

# 분산 자원 검색과 갱신을 위한 프로토콜에 관한 연구

## A Study on Protocols for Retrieval and Update of Distributed Resources

이지원, 한국교육학술정보원, jwlee@keris.or.kr

Lee Ji-Won, Korea Education & Research Information Service

### <초록>

본 연구는 분산자원 활용을 위한 Z39.50 프로토콜, ZING SRW/SRU 프로토콜, OAI 프로토콜, ZING Update 프로토콜의 개발 배경, 특징을 살펴보고, 검색과 갱신 각 기능별로 프로토콜을 비교·분석하였다. 본 연구는 분산자원을 활용하려는 도서관 및 정보서비스기관에 기초 정보를 제공하고, 각 기관에서 분산자원을 이용하여 구축하고자 하는 서비스의 대상, 종류, 환경에 적합한 프로토콜을 선택할 수 있도록 하기 위함이다.

## 1. 서론

전통적인 도서관은 물리적인 형태의 소장 장서를 중심으로 서비스를 제공하였으나, 네트워크 기술을 포함한 정보통신기술이 끊임없이 발전하면서 전자도서관, 디지털도서관, 유비쿼터스 도서관으로 그 모습을 변화하고 있으며 이용자에게 제공하는 장서와 자원의 대상과 종류를 지속적으로 확대하고 있다.

다양한 가상공간에 산재하는 많은 분산자원을 보다 효과적으로 활용하고 상호운용성을 보장하기 위하여 많은 프로토콜들이 만들어졌으며, 현재도 이와 관련된 연구와 프로젝트들이 계속 진행되고 있다.

본 연구는 분산자원의 검색과 갱신을 위한 프로토콜들의 개발 배경, 특징을 살펴보고 각 프로토콜을 비교·분석하고자 한다. 본 연구는 분산자원을 활용하려는 도서관 및 정보서비스기관에 기초 정보를 제공하고, 각 기관에서 분산자원을 이용하여 구축하고자 하는 서비스의 대상, 종류, 환경에 적합한 프로토콜을 선택할 수 있도록 하기 위함이다.

## 2. 프로토콜 개요

### 2.1 Z39.50 프로토콜

1980년대 초 LC(Library of Congress), OCLC(Online Computer Library Center), RLIN(Research Libraries Information Network), WLN(Western Library Network) 등이 상호간의 서지 레코드 검색과 레코드 전송을 위해 수행하였던 LSP(Linked Systems Project)를 기초로 개발된 Z39.50 프로토콜(이하 Z39.50)은 Z39.50-1988(제1판), Z39.50-1992(제2판)를 거쳐 1995년 7월 Z39.50-1995(제3판)가 발표되었고, 2005년 현재 최신판은 Z39.50-1995(제3판)를 개정한 Z39.50-2003이다. 이기간의 정보 검색을 위해 표준화된 통신규칙을 규정하고 있는 Z39.50은 미국 국가표준이자 국제표준(ISO 23950)으로서, Z39.50 클라이언트가 작동하는 특정 컴퓨터에서 서버 역할을 하는 다른 컴퓨터에 탐색 명령어를 보내고 그 결과를 제공받는 방식의 분산형 클라이언트/서버 환경의 정보검색을 지원하는 것이다.

Z39.50 프로토콜은 총 11가지의 기능을 정의하고 있는데, 여기에는 탐색(Search), 검색

(Retrieval), 정렬(Sort), 브라우즈(Browse) 등 정보 검색을 위한 기능 및 검색과 관련된 확장서비스(Extended-Services) 기능이 포함된다.

Extended-Services 기능은 Z39.50-1995(제3판)에 추가되었는데, 이는 데이터베이스의 탐색이나 검색과 직접 관련된 것은 아니지만 일반적으로 정보시스템에서 지원되는 기능들을 지원하기 위한 것이다. 결과 집합 저장(Persistent Result Set), 자료 주문(Item Order), 데이터베이스 갱신(Database Update) 등 7가지 태스크(task)의 확장 서비스가 정의되어 있다(National Information Standards Organization 2003).

Z39.50 프로토콜은 지난 30여년 가까이 수많은 도서관, 정보서비스 기관에서 외부 기관의 자원을 검색하고 활용하는데 적용되어 왔다. 국내에서는 국립중앙도서관, 한국교육학술정보원 등에서 Z39.50 서버를 운영하고 있으며, 주요 상용 도서관시스템에 Z39.50 클라이언트가 포함되어 있다.

## 2.2 ZING SRU/W 프로토콜

Z39.50 프로토콜이 오랜 기간 널리 사용해 온 안정적이고 완성된 프로토콜인 것은 확실하지만, 최근 웹 환경이 되면서 많은 사람들이 Z39.50 프로토콜 구현을 위한 기술 적용에 있어 최신성과 편리성이 뒤떨어진다고 생각해 왔다. 따라서 웹 기반 기술을 사용하는 개발자 및 정보제공자에게는 Z39.50의 유용성이 크게 인정을 받지 못하였고, 이러한 한계를 극복하고자 시작된 것이 ZING(Z39.50 International: Next Generation)이다. ZING의 공식 홈페이지에서는 “ZING은 Z39.50의 지적/의미적 내용을 좀 더 광범위하게 이용할 수 있도록 하고, 정보제공자, 개발자, 업체, 이용자들이 Z39.50에 보다 관심을 가질 수 있도록 하기 위해 수행되는 주도적인 성격의 여러 활동(initiatives)을 포함하는데, 이러한 활동들은 구현상의 어려움을 줄이고 현재의 Z39.50이 기여한 지적 개념은 보존하면서 진행된다”라고 밝히고 있다. 현재 5가지(SRW/SRU, CQL, ZOOM, ez3950,

ZeeRex) 분야의 연구 및 프로젝트가 수행 중에 있다.

SRW/SRU(Search/Retrieve for the Web) 프로토콜(이하 SRW/U)은 ZING 중에서 가장 활발하게 연구되고 있는 주제이며, 현재 여러 기관에서 실험적인 구현을 하고 있다. 또한 SRW/U는 ZING의 다른 영역인 CQL과 ZeeRex도 활용하고 있다. 2002년 11월 제1판이 발표되었고, 2004년 2월 제1.1판으로 개정되었다.

SRW와 SRU는 모두 데이터베이스를 탐색하고 그 결과를 보내기 위한 웹 기반 프로토콜이다. 두 가지 프로토콜에 사용되는 요청(request)과 응답(response) 파라미터들은 거의 동일하나, 클라이언트와 서버 응용 프로그램간 교환되는 질의와 결과 형식에서 차이가 있다. SRW(Search/Retrieve for the Web Service)는 SOAP(Simple Object Access Protocol) 방식의 웹 서비스 프로토콜이고, SRU(Search/Retrieve for URL Service)는 REST(REpresentational State Transfer) 방식의 웹 서비스 프로토콜이다(Morgan 2004; Sanderson 2004).

SRW/U는 Search/Retrieve, Scan, Explain 세 가지의 기능(Operation)을 제공하며, 각 기능은 요청과 응답에 대한 파라미터들이 정의되어 있다(SRW Editorial Board 2004c).

LC, OCLC, 네덜란드 국립도서관(National Library of the Netherlands) 등에서 SRW 또는 SRU 프로토콜을 통하여 데이터베이스 검색을 시험하였고, 특히 OCLC에서는 관련 소프트웨어와 소스를 다운로드받을 수 있게 공개하였다.

## 2.3 OAI 프로토콜

OAI-PMH(Open Archives Initiative-Protocol for Metadata Harvesting; 이하 OAI 프로토콜)은 개방 아카이브(archives)의 메타데이터를 수집(harvesting)하기 위한 프로토콜이다. 아카이브는 자원 특히 디지털 자원의 저장소를 의미하며, 리

포지토리(repository)라고 부르기도 한다.

1990년 초부터 등장한 다양한 아카이브는 주로 디지털 형태 학술논문의 인쇄전 자료(pre-prints)나 인쇄후 자료(post-prints)를 주제별 혹은 기관별로 저장, 관리하고 이용자에게 서비스를 제공하였다. OAI 프로토콜은 이러한 여러 아카이브들을 통합하여 서비스할 수 있도록 하는 기술적인 기반이라 할 수 있다. 1999년 산타페 회의에서 논의되기 시작한 OAI 프로토콜은 2001년 1월 제1판이 발표되었다. 제1판의 주요 특징은 6가지 명령어 집합과 기본 DC(unqualified Dublin Core)를 표준 형식으로 채택한 점이다. 2002년 6월 제2판이 발표되었고, XML 스키마의 채택, 프로토콜과 HTTP 표준과의 역할 구분 등의 개선이 이루어졌다(이수상 2004, 221-224).

OAI 프로토콜을 적용하고 있는 대표적인 해외 시스템으로는 영국의 EPrints(<http://www.eprints.org>), 미국 MIT의 Dspace(<http://www.dspace.org>)를 들 수 있으며, 국내에서는 2003년부터 KERIS에서 개발·운영중인 dCollection(국가 지식생성 및 유통 체계. <http://www.dCollection.net>)이 있다.

## 2.4 ZING Update 프로토콜

SRW/U 프로토콜이 개발되면서, 추가적인 기능들이 제안되고 이를 반영한 개선 작업들이 지속적으로 진행되었다. 이러한 활동 가운데 갱신(Update) 기능에 대한 필요성이 제기되었고, 기존의 관련된 기술들(WIKI, ATOM, REST/CRUD, WSRS, XML update)을 검토한 결과 적합한 것이 없다는 결론을 내리고, 새로운 프로토콜을 개발하기로 하였다.

2004년 6월 ZING update 프로토콜(이하 ZING update) 초안이 제안되었으며, 같은 해 10월 SRW Editorial Board Meeting에서 검토되었다. 2005년 2월부터 OCLC와 OCLC PICA에서 프로파일을 작성하고 있으며, NCC(Nederlandese

Centrale Catalogues)와 OCLC WorldCat의 상호운용성을 확보하는 방안으로 적용할 계획이다(Gatenby 2005; SRW Editorial Board 2004b).

ZING update 프로토콜은 SRW/U를 기반으로 하여 개발되고 있으나, 반드시 SRW/U를 적용하고 있어야 하는 것은 아니며 자체 HTML 인터페이스를 가진 데이터베이스 관리에도 활용할 수 있기 때문에 독립적으로 존재한다(SRW Editorial Board 2005). 따라서 SRW/U의 새로운 기능(operation)으로 포함될 것인가 별도의 프로토콜로 만들어질 것인가의 논의에서 후자의 방법으로 결정되어 개발되고 있는 것이다.

## 3. 분산자원 검색 프로토콜 특징 및 비교

분산자원 검색은 Z39.50의 출현으로 본격화되었으며, Z39.50 기반의 검색은 지난 30여년 가까이 수많은 도서관 및 정보서비스 기관에서 폭넓게 활용되었다.

그러나 웹 기반의 네트워크 환경이 일반화되면서는 Z39.50을 적용하지 않고, 여러 가지 자체(proprietary)의 메타검색 방식을 개발하여 분산자원 검색에 이용하는 것이 보다 일반적인 방법이 되었다. 이는 HTTP를 기반으로 하는 웹 환경에서는 자체 메타검색 방식이 Z39.50에 비해 구현이 보다 용이하기 때문이다. 또한 검색 대상에 대한 제약도 자체 메타검색 방식이 Z39.50 검색에 비해 적은데, 이는 Z39.50 검색은 Z39.50 서버가 구축된 시스템만 가능하지만, 자체 메타검색 방식은 웹 페이지에 검색 인터페이스를 가지는 시스템이라면 검색 대상으로 할 수 있기 때문이다.

하지만 자체 메타검색 방식은 표준 검색 프로토콜을 적용하지 않았기 때문에, 검색 대상에 대한 정보(검색 항목, 연산자, 디스플레이 항목 등)를 해당 검색 인터페이스의 개별적인 분석을 통해서 알아내야 하고 또한 이러한 정보가 변경되는 경우에도 변경된 내용을 쉽게 알 수가 없다. 따라서 개발 수준에 따라 검색 효율성에 대한 신

뢰도의 편차가 많이 존재하며, 검색 대상 정보에 대한 불확실성으로 인해 안정적인 메타검색 서비스를 제공하는데 한계가 있다.

SRW/U는 표준 검색 프로토콜이라는 Z39.50의 개념에 웹 기반 기술을 접목시켰다는 점에서 의의가 있다. 즉 자체 메타검색이 수용할 수 없었던 상호운용성을 보장하는 동시에 XML, HTTP 등의 웹 기반 기술을 사용함으로써 구현의 편리성 및 확장성을 가질 수 있게 된 것이다.

Z39.50과 SRW/U는 여러 가지 차이점이 존재하는데, SRW/U는 웹을 기반으로 하였기 때문에 Z39.50의 대표적인 특징 중에 하나인 세션(session) 중심의 고정적(stateful)인 성격을 가지고 있지 않으며, HTTP와 SOAP을 기반으로 동작하기 때문에 유동적(stateless)인 성격을 가진다는 것이다. 또한 레코드 구문(Syntaxes)도 MARC, CCF, GRS(Generic Record Syntax)-1 등 여러 가지를 지원하였던 Z39.50과는 달리 XML만을 사용한다. 그리고 한 서버에서도 다수의 데이터베이스를 가질 수 있으며, 복수의 데이터베이스를 선택하여 검색이 가능하였던 Z39.50의 복잡함을 없애고, 서버와 데이터베이스를 분리하지 않고 단일 구조로 정의하여 하나의 데이터베이스만을 접근하도록 하였다. 질의어도 사람이 쉽게 해석할 수 있는 CQL이라는 질의 언어를 적용하였으며, 질의 방식도 Z39.50의 속성 벡터 형태가 아닌 질의문 내에 색인명이 나타나는 단순 색인 형태로 바뀌었다(SRW Editorial Board 2004a).

#### 4. 분산자원 갱신 프로토콜 특징 및 비교

Z39.50 프로토콜은 확장서비스 기능을 통하여 데이터 갱신 작업을 수행한다. 확장서비스요청과 응답에는 공통적으로 사용되는 파라미터와 7가지 태스크 각각 고유한 파라미터를 사용할 수 있도록 외부형식(EXTERNAL)으로 선언된 task-SpecificParameters가 포함되어 있으며, 7가지 태스크의 고유 파라미터 또한 표준문서에 정의되어

있다.

또한 Z39.50 프로토콜 갱신 확장서비스와 관련된 프로파일로 UCP(Union Catalog Profile)가 있다. UCP는 Z39.50 프로토콜의 갱신 확장서비스를 이용하여 분산 환경에서 데이터베이스 관리를 효과적으로 할 수 있도록 갱신 관련정보를 추가적으로 정의한 것으로, 1996년 처음 개발되었으며, 1999년 국제 등록 프로파일로 승인되었다. UCP에는 서지, 전자, 소장 데이터의 갱신, 일시 정보(date/time stamp)를 이용한 동시 갱신 제어, 통합과 전역변경, 처리 결과에 따른 메시지 등이 정의되어 있다(Gatenby et al. 1999).

그러나 갱신 확장서비스와 UCP를 실제 적용한 사례는 많지 않은데, 이것은 표준 프로토콜과 프로파일이 복수 계층(확장서비스, 태스크패키지) 사용, 일괄/온라인 갱신, 단일/복수 레코드 갱신 등을 정의하고 있기 때문에 실질적으로 구현하기에 복잡하기 때문이다(Sanderson et al. 2004).

이러한 한계를 넘기 위해 SRW/U와 마찬가지로 웹 환경에 적합하고 구현이 용이한 분산자원 갱신 기능으로 제안된 것이 ZING Update이다.

ZING Update의 특징은 SRW/U와 함께 적용할 수 있고, Z39.50 프로토콜 갱신 확장서비스와 UCP의 기본 개념은 유지하면서, 구현이 용이하도록 단순화시켰다는 것이다. ZING Update는 record, info identifiers, 결과 메시지 구조, 추가 데이터(Extra Data) 처리 등을 SRW/U와 동일하게 사용한다. 또한 일괄 갱신, 복수 레코드 갱신, 태스크패키지 개념을 없앴으며, HTTP, XML, UNICODE 등을 기본으로 하기 때문에 웹 환경에 적합하다(Gatenby 2005).

ZING update 초안에 정의된 요청과 응답 파라미터는 <표 1>과 같다(SRW Editorial Board 2005).

OAI 프로토콜은 SP(Service Provider)와 DP(Data Provider)의 요청/응답 방식을 규정한 것이다. SP는 DP들로부터 메타데이터를 수확하여 검색, 브라우징, 원문제공 등과 같은 부가가치

서비스를 제공하고, DP는 자체적으로 각종 디지털 자원을 수집, 보유하여 SP의 수확 요청에 대응하여 적합한 메타데이터를 제공한다.

<표 1> Update 요청/응답 파라미터

명 칭	유 형	구분
srw:version	xsd:string	요청/응답
ucp:operation	xsd:string	요청
ucp:operationStatus	xsd:string	응답
ucp:recordIdentifier	xsd:string	요청/응답
ucp:recordVersions	xsd:string	요청/응답
srw:record		요청/응답
srw:diagnostics	sequence	응답
srw:extraRequestData	srw:extraDataType	요청/응답

SP의 요청은 HTTP 표준(GET 또는 POST 방식)을 사용하고, DP의 응답레코드는 헤더(header), 본문(body), 부가정보(about)로 구성된 XML 형식으로 전달된다. 응답레코드는 SP와 DP의 요청/응답을 수행하기 위해서 사용하는 6가지 요청 유형과 이에 따른 응답 내용은 <표 2>와 다음과 같다(이수상 2004; Lagoz et al. 2004).

OAI 프로토콜은 XML 형식을 적용하고 웹 환경을 기본으로 개발되었다는 점에서 ZING Update와 유사하다.

<표 2> OAI 프로토콜 요청 유형 및 응답 내용

요청 유형	응답 내용
Identify	리포지터리에 관한 일반정보
ListMetadataFormats	리포지터리에서 지원하는 메타데이터 형식 리스트
ListSets	리포지터리에서 지원하는 Set 구조 리스트
ListIdentifiers	리포지터리에서 수확 가능한 레코드의 식별자 리스트
ListRecords	리포지터리에서 수확 가능한 메타데이터 레코드 리스트
GetRecord	리포지터리에서 수확 가능한 특정 메타데이터 레코드

그러나 ZING Update의 경우 클라이언트의 역

할을 하는 로컬시스템에서 서버의 역할을 하는 센터시스템으로 갱신을 원하는 데이터를 보내는 Push 방식인 반면, OAI 프로토콜은 데이터를 통합하여 서비스한다는 점에서 센터시스템과 유사한 역할을 하는 SP가 단위시스템인 DP로부터 데이터를 수확해 오는 Pull 방식이다. 또한 ZING Update가 단일 레코드의 갱신(추가, 교체, 삭제)을 즉각 요청하고 이에 대한 처리결과를 바로 받아보는 온라인 대화식 방식이라면, OAI 프로토콜은 복수의 레코드를 일정(schedule)에 따라 일괄로 처리하는 방식이다. 즉 ZING update는 operation 파라미터의 값으로 create, replace, delete을 지정하여 요청하고, operationStatus 파라미터 값으로 처리결과를 응답하는 방식이지만, OAI 프로토콜은 SP가 특정 기간내의 레코드를 DP로 요청하여 해당 기간내에 추가, 교체, 삭제된 레코드 모두를 가져오는 방식인 것이다.

Z39.50 프로토콜 갱신 확장서비스, ZING Update 프로토콜을 통한 데이터 갱신은 로컬시스템에서 주체가 되어 통합 데이터를 운영하는 센터시스템으로 제공하여야 하기 때문에, 로컬기관의 자발적인 협력이 전제가 되어야 한다. OAI 프로토콜을 사용하는 경우에는 DP가 아카이브를 제대로 운영하고 있다면, 통합 데이터를 운영하는 SP는 DP의 별도 작업 없이도 가져올 수 있다는 장점이 있다. 하지만 메타데이터의 품질 측면에서 보자면, DP가 운영하는 아카이브의 성격이 메타데이터보다는 디지털 자원 제공 측면이 강조되었고, 통합 데이터를 기반으로 SP가 제공하는 서비스에 DP의 직접적인 관련이 적기 때문에 여러 SP의 메타데이터 수준이 일정하지 않으며, 따라서 DP의 경우 메타데이터 품질 관리가 쉽지 않을 수 있다.

## 5. 결론

네트워크를 통한 분산자원 활용은 디지털 도서관, 정보서비스기관의 주요한 서비스로 이미 자리

잡았으며, 그 범위와 대상은 계속 확대될 전망이다. 본 연구는 분산자원 활용을 위한 Z39.50 프로토콜, ZING SRW/SRU 프로토콜, OAI 프로토콜, ZING Update 프로토콜의 개발 배경, 특징을 살펴보고, 검색과 갱신 각 기능별로 프로토콜을 비교·분석하였다.

도서관 및 정보서비스기관은 분산자원을 활용하여 구축하고자 하는 서비스의 대상, 종류 및 개발 환경, 향후 활용 가능성을 고려하여 프로토콜 채택을 결정하여야 할 것이다. 또한 이러한 프로토콜의 기능 확대 및 개선, 적용 사례 및 NISO (National Information Standards Organization)의 MetaSearch Initiative 활동 등은 국내의 분산자원의 효과적인 상호운용을 위하여 계속 주목하여야 할 것이다.

#### 참고문헌

이수상. 2003. "학술정보 유통에 있어 OAI 프로토콜의 적용에 관한 연구." 『한국도서관정보학회지』 35(2) :219-241.

Gatenby, Jannifer, Lesley Bezear and Judith Pearce. 1999. *Union Catalogue Profile*. <<http://www.nla.gov.au/ucp/irp.html>>

Gatenby, Jannifer. 2005. *ZING Update. ZING Information Forum June 22-23, 2005*. [cited 2005. 7. 13]. <<http://www.loc.gov/z3950/agency/zing/forum-june05-output/update.ppt>>

Lagoze, Carl, Herbert Van de Sompel, Michael Nelson and Simeon Warner. 2004. *Open Protocol for Metadata Harvesting - v.2.0*. [cited 2005. 7. 16]. <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>

Morgan, Eric L. 2004. "An Introduction to the Search/Retrieve URL Service (SRU)." *Ariadne Issue* 40. [cited 2005. 7. 16].

<<http://www.ariadne.ac.uk/issue40/morgan/>>

National Information Standards Organization. 2003. *ANSI/NISO Z39.50-2003. Information Retrieval (Z39.50) : Application Service Definition and Protocol Specification*. [cited 2005. 7. 13]. <<http://www.loc.gov/z3950/agency/Z3950-2003.pdf>>

Sanderson, Rob, Matthew Dovey and Janifer Gatenby. 2004. *Updating UCP : Discussion Document*. [cited 2005. 7. 16]. <<http://www.loc.gov/z3950/agency/zing/srw/update/update.html>>

Sanderson, Rob. 2004. *A Gentle Introduction to SRW*. [cited 2004. 9. 4]. <<http://www.loc.gov/z3950/agency/zing/srw/introduction.html>>

SRW Editorial Board. 2004a. *SRW's Relationship to Z39.50*. [cited 2005. 7. 13]. <<http://www.loc.gov/z3950/agency/zing/srw/z3950.html>>

SRW Editorial Board. 2004b. *SRW Editorial Board Meeting October 2004-Meeting Report*. [cited 2005. 7. 16]. <<http://www.loc.gov/z3950/agency/zing/srw/meeting-report.html>>

SRW Editorial Board. 2004c. *SRW: Search/Retrieve Webservice Version 1.1* [cited 2005. 7. 13]. <<http://srw.cheshire3.org/SRW-1.1.pdf>>

SRW Editorial Board. 2005. *Record Update*. <<http://srw.cheshire3.org/docs/update/>>