

Z39.50 프로토콜 기반 통합 검색시스템의 설계 및 구현

이정준*, 신정호**⁰, 최석용**, 장중화***, 오용철*

*한국산업기술대학교 컴퓨터공학과, ** ㈜링크소프트, ***한국원자력연구소 신형원자로개발단
(jjlee,oh)@kpu.ac.kr, **{jhshin, sychoi}@linksoft.kaist.ac.kr, ***jhchang@kaeri.re.kr

Design and Implementation of Integrated Search System Based On Z39.50 Protocol

Jeong-Joon Lee, Jeong-Ho Ssin, Seok-Yong Choi, Jonghwa Chang, Yong-Chul Oh
Dept. of Computer Engineering, Korea Polytechnic University
LinkSoft Co., LTD
Group of Advanced Reactor Development, Korea Atomic Energy Research Institute

요 약

본 논문은 Z39.50 표준 프로토콜을 기반으로 3개 기관의 도서시스템에 기반으로 한 통합검색시스템의 설계 및 구현에 대해서 기술한다. 본 시스템은 각 기관이 독립적으로 운영하는 서로 다른 상용 데이터베이스에 맞추어 Z39.50 프로토콜을 지원하는 Z-서버 3개와 Z-클라이언트 역할을 하는 모듈을 내장한 형태로 웹 기반 검색기능을 제공하는 클라이언트로 나누어 진다. 본 시스템은 독립적으로 운영되던 시스템을 표준화된 프로토콜을 통하여 통합된 검색기능을 제공할 뿐만 아니라, 이 프로토콜을 기반으로 한 다양한 Z-클라이언트들을 통하여 외부에서 검색할 수 있도록 구현되었다. 그리고, 각 기관이 기존 데이터베이스 설계를 전혀 수정하지 않아 기존에 기관 내에서 사용되던 검색시스템도 그대로 사용할 수 있다.

1. 서론

본 논문에서는 Z39.50 표준 프로토콜을 기반으로 독자적으로 운영되던 도서 시스템을 기반으로 통합된 도서검색시스템의 설계 및 구현에 대해서 설명한다. 본 통합검색시스템은 원자력 관문국이라는 시스템에 설치되어 현재 운영되어 국내의 주요 원자력 관련 정부기관인 원자력연구소(KAERI), 원자력안전기술원(KINS), 전력연구원(KEPRI)의 자료를 통합하여 검색하는 기능을 제공한다. 즉, 각 기관이 독립적으로 운영되어 온 도서 시스템을 통합하여 3개 기관이 갖고 있는 자료를 한 화면에서 동시에 검색할 수 있는 시스템을 설계 및 개발하였다. 본 시스템은 표준화된 프로토콜인 Z39.50을 기반으로 개발하여 다른 도서시스템과의 통합검색 개발이 용이하고, 이 프로토콜을 따르는 클라이언트를 이용하면 클라이언트의 개발이 필요 없거나 매우 간편해진다. 본 서버와 함께 대중화된 웹을 기반으로 한 클라이언트를 함께 설계 및 구현함으로써 별도의 프로그램 설치 없이 통합검색이 가능하도록 하였다. 또한, 데이터베이스는 방화벽 안에, Z-서버는 방화벽 안에 위치하여 외부에서는 Z-서버를 통해서만 자료에 대한 접근이 가능하므로 보다 강화된 보안을 제공한다.

본 논문의 구성은 다음과 같다. 제 2 장에서는 Z39.50 프로토콜에 대한 설명 및 구현 사례를 설명한다. 그리고, 제 3 장에서는 통합 검색시스템의 설계에 대하여 기술한다. 제 4장에서는 통합 검색시스템의 구현에 대한 설명으로 개발환경, 개발도구, 개발화면에 대해서 기술한다. 그리고, 제 5장에서 결론을 맺는다.

2. Z39.50 프로토콜

Z39.50은 ANSI에서 지정한 표준 프로토콜로 National Information Standards Organization(NISO)에 의해 개발되어 1988년에 처음 ANSI에서 승인되었고 현재는 Z39.50 Implementors Group(ZIG)과 미의회 도서관(Library of Congress)내의 Z39.50 Maintenance Agency가 담당하고 있다[Nat95]. 버전2는 1992년에, 버전 3은 1995년에 ANSI에서 승인되었다. 현재 버전 4가 개발 중이다.

HTTP나 WAIS와 달리 Z39.50은 클라이언트와 서버간에 연결을 유지하는 세션 지향적(session oriented) 프로토콜이므로 서비스 수행에 필요한 정보를 처음 한번 인증하므로 효율적이다. 이는 클라이언트가 반복적인 결과 요구 및 교정도 가능하다. 또한, 서로 지원 가능한 서비스 종류 확인이 가능하다.

본 연구개발은 ㈜링크소프트를 통하여 한국원자력연구소의 지원을 받았다.

2.1 용어 정의

본 논문에서 사용하는 Z39.50프로토콜 관련 용어를 정리하면 다음과 같다.

- 타겟(Target) : 서비스를 처리하는 객체서버(Server): 타겟을 포함하는 응용 프로그램.
- 오리진(Origin) : 서비스를 청하는 객체.
- 클라이언트(Client) : 오리진을 포함하는 응용 프로그램.
- Z39.50 연관(Z39.50 association : Z-association): 오리진과 타겟 사이에서 이루어지는 통신 세션.

2.2 각 기능에 대한 설명

Z39.50의 주요한 서비스 기능을 정리하면 다음과 같다.

- 초기화(Initialize) 기능 : 초기화(Init) 서비스로 지원 가능한 서비스를 확인, 메시지의 크기 결정.
- 검색(Search) 기능 : 데이터베이스를 검색하여 그 검색 결과인 결과 집합을 생성.
- 추출(Retrieval) 기능 : 결과 집합에서 레코드들을 추출하여 다수의 세그먼트로 나누어서 보내는 서비스.
- 종료(Terminate) 기능 : 연관을 종료.

이 외에도 결과집합삭제(Result-set Delete) 기능, 접근제어(Access Control) 기능, 계정/자원제어(Account/Resource Control) 기능, 정렬(Sort) 기능, 열람(Browse) 기능, 확장서비스(Extended Services)기능, 설명(Explain)기능 등이 정의되어 있다.

2.3 구현 사례

본 절에서는 지금까지 개발된 Z39.50 서버들 중에서 가장 널리 알려진 Isite [CNI94], Zebra[Ind98]를 소개하고, 각 시스템들이 제공하는 기능과 특징에 대해 설명한다.

가) Isite

Isite는 미국의 분산 정보 관리 시스템 개발 기관인 CNIDR (The Clearinghouse for Networked Information Discovery and Retrieval)에서 개발한 Z39.50 서버와 검색 시스템이 결합된 시스템이다[CNI94]. Isite는 크게 Z39.50 서비스를 처리하는 Z39.50 서버, 임의의 검색 시스템에 대한 접근을 일반화 하기 위해 정의된 검색 API, 그리고 CNIDR에서 자체 개발한 검색 시스템인 Isearch로 구성되어 있다. Isearch는 텍스트 기반 검색을 지원하고, 관련 문서의 서열화(relevance ranking)를 제공한다.

나) Zebra

Zebra는 덴마크의 컨설팅 및 소프트웨어 개발 업체인 Index

Data에서 개발한 Z39.50 서버이다[Ind98]. 이 서버는 자체 개발한 검색 시스템을 가지고 있다. 이 검색 시스템은 대용량의 데이터베이스를 구축할 때 자동적으로 여러 디스크에 나누어 저장할 수 있고, 데이터를 ASCII 형태 뿐만 아니라 SGML, MARC같은 형태로도 저장할 수 있다. 또한 검색 시 관련 문서의 서열화를 지원한다. Zebra는 멀티 프로세스 방식으로 구현되어 있으며, 현재 초기화, 검색, 전송, 스캔 서비스만을 지원하고 있다.

3. 통합검색시스템의 설계

3.1 기존 검색시스템

국내의 주요 원자력관련 정부기관으로 원자력연구소(KAERI), 원자력안전기술원(KINS), 전력연구원(KEPRI) 각 기관이 별도로 구성한 검색시스템을 통해 상이한 검색 서비스를 받으며 중복자료 등에 대한 검색이 용이하지 않다.

3.2 통합검색시스템의 구성

Z39.50 프로토콜을 이용하여 기존의 시스템에 맞추어 통합하였다. 그림 1과 같이 각 상용 데이터베이스 서버의 특성에 맞추어 세가지 Z-서버를 개발 및 연동하고, 웹 서버와 연동될 수 있도록 Z-클라이언트를 개발하였다.

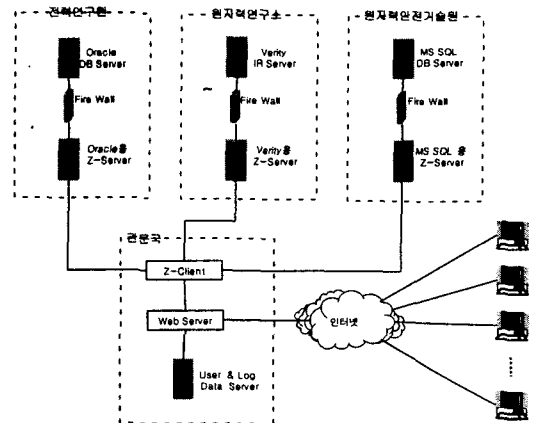


그림 1. 통합검색시스템의 구성.

3.3 Z-서버의 설계

그림 2와 같이 Z-서버는 크게 타겟데몬(Target Daemon: TD)과 타겟레이브(Target Slave:TS) 프로세스로 구성된다. 특정 포트를 리스너하는 TD가 Client요청에 따라 TS를 실행하여 Z-Client와 세션이 연결되어 검색기능을 실행하도록 한다. TS는 크게 Z39.50 프로토콜 처리모듈 ZPM과 데이터베이스 연동모듈 DIM으로 구성된다. ZPM은 인코딩-디코딩 하는 역할을 하며, DIM은 파싱트리 형태 질의를 해당 특정

서버시스템 질의로 변경하는 역할을 한다. DIM은 서비스시스템에 종속적이므로 해당 서버마다 모듈의 변경이 필요하다.

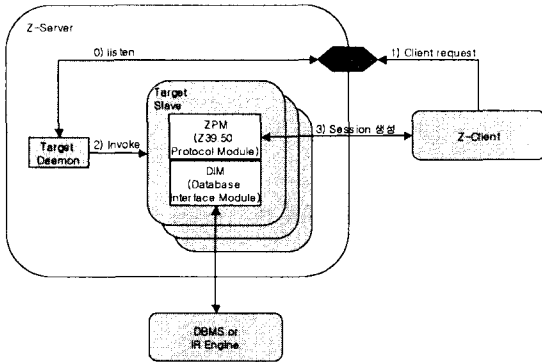


그림 2. Z-서버의 설계.

3.4 Z-클라이언트의 설계

통합검색 서비스를 웹기반으로 제공하기 위하여, 그림 3과 같이 Z-클라이언트는 모듈형태로 JSP-엔진에 플러그인되어 웹서버가 JSP-엔진을 경유하여 사용할 수 있도록 구성되었다. Java Beans 형태의 오픈소스인 JZkit을 기반으로 구현하고, Z39.50프로토콜로 인코딩되어 전달받은 질의결과를 디코딩하여 XML형태로 결과를 전달한다

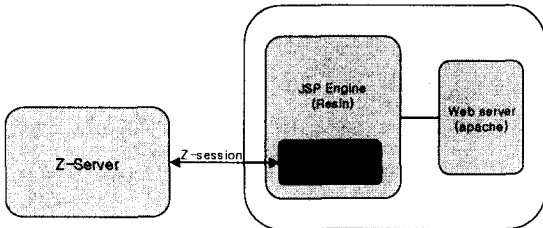


그림 3. Z-클라이언트의 설계.

4. 통합검색시스템의 구현

Pentium-IV 사양의 PC 서버에 리눅스 7.2를 기반으로 다음과 같은 개발도구를 이용하여 개발하였다.

4.1 개발도구

Z-서버는 CrossNet Systems사의 ZedKit 버전 2.2를 이용하고, 데이터베이스의 검색엔진과의 연동은 각 제조사가 제공하는 C-API를 이용하였다. Z-Client는 JZKit Java 모듈을 사용하였다. 그리고, JDK 1.2.2와 JSP 엔진으로 Resin1.2.10을 웹서버로는 Apache 1.3.23을 로그저장 DB는 MySQL 3.23을 사용하였다.

4.2 개발 화면

본 절에서는 지면 관계상 메인 검색화면을 대표로 예로 들어 설명한다. 그림 4와 같이 검색할 기관 선택, 검색할 자료유형을 선택하고 제목, 보고서번호, 발행연도 등의 검색조건들을 입력하여 검색할 수 있는 화면을 제공한다.

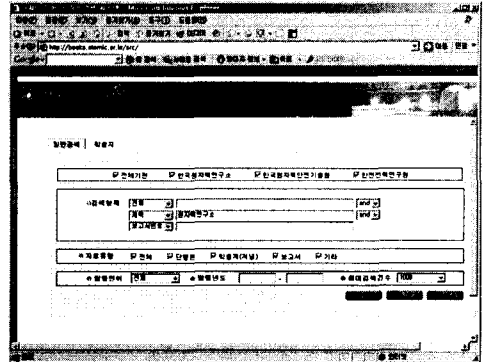


그림 4. 통합도서검색을 위한 메인 검색화면.

5. 결론

본 논문에서는 Z39.50프로토콜을 기반으로 서로 다른 상용 데이터베이스를 기반으로 독립적으로 운영되어온 3개 기관을 통합한 검색시스템의 설계 및 구현에 대해 설명하였다. 본 시스템은 표준화된 Z39.50 프로토콜을 기반으로 서버를 설계 및 구현했으므로, 표준화된 검색기능을 각각의 시스템에 부여하였다. 따라서, 이 프로토콜을 따르는 모든 클라이언트에게 검색기능을 제공한다. 그리고, Z39.50프로토콜 기반 클라이언트 모듈을 웹 환경과 통합된 형태로 제공하여 사용자는 별도의 프로그램 설치 없이 통합검색시스템을 사용할 수 있도록 하였다.

참고문헌

[Nat95] National Information Standards Organization, *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*, NISO Press, <http://lcweb.loc.gov/z39.50/agency>, 1995.
 [CNI94] CNIDR, *The Isite Information System Version 1.05*, CNIDR, <http://vinca.cnidr.org/software/Isite/Isite.html>, 1994.
 [Ind98] Index Data, *Zebra Server - Administrator's Guide and Reference*, Index Data, <http://www.indexdata.dk/zebra/>, 1998.