데이터가 존재하는 지역 저장소를 파악할 수 있었다.