

# 인터넷상의 온톨로지간의 P2P 질의처리 방안

김병곤\*, 오성균\*\*

## 요 약

단순한 형태의 네트워크 시스템에서는 질의가 발생하면 질의를 효율적이고 정확하게 처리하기 위하여 연결된 모든 피어로 질의를 전송한다. 그러나, 이러한 처리방식은 전송 대역폭을 낭비하게 되고, 각 피어의 효율을 감소시킨다. 이를 극복하기 위하여, 질의처리 기법뿐만 아니라 질의를 적절한 곳으로 전송하기 위한 라우팅 기법이 필요하다. 더구나 네트워크의 환경이 P2P 환경으로 구성되어 있는 경우엔, 효율적인 라우팅을 위해서는 네트워크를 구성하는 피어들을 여러개의 그룹으로 묶어 클러스터를 구성하는 것이 효율적이다. 이때 랜덤하게 클러스터를 구성하는 것 보다 유사성에 기준을 두고 클러스터를 구성하는 것이 중요한 요소가 된다. 이 논문은 특히 갈수록 중요성을 더해가는 시맨틱웹을 구성하기 위한 온톨로지 환경에서의 P2P 질의 처리를 위한 클러스터링 기술을 제안한다. 클러스터링을 하기 위한 유사성 측정요소를 제안하며, 질의를 P2P 환경에서 처리하기 위한 클러스터 인덱스 구조, 질의처리 단계 등을 제안한다.

## P2P query processing method between ontologies in internet environment

Byung-gon Kim\*, Sung-Kyun Oh\*\*

## Abstract

In simple topology in network system, query should be delivered to all linked peers for query processing. This causes waste of transmission band width and throughput of each peer. To overcome this, as well as query processing strategy, efficient routing technique to deliver query to proper peer is needed. For efficient routing, clustering of peers in P2P networks is important. Clustering of P2P network bases on that combines peers that have similar characteristics in same cluster reduces quantity of message in network than assign peer for cluster randomly. In this paper, we propose clustering techniques for ontology based P2P query processing. Similarity measure point, cluster index structure, and query processing steps in ontology based P2P cluster environment are proposed.

Key Words : Ontology, P2P, Query Processing, RDF, OWL

## 1. 서론

웹 자원의 수는 폭발적으로 계속 증가하고 있고 지역적으로도 분산되어 있는 웹 환경을 고려할 때, 웹 환경에서의 데이터 관리에 대한 연구

의 중요성이 증대되고 있다. 또한 차세대 인터넷은 지금까지의 데이터베이스를 이용한 단순한 검색의 수준을 벗어난 온톨로지 등을 이용한 시맨틱 웹 개념의 지능형 웹 시스템으로의 변화를 시도하고 있다. 시맨틱 웹을 이용한 차세대 웹 환경은 각 호스트마다 온톨로지를 구축하여 호스트간의 메시지 교환을 통하여 사용자에게 정확하고 다양한 서비스를 제공하고 있다. 하지만, 기구축된 온톨로지들은 동음이의어와 같이 동일한 개념에 대한 용어상의 차이 혹은 기술언어의 차이 그리고, 데이터 모델의 차이를 기계가 자동으로 이해하고 식별할 수 없기 때문에 의미적으로는 동일한 온톨로지임에도 불구하고 서로 다른 온톨로지로 인식될 수밖에 없다. 또한, 서비

※ 제일저자(First Author) : 김병곤  
접수일:2009년 03월 26일, 완료일:2009년 06월 11일  
\* 부천대학 e-비즈니스과  
bgkim@bc.ac.kr  
\*\* 서일대학 소프트웨어과  
■ 본 논문은 2008년도 서일대학 학술연구비에 의해 연구되었음

스를 제공하는 호스트는 물리적으로 분산되어 있고 서비스를 제공하는 회사나 기관들도 논리적으로 분산되어 있는 것이 일반적이기 때문에 단일 시스템에서의 서비스가 아닌 분산 환경을 반드시 고려하여야만 한다.

기존의 온톨로지간의 질의 통합처리는 온톨로지들을 의미적으로 통합하는 통합스키마를 구성하는 방식 즉 온톨로지의 스키마를 매핑, 정렬, 합병, 통합, 결합 등을 통하여 논리적으로 통합하여 유지하는 방식의 연구가 주를 이루었다. 또한 미들웨어를 이용한 전역 뷰 혹은 지역 뷰를 생성하여 사용자에게 정보를 제공하는 형태의 연구가 진행되었으며, 전역 뷰를 생성하기 위하여 지역 피어에 존재하는 상이한 데이터 모델의 변환 규칙을 정의하는 연구와 사용자의 질의를 지역 피어에서 처리하기 위한 질의 재형성 등의 알고리즘에 대한 연구가 이루어 졌다[1,2,3,4]. 이러한 연구들은 각 온톨로지의 스키마들을 효율적으로 통합하는 통합스키마를 통한 일관된 질의 처리가 가능하다. 그러나 계속적으로 변화하는 동적인 인터넷 환경에서의 온톨로지들을 일관성 있게 관리하는 통합 스키마를 유지하는 것은 현실적으로 불가능하며 이를 극복하기 위해서는 통합처리 방식이 아닌 P2P 방식의 질의 처리 방식으로의 전환이 필요하다고 할 수 있다.

P2P 환경에서의 온톨로지간 메시지 교환을 통한 통합 질의처리 환경은 P2P 네트워크상에서 분산된 데이터들의 효율적인 관리와 질의 처리를 목적으로 구성된다. 그러나, P2P 네트워크 시스템에서 단순한 토폴로지를 사용하게 되면 질의가 발생하였을 때 연결되어 있는 모든 피어에 질의가 전달되어야 하며 이는 각 피어의 전송 대역폭과 처리 능력의 낭비를 초래한다. 이를 극복하기 위해서는 피어 내에서의 질의 처리 전략 뿐만 아니라, 질의를 필요한 피어로 전달하기 위한 효율적인 라우팅 기법이 필요하다. 또한 피어들간의 그룹핑을 통한 클러스터링을 이용하여 라우팅을 수행하면 효율적인 질의 처리가 가능하다. 그러므로, 구성된 라우팅 인덱스와 클러스터링 기법을 이용하여 각 피어에서 발생한 질의를 P2P 방식으로 처리하는 질의 처리 기법에 대한 연구가 필요하다.

본 연구는 위에서 설명한 바와 같이 P2P 환경에서 각 피어에 온톨로지가 구축되어 있는 경우

에 효율적인 질의 처리를 위한 피어간의 클러스터링 기법을 소개하고 각 클러스터내에서의 인덱스 구성과 이를 이용한 질의처리 방안을 제시한다.

## 2. 관련연구

### 2.1 시맨틱 웹 표현 기술 분석

시맨틱 웹은 웹 상에 정보의 리소스들이 서로 의미적 연결을 가지고 있고 인간과 기계 모두가 쉽게 문서를 이해할 수 있도록 지원해야 한다. 초기의 연구들에서는 XML을 기반으로 한 연구들이 많이 진행 되었으나, XML의 구조적인 한계와 표현능력의 한계로 인하여 점차 RDF, RDF/S, OWL등의 시맨틱 웹 전용언어로의 연구로 변화하고 있다.

온톨로지를 표현하기 위해 W3C에서 제안된 RDF는 웹상에 존재하는 자원들의 메타데이터를 기술하기 위한 언어이다. RDF는 XML 문법을 따르고 있으며, 기본적으로 자원-속성-값의 기본 구조를 이용하여 자원의 메타데이터를 기술한다[5]. RDF/S는 RDF를 기술하기 위해 필요한 용어들의 의미와 용어들 간의 의미적 관계를 기술하기 위해 W3C에서 제안한 언어이다. RDF/S를 이용해 정의된 용어와 용어들 간의 관계는 RDF로 기술된 메타데이터의 의미를 좀 더 명확하게 표현할 수 있다.

OWL은 기존 RDF/S를 이용한 온톨로지의 표현을 확장 및 강화하기 위하여 개발되고 있다. 또한, 표현이 대상이 되는 클래스, 속성, 인스턴스 간의 복잡한 관계에 대한 기술이 가능하기 때문에 RDF/S를 이용한 질의 유형 보다 확장된 형태의 다양한 질의 처리가 가능하다[6]. 즉, 온톨로지의 다양한 표현에 대한 보다 많은 제약사항을 기술하기 때문에 클래스와 구성원간의 관계를 집합이란 개념을 통하여 온톨로지로는 정의되지 않는 사실들에 대한 논리적인 추론이 가능하기 때문에 최근 이를 이용한 다양한 연구가 진행되고 있다.

### 2.2 시맨틱 웹 P2P 기술 분석

P2P 네트워크는 Napster나 Gnutella와 같은 단순 시스템에서 CAN, CHORD와 같이 분산 인

텍스를 사용하는 더욱 발전된 형태의 시스템으로 진화되어 왔다[7,8]. 그러나 이러한 시스템들은 아직도 더욱 복잡한 형태의 메타데이터나 질의를 처리하는데 많은 한계점을 지니고 있다.

인터넷이 더욱 일반화되고 시맨틱웹이 현재의 인터넷 환경의 대안으로 연구되어 지고 있는 상황에서 시맨틱웹상에서의 P2P 네트워크에 대한 연구는 더욱 중요성을 더해가고 있다. 시맨틱웹의 중요한 측면은 RDF스키마, OWL과 같은 언어들을 사용하여 웹상의 자원들에 대한 온톨로지를 구성하고, 이를 이용하여 컴퓨터시스템간의 데이터 교환이 자율적으로 발생한다는 것이다.

현재의 인터넷 환경에서 메타데이터를 지니는 온톨로지는 단지 하나의 피어의 한 서버에만 존재하는 것이 아니다. 즉, 동일한 자원에 대한 메타데이터가 여러 피어에 분산되어 저장될 수 있다. 그러나, 현재 연구되어진 P2P 시스템들은 단순한 파일이름과 같은 제한된 형태의 메타데이터를 지원하므로 좀 더 복잡한 형태의 질의를 처리하지 못하는 것이 현실이며, 이를 극복하기 위하여 스키마를 지니는 온톨로지를 적용한 P2P 네트워크 시스템의 개발이 필요하다. 또한 네트워크간의 상이함을 고려하여 디자인하여 서로 다른 네트워크간의 결합이 용이하도록 디자인하여야 한다.

스키마기반의 P2P 네트워크 시스템의 예로 E dutella시스템이 있다. 이 시스템은 P2P 네트워크를 통한 분산 디지털 자원들에 대한 접근을 목표로 한다. 메타데이터를 표현하기 위하여 RDF와 RDF 스키마를 사용하며, 질의어로는 RDF-QEL을 사용한다[9].

Gloserv 시스템은 OWL DL을 사용하는 P2P 전역 서비스 검색 시스템으로서 광역 네트워크와 지역 네트워크에서 모두 동작하도록 설계되었다. 넓은 영역에 걸쳐 있는 서비스들은 OWL 온톨로지를 통하여 효율적으로 표현되고, CAN P2P 네트워크에 연결된 노드들을 걸쳐서 서비스를 검색하도록 동작한다. Gloserv 서버는 서비스 분류(Service classification) 온톨로지, 시소러스(Thesaurus) 온톨로지, CAN 조사테이블(look up table)로 구성된다. 서비스분류 온톨로지는 전체 서비스 목록을 바탕으로 서비스들을 계층적으로 분류한 정보를 지닌다. 시소러스 온톨로지는 서비스분류 온톨로지에 나타난 단어들에 대

한 유사어 사상 정보를 지니고, 서비스가 존재하는 올바른 서버를 찾는 데 정확도를 높이는데 사용된다. CAN 조사테이블은 CAN 환경의 P2P 네트워크에서 각 서버의 관련된 클래스들을 연결시켜주는 역할을 한다[10].

이러한 연구들은 P2P 네트워크상에서 분산된 데이터들의 효율적인 관리와 질의 처리를 목적으로 구성되었다. 그러나, P2P 네트워크 시스템에서 단순한 토폴로지를 사용하게 되면 질의가 발생하였을 때 연결되어 있는 모든 피어에 질의가 전달되어야 하며 이는 각 피어의 전송 대역폭과 처리 능력의 낭비를 초래한다. 이를 극복하기 위해서는 피어 내에서의 질의 처리 전략뿐만 아니라, 질의를 필요한 피어로 전달하기 위한 효율적인 라우팅 기법이 필요하다. 라우팅을 위해서는 각 피어에 라우팅 인덱스가 필요하다. 라우팅 인덱스는 자신과 연결되어 있는 피어들에 대한 메타데이터 정보를 지니게 된다. 라우팅 인덱스의 구조는 어느 한 피어의 지역 스키마에 의존하지 않고, 네트워크 전체에서 사용되는 일정한 형태의 구조를 지니도록 설계하여야 한다.

이와 관련된 중요한 기술 중의 하나가 클러스터링이다. P2P 네트워크의 클러스터링은 비슷한 특성을 지니는 피어들을 가까운 위치에 결합하는 것을 기본으로 한다. 이러한 방법은 무작위로 피어를 클러스터에 배정하는 것보다 네트워크내에서의 메시지의 양을 줄여준다. 어떤 종류의 유사성 측정을 통하여 네트워크를 분할할지가 중요한 관건이다.

본 연구에서는 네트워크상에 존재하는 온톨로지를 지니는 피어가 추가될 때의 클러스터링 방법으로 온톨로지 스키마의 유사도에 근거하여 가장 유사한 스키마를 지니는 클러스터에 배정하는 방법을 사용한다. 이 방법은 스키마를 고려하지 않은 기존의 방법에 비하여 클러스터간의 질의 전달을 최소화하여 전체 시스템의 처리 능력을 향상 시킨다.

클러스터가 결정된 후, 클러스터를 기반으로 한 질의 처리 방법은 클러스터에 존재하는 피어들에 관한 인덱스의 위치에 따라 달라진다. [11]에서는 클러스터에 슈퍼피어를 유지하는 방법을 사용하였다. 각 슈퍼피어는 각 피어에서 사용되는 메타데이터를 지니며, 등록시에 각 피어는 슈퍼피어에게 자신의 메타데이터 정보를 제공한다.

인덱스는 항상 최신의 정보를 지니도록 하며 이를 위하여 인덱스의 업데이트는 변동사항이 있을 때 마다 갱신된다. 이때, 피어들은 슈퍼피어에 임의로 분산되어 연결되어 있으므로 대부분의 질의들은 모두 슈퍼피어에 보내져야 하므로, 처리 특성상 슈퍼피어로의 전송량이 증가하여 슈퍼피어의 부하가 전체 시스템 성능에 영향을 줄 가능성이 존재한다.

본 연구에서는 클러스터내의 질의 전송에 대한 부하의 편중 현상을 줄이기 위하여 클러스터내의 피어들의 인덱스 정보를 각 피어내에 유지하는 피어인덱스방법을 제안하였다. 또한, 클러스터간의 연결 노드 역할을 하는 인터페이스 피어를 유지하도록 하였다. 이 방법은 질의가 클러스터내에서 처리되는 경우에 더욱 효율적으로 운영된다. 다음 장에서는 제안된 처리방법의 구체적인 내용을 제시하였다.

### 3. 온톨로지 P2P 질의 처리

#### 3.1 P2P 클러스터 구성

P2P 네트워크에서 질의가 발생하였을 때 모든 피어에 질의를 전송하는 것은 효율적이지 않다. 그러므로, 온톨로지가 연결되어 있는 P2P 네트워크에서의 효율적인 질의 전송을 위하여 클러스터의 개념이 필요하다. 각 클러스터에서는 인덱스를 사용하여 자기 클러스터에 소속되어 있는 피어들을 관리하고, 클러스터에서 얻을 수 없는 정보는 클러스터간의 정보 교환을 통하여 해결하는 방법이 필요하다.

가장 단순한 클러스터 구성 방법은 새로운 피어가 네트워크에 발생하면 무작위에 의한 클러스터에 배정하는 방법이 있다. 그러나 일반적으로 P2P 네트워크의 클러스터는 비슷한 특성을 지니는 피어들을 가까운 위치에 결합하는 것을 기본으로 한다. 이러한 방법은 무작위로 클러스터에 피어를 배정하는 것보다 네트워크내에서의 메시지의 양을 줄여준다. 이와 같이 어떤 종류의 유사성 측정을 통하여 네트워크를 분할할지가 중요한 요소이다.

온톨로지로 구성된 네트워크를 클러스터링 할 때에는 질의의 특성이나 각각의 스키마간의 관계를 고려하는 방법이 사용될 수 있다. 질의의

특성을 고려하는 방법의 경우에 이중의 P2P 네트워크에서 질의의 특성이 미리 정의 되는 것은 어렵기 때문에 빈도 측정 알고리즘과 같은 방법을 사용할 수 있다. 이를 통하여 가장 많이 사용되는 스키마, 프로퍼티, 값 등을 분석하여 질의가 전송될 클러스터를 결정하거나 두개의 피어가 같은 클러스터에 존재하게 할지 등의 문제를 해결하는데 사용될 수 있다. 그러나 통계적인 빈도 측정은 계속적인 데이터의 축적이 필요하며 유동적인 질의 패턴에는 적용하기가 어려운 단점이 있다.

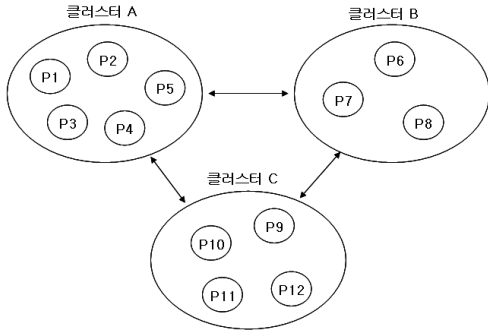
본 연구에서 제안하는 클러스터 구성 방법은 온톨로지의 특성인 구조적 관계에 기반한 클러스터링 기법이다. 즉, 각 피어의 온톨로지가 지니는 스키마내의 클래스, 프로퍼티, 값 등에 대한 피어들간의 관계를 파악하여 피어들간에 관계가 성립되면 클러스터에 추가하며 각 피어의 인덱스에 관계된 내용들을 등록하는 것이다. 다음은 클러스터에 새로운 피어가 추가되는 경우의 유사도를 측정하기 위한 요소들을 나열한 것이다.

<표 1> 유사도 관계 측정요소

분류	내용	우선 순위
스키마	새로운 피어가 클러스터의 피어들과 동일한 스키마를 사용하는 경우	1
클래스	새로운 피어에 클러스터의 피어내에 동일하거나 동치인 클래스가 존재하는 경우	2
	새로운 피어에 클러스터의 피어내에 서브클래스관계의 클래스가 존재하는 경우	3
프로퍼티	새로운 피어에 클러스터의 피어내에 동일하거나 동치인 프로퍼티가 존재하는 경우	4
	새로운 피어에 클러스터의 피어내에 서브프로퍼티관계의 프로퍼티가 존재하는 경우	5

<표 1>에 나열된 요소들은 순서에 따라 우선순위가 있으며, 한 피어는 여러개의 클러스터 중에서 가장 높은 유사도가 있는 클러스터에 소속되도록 한다. 클러스터가 결정되면 각 피어는 피어에서 사용하는 메타데이터 스키마를 클러스터내

의 피어들에게 알려 등록한다.



(그림 1) 클러스터 구성 개념

(그림 1)은 유사도 관계 측정에 의하여 구성된 클러스터의 개념을 보여준다. 모든 클러스터는 자신만의 규칙 즉, 어떤 피어들이 클러스터에 포함될 수 있는지를 표현하는 규칙들을 지닌다. 각 규칙들은 이벤트, 제한사항 등의 형태로 구성되며 <표 2>에서 제시된 내용들을 확인하여 결정한다.

### 3.2 클러스터 인덱스

#### 3.2.1 피어 인덱스

클러스터내의 피어는 클러스터내의 다른 피어들에 대한 메타데이터를 지니도록 한다. 등록시에 각 피어는 클러스터의 각 피어에게 자신의 메타데이터 정보를 제공한다. 인덱스는 항상 최신의 정보를 지녀야 하며 이를 위하여 인덱스의 업데이트는 변동사항이 있을 때 마다 수행된다. 피어 인덱스는 클러스터의 다른 피어들의 정보를 지닌다. 피어들의 사용 스키마, 클래스, 프로퍼티 등이다. <표 2>에 피어 인덱스의 내용들이 기술되어 있다. 각 클러스터에서는 질의에 존재하는 요소들에 대하여 질의에 응답이 가능한 피어를 결정하기 위하여 피어 인덱스의 내용들과 일치하는 지를 판단한다. 일치하는 내용이 있는 피어가 결정되면 해당 피어는 해당 질의에 응답하도록 한다.

<표 2> 피어 인덱스 기본내용

종 류	내 용
스키마인덱스	피어들은 서로 같거나 다른 스키마를 지원할 수 있고, 이 스키마들은 네임스페이스 등을 이용하여 서로 구분할 수 있다. 피어 인덱스는 각 피어가 지원하는 스키마의 구분자를 지닌다.
클래스인덱스	피어들은 자신들의 콘텐츠를 구성하기 위하여 스키마의 일부 클래스만을 선택할 수 있다. 이러한 경우에 인덱스에서는 해당 피어들을 서술하기 위하여 피어ID와 클래스ID로 구성된 클래스의 리스트를 지닌다.
프로퍼티인덱스	피어들은 자신들의 콘텐츠를 구성하기 위하여 스키마의 일부 프로퍼티만을 선택할 수 있다. 이러한 경우에 인덱스에서는 해당 피어들을 서술하기 위하여 피어ID와 프로퍼티ID로 구성된 프로퍼티들을 지닌다.

피어 인덱스는 <표 2>에서 나타난 내용과 더불어 스키마간의 이행성과 상속성을 유지할 필요가 있다. 서로 다른 피어의 클래스나 프로퍼티가 동치관계에 있을 때 클래스나 프로퍼티들 간에는 이행성 규칙을 적용하여 수행 할 수 있다. 즉, 클래스 A와 B가 동치관계이고, 클래스 B와 C가 동치관계이면 클래스 A와 C는 동치관계이다. 여러 피어의 온톨로지 구성요소들에 대하여 이행성 규칙을 적용하면 동일클래스그룹이 생성된다. 동일클래스그룹이란 온톨로지의 구성 요소들간에 이행성 규칙에 의하여 생성된 동치 요소들의 집합이다. 또한, 클래스나 프로퍼티간의 상하관계를 이용하여 질의의 결과를 요구할 수도 있다. 다음은 클러스터내의 피어들간의 구성요소들의 관계를 추출하기 위한 규칙들을 나열한 것이다.

<표 3> 관계규칙

규칙	내용
관계규칙1	이행성 규칙에 의하여 클래스간의 동일클래스그룹이 생성되었을 때, 프로퍼티 P가 서로 다른 동일클래스그룹에 속한 클래스를 도메인과 레인지로 지닌다면, 프로퍼티 P를 도메인과 레인지로 지니는 이행성 그룹의 모든 클래스는 프로퍼티 P의 도메인과 레인지가 될 수 있다.
관계규칙2	동일클래스그룹의 한 클래스의 인스턴스는 그룹내의 다른 클래스의 인스턴스로 간주될 수 있다.
관계규칙3	이행성 규칙에 의하여 프로퍼티간의 동일클래스그룹이 생성되었을 때, 동일클래스그룹내의 하나의 프로퍼티 P의 도메인과 레인지에 해당하는 클래스는, P가 속한 프로퍼티 동일클래스 그룹내의 다른 프로퍼티들에 대하여 도메인과 레인지로 동일하게 적용된다.
관계규칙4	역관계의 역관계는 동치이다. 그리고 역관계 요소의 동일클래스그룹에 포함시킬 수 있다.
관계규칙5	두개의 프로퍼티가 상/하위 프로퍼티 관계이면, 하위 프로퍼티의 도메인과 레인지 클래스는 상위 프로퍼티의 도메인과 레인지 클래스와 각각 상/하위 클래스의 관계가 성립한다.
관계규칙6	프로퍼티 P의 도메인/레인지 클래스의 실제 인스턴스들은 도메인과 레인지의 상위 클래스들의 각각의 인스턴스로 간주될 수 있다.

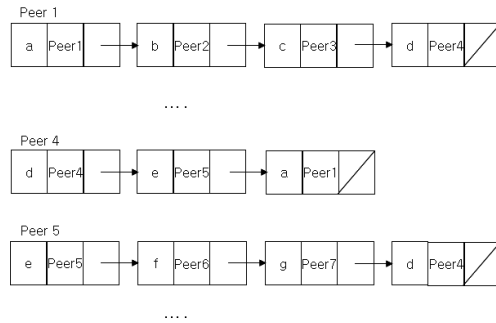
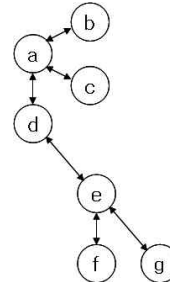
<표 3>에 제시된 관계규칙들을 적용하면 피어가 속한 클러스터내의 다른 피어의 온톨로지들과의 관계를 파악할 수 있다. 여기서 파악된 내용들은 질의 처리시에 사용된다. P2P 환경에서 질의와 연관된 피어의 온톨로지에 대한 효율적인 질의 처리를 위하여 저장구조가 요구된다. 본 논문에서는 각 피어의 온톨로지의 구성요소별로 다음과 같은 연결노드 구조를 가지는 리스트를 구성하고 저장한다.

클러스터 내의 요소들간의 이행성과 상속성을 표현하기 위해서는 피어 인덱스 내에 다음과 같은 내용이 필요하다.

Class name or Property name	Peer No.	Next pointer
-----------------------------	----------	--------------

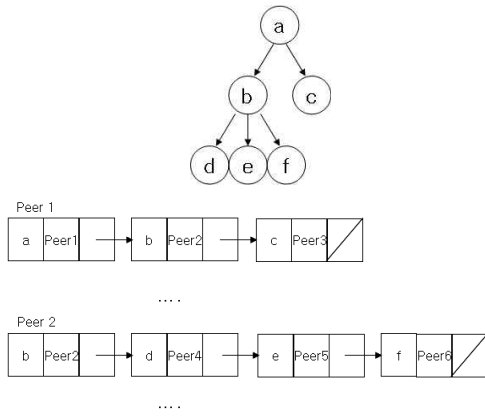
(그림 2) 이행성과 상속성 표현을 위한 연결노드 구조

(그림 2)는 이행성과 상속성을 표현하기 위한 연결노드의 구조를 나타낸다. (그림 3)은 아래와 같이 그래프로 표현된 이행성관계를 질의를 처리하기 위한 노드연결 예를 보여준다.



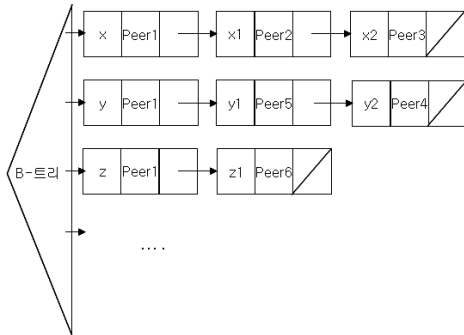
(그림 3) 이행성에 대한 피어별 연결노드 예

(그림 3)에서 피어1의 a노드는 자신과 동치인 b, c, d노드에 대한 연결노드를 지닌다. 또한, 노드 d를 지니는 피어 4에는 노드 d와 연결된 a, e노드에 대한 연결 노드만을 지닌다. 이는 만약, 피어 1에서 질의가 발생하고 이 질의가 노드 a와 관련 있는 질의이면, 이 질의는 연결노드를 통하여 피어1에서 피어 4, 5, 6, 7과 같은 동치노드들이 있는 피어로 전파되어 처리된다. 피어 4, 5, 6, 7에서는 피어 1과 마찬가지로 자신과 연관된 온톨로지 요소들에 대한 연결노드를 지니므로 자신의 피어에서 처리할 부분과 다른 피어로 전송할 부분을 다시 생성 할 수 있다.



(그림 4) 상속성에 대한 피어별 연결노드 예

(그림 4)는 상단의 상속성 그래프에 대한 피어별 연결노드의 예를 보여준다. 연결성 그래프와 마찬가지로 현재 발생한 질의에 의하여 어느 피어로 질의가 전이될 것인지 판단하기 위하여 이를 바탕으로 규칙들을 적용한다. (그림 4)의 연결노드는 하위클래스에 대한 정보를 지니며, 상위 클래스에 대한 정보도 별도의 연결노드로 유지한다.



(그림 5) 피어 온톨로지 상속성, 동치성 통합인덱스

(그림 5)는 (그림 3)과 (그림 4)에서 표현된 온톨로지의 클래스나 프로퍼티별 연결노드를 접근하기 위한 통합 인덱스를 표현하였다. 피어에서 질의가 발생하면 질의를 파싱하여 질의처리에 필요한 온톨로지 요소를 파악하고 <표 2>와 <표 3>의 내용과 (그림 5)의 통합인덱스를 통하여 피어들의 정보에 접근하게 된다.

3.2.2 인터페이스 피어와 클러스터인덱스

P2P 질의 처리에서는 클러스터내에서 질의가 처리되는 경우에는 클러스터간 전송이 필요 없지만, 다른 클러스터로의 정보 전송이 요구되는 경우에 다른 클러스터로의 질의를 전송하는 것이 중요하다. 이를 위하여 인터페이스피어가 필요하다. 인터페이스 피어는 클러스터의 인터페이스 피어들을 지니는 클러스터인덱스를 지닌다.

인터페이스 피어는 클러스터에 있는 피어중에서 하나를 선정하며 일반적으로 가장 먼저 등록된 피어를 선정하는 방법이 가장 간단한 방법이 될 수 있다. 인터페이스 피어는 말 그대로 클러스터간의 인터페이스 즉 연락창구 역할을 하며, 다른 클러스터들에게 자신의 클러스터가 지닌 정보를 보여주고, 요구된 질의를 클러스터내의 적절한 피어로 보내주는 역할을 수행한다.

질의는 클러스터 인덱스를 바탕으로 이웃 클러스터의 인터페이스 피어로 전송되며 피어 인덱스에 의하여 연결된 소속 피어로 다시 전달된다.

클러스터 인덱스와 피어 인덱스의 수정은 연결된 피어들의 수정메시지에 의하여 수행된다. 이것은 하나의 피어는 임의의 하나의 클러스터에 연결되며 다음과 같은 등록 프로세스를 따른다. 클러스터에 새로운 피어가 등록되면 해당 피어의 스키마에 대한 정보를 멤버들에게 보낸다. 멤버들은 피어인덱스의 엔트리들과 내용을 비교하여 반영한다. 즉, 피어인덱스에 새로운 항목이 하나 추가되면 피어는 이와 연결된 네트워크에 이에 대한 사항을 전송하여 수정한다.

3.3 질의처리과정

분산된 온톨로지 환경에서의 P2P 질의 처리 과정은 다음과 같다.

단계 1 : 질의 분석

한 피어의 온톨로지를 기반으로 한 지역질의가 발생하면, 지역 피어의 질의처리기는 질의를 트리플 단위로 파싱하여 클래스, 프로퍼티별로 분석하여 먼저 클러스터내의 다른 피어의 정보를 지니는 피어인덱스와 비교 검색할 준비를 한다.

단계 2 : 피어인덱스 검색

파싱한 질의에 대해 피어인덱스를 사용하여 클러스터내의 클래스와 프로퍼티의 내용을 검색한다. 검색된 결과를 바탕으로 자신의 피어에서 처리가 가능한 경우에는 자체 질의 모듈로 질의를 처리하고 종료한다. 다른 피어로의 질의 전송

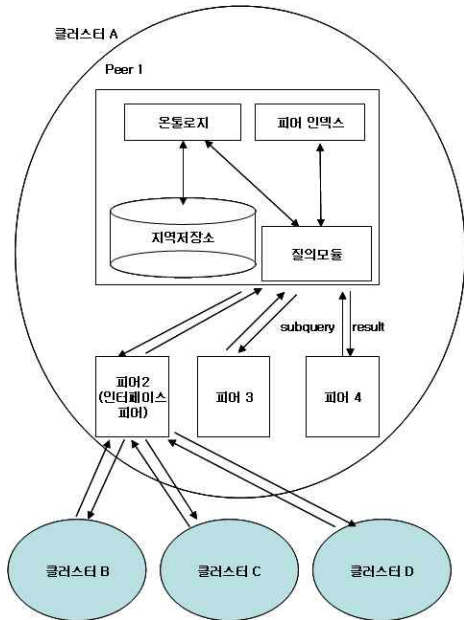
이 요구되는 경우에는 각 피어에 전송될 별도의 질의를 산출한다.

단계 3 : 클러스터인덱스 검색

소속된 클러스터내에서 질의의 처리가 수행되지 못한 경우에는 인터페이스피어의 클러스터인덱스를 이용하여 외부 클러스터의 인터페이스피어로 질의가 전송된다.

단계 4 : 부분질의의 전송 및 결과 취합

단계 2, 3을 통해 각 지역 피어로 부분질의가 전송되면 지역 피어는 전송받은 질의를 실행하여 결과 데이터를 추출한다. 이때, 각 지역 피어는 질의 발생 피어에서 수행한 단계 1, 단계 2의 질의처리를 마찬가지로 수행한다. 최종적인 부분질의의 처리가 끝나면 결과 데이터는 다시 호출피어로 전송되고, 호출피어는 전송받은 결과를 취합하여 결과를 반환한다.



(그림 6) 질의처리 구성도

(그림 6)은 질의처리개념을 구성도로 표현한 것이다. 각 피어별로 지역저장소가 존재하며 이는 XML, DB등의 각각의 구조로 데이터가 저장되어 있다. 데이터에 대한 데이터를 표현하는 온톨로지는 OWL로 구성되어 있다. 앞에서 설명한 4단계의 질의처리를 위하여 질의모듈이 있으며, 질의모듈은 온톨로지와 피어인덱스를 참조하여

부분질의를 만들어 전송하고 취합하는 역할을 한다.

#### 4. 결론 및 향후 과제

인터넷의 환경이 텍스트 중심의 단순검색에서 웹2.0 기술을 접목한 의미검색으로의 변화가 일어나면서, 온톨로지와 같은 시맨틱웹 기술을 사용하여 데이터를 구축하는 연구가 활발해지고 있다. 본 논문은 온톨로지 기반의 데이터가 여러 곳에 분산되어 있는 경우에 사용자에게 의하여 요구된 질의를 피어들간의 P2P 방식을 이용하여 처리하는 효율적인 방법을 제시하였다. 먼저 모든 피어들로의 질의 전송을 피하기 위하여 클러스터를 구성하는 방법을 제시하였다. 이를 위하여, 온톨로지의 특성인 구조적인 연관성을 바탕으로 클러스터를 선정하는 방법을 택하였다. 즉, 각 피어의 온톨로지가 지니는 스키마내의 클래스, 프로퍼티, 값 등에 대한 피어들간의 관계를 파악하여 피어들간에 더욱 높은 유사도가 성립되면 클러스터에 추가하는 방식을 택하였다.

구성된 클러스터내에서의 질의 처리를 위하여 피어인덱스를 구성하는 방법과 클러스터간의 메시지 전송을 위하여 인터페이스 피어를 선정하고 클러스터인덱스를 유지하는 방법을 제시하였다. 클러스터 인덱스는 각 피어마다 각각 유지하는 피어인덱스 방식을 선택하여 질의 처리 특성상 각피어의 부하가 전체 시스템 성능에 영향을 줄 가능성을 줄이고자 하였다.

현재 제안된 시스템의 성능을 평가하기 위하여 P2P 환경에서의 온톨로지를 구축하고 있으며, 이를 통하여 클러스터 구성방법에 따른 효율성을 비교하고, 클러스터내에서의 질의 처리를 위하여 피어인덱스를 구성하는 방법과 다른 연구들과의 정량적인 성능 비교를 수행할 예정이다.

지금까지의 온톨로지 통합에 관한 연구들은 스키마 통합 측면에서의 연구가 주로 진행되었으나, 본 연구는 분산환경에서의 질의를 P2P 방식으로 처리하도록 하였고, 효율적인 피어들의 관리와 질의처리가 가능하도록 구성하였다. 이 연구방법을 이용하여 웹2.0 시스템 구성에 적용 가능하며 차세대 검색시스템의 P2P 처리에 적용할 수 있다.



참고문헌

[1] Christophides, V. et al, "The ICS-FORTH SWIM: A Powerful Semantic Web Integration Middleware", In Proceedings of the First International Workshop on Semantic Web and Databases (SWDB), Co-located with VLDB 2003.

[2] G. Karvounarakis, et al, "Querying the Semantic Web with RQL", Computer Networks, pp. 617-640, August, 2003.

[3] Longbing Cao, Dan Luo, Chao Luo, and Li Liu, "Ontology transformation in multiple domains", LNAI3339, pp. 985-990, 2004

[4] Huiyong Xiao and Isabel F. Cruz. "Ontology-based Query Rewriting in Peer-to-Peer Networks.", In Proceedings of the 2nd International Conference on Knowledge Engineering and Decision Support (ICKEDS 2006), pp. 11-18, 2006.

[5] RDF Vocabulary Description Language 1.0 : RDF Schema, <http://www.w3.org/TR/rdf-schema>.

[6] W3C, OWL Web Ontology Language Overview, <http://www.W3C.org/TR/owl-features/>

[7] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A scalable content addressable network. In Proceedings of the 2001 Conference on applications, technologies, architectures, and protocols for computer communications. ACM Press New York, NY, USA,

[8] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In Proceedings of the 2001 Conference on applications, technologies, architectures, and protocols for computer communications. ACM Press New York, NY, USA, 2001.

[9] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmér, and T. Risch. EDUTELLA: a P2P Networking Infrastructure based on RDF. In WWW 11 Conference Proceedings, Hawaii, USA, May 2002.

[10] K. Arabshian, H. Schulzrinne, D. Trossen, and D. Pavel. Gloserv: Global service discovery using the owl web ontology language. In The IEE International Workshop on Intelligent Environments, University of Essex, Colchester, UK, June 2005.

[11] W Nejdl, M Wolpers, W Siberski, C Schmitz, M Schlosser, I Brunkhorst, A Lser, "Super-peer-based

routing and clustering strategies for rdf-based peer-to-peer networks", WWW2003, May 2003, Budapest, Hungary.



김 병 군

1990년 : 홍익대학교 공과대학 전  
자계산학과 이학사

1992년 : 홍익대학교 공과대학 전  
자계산학과 이학석사

2001년 : 홍익대학교 공과대학 전  
자계산학과 이학박사

1992년~1998년 : 국방과학연구소 연구원

2001년~현재 : 부천대학 e-비즈니스과 조교수

관심분야 : 다차원 인텍싱, 데이터웨어하우스, 시맨틱 웹 등



오 성 군

1981년 : 홍익대학교 이공대학 전  
자계산학과 이학사

1984년 : 연세대학교 산업대학원  
전자계산학과 공학석사

1999년 : 홍익대학교 공과대학 전  
자계산학과 이학박사

1987년~현재 : 서일대학 소프트웨어과 교수

관심분야 : 능동데이터베이스, XML모델링, 소프트웨어공학