

시맨틱 웹 기반의 분산 시스템을 위한 질의 변환 및 인덱싱 기법

(Query Rewriting and Indexing Schemes for Distributed Systems based on the Semantic Web)

채 광 주[†] 김 연 희^{**}
(Kwang Ju Chae) (Youn Hee Kim)

임 해 철^{***}
(Hae Chull Lim)

요약 시맨틱 웹의 기반이 되는 온톨로지는 기술 언어로 OWL이 발표되면서, 웹 리소스의 의미를 더욱 다양하게 기술할 수 있는 강한 표현력을 갖추어 가고 있다. 그리고, 시맨틱 웹의 개념이 널리 인식되면서 정보의 양이 더욱 많아지고 온톨로지가 지역적으로 분산되어 구축됨에 따라, 분산 환경에서 원하는 데이터를 포함하고 있는 지역 저장소를 빠르게 검색하는 것이 전체 시스템의 성능에 중요한 영향을 미치게 되었다. 따라서 본 논문에서는 첫째, 분산된 온톨로지 환경에서 사용자가 원하는 데이터가 위치하고 있는 지역 저장소를 빠르게 검색하기 위한 인덱싱 구조를 제안한다. 둘째, 분산 환경을 지원할 수 있는 OWL의 다양한 표현을 이용하여 질의를 확장하기 위한 질의 변환 기법을 제안한다. 본 논문에서 제안한 기법을 통해 OWL의 다양한 표현을 활용하는 것이 가능하고, 시맨틱 웹 환경의 모든 질의 유형에 대해 데이터가 존재하는 지역 저장소를 빠르게 파악할 수 있다.

· 이 논문은 제34회 추계학술대회에서 '시맨틱 웹 기반의 분산 시스템을 위한 질의 변환 및 인덱싱 기법'의 제목으로 발표된 논문을 확장한 것이다

[†] 학생회원 : 홍익대학교 컴퓨터공학과
ckj4300@gmail.com

^{**} 정 회 원 : 부천대학 e-비즈니스과 전임 강사
ellyblue06@gmail.com

(Corresponding author임)
^{***} 종신회원 : 홍익대학교 컴퓨터공학과 교수
lim@cs.hongik.ac.kr

논문접수 : 2007년 12월 14일
심사완료 : 2008년 8월 5일

Copyright©2008 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 컴퓨팅의 실제 및 레터 제14권 제7호(2008.10)

키워드 : 시맨틱 웹, 온톨로지, 분산 시스템, 인덱싱, 질의 변환 전략

Abstract Ontology plays an important role of the Semantic Web to describe meaning and reasoning of resources. Ontology has more rich expressive power through OWL that is a next standard representation language recommended by W3C. As the Semantic Web is widely known, an amount of information resources on the Web is growing rapidly and the related information resources are placed in distributed systems on the Web. So, for providing seamless services without the awareness of far distance, efficient management of the distributed information resources is required. Especially, searching fast for local repositories that include data related to user's queries is important to the performance of systems in the distributed environment. In this paper, first, we propose an index structure to search local repositories related to queries in the distributed Semantic Web. Second, we propose a query rewriting strategy to extend given user's query using various expression of OWL. Through the proposed index and query strategy, we can utilize various expressions of OWL and find local repositories related to all query patterns on the Semantic Web.

Key words : Semantic web, Ontology, Distributed System, Index, Query rewriting strategy

1. 서론

기존의 웹 환경에서는 데이터를 검색하기 위해 사용자가 수동적으로 하이퍼링크를 따라가며 원하는 결과를 찾을 때까지 문서를 검색해야 하는 어려움이 있었다. 하지만 차세대 웹으로 인식되고 있는 시맨틱 웹 환경에서는 지능적인 소프트웨어 에이전트가 웹 리소스간의 의미를 파악하여 자동으로 원하는 결과를 찾을 수 있다. 이를 위해, 웹 리소스가 가지고 있는 의미를 컴퓨터가 처리할 수 있는 형태로 기술한 것을 온톨로지라고 하며 온톨로지를 기술하기 위한 언어 중 OWL은 가장 최근에 시맨틱 웹 표준 언어로 권고되었고 기존의 언어보다 다양한 표현력을 제공한다[1]. 따라서 앞으로 시맨틱 웹을 보다 발전시키기 위해서는 웹 리소스의 의미를 다양하게 기술할 수 있는 OWL에 기반한 연구가 필요하다.

한편, 웹 리소스의 수는 기하급수적으로 증가하고 있고 지역적으로 점차 분산되고 있는 웹의 특성을 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. 이러한 분산 시스템의 성

능을 결정하는 요인 중에 중요한 것이 질의에 적합한 데이터를 포함하고 있는 지역 저장소를 빠르게 검색하는 것이라 할 수 있다. 특히 지역 저장소의 수가 많을수록 그 중요성은 더욱 커지게 된다. 따라서 질의 처리에 적합한 데이터를 포함하고 있는 지역 저장소를 빠르게 검색하기 위한 기법에 대한 연구가 필요하다.

특히, OWL은 분산되어 있는 웹 리소스간의 의미적 관계를 기술할 수 있는 추가적인 표현이 존재하기 때문에 분산 환경에 적합한 언어라 할 수 있다. 그러므로 분산 환경을 고려한 OWL의 연구는 시맨틱 웹을 발전시키고 차세대 웹 시스템을 개발하기 위해 반드시 필요한 연구이다.

본 논문에서는 첫째, OWL의 다양한 기능을 활용하여 사용자에게 보다 정확한 검색 결과를 반환해줄 수 있도록 사용자가 제시한 초기 질의를 변환하기 위한 기법을 제안한다. 둘째, 분산된 시맨틱 웹 환경에서 질의에 적합한 데이터를 포함하고 있는 지역 저장소를 빠르게 검색하기 위한 인덱싱 기법을 제안한다.

2. 관련연구

시맨틱 웹 환경에서는 <리소스-속성-값>의 트리플 구조로 리소스가 가지는 의미를 기술하기 때문에 시맨틱 웹 데이터의 질의 처리는 이러한 트리플 구조를 이용하는 것이 일반적이다. 한편, 분산 환경에서 데이터를 빠르게 검색하게 위한 기존의 연구들은 대부분 질의에 적합한 데이터를 어느 지역 저장소가 가지고 있는지를 탐색하는 과정에 초점이 맞추어져 있다. 특히, 분산 시맨틱 웹 환경에서 효율적인 질의 처리를 지원하기 위한 인덱스 구조에 관한 연구가 현재까지 많이 진행되어 왔다[2-5]. 본 절에서는 분산 시맨틱 웹 환경에서 질의 처리에 적합한 데이터를 포함하고 있는 지역 저장소를 빠르게 검색하기 위한 인덱스에 대한 기존의 연구 결과를 살펴본다.

속성 인덱스는 트리플 구조에 기반한 질의 유형 중에서 질의 조건으로 속성이 주어진 경우의 효율적인 처리를 목적으로 한다[2]. 속성 인덱스는 각 속성에 대해 지역 저장소가 해당 속성을 몇 개 포함하고 있는 지에 대한 정보를 유지하여 질의 조건으로 속성이 주어진 경우 관련된 정보를 저장하고 있는 지역 저장소를 빠르게 검색하고자 한다. 그러나 속성 인덱스는 각 지역 저장소마다 독립적으로 존재하기 때문에 질의 처리 시 모든 지역의 속성 인덱스를 확인해야 하는 문제점이 존재한다. 따라서 본 논문에서는 미들웨어 구조를 이용해 질의 처리에 필요한 지역 저장소에만 선택적으로 질의를 전송하여 보다 효율적인 처리가 가능한 인덱스 구조를 제안한다.

경로 기반 인덱스는 분산 환경에 보다 적합하도록 속성 인덱스를 확장한 것으로 각 지역 저장소와는 별도로 인덱스를 유지함으로써 각 속성 정보와 관련된 지역 저장소만

을 선별할 수 있다[3]. 그러나 속성 인덱스와 마찬가지로 경로 기반 인덱스는 속성이 질의 조건으로 명시되어 있는 질의만을 처리 대상으로 한다는 한계가 있다. 따라서 본 논문에서는 속성 기반의 질의 유형 외에도 다양한 질의 유형의 처리를 지원하기 위한 기법을 제안한다.

YaCy 인덱스는 시맨틱 웹 데이터에 대한 다양한 질의 형태를 지원하는데 목적을 두고 있다[4]. YaCy 인덱스는 트리플 구조에 기반하여 표현이 가능한 모든 질의 유형별로 관련된 정보를 유지한다. 그러나 YaCy 인덱스는 각 지역 저장소마다 독립적으로 유지되며 온톨로지가 따로 존재하지 않는 메타데이터만을 대상으로 한다. 따라서 각 저장소에서 추출된 리소스가 어느 클래스의 인스턴스인지를 파악할 수 없어 클래스만이 질의 조건으로 주어지면 처리가 불가능하다. 이러한 문제점을 해결하기 위해 본 논문에서는 온톨로지와 메타데이터를 함께 고려하는 모든 질의 유형을 지원할 수 있는 인덱스 구조를 제안한다.

특히, 기존 연구 결과들은 대부분 시맨틱 웹 데이터의 기술 언어로 RDF와 RDFS를 이용하기 때문에 분산되어 있는 웹 리소스간의 다양한 의미적 관계를 표현하지 못하는 한계점이 있다. 따라서 본 논문에서는 분산 환경에 적합한 OWL을 기반으로 하고, OWL의 다양한 표현력을 이용하는 질의 변환 알고리즘을 제안한다.

3. 분산 환경을 지원하는 OWL의 요소와 표현력

OWL은 웹 리소스의 의미를 다양하게 기술할 수 있고 분산되어 있는 리소스간의 의미적 관계를 기술할 수 있는 표현력을 제공한다. 그러므로, 향후 분산 시맨틱 웹 환경에서 사용자의 질의를 효율적으로 처리하기 위해서는 OWL을 고려한 연구가 필요하다. OWL을 기반으로 하는 시맨틱 웹 환경에서 사용자는 이러한 표현력을 활용하여 보다 다양한 질의를 처리할 수 있어야 한다.

표 1 OWL의 다양한 표현

분류	표현
Hierarchy	subClassOf, subPropertyOf
Property Characteristics	transitive, symmetric, functional, inverseOf, inverseFunctional
Property Restriction	allValuesFrom, someValuesFrom, cardinality, hasValue
Ontology Mapping	equivalentClass, equivalentProperty, sameAs, differentFrom, allDifferent
Complex Class	intersectionOf, unionOf, complementOf, oneOf, disjointWith

표 1은 OWL이 가지고 있는 다양한 표현들의 분류를 보여주고 있다. 특히, 표 1의 내용 중 온톨로지 매핑 관련 표현들을 통해 분산되어 있는 리소스간의 의미적 관계를 기술할 수 있다. 그러나 기존의 OWL 관련 연구에

서는 계층 관계와 같은 일부 표현만을 활용하고 OWL의 다양한 표현들을 모두 고려하지 않는 문제점이 존재한다. 따라서 본 논문에서는 OWL의 다양한 표현력을 이용하여 사용자 질의에 대해 보다 정확한 검색 결과를 제공할 수 있도록 사용자가 제시한 초기 질의 내용을 효과적으로 변환하기 위한 알고리즘을 제안한다.

4. 시맨틱 웹 기반 분산 시스템

4.1 시스템 구성

본 논문에서 제안한 시스템은 크게 미들웨어와 지역 저장소로 구성된다. 미들웨어는 지역적으로 분산되어 있는 저장소를 통합하여 사용자에게는 논리적으로 하나의 저장소에 접근하는 것과 같은 서비스를 제공한다. 특히, 본 논문에서는 미들웨어의 두 가지 구성 요소에 초점을 맞추고 있다. 첫째는 질의 변환 기법을 수행하는 질의 변환 모듈이며, 둘째는 적합한 데이터를 포함하는 지역 저장소를 빠르게 검색하기 위한 중재 인덱스이다. 그림 1은 본 논문에서 제안한 전체 시스템의 구성도를 보여준다. 그림 1에서 전역 스키마 저장소는 각 지역 저장소의 스키마를 통합하여 구축한 하나의 전역 스키마 정보를 저장한다. 그리고 질의 전송기는 중재 인덱스를 이용해 질의 처리 대상으로 선택된 각 지역 저장소에 질의를 전달하는 역할을 담당한다. 지역 저장소에는 각 지역에 분산되어 있는 리소스에 대한 정보가 저장된다.

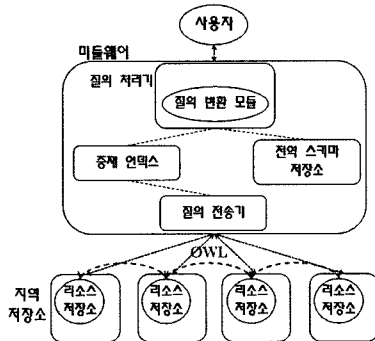


그림 1 시스템 구성도

4.2 전역 스키마 저장소와 중재 인덱스

본 논문에서 제안한 시스템의 전역 스키마 저장소는 OWL의 각 분류마다 독립적인 테이블을 유지함으로써 OWL로 표현된 모든 정보를 저장할 수 있으며, 각 표현에 대한 정보가 어떤 테이블에 존재하는지를 쉽게 파악할 수 있다. 그림 2는 본 논문에서 제안한 전역 스키마 저장소의 구조를 보여준다.

그림 1의 분산 시스템에서는 전역 스키마 저장소만으로 질의에 적합한 데이터를 포함하고 있는 지역 저장소를 찾을 수 없다. 따라서 본 논문에서는 질의 처리에 적

합한 데이터를 포함하고 있는 지역 저장소를 빠르게 검색하기 위해 지역 저장소에 대한 정보를 저장하고 있는 중재 인덱스를 제안한다. 본 논문에서 제안한 중재 인덱스는 클래스 인덱스, 속성 인덱스, 리소스-클래스 인덱스로 구성된다. 속성 인덱스는 속성이 질의 조건으로 주어진 질의 처리에 적합한 지역 저장소를 빠르게 검색하는 것을 목적으로 한다. 그리고, 클래스만이 질의 조건으로 주어진 경우 적합한 지역 저장소를 빠르게 식별할 수 있도록 별도의 클래스 인덱스를 유지한다. 클래스 및 속성 인덱스를 위한 노드 구조는 클래스나 속성의 이름, 데이터가 존재하는 지역 저장소의 식별자와 관련된 OWL 표현에 대한 정보를 포함한다. OWL 표현에 대한 정보는 질의 변환을 수행할 때 참조된다. 질의 변환에 대해서는 다음 절에서 자세히 설명한다.

그림 3은 중재 인덱스의 노드 구조를 보여준다.

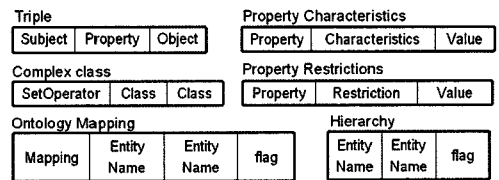


그림 2 전역 스키마 저장소의 구조



그림 3 중재 인덱스의 노드 구조

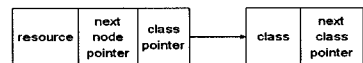


그림 4 리소스-클래스 인덱스의 노드 구조

그러나 리소스만이 질의 조건으로 주어진 경우는 클래스나 속성 인덱스를 통해서도 적합한 지역 저장소를 검색할 수 없다. 예를 들어 사용자가 특정 리소스에 대한 모든 정보를 검색하고자 하는 경우, 앞서 제안한 전역 스키마 저장소나 클래스 및 속성 인덱스에는 해당 리소스가 어느 클래스의 인스턴스인지에 대한 정보가 저장되어 있지 않기 때문에 적합한 지역 저장소를 검색할 수 없다. 따라서 모든 질의 유형을 고려하여 적합한 지역 저장소를 효율적으로 검색하기 위해서는 리소스와 클래스간의 관계를 빠르게 검색할 수 있는 리소스-클래스 인덱스가 추가적으로 요구된다. 리소스-클래스 인덱스의 노드 구조는 리소스 이름과 클래스 이름을 표현한다. 따라서 리소스만이 질의 조건으로 주어진 경우, 리소스-클래스 인덱스를 통해 리소스가 어느 클래스의 인스턴스인지 검색할 수 있고, 검색된 클래스 이름을 이용해서 클래스 인덱스를 검색하면 질의에 적합한 지역 저

표 2 질의 변환을 위한 기본 연산

연산	설명
s(q), p(q), o(q)	질의의 subject, property, object를 각각 반환한다. 변수일 경우, null을 반환한다.
type(x)	리소스-클래스인덱스를 탐색하여 리소스 x의 클래스를 반환한다.
getExpr(x)	중재 인덱스를 탐색하여 x란 클래스나 속성이 가지고 있는 OWL의 표현 정보를 반환한다. 없을 경우, null을 반환한다.
generateQ(s, p, o)	주어진 s, p, o로 (s, p, o)의 질의를 생성한다.
addQ(q1, q2, x)	질의 q1에 q2를 추가하여 확장한다. x는 and 또는 or를 나타내는 논리연산자이다.

장소를 파악할 수 있다. 그림 4는 리소스-클래스 인덱스의 노트 구조를 보여준다.

4.3 질의 변환 알고리즘

OWL의 다양한 표현을 활용하여 사용자에게 보다 정확한 검색 결과를 반환하기 위해서는 질의 변환 알고리즘이 필요하다. 즉, OWL로 표현된 온톨로지를 이용하여 사용자가 제시한 질의에 내포되어 있는 개념 및 의미적 관계를 찾아야만 보다 정확하고 풍부한 질의 결과를 획득할 수 있다. 예를 들어, “모든 사람을 검색하라.”라는 질의의 처리가 요청되었다고 가정하자. 이러한 질의는 (?x, type, Person)의 트리플 구조로 표현할 수 있다. 만약, “Person” 클래스가 “Male” 클래스, “Female” 클래스와 “unionOf” 관계를 맺고 있다면, “Person” 클래스는 “Male” 클래스와 “Female” 클래스의 합집합이기 때문에, “Male” 클래스의 인스턴스와 “Female” 클래스의 인스턴스는 모두 “Person” 클래스의 인스턴스가 된다. 따라서 보다 정확한 결과를 얻기 위해서는 단순 질의 (?x, type, Person)를 {(?x, type, Person); (?y, type, Male); (?z, type, Female)} 형태로 확장할 필요가 있다. 이와 같이 OWL의 표현력을 활용한 질의 변환을 통해 사용자에게 보다 정확한 결과를 반환할 수 있다. 이러한 질의 변환을 수행하기 위해서는 OWL의 표현을 클래스 관련 표현과 속성 관련 표현으로 구분할 필요가 있다. 이것은 본 논문에서 제안한 중재 인덱스가 클래스 인덱스와 속성 인덱스로 구성되어있기 때문이다. 질의 변환을 수행할 때, 중재 인덱스 노트에 대한 접근을 통해 관련된 OWL 표현을 쉽게 파악할 수 있고, 질의 변환 수행 시 중재 인덱스를 활용함으로써 빠르게 질의 변환을 수행할 수 있다.

표 2는 질의 변환 알고리즘을 위한 기본적인 연산들을 설명하고 있다. 그림 5는 질의 변환 알고리즘 중 “subClassOf” 표현을 위한 질의 변환 알고리즘을 보여준다. 그림 5의 첫번째 줄에서는 주어진 질의에서 주어, 술어, 목적어를 추출한다. 만약, 추출한 정보가 검색의 대상이 되는 변수일 경우는 널 값이 반환된다. 3~6번째 줄에서는 주어가 리소스일 경우, 리소스-클래스 인덱스를 탐색하여 관련 클래스를 파악하고, 만약 주어가 변수인 경우는 질의가 (?x, type, 클래스이름) 유형

인지를 확인하여, 해당 클래스를 파악한다. 그리고 나서 중재 인덱스의 클래스 인덱스를 탐색하여 해당하는 노트의 표현 정보를 알 수 있다. 8~14번째 줄에서는 앞서 검색된 표현 정보가 “subClassOf”일 경우 그에 맞는 질의 변환을 수행한다. 10번째 줄에서는 그림 3에서 제시한 전역 스키마 저장소의 “Hierarchy” 테이블을 탐색하여 해당 클래스가 어느 클래스와 “subClassOf” 관계인지를 파악한다. 11 ~12번째 줄에서는 10번째 줄에서 파악한 클래스의 인스턴스를 구하기 위해 새로운 변수를 정의하여 새로운 질의를 생성한다.

```

// 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 5 “subClassOf” 표현에 대한 질의 변환 알고리즘

13번째 줄에서는 기존의 질의에 새로 생성된 질의를 추가하여 최종적으로 질의를 확장한다. 질의 변환 알고리즘은 더 이상 새로운 표현 정보를 추출할 수 없을 때까지 반복적으로 수행된다. 예를 들어, “Hierarchy” 테이블을 탐색하여 파악한 클래스가 다른 클래스와 “subClassOf” 관계를 맺고 있다면 새로 파악된 클래스에 대해서 다시 변환 과정을 반복한다.

5. 실험 결과

본 논문에서는 유전자에 대한 정보를 OWL로 기술한 유전자 온톨로지(Gene Ontology)를 실험 데이터로 이용한다. 유전자 온톨로지는 많은 정보를 포함하고 있으며 OWL의 다양한 표현력이 활용되어 있기 때문이다.

표 3 3가지 데이터 유형

	파일 크기	트리플 개수
데이터1	120KB	500개
데이터2	513KB	2000개
데이터3	1.2MB	5000개

본 논문에서는 실험 데이터를 트리플 문장의 개수에 따라 표 3과 같이 세 개의 데이터 파일로 나누어 실험을 진행하였다.

실험은 크게 본 논문에서 제안한 인덱스 구조와 질의 변환 기법을 함께 적용한 경우와 제안 인덱스 구조만을 적용한 경우로 나누어 진행하였다. 그리고 반환된 질의 결과의 개수와 질의 처리 시간을 기준으로 실험 결과를 분석한다.

표 4는 실험에 사용된 두 가지 질의 유형을 보여준다. 첫 번째는 속성이 조건으로 주어진 경우이고 두 번째는 클래스 타입만이 조건으로 주어진 경우이다.

표 5는 각 질의의 처리 결과로 반환된 리소스의 개수를 보여준다. 본 논문에서 제안한 질의 변환 기법을 적용하면 그렇지 않은 경우에 비해 반환되는 리소스의 개수가 거의 2배가 되는 것을 표 5에서 확인할 수 있다. 즉, 제안한 질의 변환 기법을 적용하면 계층 관계나 UnionOf, 동의어 관계 등 다양한 의미적 관계를 이용하여 보다 풍부하고 정확한 정보의 검색이 이루어지게 된다. 특히, 두 번째 질의의 경우는 기존 인덱스 연구에서 고려하지 않은 질의 유형이다. 따라서 질의 변환 기법을 함께 적용한 본 논문의 인덱스 구조는 질의 변환 기법을 고려하지 않는 기존 인덱스의 한계점을 보완한 것으로 볼 수 있다.

표 6은 각 질의를 처리하는데 소요되는 시간을 보여

표 4 실험에 사용된 질의 유형

	질의 내용
질의1	[GO_0032433]과 동일한(synonym) 유전자 검색
질의2	[GO_0032433]타입의 모든 유전자 검색

표 5 질의별 반환된 결과 데이터 수

	반환된 결과 리소스 수		
	데이터	변환 기법 적용	변환 기법 제외
질의1	데이터1	57	24
	데이터2	309	189
	데이터3	590	270
질의2	데이터1	251	98
	데이터2	954	412
	데이터3	1948	1147

표 6 질의별 처리 시간

	질의 처리 시간		
	데이터	변환 기법 적용	변환 기법 제외
질의1	데이터1	0.43ms	0.30ms
	데이터2	49.62ms	38.10ms
	데이터3	192.63ms	172.02ms
질의2	데이터1	0.69ms	0.33ms
	데이터2	149.20ms	120.77ms
	데이터3	690.29ms	559.46ms

준다. 본 논문에서 제안한 질의 변환 기법을 적용하더라도 적용하지 않았을 때 보다 처리 시간이 크게 증가하지 않는다는 것을 표 6에서 확인할 수 있다. 따라서, 표 5에서 제시된 질의 처리 결과 리소스의 수를 고려할 때 본 논문에서 제안한 질의 변환 기법은 활용 가치가 있다고 볼 수 있다.

6. 결론 및 향후 연구

시맨틱 웹의 기반이 되는 온톨로지는 웹 리소스의 의미를 더욱 다양하게 기술하기 위한 표현력을 갖추어 가고 있다. 또한 정보의 양이 더욱 많아지고 온톨로지가 지역적으로 분산되어 구축됨에 따라 향후 시맨틱 웹 기반 분산 환경에서 원하는 데이터를 포함하고 있는 지역 저장소를 빠르게 검색하는 것은 시스템의 전체 성능을 결정하는 중요한 요인이 될 것이다. 따라서 본 논문의 연구는 다음과 같은 두 가지 측면에서 시맨틱 웹의 발전을 위해 기여할 수 있다.

첫째, 분산된 온톨로지 환경에서 질의에 적합한 데이터가 위치하고 있는 지역 저장소를 빠르게 검색하기 위한 중재 인덱스 구조를 제안하였다. 둘째, OWL의 다양한 표현을 이용하여 질의를 변환하기 위한 기법을 제안하였다. 본 논문에서 제안한 질의 변환 기법을 적용하면 OWL의 다양한 표현을 이용해서 사용자의 질의에 대해 보다 정확하고 풍부한 검색 결과를 반환하는 것이 가능하다. 그리고 본 논문에서 제안한 중재 인덱스를 이용하면 모든 질의 유형에 대해 적합한 데이터가 존재하는 지역 저장소를 쉽게 파악할 수 있었다.

향후에는 인덱스 공간을 보다 효율적으로 사용하기 위한 개선 방안과 질의 변환 알고리즘의 효율성을 향상시키기 위한 추가적인 연구가 필요하다.

참고 문헌

- [1] OWL Web Ontology Language Guide, <http://www.w3.org/TR/owl-guide/>.
- [2] G. Adamku, H. Stuckenschmidt, Implementation and Evaluation of a Distributed RDF Storage and Retrieval System, In Proc. of The International Conference on Web Intelligence, 2005.
- [3] H. Stuckenschmidt, R. Vdovjak, G. Houben and J. BroekstraIndex, Index Structures and Algorithms for Querying Distributed RDF Repositories, In Proc. of World Wide Web Conference, 2004.
- [4] E. Demidova and W. Nejdl, Integrating RDF Querying Capabilities into a Distributed Search Infrastructure, In Proc. of Asian Semantic Web Conference (ASWC), 2006.
- [5] A. Harth, S. Decker, Optimized Index Structures for Querying RDF from the Web, In Proc. of the 3rd Latin American Web Congress, 2005.