情報知識学会第 6 回研究報告会講演論文集 (1998.5.23)

Z39.50 プロトコルによる日本語書誌データ情報検索システム

○江草由佳
真野泰久
宇陀則彦
石塚英弘

Information Retrieval System for Japanese Bibliographic Data based on Z39.50 protocol

○ Yuka Egusa
Yasuhisa Mano
Norihiko Uda
Hidehiro Ishizuka

要旨

This paper reports an information retrieval system for Japanese bibliographic data based on Z39.50 protocol. The system consists of a client which can access to any other Z39.50 servers, and of a server which provides Japan/MARC data to any other clients. The server facilitates AND/OR operation between results-sets as well as basic retrieval such as AND, OR, and truncations. The client has a connection window which displays 15 servers, and a retrieval one which displays number of results and data of ones.

1 はじめに

従来の書誌データ情報検索システムは、DIALOG に代表される多数のデータベースを持つ集中型システムであったが、欧米ではインターネットの普及によりネットワーク上に多くの書誌データ情報検索システムが分散して存在する状況が出現した。そして、クライアント / サーバ方式による標準情報検索プロトコル:Z39.50 が規定され、普及するようになった。データベースを持つサーバと検索ユーザが使用するクライアントが Z39.50 を用いてコミュニケートすることによって、サーバ側から見れば、インターネットを通じてどこからでもユーザのアクセスを受け入れることができ、またユーザーにとっては使い慣れたクライアントで多くのサーバのデータベースに検索できるようになる。現在、欧米では、Z39.50 を用いた検索システムは広く使われており、Z39.50 関連のプロジェクトはアメリカ議会図書館の Z39.50 管理機構 [1] に登録されているだけで 105 ある。その中の、地球観測衛星データの書誌検索データおよび観測データ注文を行なうシステムについては文献 [2] に紹介がある。

一方、日本ではまだ Z39.50 が注目され始めた段階で、システム構築例は少ない。Z39.50 の紹介や欧米の事例の紹介としては文献 [3][4][5] などがある。また、システム構築例としては文献 [6][7][8][9] がある。安達は、学術情報センターの電子図書館システム：NACSIS-ELS[6] の中の検索機能に Z39.50 を用いた。日本語による検索と結果表示ができる。ただし、98 年 2 月現在、クライアントから選択できるサーバは NACSIS-ELS のみである [10]。

安齋ら [7] は、 WWW CGI-to-Z39.50 インタフェースを開発し、 WWW ブラウザから日本語で検索できる日本語書誌情報サーバを試作した。原 [8] は、 WWW の Z39.50 ゲートウェイを利用した基本サーバを構築した。クライアントは Isite ソフトウェアパッケージの附属簡易クライアントを用いている。日本語書誌データの検索例は報告されていない。堀江 [9] は Z39.50 ゲートウェイによる JOISwithSTN システムを開発した。ここで STN(The Scientific & Technical Information Network) は英語で書かれた欧米の科学技術分野のデータベース検索システムである。

そこで、本研究では日本語を扱い、かつ、他のサーバ / クライアントから接続できる Z39.50 に基づいたサーバとクライアントを作成した。

2 Z39.50

Z39.50[11] はクライアント / サーバモデルに基づいた情報検索のための通信プロトコルである。 Z39.50 はクライアントサーバ間でやり取りする値を PDU(Protocol Data Unit: プロトコルデータ要素) として転送する。クライアント / サーバ間のやりとりはまずクライアントから Request PDU が出され、それに対しサーバは Response PDU を返すという形でコミュニケーションをとる。

Z39.50 の基本的な機能は接続 (Initialize)、検索 (Search)、返戻 (Present) の 3 つである。接続機能は Z39.50 セッションを開始するためのもので、プロトコルのバージョン、サポートする機能、クライアントの情報、サーバの情報などをやりとりする。検索機能は、実際に検索を行なうもので、検索質問、検索質問の構造、データベース名、ヒットした件数、検索の状態 (検索に成功したかどうか) などをやりとりする。返戻機能は検索結果からレコードを返戻するもので、返戻する数、返戻結果の形式 (USMARC、 SUTRS)、エレメントセットネーム (簡易、詳細)、実際のレコードデータなどをやりとりする。

システム毎に異なる検索質問の構造や検索結果の構造は、 1 つのクライアントから、様々なサーバへのアクセスを妨げていた原因であるが、これは Z39.50 を使用することにより解決される。

まず、 Z39.50 では従来システム毎に異なっていたアクセスポイントの構造を、包括的なスキーマを定義し、実装する際にこれを実際のアクセスポイントにマッピングすることで対処している。書誌情報に対するスキーマは 2 つの数字の組合せで表される bib-1 アトリビュートで定義されている。つまり、この bib-1 で書名、著者名、前方一致、後方一致、 word、 phrase など様々な検索データの特性を表現できる。また、検索結果は、 MARC 形式やプレインテキストである SUTRS 形式など様々な形式を用いることにより対応している。

ユーザは Z39.50 に基づいた検索クライアントを使うことにより、システム固有のインタフェースに習熟する必要なしに、また、システムの違いを意識することなしに Z39.50 を用いた検索サーバにアクセスし、検索できる。また、 WWW の HTTP と異なり、 Z39.50 ではセッションを維持したまま検索ができるため、検索履歴を利用した検索も可能である。

3　システムの概要

図 1: 検索システムのイメージ

本研究では日本語書誌データベースを検索するためにサーバとクライアント二つのシステムを UNIX ワークステーション上で開発した。図 1 にシステムの概念図を示す。左がクライアント、右がサーバを表す。サーバはデータに日本語の Japan/MARC を持ち、 Z39.50 を介した検索要求に答える。またクライアントは検索語の日本語入力、検索結果の日本語表示を可能とし、 Z39.50 を用いた様々なサーバにアクセスし検索することができる。また同様に、サーバも Z39.50 に基づいた様々なクライアントからの検索要求を受けることができる。

なお、クライアント・サーバともに、日本語文字コードには EUC コードを使用した。また、システム構築の際に利用したシステム構築ツール、プログラミング言語は EUC コード対応のものを用いた。

4　サーバの構成と機能

サーバーはプロトコルエンジン、コントロールモジュール、データベースエンジンと 3 つの部分で構成される。

- プロトコルエンジンは、 PDU の読み書きを行ない、クライアントとのセッションを維持する。本システムでは Index Data 社が開発した Z39.50 に基づいた検索システムを作るためのツールキット YAZ[12] を利用した。 YAZ は C 言語で書かれたプ

ログラム群からなり、 Z39.50 の定義するデータ構造を忠実に再現している。そこで、クライアント / サーバ間でやりとりされる PDU のデータ構造の解析のために YAZ を利用した。

- コントロールモジュールはプロトコルエンジンが処理した PDU の内容に応じて処理を行なう。例えば、クライアントからの検索要求の場合、データベースエンジンに検索を要求する。
- データベースエンジンは実際の検索要求や返戻要求に対してデータベースにアクセスし、結果を返す。本研究ではデータベースエンジンに DBMS Oracle7[13] を用いた。

サーバーはクライアントからの要求をプロトコルエンジンで受け取ると、受け取った PDU を解析しコントロールモジュールに渡す。コントロールモジュールは、クライアントからの要求が、検索や返戻であれば SQL に変換して Oracle7 にアクセスする。 Oracle7 は SQL によって得られた検索結果をコントロールモジュールに返す。コントロールモジュールはプロトコルエンジンに返し、プロトコルエンジンは、 PDU に変換しクライアントに送る。検索結果レコードはプレインテキストである SUTRS 形式でクライアントに渡される。

サーバの機能には、接続機能、検索機能、返戻機能がある。接続機能は、クライアントとの接続をする機能である。複数のクライアントを同時に接続できる。検索機能では、クライアントからの検索要求に答える機能である。サポートする検索は、書名と著者名をアクセスポイントとした検索、検索履歴を利用した検索である。返戻機能ではクライアントに検索レコードを返戻する機能である。返戻するレコード形式は SUTRS である。

5　クライアントの構成と機能

クライアントはプロトコルエンジンと GUI で構成される。本研究で開発した検索クライアントは YAZ を利用してプロトコルエンジンを、 Tcl/Tk を用いて GUI 部分を開発した。ユーザーからの入力は、プロトコルエンジンが適切な PDU に変換しサーバに渡す。サーバからの結果を受けとると、受けとった PDU を解析して、ウインドウに表示する。

クライアントの機能には、接続機能、検索機能、返戻機能がある。接続機能はサーバと接続する機能である。検索機能はサーバに検索要求を出す機能である。著者や書名などのアクセスポイントを指定した検索、前方一致や後方一致、履歴を利用した検索などができる。返戻機能はサーバに返戻を要求する機能である。レコード数を指定した返戻、履歴を利用した返戻、エレメントセットネーム (詳細、簡易) を指定した返戻などがある。また、結果レコードの形式は、 SUTRS, USMARC など 8 種類ある。

図 2: 接続ウインドウ

図 3: 検索ウインドウ

6 システムの実行例

図 2は接続ウインドウの画面である。接続ウインドウではサーバとの接続を行なう。クライアントには 15 サーバの情報 (IP アドレス、ポート番号、データベース名) を登録しており、接続ウインドウにそのサーバのリストが表示される。ユーザはそのリストから接続したいサーバをクリックして選択する。図 2は、本研究で開発した Japan/MARC サーバを選択しているところである。また、キーボードからサーバの情報を入力することでリストにないサーバに接続することもできる。

図 3は検索ウインドウの画面である。検索ウインドウでは検索をしたり、検索結果の表示を行なう。検索式は検索式入力フォームにキーボードから入力する。検索ウインドウの初期状態は検索式入力フォームの色が変わっている。これは、ユーザが次に入力すると予想されるボタンやフォームは色が変わり強調されるようにしたからである。

検索式は検索式入力フォームにキーボードから入力する。検索式の書式は YAZ とおなじものを採用した。アクセスポイントを利用した検索や検索履歴を利用した検索が行なえる。検索式を入力し search ボタンをクリックすると検索結果集合番号、検索式、ヒット件数が検索結果表示部分に表示される。

返戻結果の表示では、 B、 F ボタンをクリックすることで簡易表示 (Brief)・詳細表示 (Full) の選択、返戻形式メニューで USMARC や SUTRS などのレコード形式の選択が行なえる。 show ボタンをクリックすると返戻表示部分に直前の検索結果のレコードが表示される。また、返戻入力フォーム (左から : 検索結果集合番号、開始レコード番号、終了レコード番号) にキーボードから入力することで、以前に検索したレコードの返戻や、結果集合の一部のレコードの返戻を行なうことができる。

図 4は本システムのサーバを検索しているところである。検索を順に追うと、

1. 「図書」を検索　ヒットした件数が 68 件
2. 「情報」を検索　ヒットした件数が 52 件

図 4: 検索ウインドウ：Japan/MARC 検索サーバ

図 5: 検索ウインドウ：サーバ名：Wisconsin-madison

3. 1 番目と 2 番目の検索結果集合の AND 検索　ヒットした件数が 2 件

3 番目の検索式の"@set 1" は 1 番目の検索集合を"@set 2" は 2 番目の検索集合を指している。 @and はその AND をとるという意味である。返戻結果表示部分は 3 番目の検索結果集合を表示しているところである。

図 5は Wisconsin Madison 大学のサーバに検索しているところである。検索を順に追うと、

1. 「書名に library」を検索 ヒットした件数が 5660 件
2. 「著者名に Lancaster」を検索 ヒットした件数が 480 件
3. 1 番目と 2 番目の検索結果集合の AND をとる検索 ヒットした件数が 9 件

検索式のアクセスポイントの指定は bib-1 を元に指定しており、たとえば 1 番目の検索式「@attr 1=4 library」の"@attr 1=4" は書名を表す。返戻結果表示部分は 3 番目の検索結果集合を表示している。

7 考察

Z39.50 は異なる検索システムを相互に接続し、利用とすることを目的としている。そのため、 Z39.50 クライアント、サーバともに一つだけではなく、他の Z39.50 サーバとクライアントに接続できる必要がある。

本研究で開発したサーバ・クライアントは Z39.50 の定義するクライアント / サーバモデルに忠実に開発しているので、サーバ・クライアントそれぞれ単独に利用可能である。そのため、クライアントは本システムの Japan/MARC サーバはもちろん、他の Z39.50 サーバにアクセスし検索することができる。また同様にサーバも本システムのクライアントからだけではなく、他の Z39.50 クライアントからの検索要求に答えることができる。

日本語を扱う時には、クライアント/サーバ間で扱っている文字コードが異なると、お互いに検索、返戻時に認識できないコードを受け取ることになり、表示や、検索が行なえない不都合が生じるが、Z39.50 では ISO-2022、ISO-10646(Unicode) もしくはクライアント/サーバ間で合意した文字コードを使うことができ [14]、文字コードの折衝は接続の時に行なう [15] と規定している。本システムでは EUC コードを用いたため、開発したクライアントは EUC コードを扱えるサーバならば日本語で検索することができる。また、同様にサーバも EUC コードを用いるクライアントからの日本語を用いた検索に対応できる。

最近、書誌データを図書からマルチメディア資料に拡張した「メタデータ」が注目されている。Z39.50 システムはメタデータとの関連においても、ますます重要になってくると思われる。

8 おわりに

本研究では日本語書誌データを扱った Z39.50 検索サーバと日本語入力・日本語表示ができる Z39.50 検索クライアントをそれぞれ構築した。

謝辞

本研究用のデータとして Japan/MARC を使うことを許可してくださった国立国会図書館の関係各位に深く感謝いたします。

参考文献

[1] Z39.50 Maintenance Agency. Z39.50 Register of Implementors <URL:http://lcweb.loc.gov/z3950/agency/register/entries.html>

[2] 祖父江 真一, 新井 康平, 落合 治. 地球観測衛星データ用の相互運用プロトコル. 情報処理. Vol.39, No.3, p.222-228(1998.3)

[3] 牛崎 進.Z39.50—IR サービスの通信プロトコル —. カレントアウェアネス. No.175, p.4(1994)

[4] 上田　修一. Z39.50 の可能性と問題点. 三田図書館・情報学会研究大会予稿集 (1996.11)

[5] 特集：Z39.50: 情報の科学と技術. Vol.48, No.3, p.125-180(1998)

[6] 安達 淳. 電子図書館成立の条件. ディジタル図書館. No.1, p.23-27(1994)

[7] 安齋 宏幸, 山本 毅雄, 石塚 英弘. Z39.50 を用いた日本語書誌情報サーバの試作. 情報処理学会情報学基礎研究会報告．Vol.96, No.116, p.9-16(1996.10)

[8] 原 修."情報検索プロトコル ANSI/NISO Z39.50 に対応した基本サーバの構築". 学術情報センター・セミナー研究レポート. 文部省学術情報センター編. 東京, 文部省学術情報センター,1997,p.1-40

[9] 堀江 隆.JOISwithSTN システムの開発 —Z39.50 によるゲートウェイ—, 情報管理. Vol.40, No.8, p.692-701(1997.11)

[10] 学術情報センター. 学術情報センター電子図書館サービス専用クライアント・ソフトウェア 3.0 利用の手引.
<URL:http://www.nacsis.ac.jp/els/man/txt-j.html>

[11] ANSI/NISO Z39.50-1995.Information Retrieval (Z39.50) : Application Service Definition and Protocol Specification
<URL:http://lcweb.loc.gov/z3950/agency/>

[12] Index Data.Index Data homepage.
<URL:http://www.indexdata.dk/>

[13] Eric Armstrong ほか. Oracle7 Server 概要. 第 6 版. 東京, 日本オラクル株式会社, 1995, 299p.(6693-70-1292.ja6) 他 38 冊

[14] Z39.50 Maintenance Agency. Z39.50 and multi-national/multi-lingual environments
<URL:http://lcweb.loc.gov/z3950/agency/indexing/makx.html>

[15] Z39.50 Maintenance Agency. Character Set and Language Negotiation
<URL:http://lcweb.loc.gov/z3950/agency/defs/oid/charsets.html>

図書館情報大学
(University of Library and Information Science)