

## 電子図書館連携のためのZ39.50ゲートウェイの検討

爰川 知宏<sup>†</sup>, 傳田 壮志<sup>††</sup>, 西野 正和<sup>†</sup>, 萱野 忠<sup>†</sup>

<sup>†</sup>NTTマルチメディアネットワーク研究所

<sup>††</sup>九州工業大学 情報工学部

あらまし 電子図書館においては、膨大かつ多様なコンテンツを扱うために、図書館間の連携は非常に重要である。現在、電子図書館システム連携のための分散プラットフォーム技術の検討がされている。一方で、米国の図書館を中心に、ANSI標準の検索プロトコルであるZ39.50を採用した広域分散対応の書誌検索サービスが広く普及しつつある。本報告では、分散プラットフォームをベースとした電子図書館とZ39.50システムとの共存、相互接続を行うための要求条件及び実現方式を、当研究グループにて検討を進めているネットワーキングライブラリNetLibraをベースに検討を行った。

## Z39.50 Gateway for Co-operation among Digital Libraries

Tomohiro KOKOGAWA<sup>†</sup>, Takeshi DENDA<sup>††</sup>, Masakazu NISHINO<sup>†</sup>,  
and Tadashi KAYANO<sup>†</sup>

<sup>†</sup>NTT Multimedia Networks Laboratories

<sup>††</sup>Faculty of Computer Science and System Engineering, Kyushu Institute of Technology

E-mail: {koko, nishino, kayano}@nttmhs.tnl.ntt.co.jp

**Abstract** It is important for digital library systems to co-operate among libraries, because huge varieties and huge numbers of contents have to be managed. Recently distributed networking platform for digital libraries are discussed. many of contents retrieval system in libraries in the US. are based on ANSI Z39.50 protocol which is one of the standard system for information retrieval. In this paper we discussed about requirements and implementation methods for co-operation among digital libraries platforms such as NetLibra and Z39.50-based information retrieval systems.

## 1. はじめに

インターネットの普及やマルチメディア技術の発展を背景に、大学などを中心に電子図書館に関する研究や試みが進められている。World Wide Web (WWW) 上のサービスとして、新着図書やお知らせ等の情報提供、書誌の検索サービス、Q&Aサービスなどの提供はもとより、コンテンツそのものの電子化により、幅広い利用者が来館することなく必要な文献を自宅のパソコン等から閲覧できるな環境ができつつある。さらには、他のインターネット/イントラネットサービスと連携することで、図書館員や司書の情報交換や業務の効率化も考えられる。

しかし、これら電子図書館の構築に関するアプローチは各組織が個別に進めているのが現状である。図書館で扱うコンテンツは莫大であり、大学や自治体クラスの大規模図書館では蔵書数が100万冊を越えることも少なくない[1]。本格的な電子図書館サービスを提供するためには、図書館間の連携は必須事項である。それに対して、図書館間の相互接続や連携のための分散プラットフォーム技術の検討がされつつある[2]。我々の研究グループにおいても、分散プラットフォーム技術を用いたネットワーキングライブラリ NetLibra を提案し、実装・評価を進めている[3]。

一方で、米国を中心に、Z39.50 プロトコル [4] を用いた検索サービスが普及しつつあり、図書館における事実上の標準となっている。これは広域分散検索に適したものであるが、多国語対応や他のサービスとの連携など、解決すべき課題も多い。

本検討では、図書館間の相互連携技術の一つとして、今後主流となると考えられる分散プラットフォーム技術と Z39.50 プロトコルを用いた広域分散検索技術の相互連携方式についての基礎検討を行った。

## 2. Z39.50とその現状

### 2.1 Z39.50 とは

Z39.50 は ANSI (American National Standards Institute) 標準である "Information Retrieval Service Definition and Protocol Specification" の識別番号である。クライアント-サーバ環境で、複数の情報サーバに対する検索を支援する。

ANSI の認可のもとで、NISO (National Information Standards Organization) が情報検索や流通の分野での標準化の一部として Z39.50 を整備し、1988 年以降は、米国家議会図書館の The Network Development and MARC Standards Office が管理している。現在では ISO/DIS 10162 及び 10163 (探索及び検索プロトコル) として国際標準となっている。

WWW による情報検索サービスでは、stateless な HTTP プロトコルをベースにしているため、検索結果集合の蓄積やそれらの間でのブール演算、ネットワーク上の複数のデータベースに対する同時検索などが行えない。また、書誌情報を提供するサーバがインターネット上に多数あるにも関わらず、それらを一つの分散情報システムとして利用することはできない。Z39.50 はクライアント-サーバ間で検索セッションを維持 (stateful) することで、対話的な検索処理を実現する。また、検索質問、検索語、アクセスポイントの指定法、レコードの返戻、データの異機種間のやりとりを規定することにより、複数の情報サーバを一つの情報システムとして利用することができ、ユーザの熟練レベルに応じたインタフェースを装備することが可能となる。

これらの特徴や状況により、米国では図書館の書誌検索サービスにおける事実上の標準プロトコルとなっている。一方、日本においては、Z39.50 ベースの日本語書誌検索サーバの試作[5] などの個別の取り組みはあるものの、言語構造の問題 (マルチバイトコード、漢字フォント、外字、単語抽出など) や、米国の USMARC にあたる書誌レコードの標準

化の遅れなどが、普及への課題として残っている。また、専用のクライアントソフトウェアを必要とすることから、利用環境や他サービスとの連携が WWW ベースの検索サービスに比べて制限される。

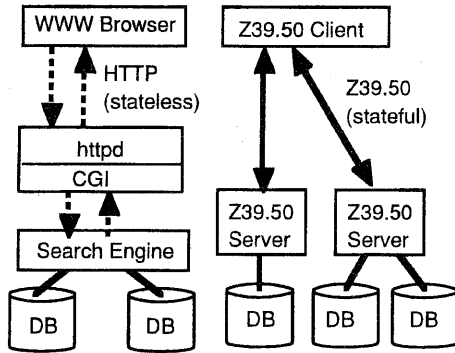


図1 WWW と Z39.50 検索システム

## 2.2 ツール・開発環境

現在、Z39.50プロトコルをベースにしたシステムを構築するための環境（クライアント、サーバ、ツールキットなど）が数多く提供されている。代表的なものとして、

- ・ ZedKit for UNIX  
(<http://www.roadrunner.crxnet.com/>),
- ・ YAZ (<http://www.indexdata.dk/>),
- ・ Isite (<http://www.cnidr.org/>)

がある。

## 2.3 Z39.50 の機能

Z39.50 における Request と Response 処理としては、以下の12種類が規定されている。

- \* Initialize (初期化, セッション開始)
- \* Search (検索)
- \* Present (参照)
- \* Delete Result Set
- \* Access Control
- \* Resource Control
- \* Resource Report
- \* Scan
- \* Sort
- \* Segment

\* Extended Services

\* Close (セッション終了)

2.2 において示した YAZ を用いた場合、標準で用意されているデモプログラムにおいては、一通りの機能をサポートしており、コマンドライン上での Z39.50 クライアントとして動作する。YAZ のクライアントコマンドの動作例を図2 に示す。図2における open, find, show, close コマンドは、それぞれ Z39.50 における Initialize, Search, Present, Close 機能に相当する。情報検索システムの構築には、最低限の機能として、ここに挙げた4つの機能の実装が必要である。

```

Z> open tcp:myhost.domain.com:9999
Z> find @and "library" "science"
Sent searchRequest.
Search was a success.
Number of hits: 16, setno 2
records returned: 0
Z> show
Sent presentRequest (1+1).
Records: 1
[DUMMY]Record type: USmarc
001 11224466
003 DLC
005 000000000000000.0
008 910710c19910701nju      00010 eng
010 $a 11224466
040 $a DLC $c DLC
050 00 $a 123-xyz
100 10 $a Jack Collins
245 10 $a How to program a computer
260 1 $a Penguin
263 $a 8710
300 $a p. cm.
301
Z> close

```

図2 YAZ を用いた Z39.50 クライアント動作例

## 3. 電子図書館連携の検討

### 3.1 電子図書館連携の要求条件

現在、日本では大学や大規模図書館を中心に、電子図書館サービスの構築・実

験提供が行われている。これらは書誌検索のレベルからコンテンツ表示まで様々であるが、外部への公開を意識したものは、基本的に WWW サービスとして実装されている。WWW ブラウザは幅広いプラットフォームに普及しており、標準的な HTML 書式に従う範囲では、プラットフォーム依存をほとんど考慮せずにサービス提供が可能である。したがって、既存の電子図書館と Z39.50 システムとの連携には、WWW ブラウザ上でのサービス統合が必要である。Z39.50 システムは専用のクライアントソフトウェアが必要であり、既存の WWW ブラウザは Z39.50 に対応していないことから、何らかのインタフェース変換（ゲートウェイ）方式が必要である。

Z39.50 ゲートウェイに対する要求条件は以下の通りである。

- (1) WWW ブラウザを介しての Z39.50 サーバへのアクセスを実現
- (2) Z39.50 の特徴である、ブール演算を用いた複雑な検索要求や、複数のデータベースに対する情報検索への対応
- (3) 既存の（電子）図書館サービスとの親和性（メニューの統一、入力項目の統一など）

### 3.2 Z39.50 / WWW ゲートウェイ

Z39.50 と WWW を結ぶゲートウェイ方式は幾つか考えられている。最大の課題は、Z39.50 が stateful なプロトコルであるのに対し、HTTP は stateless なプロトコルであることである。1つの方法として、図3に示すように、CGI プログラムを Z39.50 クライアントに見立て、CGI-Z39.50 サーバ間で stateful なコネクションを維持する。また、WWW ブラウザから入力された利用者情報等を用いてセッション ID を発行し、CGI-Z39.50 間の接続と対応させることで、WWW ブラウザ-Z39.50 サーバ間で擬似的なセッションを実現する。これにより、利用者は通常の WWW サービスと同様のインタフェースより情報検索サービスを利用でき、さらに、Z39.50 の特徴である、広域分散検索

や、ブール演算などを伴う複雑な検索を WWW ブラウザ上から利用できる。この方式は 2.2 で示した Isite などでも採用されている。

この方式の問題点は、セッション終了の検知が困難なことである。WWW サービスは本質的に stateless なため、検索→絞り込み検索といった一連の操作の途中で他のサービスへ利用者が移ったとしても、サーバ側でそれを認識するのは困難である。したがって、サービスメニュー中で終了ボタンを設けるなどして、明示的にセッション終了を表現するか、タイマーを設けて、一定時間新たな入力がない場合にセッションを強制終了する、などの対策が必要となる。

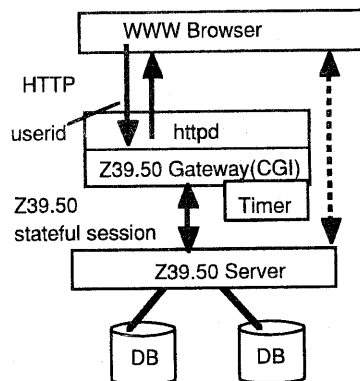


図3 Z39.50 / WWW ゲートウェイ

## 4. Z39.50 / CORBA連携に関する検討

### 4.1 電子図書館と分散プラットフォーム

1章で述べた通り、大規模な図書館システムの構築の観点から、CORBA [6]等の分散オブジェクト技術を用いたプラットフォーム技術が注目されている。分散プラットフォーム技術を用いた電子図書館アプローチは、

- ・接続図書館の分散・自律運用の実現
- ・検索系だけでなく、図書館システム全体に対して、ネットワーク機能を隠蔽、共用化
- ・著作権処理、セキュリティなどの付加

価値機能、管理機能をミドルウェア化してサービス横断的に提供などの特徴をもつ。

分散プラットフォーム技術を用いた電子図書館の例として、ネットワーキングライブラリ NetLibra[3] においては、他の電子図書館プロジェクトと同様、WWW サービスとしてのサービス提供を目ざしている。CORBA とのインタフェースに関しては、3.2 と同様の方式で、CORBA / WWW ゲートウェイを設けている。

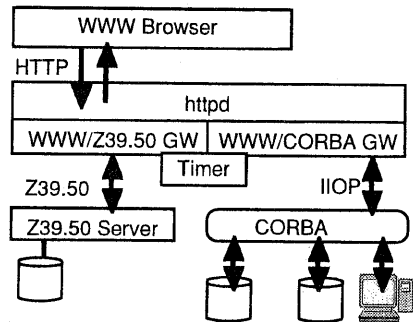
#### 4.2 Z39.50-CORBAゲートウェイの検討

Z39.50 を用いた検索システムとネットワーキングライブラリ NetLibra の連携について検討する。課題となるのは、Z39.50 ベースの検索システムと分散プラットフォーム (CORBA) の協調である。分散検索についてはそれぞれ別個の方式をとっているため、一方に取り込むか並立させるかを検討する必要がある。図4 にZ39.50 と CORBA プラットフォームの連携例を示す。

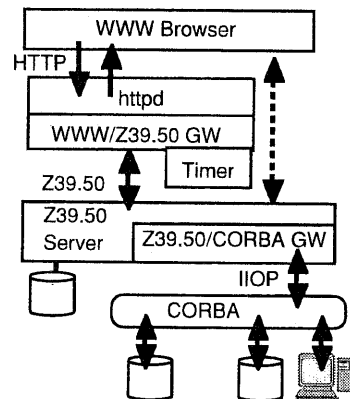
(a)の方法は、単純に両システムを並立して設置したものであり、それぞれのシステムアーキテクチャをほとんど変更することなく、実装が可能となる。どの程度連携させるかは、WWWサービス(CGI)レベルの実装方針に依存するが、横断的な検索や管理を実現しようとする、共通の Query に対する検索式の変換や、セッション管理の点で、サービスアプリケーションレベル(CGI)での負担が大きく、機能制限事項も大きくなる。現在、(a)に基づいてシステムの実装を進めている。

(b)の方法は、CORBA 上のデータベースアクセスを Z39.50 サーバを介して接続することで、システム全体を Z39.50 で統一したものである。プラットフォーム側で Z39.50 / CORBA ゲートウェイを作っておけば、どの Z39.50 クライアントからも接続可能であり、WWW とのゲートウェイを用いた場合でもクライアントの汎用性は高くなる。一方で、検索サービス以外の CORBA 上のサービスとの連携は

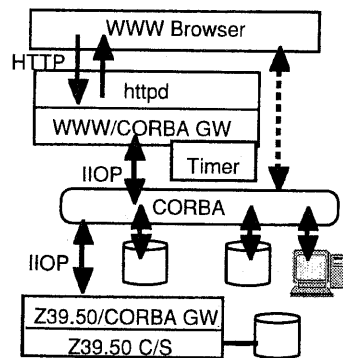
(a) と併用せざるを得なくなり、図書館システム全体としては、構成が煩雑になる。



(a) ゲートウェイの並立化



(b) Z39.50 での検索系の統一



(c) CORBA への Z39.50 の取り込み

図4 Z39.50 / CORBA 連携例

(c)の方法は、CORBAプラットフォームにZ39.50システムを載せてしまうものである。これにより、Z39.50サービスを他のCORBA上のサービスと等価的に扱え、データベースの種類に依存しない情報検索環境が構築できる。その反面、CORBAを介することで、Z39.50にて実装されている機能に制限が付き、Z39.50が持つ本来の特徴を生かせない可能性がある。

#### 4.3 Javaを用いた連携

WWWとのゲートウェイを用いたシステムアーキテクチャは、汎用性の確保と開発コストの削減という面において有効な手段である。しかし、HTTPとZ39.50の接続方式の違いを吸収するために、制限が生じるアーキテクチャにならざると得ない。また、statelessなHTTPを前提としているがために、検索要求の中断などの処理に問題が生じる。

この課題の一解決手法として、Javaを用いた方法が考えられる。この手法では、まず最初にWWWブラウザにJavaアプレットをダウンロードし、このアプレットにゲートウェイ機能を持たせることにより、ブラウザとサーバの間にIIOPないしZ39.50によるstatefulな接続を直接実現できる。ゲートウェイの作成に際しては、ファイアウォールなど技術的な問題が多数存在するが、一部CORBAとJavaのインタフェースは製品化され始めており、今後有効な解決手段となると考えられる。

#### 5. まとめと課題

本検討では、検索プロトコルの標準となりつつあるZ39.50と電子図書館システム連携技術に関する検討を行った。Z39.50プロトコルとその現状(特徴、普及状況、開発環境など)についてまとめ、電子図書館システムとの連携のための要求条件を整理した。分散プラットフォーム技術との連携の例として、ネットワーキングライブラリNetLibraとの連携

のためのZ39.50ゲートウェイの実現方式の提案と比較を行った。現在一部実装を進めており、既にサービスが行われているZ39.50をベースとした検索サービスとの連携等を含めた評価を進めていく予定である。検索系に優れた機能とインタフェースを持つZ39.50と、検索サービスだけでなく幅広い関連サービスとの連携の可能性を持つCORBAプラットフォームの連携は、どちらを主体とするかは一長一短がある。技術的な問題だけでなく、サービスとしての要求条件や、今後の両技術の普及、発展の方向性も含めて検討を続けていく必要がある。

#### 謝辞

本研究の一部は平成10年8月のNTT学外実習プログラムの一環として行ったものである。実習の機会を与えていただきました。マルチメディアネットワーク研究所ネットワーク研究部の鈴木孝至部長、および高機能ネットワーク研究グループ吉開範章グループリーダーに感謝します。

#### 参考文献

- [1] "図書館はいま ー白書・日本の図書館1997",
- [2] 石本, 福嶋, "次世代電子図書館プロジェクトの概要", デジタル図書館 No. 10, pp.57-59, 1997.
- [3] 萱野, 西野, 関, 爰川, "分散ネットワーク電子図書館 NetLibra の提案", 第 57 回情処全大, 1998.
- [4] Library of Congress, "ANSI/NISO Z39.50-1995 Information Retrieval (Z39.50) : Application Service Definition and Protocol Specification". (<http://lcweb.loc.gov/z3950/agency/>)
- [5] 安齋, 山本, 石塚, "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, pp.9-16, 1996.
- [6] Object Management Group, "The Common Object Request Broker; Architecture and Specification", CORBA Version 2.0, 1995. (<http://www.omg.org/>)