분산 환경에서 RDF를 이용한 효율적인 파일 검색 시스템

엄영현°, 국윤규, 정계동, 최영근
광운대학교 컴퓨터과학과
{class76°, ykkook, gdchung, ygchoi}@kw.ac.kr

An Efficient File Retrieval System using RDF in Distributed Environment

YoungHyun Eum° YounGyou Kook, GyeDong Chung YoungKeun Choi
Dept of Computer Science, Kwang-woon University

요 약

컴퓨팅 환경과 멀티미디어 환경이 발전함에 따라 P2P 파일 공유 시스템이 증가하고 있다. 각각의 P2P 파일 공유 시스템들은 서로 다른 애플리케이션을 제공하며 여러 가지 기준에 따라 그 특징 및 장단점을 가지고 있다. 같은 속성의 파일을 공유하는 서비스에서는 상관없지만, 일반적으로 모든 형식의 파일을 공유하는 서비스에서는 사용자의 기호에 따라 분산된 peer마다 사용하는 애플리케이션이 다르기 때문에 같은 그룹 peer에서 원하는 파일이 없을 경우 다른 그룹 peer에 대한 검색이 이루어지는 불편함이 있다. 따라서 본 논문은 웹을 통한 파일 공유 서비스를 제시하여 파일 공유에 대한 상호 운용성을 높이고, 저장 방식 및 검색 성능을 향상시키기 위하여 RDF를 이용한 효율적인 파일 검색 시스템을 제안한다.

1. 서 론

컴퓨팅 환경과 WWW(World Wide Web)의 출현으로 인터넷이 활성화됨에 따라 웹 중심의 서비스를 제공하는 서버와 이 서비스를 이용하는 클라이언트 환경이 발전하였지만 기존의 서버 중심의 인터넷 환경에서는 클라이언트의 자원을 효율적으로 활용할 수 있는 방법이 부족했다. 그래서 이러한 환경에 맞추어 클라이언트 기반의 파일 공유 시스템이 등장하였으며 애플리케이션을 통한 서비스가 제공되고 있다.

P2P 파일 공유 서비스는 같은 형식의 파일 공유[5][6]에선 효과적이지만, 일반적으로 여러 형식의 파일 공유에선 사용자의 기호에 따라 분산된 peer마다 사용하는 애플리케이션이 다르기 때문에 같은 그룹 peer에서 원하는 파일이 없을 경우 다른 그룹 peer에 대한 검색을 해야함으로 사용자 인증 및 애플리케이션의 설치를 해야만 하는 불편함이 있다.

본 논문에서는 기존의 애플리케이션 기반 파일 공유 서비스[7][8]를 웹 서비스 모델로 제시하고, RDF를 이용하여 각각의 웹사이트에 저장된 공유 파일 데이터베이스 정보를 통합하고 검색하는 시스템을 제안한다.

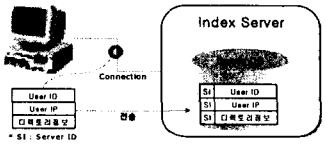
파일 공유 서비스를 하기 위한 웹사이트는 각각의 클라이언트의 공유된 데이터에 대한 디렉터리 및 파일에 대한 메타 정보를 웹사이트의 데이터베이스에 저장 하고, 파일 검색 시스템은 메타 정보를 처리 기반으로 사용되는 RDF(Resource Description Framework)[1]를 이용하여 각 사이트가 제공하는 메타 정보들을 XML 데이터베이스로 구성하여 검색 서비스를 제공하는 기법을 제시한다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구에 대해 설명한다. 3장은 파일 검색 시스템에 대해 설명한다. 4장은 시스템 평가를 설명하고 마지막으로 5장에서 결론을 맺는다.

2. 관련 연구

2.1 P2P 서비스 시스템

P2P 서비스 시스템들간에 상호운용성을 높이기 위하여 웹을 통한 파일 공유 서비스 서버는 인덱스 서버이다. peer는 사용자 인증을 위한 ID와 접속한 IP와 공유 디렉토리내의 파일에 대한 메타정보를 인덱스 서버의 데이터베이스에 저장하는 것을 모델로 한다. peer들은 네트워크에 접속을 하였을 시에 일정시간 간격으로 자신의 상태정보(KeepAlive)를 서버에게 통보하고, 사용자 ID, 접속 IP, 디렉토리내의 파일 정보들을 데이터베이스에 저장을 한다. 인덱스 서버의 그림은 다음과 같다.

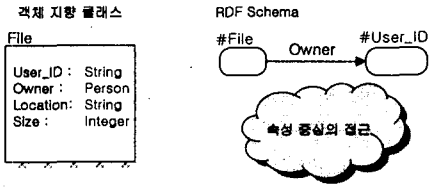


[그림 1] P2P 인덱스 서버

2.2 RDF(Resource Description Framework)

RDF(Resource Description Framework)는 XML의 문제점을 해결하고 의미에 초점을 맞추기 위해 제시된 기반구조이다. RDF는 자원을 기술하는 속성명이 고유한 URI로 표현되어 각 네임스페이스에서 정의된 속성을 사용함으로써 의미충돌을 예방할 수 있다. 또한 기술 내용이 간단하기 때문에 자원을 기술하고 검색하는 것이 쉽고, 웹 자원을 자동적으로 분류할 수 있어 검색 능력을 향상시킬 수 있다.

RDF Schema[2]는 RDF에 쓰여진 어휘들의 정의를 위한 메타언어를 역할하며, RDF 문장에서 쓰이는 어휘 사이의 관계를 의미적으로 정의 내리기 위해 사용한다. 속성과 클래스의 정의, 클래스와 클래스 간의 관계, 속성과 속성 간의 관계 등을 정의할 수 있는 언어이며, 사람이 이해하는 동시에 기계 처리가 가능한 형태로 메타데이터의 속성과 클래스 간의 관계 표현을 해결한다. 객체 지향 클래스와 RDF Schema와의 관계 그림은 다음과 같다.



[그림 2] 객체 지향 클래스와 RDF Schema

객체 지향 클래스의 표현 방식을 RDF Schema는 Property중

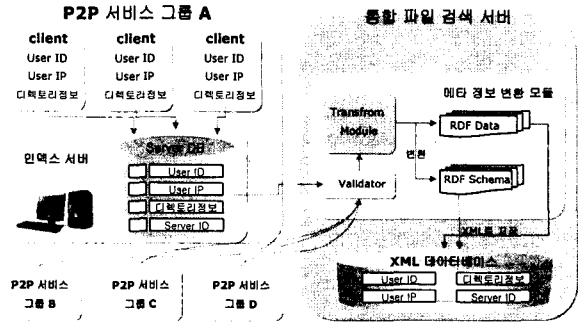
심으로 각 속성의 특성을 정의하고 제한하는 기능을 하기 때문에, 같은 용어가 다른 의미로 쓰이는 중복성을 막기 위해 XML의 Namespace 기법을 이용한다.

그러므로, RDF Schema는 사람이 이해하는 동시에 기계 처리가 가능한 형태로 메타데이터의 속성과 클래스 간의 관계 표현을 해결한다.

3. 파일 검색 시스템

3.1 시스템 구조

본 논문에서 제안하는 RDF를 이용한 파일 검색 시스템의 구조는 그림3과 같다.



[그림 3] 전체 시스템 구성도

3.1.1 메타 정보 변환 모듈

메타 정보 변환 모듈은 각각의 사이트에 존재하는 서로 다른 메타 DB 정보들을 RDF 데이터 및 RDF Schema로 기술하기 위하여 XML형식으로 변환하기 위한 모듈이다.

각각의 특성을 가진 여러 P2P 서비스 그룹의 메타 DB 정보들은 peer의 접속과 동시에 통합 파일 검색 서버의 validator로 보내진다. 메타 DB 정보의 형식은 파일, XML, 레코드 셋의 3가지로 분류하고, 파일 및 레코드 셋의 경우 RDF 로 기술하기 위하여 XML로 바꿔야 하므로, 이러한 사실을 변환 모듈로 전송한다. 이러한 변환 및 생성 과정은 다음의 알고리즘에서 보여준다.

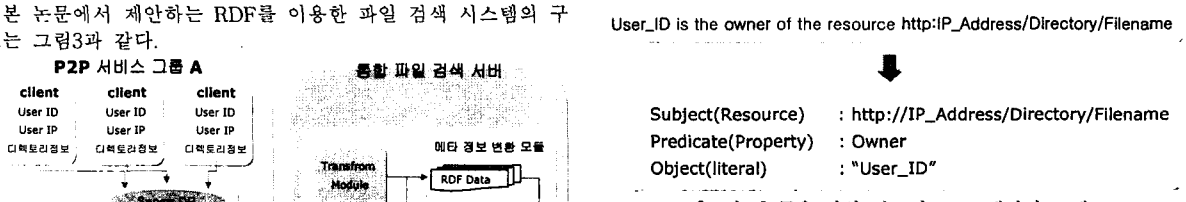
```
/* validate 객체 생성 및 호출 */
Create Validate();
Call Validate();
/* validate 안에 메타데이터를 읽음 */
Read Validate(metadata);
/* 입력 메타데이터 */
/* Default */
if (metadata.type = XML) {
  Make metadata to RDF Data Model in XML Database;
} /* File Type */
else if (metadata.type = File) {
  Transform metadata to XML;
  Make XML to RDF data Model in XML Database;
} /* Record Set */
else if (metadata.type = Record_Set) {
  Transform metadata to XML;
  Make XML to RDF data Model in XML Database;
}
```

[리스트 1] 메타 DB 정보 변환 알고리즘

3.1.2 RDF 데이터 모델

RDF 데이터 모델[1]은 객체들의 관계를 정의하는 모델로서 Resource, Property, Statement의 triple로 구성된다. 각각의 리소스들 간에는 관계를 설정 가능하며 URI를 사용하여 식별 가능한 대상들로 문장 구성하며, 구성된 문장은 노드와 화살표의 그래프로 표현가능하다.

제안된 P2P 서비스 시스템에서 메타 정보는 사용자 ID,접속 IP, 공유된 디렉토리 파일정보인데 이를 RDF 데이터 모델로 표현하면 다음 그림4와 같다.



[그림 4] 공유 파일 정보의 RDF 데이터 모델

RDF 데이터 모델은 XML을 이용해 구문을 표현하여 기계가 이해할 수 있는 형태로 처리된다. 그림 6의 모델을 XML 구문으로 기술하면 리스트 2와 같다.

```
<?xml version ="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<rdf:Description
      about ="http://IP_Address/Directory/Filename">
    <Owner>User_ID</Owner>
    </rdf:Description>
</rdf:RDF>
```

[리스트 2] RDF 데이터 모델의 구문

각 P2P 서비스에서 인덱스 서버의 DB 메타 정보에서 UserID를 하나의 의미 정보로 통일하기 위한 RDF schema 클래스 표현은 다음 리스트3과 같다.

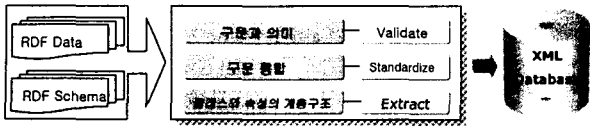
```
<rdf:RDF  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
<rdf:Description rdf:ID="UserID">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
<rdfs:subClassOf
 rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
</rdf:Description>
<rdf:Description rdf:ID="UserSerial">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
<rdfs:subClassOf rdf:resource="#UserID"/>
</rdf:Description>
<rdf:Description rdf:ID="Username">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
<rdfs:subClassOf rdf:resource="#UserID"/>
</rdf:Description>
</rdf:RDF>
```

[리스트 3] UserID에 대한 RDF schema 클래스 표현

UserID뿐만 아니라, 서로 다른 DB의 메타 정보에 대한 여러 가지 표현을 RDF schema 클래스로 의미 있는 정보로 표현함으로써, 효율적인 파일 검색을 할 수 있도록 해준다.

3.2 XML 데이터베이스 시스템

본 논문에서 제안된 RDF 데이터를 저장 관리하는 시스템으로 XML 데이터베이스를 이용한다. 이 시스템은 RDF description 데이터와 RDF schema 데이터가 들어오면 구문과 의미에 대한 검증 및 여러 형태의 RDF/XML 구문[3]을 저장 공간의 효율성과 검색의 용이성을 위해 하나의 구문으로 통합하는 작업을 수행한다. 그리고 RDF Schema 데이터의 클래스와 속성의 계층 구조 및 속성의 domain, range 제약에 대한 정보를 뽑아서 별도의 XML 문서로 저장하여 검색의 성능을 향상시키기 위해 사용되며 그림은 다음과 같다.

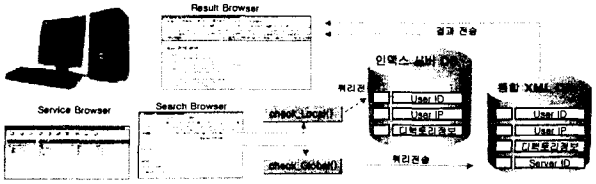


[그림 5] RDF Data & RDF Schema 저장 과정

또한 RDF 질의 언어인 RQL[4]을 이용하여 RDF Description 데이터뿐만 아니라, RDF Schema 데이터도 검색할 수 있도록 함으로써, RDF 데이터 모델에 맞게 원하는 데이터를 쉽게 검색할 수 있다.

3.3 검색 알고리즘

검색 서비스는 하나의 서비스 그룹내의 인덱스 서버 데이터베이스 검색을 하는 로컬검색과 로컬에서 파일 찾지 못한 경우, 통합 파일 검색 시스템의 데이터베이스를 검색하는 고급검색으로 나누어 서비스하며 그림과 알고리즘은 다음과 같다.



[그림 6] 검색 과정

```
/* Local 객체 생성 */
Create localCheck;
Create localQuery;
/* Global 객체 생성 */
Create globalCheck;
Create globalQuery;
/* query 선택에 따른 객체 생성 및 query 전송 */
/* localQuery 전송 */
if (localCheck = CHECKED) {
   Transfer localQuery to Local Database;
} /* globalQuery 전송 */
else if (globalCheck = CHECKED) {
   Transfer globalQuery to XML Database; }
```

[리스트 4] query 전송 알고리즘

4. 시스템 평가

각 P2P 서비스는 서로 다른 어플리케이션 및 이기종 DB 사용으로 상호 운용성이 떨어지며, 검색 성능에 있어서 찾고자 하는 파일을 얻지 못하는 경우가 많지만, 본 논문에서 제안한 시스템에서는 웹 사이트를 통한 서비스로 상호운용성을 높이고, RDF를 이용한 통합 파일 검색 시스템을 통해 서로 다른 DB간에도 의미성을 통일하여 검색성능을 향상시켰다는 강점이 있다.

[표 1] 각 P2P 서비스 별 비교분석

구 분	Hybrid 형 P2P 시스템	순수형 P2P 시스템	본 시스템
어플리케이션 유/무	유	유	무
네크워크 견고성	견고하지 못함	견고함	견고함
상호 운용성	불가능	불가능	가능
검색 효율성	조금 효율적	비효율적	효율적

5. 결론

본 연구에서는 분산 환경에서 RDF를 이용한 효율적인 파일 공유 검색 시스템을 제안하였다. P2P 서비스 애플리케이션을 웹사이트를 통해 서비스하고 각각의 웹사이트에 저장된 공유 파일에 대한 메타 DB 정보를 RDF 데이터 모델로 기술하고, 저장방식의 효율성과 검색을 향상시키기 위하여 XML 데이터베이스 시스템을 사용하였으며 RDF Schema를 이용하여 메타 정보에 대한 의미의 중복성을 피함으로써 검색의 효율성을 제공한다.

향후 과제로는 웹을 통한 서비스 뿐만 아니라, 어플리케이션 기반의 서비스도 효율적인 검색이 가능한 프레임워크 연구가 이루어져야 할 것이며, 분산된 이기종 메타 정보의 논리적인 추론이 가능하게 하는 RDF와 RDF Schema를 기반으로 두 언어에 부족한 모델링 요소를 강화하는 연구가 필요하다.

6. 참고 문헌

[1] O. Lassila and Ralph Swick, "Resource Description Framework(RDF) Model and Syntax Specification," W3C Recommendation, Feb. 1999; available online at http://www.w3.org/TR/REC-rdf-syntax/.

[2] Dan Brickley, R.V. Guha, Editors, Resource Description Framework(RDF) Schema Specification 1.0, World Wide Web Consortium Candidate Recommendation, 27 March 2000. This version is http://www.w3.org/TR/2000/CR-rdf-schema-20000327/. The latest version of RDF Schema is http://www.w3.org/TR/rdf-schema/.

[3] D. Beckett. RDF/XML Syntax Specification (Revised). W3Cp Working Draft 20

[4] G. Karvounarakis, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl. RQL:A Declarative Query Language for RDF. WWW20

[5] 소리바다, http://www.soribada.com/

[6] Napster. http://napster.com.

[7] Gnutella, http://gnutella.wego.com.

[8] eDonkey2000, http://www.edonkey2000.com.