

# XML을 이용한 CNRP기반의 통합 검색시스템 설계

홍승우<sup>o</sup> 유영호 김형수 김정석  
부산대학교 전자계산학과  
(swhong, yhyou, hskim, gimgs)@asadal.cs.pusan.ac.kr

## Design of Integrated Search System based on CNRP Using XML

Seung-Woo Hong<sup>o</sup>, Yong-Ho Ryu, Hyung-So Kim, Kyung-Suk Kim  
Dept. of Computer Science, Pusan University

### 요약

웹이 빠른 속도로 증가함에 따라 사용자들은 점차 원하는 리소스를 얻기가 힘들어지고 있다. 비록 기존의 검색엔진이 도움을 주고 있지만, 하나의 서버가 방대하고 다양한 종류의 웹 리소스들을 모두 구분해 서비스하는 것은 점차 그 한계를 드러내고 있다. 본 논문에서는 각 검색서버들이 분류화된 이름공간내에서 검색서비스를 하고 이들을 하나의 질의 인터페이스로 통합할 수 있는 방안을 제안하고, 그 시스템을 설계하고자 한다. 제안된 시스템은 XML을 사용하는 CNRP(Common Name Resolution Protocol)을 기반으로 하여 여러 검색 서버의 결과를 쉽게 통합하고, 검색된 데이터를 일반 응용프로그램에서도 사용할 수 있으며, 동적으로 생성되는 질의 인터페이스를 통해 어떠한 검색 시스템으로도 확장이 가능하다.

### 1. 서론

웹이 방대해지고, 웹 리소스에 대한 URI(Uniform Resource Identifier)가 갈수록 복잡해짐에 따라 사용자들은 복잡하고 친숙하지 않은 URI 보다는 일상생활을 통해 쉽게 연상되는 일반적인 단어, 즉 회사이름, 브랜드 이름, 제품 이름, 사람 이름, 책제목 등의 보통이름(Common Name)을 통해 원하는 웹 리소스를 얻고 싶어한다[1,5]. 일반적으로 기존의 검색엔진이 이러한 단어들을 통해 연관된 웹 리소스를 찾을 수 있도록 서비스하고 있지만, 단순히 텍스트 내용을 기반으로 하기 때문에 검색결과가 기대한 것이 아닐 경우가 많거나 결과가 너무 많은 등, 방대하고 다양한 종류의 웹 리소스를 하나의 서버가 모두 분류하고 서비스하기에는 그 한계를 드러내고 있다. 또한 검색된 결과는 검색결과에 불과해 다른 형태의 데이터로 전혀 재 사용할 수 없다[4].

따라서 본 논문에서는 여러 검색 서버가 각각의 전문화된 이름공간(Namespace)내에서 검색 서비스를 하고 이를 하나의 질의 인터페이스로 통합할 수 있는 시스템을 제안한다. 제안하는 시스템은 메타프로퍼티기반의 검색 프로토콜인 CNRP를 이용하여 통신하며, XML을 기반으로 하기 때문에 결과 통합이 용이할 뿐만 아니라 질의 결과를 일반 응용프로그램에서도 사용할 수 있다. 또한 서버의 서비스 스키마에 따라 동적으로 생성되는 질의 인터페이스를 통해 사용자 하여금 편리함을 제공하고 어떠한 검색시스템에 대해서라도 쉽게 확장이 가능하다. 본 시스템을 이용하여 전문화된 분야에서 원하는 리소스를 정확하게 검색할 수 있도록 함으로서 구조화된 통합서비스를 제공할 수 있을 것이다.

### 2. 관련연구

#### 2.1 CNRP(Common Name Resolution Protocol)

CNRP는 최근 IETF(Internet Engineering and Task Force)에서 제안된 것으로 단순히 보통이름(Common Name)을 웹 리소스의 URI로 풀이(Resolution)하기 위한 서버와 클라이언트간의 재해교환 프로토콜을 명시하고 있다[2,3]. 프로토콜 명세는 XML DTD로 정의되어 있으며, 서버와 클라이언트는 명시된 XML 태그로 데이터를 CNRP 객체화하고 이를 서로 주고받음으로서 통신하게 된다. 이러한 객체들은 해당 리소스에 대한 정보를 기술하는 메타프로퍼티를 포함할 수 있어 보통이름을 웹 리소스로 풀이할 때 더욱 정밀하게 해 준다. 즉 CNRP는 메타 데이터를 이용한 간단한 검색 프로토콜이라 할 수 있다.

#### 2.2 CNRP 객체모델

다음은 CNRP에서 사용되는 객체들이다.

- <servicequery>
- <query> : <id>,<commonname>,<property>\*
- <result> : <service>\*,<resource>\*,<referral>\*,<error>
- <service> : <serviceURI>,<servers>,<description>,<property>\*,<propertieschema>,<queryschema>,<resourceschema><serviceschema>
- <resource> : <commonname>,<id>,<resourceURI>,<description><property>\*

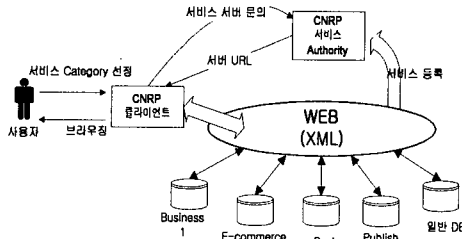
CNRP서버와 클라이언트는 질의/응답의 메커니즘을 사용하며, 질의종류에는 초기 서비스질의와 일반 질의가 있다. 즉 클라이언트에서 서버에 서비스 초기화 질의<servicequery>를 보내고 서버는 <service> 객체에 질의에 필요한 스키마 정보를 보낸다. 스키마

정보를 받은 클라이언트는 그 스키마 정보를 참조하여 해당 서버에 적합한 질의를 만들어 보내고, 서버는 해당 결과 <resource>를 반환하는 것이다. 본 논문에서 제안하는 시스템은 이러한 CNRP 질의/응답 메커니즘을 응용할 것이다.

### 3. CNRP 기반 통합 검색 시스템 설계

#### 3.1 시스템구조 및 검색과정

사용자에게 검색서비스를 제공하기 위해 클라이언트, CNRP서비스등록소(ServiceAuthority) 그리고 서버들이 상호통신하며 [그림 1]과 같은 구조를 가진다. CNRP서비스등록소는 서비스 서버의 목록을 관리하는 역할을 하며, 클라이언트는 사용자에게 질의 및 서버로부터 제공받은 검색결과를 브라우징할 수 있는 인터페이스를 제공한다.



[그림 1] CNRP기반 통합 검색 시스템 구성도

다음은 [그림 1]과 같은 구조에 따라 전체 시스템이 동작하는 검색과정을 나타내었다.

- ① 각 검색서버들은 CNRP서비스등록소에 자신들의 서비스 종류를 등록한다.
- ② CNRP서비스등록소는 이들의 서비스종류를 분류하여 종류(Category) 목록을 만들고 각 서버에 대한 ServiceURI를 관리한다.
- ③ 클라이언트는 최초 실행시 CNRP서비스등록소로부터 종류 목록을 가져와 이를 사용자에게 브라우징한다.
- ④ 사용자가 원하는 종류를 선택하면, 클라이언트는 CNRP 서비스등록소로부터 ServiceURI를 가져오고 해당 서버들에게 스키마정보를 위한 <servicequery/>를 보낸다.
- ⑤ 각 해당 서버는 자신들에게 질의시 필요한 항목에 대한 스키마 정보를 <service> 객체에 담아 클라이언트에 전송한다.
- ⑥ 스키마정보를 받은 클라이언트는 이를 통합하여 질의에 필요한 사용자 폼을 생성한다.
- ⑦ 사용자는 폼을 작성하고, 클라이언트는 이를 <query> 객체에 담아 각 서버로 전송한다.
- ⑧ 서버들은 질의에 대한 리소스를 검색하고 그 결과를 <result> 객체에 담아 클라이언트로 전송한다.
- ⑨ 클라이언트는 결과를 통합하고 사용자에게 브라우징한다.

#### 3.2 메타 프러퍼티의 확장

CNRP의 질의 및 결과 오브젝트는 프로퍼티들의 (<property>) 리스트로 구성되어 지며, 아래와 같은 몇 가지의 기본적 프로퍼티를 제시하고 있지만, 구현상의 유연성을 위해 확장이 가능하도록 하였다. 따라서 본 논문에서는 다음과 같이 기본적인 프로퍼티의 타입을 확장하고 또한 새로운 프러퍼티를 정의한다.

[표 1] 메타프로퍼티의 확장

	이름	기본타입	확장타입
CNRP 기본	language	RFC1766	
	gcography	ISO3166-1, GPS, freeform. ISO6707	
프로퍼티	category	NAICS, freeform	valuereference
확장 프로퍼티	region	postcode, city, province, freeform	
	date	yy/mm/dd/HH/MM	
프로퍼티	price	us, kr, freeform	

#### (1) 확장 정의된 프로퍼티

기본 프로퍼티만으로는 검색 시스템에 적용이 어려우므로 [표 1]과 같이 region, date, price 의 새로운 프로퍼티와 그 기본타입을 정의한다. 다음은 그 사용 예이다.

```
<query> ----- (보기 1)
<commonname>극장</commonname>
<property name="region" type="city">부산</property>
</query>
```

#### (2) "valuereference" 타입

서비스의 종류에 따라 프로퍼티에 대한 고유한 값의 범위를 가지고자 할 때 사용될 수 있다. 이때는 서비스스키마에 사용될 수 있는 값의 범위를 명시해야 한다. 클라이언트는 이를 질의의 폼에 선택리스트로 나타낼 것이다.

```
<service> ----- (보기 2)
.....
<propertySchema>
  <propertyDefinition id="1">
    <propertyName>category</propertyName>
    <propertyType>valuereference</propertyType>
    <valueList>도서,CD,MP3</valueList>
  </propertySchema>
</service>
```

#### (3) 기타 사용자정의 프로퍼티

- 각 서버의 서비스에 따라 자신만의 프로퍼티를 만들 수 있지만, 값의 타입은 "valuereference"이거나 단순텍스트로 처리하는 "freeform" 또는 숫자와 같이 잘 알려진 것이어야 한다.

### 3.3 CNRP 서비스 등록소

CNRP서비스등록소는 검색서버의 서비스 종류에 따라 보통이름의 이름공간을 분류 및 관리하고 서비스 서버의 URI를 클라이언트에 제공한다. 다음 [표 2]에 가능한 서비스분류를 정의하였으며 서버가 다양해질 수록 확장해 나갈 것이다.

[표 2] 보통이름 이름공간 분류

이름공간	보통이름	제공가능 리소스
비즈니스검색	회사이름, 상품이름	URI, e-mail, 주소, 주요상품 등
개인정보검색	성명, 전화번호, e-mail주소	적원정보, 개인정보 등
쇼핑몰상품검색	상품명, 상표이름, 상품종류 등	제공 가능한 모든 상품
출판물검색	출판물타이틀, 저자 및 가수이름	도서, CD, MP3 등
유흥정보검색	극장, 영화, 가수이름, 각종 이벤트	영화,극장,권서트,TV쇼 등
음식주점검색	각종 음식점이름	주변가능 음식점
뉴스기사검색	뉴스헤드라인	뉴스기사

### 3.4 CNRP 검색 클라이언트 및 서버 설계

아래의 [그림 2]는 CNRP를 기반으로 하는 보편이름 검색시스템의 클라이언트와 서버의 시스템 설계도이다.

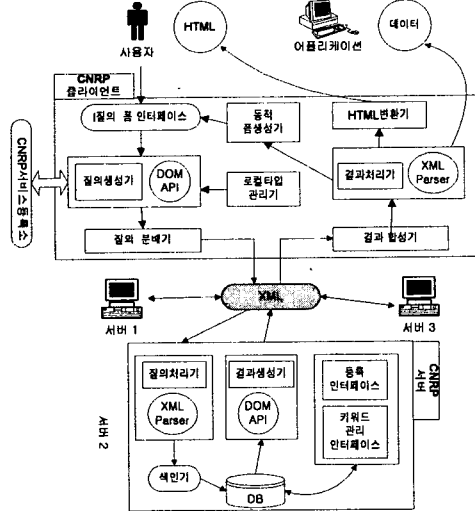


그림 2 CNRP기반의 검색 클라이언트/서버 구조

#### 3.4.1 CNRP 검색 클라이언트

##### (1) 질의 품 생성

질의생성기에 의해 전송된 <serviequery/>를 받은 각 서버가 스키마정보를 보내오면 결과합성기와 결과처리를 거쳐서 "동적 품 생성기"가 <service>객체내의 <propertySchema>를 참조하여 품을 생성한다. 3.2의 (보기 2)의 서비스 스키마를 받았다면 다음과 같은 질의 품을 생성할 것이다.

CommonName	조성모	Category	MP3
			CD
			도서

[그림 3] 품 생성 예 ]

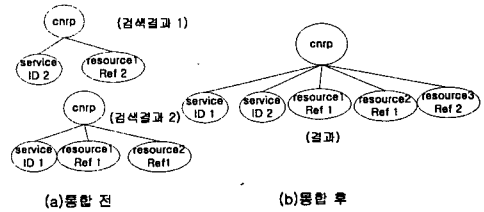
##### (2) 질의

폼에서 [그림 3] 같이 질의를 입력하면 "질의생성기"는 다음과 같은 질의를 생성하고 질의분배기는 생성된 질의를 각 해당서버로 전송한다.

```
<query> ----- (보기 3)
<commonname>조성모</commonname>
<property name="category" type="system">MP3</property>
</query>
```

##### (3) 질의결과 통합

질의를 받은 해당 서버들이 검색결과를 보내오면 결과 합성기에서는 검색결과 XML문서를 다음과 같이 DOM 트리를 사용하여 XML문서를 병합한다. 각 문서의 객체 트리는 최상위에 <cnrp>가 있고 자식 노드로 서버의 ID를 가지고 있는 <service>객체 하나와 검색결과를 나타내는 <resource>객체들이 있다. 각 <resource> 객체내의 <serviceRef ref="서버 ID">이 자신이 어느 서버의 검색결과인가를 나타낸다. 따라서 DOM 트리를 사용하여 [그림 4]와 같이 간단하게 통합할 수 있다.



[그림 4] 질의결과 XML통합 방법

#### 3.4.2 CNRP 검색 서버

[그림 2]에서 나타나는 것과 같이 서버에서는 XML로 받은 질의를 파싱하여 키워드를 추출한 후 데이터베이스에서 해당 리소스를 찾는다. 결과 리소스는 결과 생성기에 의해 CNRP 검색 오브젝트로 생성된 후 클라이언트로 전송한다. 즉 CNRP 질의 XML문서를 파싱하고 결과를 다시 CNRP 결과 XML문서로 만들 수 있는 모듈만 있으면, 어떠한 데이터베이스라도 CNRP 검색 서비스를 할 수 있다.

### 4. 결론 및 향후 연구방향

본 논문에서는 메타프로퍼티를 이용하는 검색 프로콜인 CNRP를 이용하여 여러 검색서버가 각각의 전문화된 이름공간내에서 검색서비스를 하고 이를 하나의 질의 인터페이스로 통합할 수 있는 시스템을 설계해 보았다. 제안하는 시스템은 전문화된 검색서버를 통해 사용자에게 정확한 리소스를 제공하고 또한 XML로 이루어진 질의결과는 데이터로서 사용할 수 있는 등 XML의 장점을 그대로 가질 수 있다. 뿐만 아니라 여러 검색 서버의 결과를 쉽게 통합할 수 있으므로 본 논문에서 제안한 시스템을 사용한다면 현재 웹에 무수히 흩어져 있는 전문화된 검색 시스템을 쉽게 통합할 수 있을 것이다.

향후 연구과제로서 실제로 시스템을 구현해 보고, XML로 얻어진 데이터를 라이브러리 형식으로 일반 어플리케이션 프로그램에 제공해 줄 수 있는 인터페이스를 설계해야 할 것이다.

#### 참고문헌

- [1] Nicolas Popp, Michael Mealling, Larry Masinter, Karen Sollins "Context and Goals for Common Name Resolution" draft-ietf-cnarp-goals-01, April 2000
- [2] Moseley, M., Mealling, M. and N. Popp, "CNRP PROTOCOL SPECIFICATION", Internet-Draft draft-ietf-cnarp-protocol-02, February 2000.
- [3] Mealling, M., "A URI Scheme for the Common Name Resolution Protocol", draft-ietf-cnarp-uri-02 December 1999.
- [4] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, BCP 14, March 1997.
- [5] RealNames Corporation "RealNames Internet Keyword Interface" http://www.internetkeywords.org May 2000.
- [6] 탁우현, 유영호, 이종환, 김경석, "멀티데이터베이스에서 XML을 이용한 스키마 통합" 한국정보과학회, 99 가을 학술 발표 논문집(1), vod 26(2)